Article:

Kleanthous, S and Dimitrova, V (2013) Adaptive notifications to support knowledge sharing in virtual communities. User Modeling and User Adapted Interaction, 23. 287 - 343. ISSN 0924-1868

https://doi.org/10.1007/s11257-012-9127-y

Adaptive Notifications to Support Knowledge Sharing in Close-Knit Virtual Communities

Styliani Kleanthous Loizou[1] and Vania Dimitrova
Cyprus University of Technology, Cyprus
School of Computing, University of Leeds, UK
Email: Styliani.kleanthous@googlemail.com, vania@comp.leeds.ac.uk

Social web-groups where people with common interests and goals communicate, share resources, and construct knowledge, are becoming a major part of today's organisational practice. Research has shown that appropriate support for effective knowledge sharing tailored to the needs of the community is paramount. This brings a new challenge to user modelling and adaptation, which requires new techniques for gaining sufficient understanding of a virtual community and identifying areas where the community may need support. The research presented here addresses this challenge presenting a novel computational approach for community-tailored support underpinned by organisational psychology and aimed at facilitating the functioning of the community *as a whole* (i.e. as an entity). A framework describing how key community processes - transactive memory, shared mental models, and cognitive centrality - can be utilised to derive knowledge sharing patterns from community log data is described. The framework includes two parts: (i) extraction of a community model that represents the community based on the key processes identified and (ii) identification of knowledge sharing behaviour patterns that are used to generate adaptive notifications. Although the notifications target individual members, they aim to influence individuals' behaviour in a way that can benefit the functioning of the community as a whole. A validation study has been performed to examine the effect of community-adapted notifications on individual members and on the community as a whole using a close-knit community of researchers sharing references. The study shows that notification messages can improve members' *awareness* and perception of how they relate to other members in the community. Interesting observations have been made about the linking between the physical and the virtual community, and how this may influence members' awareness and knowledge sharing behaviour. Broader implications for using log data to derive community models based on key community processes and generating community-adapted notifications are discussed.

Keywords: Community modelling, Adaptive support for knowledge sharing, Virtual communities.

Declaration
This paper or a similar version is not currently under review by a journal or conference, nor will it be submitted to such within the next three months. This paper is void of plagiarism or self-plagiarism as defined in Section 1 of ACM's Policy and Procedures on Plagiarism.

1 Introduction

Social environments foster interpersonal ties between people and are becoming the primary medium where humans share knowledge (Wellman, 2001). With the advancement of the internet, social environments are being built online and constitute an important part of peoples' lives and organisational practices (Preece et al., 2003). Virtual communities (VCs) (spanning from large, loosely structured[2] to small, close-knit[3]) are one of the most effective environments for sharing knowledge (Wenger, 2000). We focus here on close-knit virtual communities that may exist either in organisational or educational context and possess the following characteristics: *common purpose, identified by the participants or a facilitator; commitment to the sharing of information and generation of new knowledge; shared resources; interaction and collaboration; equal membership inside the community*. Close-knit VCs are becoming very popular because they offer opportunities for

[1] The work reported here is based on the PhD studies of the author conducted at the University of Leeds, UK
[2] Examples include Delicious communities: http://www.delicious.com/
[3] Private communities built on CiteULike reflect this kind of community: http://www.citeulike.org/

knowledge sharing and collaboration, and are considered to be one of the most inexpensive and successful ways of managing knowledge in an organisation (Nonaka et al., 2000). In academic settings, researchers across the globe are coming together to collaborate on large or small projects, organise academic conferences or just share resources and develop collective knowledge (Puntambekar, 2006; Paletz and Schunn, 2010).

Studies show that for VCs to be successful, more than just people and technology is needed – intelligent support for the effective functioning of the community is paramount (Fischer and Ostwald, 2001). This requires a good understanding of VCs focusing on processes that have to be supported to enable effective knowledge sharing and add value for all members (McDermott, 2000). Researchers in Computer Supported Cooperative Work have argued that organisation and psychology theories should be followed as a solid foundation for enabling effective knowledge sharing and working together in communities (Barley et al., 2004). The implication of this argument for personalisation and adaptation technologies for teams and close-knit communities drives the research presented in this paper. We explore how to utilise organisation psychology to develop a *holistic community adaptation approach* that gains an understanding of a close-knit virtual community and provides support tailored to the needs of the community as a whole.

Following (Mohammed and Dumville, 2001; Ilgen et al., 2005), processes that can have an impact on the collective knowledge sharing and are important for the effective functioning of teams and close-knit communities have been identified:

- *Transactive Memory* (TM): Members are aware how their knowledge relates to the knowledge of others;
- *Shared Mental Models* (SMM): Members develop a shared understanding of the key processes and the relationships that occur between them;
- *Cognitive Centrality* (CCen): Members who hold strong relevant expertise can be influential; members of effective communities gradually move from being peripheral to becoming more central and engaged in the community.

Based on TM, SMM and CCen, this research aims to support members of a close-knit community to answer questions like: "Who knows about subject S?", "What are others doing in this community?", "Who shares the most valuable resources in this community?", "Whose knowledge is important to me?", "Do others in this community know what I know?", "To whom is my knowledge important?". Studies stress that the outcome of a member's actions in the community can influence others' actions (Schmidt, 2002). 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. Discovering patterns that promote the development of a TM system, the establishment of SMM, and exploit CCen, can facilitate knowledge sharing in the virtual community (VC) (Ilgen et al., 2005).

The paper presents a novel holistic personalisation approach for supporting close-knit VCs and addresses three research questions:

- *How to extract a computational model to represent the functioning of a community as a whole by using semantically enhanced system log data?*
- *How can user modelling, adaptation and personalisation techniques be utilised to support processes which are important for the functioning of close-knit virtual communities?*
- *Can adaptive support, driven by community processes, affect the functioning of the community?*

To address the above questions, we have developed a computational framework for community-based adaptive support for knowledge sharing. This involves the extraction of a community model, and its use for identifying community knowledge sharing behaviour patterns. These patterns are then used as input to algorithms for generating adaptive notifications that target individual members but aim at improving the functioning of the community as a whole. An evaluation study is conducted to validate the framework and examine what effect the generated notifications have on individual members and on the functioning of the community as a whole.

The research presented here contributes to an emerging trend in adaptation and personalisation developing intelligent techniques to provide adaptive support for team collaboration (Paramythis et al., 2011). 'Traditional' approaches of personalisation in social systems use the social context in order to enhance adaptation mechanisms aimed at benefiting individual members. In contrast, we propose an adaptation approach aimed at benefitting a close-knit community *as a whole* (which in turn will benefit the individual members). Such *holistic personalisation approach* allows for social processes to be taken into account, and the personalised support (in terms of notifications) to target individual community members, but aim at improving the knowledge sharing and functioning of the whole community as an entity.

The paper makes the following contribution to research in personalisation in social environments, focusing on adaptive support for knowledge sharing in close-knit communities:

- a novel semantic-enriched mechanism for extracting a community model based on log data analysis and underpinned by three key community processes (TM, SMM and CCen);
- an original mechanism for discovering community knowledge sharing patterns by mapping the three community processes (TM, SMM, and CCen) to properties in graphs representing community relationships; and applying this mechanism for generating community-adapted notifications;
- an experimental evaluation study which examines benefits and drawbacks of community-adapted notifications and derives implications for future work in socio-personalisation.

The next section will compare our approach with related work, pointing at key differences and justifying the major contribution of the research presented in this paper. Section 3 provides an overview of the proposed computational framework for community adapted support. Section 4 presents the community model extraction mechanism. A graph-based mechanism for extracting community knowledge sharing behaviour patterns is described in Section 5, and the use of these patterns for adaptive notification generation is presented in Section 6. Section 7 presents an evaluation study describing the experimental design and the main findings, followed by a discussion on the lessons learnt and revisiting the research questions to conclude the paper.

2 Related Work

Our approach relates to existing work in three research areas. Firstly, we will compare to similar user modelling approaches. Then, a comparison will be provided to social network techniques that use graph based algorithms for identifying connections based on people's interactions in social settings. Finally, we will discuss relevant approaches for intelligent support in social environments.

2.1 User Modelling Approaches

Modelling virtual communities has recently become popular in different research areas. For the purpose of this research, approaches from both user modelling and social network areas are considered as important and discussed in this section. In user modelling, modelling a group provides the grounds for generating group recommendations (Masthoff, 2004); in social networks, community modelling aids the discovery of relationships between people or communities (Lin et al., 2008).

Discovering Connections: A fairly simple and elegant community model is presented in (Cheng and Vassileva, 2006). It is based on a list of topics based on the resources VC members are sharing. A reward factor is calculated to measure the relevance of each contributed resource to the current topic the VC is working on. Each member has an individual user model consisting of the reputation measure of that member in the VC (Cheng and Vassileva, 2006). An earlier work in the same group presented a more elaborate relationship model (Bretzke and Vassileva, 2003), which is the closest to ours. Users' interests are modelled in (Bretzke and Vassileva, 2003) considering how frequently and how recently users have searched within 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 along with an ontology representing the community context, and derives a semantically relevant list of interests for every user.

Modelling Interests: A different approach is followed by Tian et.al. (2001) where the community model represents the interaction activities that happened in the VC. All interactions are associated with a core lexicon which represents the interests of people in the VC. User interests are modelled according to the interactions each user is performing in the VC and associated to the core lexicon of the VC. Shared interests or relationships are also modelled based on social interaction activities of users, and are linked to the VC lexicon. The approach presented in this paper also models user interests based on resources members are uploading or downloading. However, we exploit semantic enrichment of the uploading/downloading activities by using, in addition to the resource key words, concepts extracted from an ontology. For example, we use semantically-enriched data to extract interest similarity between community members.

User interests have been extensively studied also in (Davies et al, 2003) - an approach where user interests are extracted as keywords from the user profiles and other web content shared by a user in the community is presented. An ontology is then accessed, where associations are derived with ontology concepts and further recommendations are made to users. Interests are also used in finding relationships between users or connections in social graphs. Li et al (2008) are extracting interests based on tags users ascribe to items posted online. Relationships/associations between users are derived based on their tags. This is similar to the interest relationship model used in our work - both approaches consider that members can be connected by interest similarity even when they have not read any resources uploaded by each other.

Modelling Expertise: Interests of users are usually associated with expertise, especially in social network research (Song et al., 2005; Fu et al., 2007; Lin et al., 2007; Zhang et al., 2007). Zhang et al. (2007) extract shared interests in a discussion based on posting/replying threads. Based on the discussion topics a member of the community is contributing to, his/her interests and expertise are extracted; subsequently, user interest relationships are obtained. Fu et al. (2007) are following a similar method but are mining email communication networks. Relationships are inferred according to the expertise/interests of members, which are extracted from communication recorded in their email conversations. Modelling expertise relations plotted as graphs is also explored by Song et al.(2005). A relational network is extracted according to people's publications. The expertise/interests of a person are obtained by his/her previous publications; and two people are considered related if they have publications in the same research area. The work presented here adds to the above approaches. Our approach does not identify only expertise, but also derives a person's influence in the VC based on the relationships he/she has developed with others, which can benefit the VC as a whole.

Community Graph Models: Existing research employed graph theory to model communities and relationships between members (Hubscher and Puntambekar, 2004; Kay et al., 2006) or members' interactions in general (Falkowski et al., 2007; Falkowski and Spiliopoulou, 2007). In (Hubscher and Puntambekar, 2004), the individual user model represents the conceptual understanding of a user, based on which a graph network is constructed. Similarities are then extracted according to a user's conceptual understanding, and group models are derived based on the distance between members in a graph. Kay et al. (2006) uses the notion of an interaction network to represent relationships between users in a learning community. Two members are related if they have modified the same resource; hence, they appear connected in the interaction graph. Falkowski et al. (2007) consider the exchange of messages as an interaction between two users, represented in a graph. A relationship between two users exists if they have engaged in some message exchange (Falkowski and Spiliopoulou, 2007). Our work also follows a graph-based approach to model a community. The key contribution of the approach presented here to graph community modelling is the consideration of semantic relationships in addition to the interactions between users, i.e. an edge connecting two members represents their semantic similarity to each other, and the relevance of this link to the community's domain.

2.2 Graph-based Approaches for Social Network Analysis

Graph mining has been used in social network analysis primarily for monitoring information flow and improving communication in organisations (Chakrabarti and Faloutsos, 2006). In social networks, graphs represent individuals as nodes, and edges are their interconnections, which can represent business relationships, email conversations or, as in this research, semantic relationships based on knowledge sharing. Structural patterns in social networks refer to mathematical attributes of graphs that can be recognised in a network (e.g. cliques, degree centrality, structural equivalence, and structural holes). These are quantitative approaches that do not consider any semantics and thus have yet to be applied to the investigation of social "roles" (e.g. newcomers, oldtimers) or social "power" (e.g. central or peripheral members). A review of different methods used for extracting communities (sub-networks of people structurally connected together in a graph) from large networks over the web is given in (Chakrabarti and Faloutsos, 2006). These approaches focus on structural attributes of graphs which are linked to patterns that can be identified in peoples' interactions (Khan and Shaikh, 2006; Lo and Lin, 2006; De Choudhury et al., 2007; Viermetz and Skubacz, 2007; Kunegis et al., 2009). We review them below discussing common features with our approach.

Measuring Centrality: Viermetz and Skubacz (2007) apply text mining techniques to email conversations to extract patterns/networks of people and their relationships. Keywords are extracted from email communications to form vectors, each vector represents a single message. From this, a network of email messages is built. Clusters of similar messages are found using DBScan (Ester et al., 1996). Network segmentation combines the messages for each topic cluster into a sub-graph of the messages extracted by the actors involved. The centrality of each actor is measured according to how central the topics of the messages exchanged by this actor are. Compared to our work, it can be noted that the text analysis does not consider the relevance of each email message to the overall community. Thus, the extracted centrality measure considers only the keywords associated with a user and does not take into account the importance of these keywords to the community as a whole. In addition, our approach measures centrality considering also the relevance of a resource or a relationship between two members to the rest of the community by using an ontology representing the community domain.

Extracting Patterns of Behaviour: A publication network is used in (Khan and Shaikh, 2006) to extract predefined algebraic functions that represent social relationships in a network. The patterns developed have been applied to an existing publication network to extract reviewers for a specific paper. The algebraic functions deal with structural functions of graphs and have no added semantics. For example, in the extracted network a binary 1 represents that an edge exists between two nodes and a 0 denotes no edge between the two nodes. This does not represent any semantic connection between the two nodes in the network. A different application discussed by Khan and Shaikh (2006) is to identify who will be infected from the nearest network of a person if that person is infected with some kind of virus. In order to discover who will need immunisation, algebraic functions are applied to the network of that person using structural characteristics of the social network graph (matrices and sets operations). In contrast, our approach not only takes into account structural characteristics of the graphs, but also links these characteristics to patterns related to social processes in close-knit VCs.

A recommendation tool developed in (Lo and Lin, 2006) suggests friends to community members based on exchange of messages. Sending messages to each other is the connection that exists between two community members in their network. No semantic information representing the message content is considered. In our work, semantics is taken from the keywords of the resources shared by members. In addition, we also consider different connection types between two members (e.g. if two members read, upload similar resources, or have resources shared with each other).

All the above approaches have developed some pattern algorithms and employed them to analyse people networks. Generally, we follow the same approach. However, our approach focuses on modelling 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 domain. This, combined with the theoretical underpinning, forms our contribution to social network analysis, i.e. a graph-based approach for qualitative analysis to automatically detect relevant interaction patterns.

2.3 Relevant Techniques for Virtual Communities

There is a growing interest in providing intelligent support for teams, groups and communities.

Visualization Techniques: Visualization techniques are among the most popular methods that can be employed to present group and community models in a graphical way, to help groups function more effectively (Kay et al., 2006; Upton and Kay, 2009; Ardissono et al., 2011), to motivate community participation (Cheng and Vassileva, 2006), and to make members aware of reciprocal relationships (Sankaranarayanan and Vassileva, 2009). The key limitation of visualization 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 as a whole and what activities they can engage in. In contrast, we analyse a community model to automatically detect problematic cases which can be used to decide when and how to intervene, offering support to improve the knowledge sharing processes in the community.

Identifying Interest Clusters: User browsing behaviour has been used as a source for identifying clusters of users with similar interests (Pierrakos and Paliouras, 2010). Such clusters play the role of community models, and can be used for personalisation by narrowing down the search space and choosing the category most related to the user's interests. The community web directory modelling approach is applicable to a large, loosely structured community, and this enables the use of clustering and probabilistic modelling techniques. In contrast, we consider close-knit communities, where the interests are more compact (i.e. there exists a common community context, which we represent as an ontology) and relationships between users develop based on uploading and downloading of resources related to the community context. Hence, instead of machine learning techniques, we consider semantic augmentation and similarity, which allows us to model relationships at a finer granularity.

Identifying Expertise: Different tools and algorithms have been developed to support people in locating expertise on a specific subject inside groups or VCs (Shami et al., 2007; Zhang et al., 2007; Pal et al., 2011). Our work contributes to this research stream. In addition to identifying the interests and expertise of community members, we detect possible connections between members which have not been exploited in the community. This is used in notification messages to encourage cognitively central and peripheral members to engage in interactions beneficial for the VC.

Intelligent Group Interventions: The closest to our approach is research on intelligent interventions to support groups or communities, which are usually used to augment existing collaboration environments. For example, (Baghaei and Mitrovic, 2007) extend a constraint-based tutoring system with intelligent feedback, based on analysis of students' activities and comparing them to an ideal model of collaboration, which is intended at making the collaboration process more effective. Collective bookmarking systems have been extended with algorithms for detecting and promoting cognitively central menbers (Bretzke and Vassileva, 2003; Farzan et al., 2009). Most recently, (Ardissono and Bosio, 2012) propose extending collaboration environments to improve user awareness by providing context-aware notifications delivered on the basis of the user's current activities and taking into account collaboration context. Our work is positioned in this research stream and makes a unique contribution by examining how social processes (transactive memory, shared mental models, and cognitive centrality) seen as important for the functioning of a close-knit VC can inform the generation of notification messages. The key novelty of our work is that we consider semantics between relationships and suggest community interventions aimed at improving the functioning of the VC as an entity taking into account TM, SMM, and CCen.

In the next section we will present the computational framework developed in this research and discussed its main components.

3 Computational framework for community adapted support

The main hypothesis of this research is that providing adaptation tailored to the community as a whole by promoting the building of TM, development of SMM, and identifying CCen inside the community can improve the functioning of a close-knit VC. Based on this, and following the general architecture of user-adaptive systems presented in (Jameson, 2003), a framework for providing intelligent support to virtual communities has been defined (see Figure 1). It includes: (i) acquisition of a community model that represents the whole community and focuses on aspects related to Transactive Memory (TM), Shared Mental Models (SMM), and Cognitive Centrality (CCen); (ii) application of the community model to extract patterns of knowledge sharing behaviour that are influenced by TM, SMM and CCen and (iii) using the detected patterns to offer adaptive notifications and improve the functioning of the community.

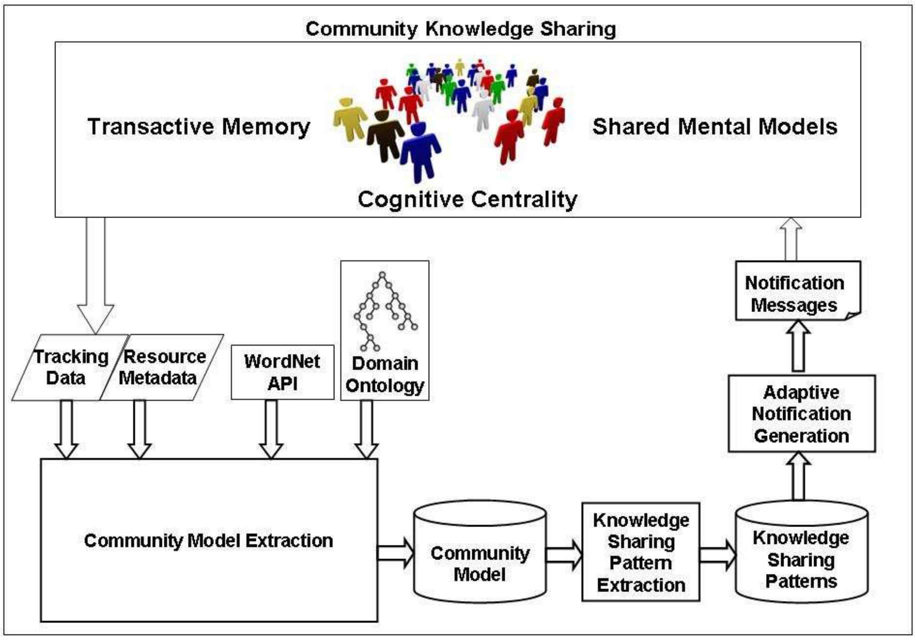

Figure 1 Computational framework to provide holistic support to close-knit VC. A Community Model is extracted from formalised system tracking (log) data. Knowledge sharing patterns are defined based on TM, SMM, and CCen. Using the extracted patterns, adaptive notifications are generated.

Community Model Extraction: Extracting a community model involves identifying the input data, formalising the input data and defining the community model components. The main input for our framework is tracking data from a knowledge management system the VC uses. We consider generic tracking data common in most knowledge management system. This includes information about: members, reading/uploading resources, folder creation/deletion, and resource rating. In addition to the tracking data, we consider metadata which can be extracted from the resources people are sharing. Metadata is an important addition to the input data since they provide the semantic information about a resource that is missing from the tracking data. Dublin Core is selected over the IEEE Metadata element set since it is much simpler in structure. To further semantically empower the community model acquisition algorithms, an ontology representing the main topics of the community is employed. The ontology represents the vocabulary relevant to the main area of interest of the community under study. Furthermore, WordNet is used to measure the semantic importance of a resource to the VC and the semantic similarity between members and resources. A description of the community model extraction mechanism is provided in Section 4.

Knowledge Sharing Patterns Extraction: In order to provide intelligent support to VC members, relevant situations/patterns have to be discovered. The extracted community model is used as an input for the mechanism for automatic detection of knowledge sharing patterns related to TM, SMM, and CCen. The community model includes several relationships represented as graphs. Graph based algorithms have been defined to automatically detect pre-defined patterns from the community relationship graphs. Detected patterns relate to lack of awareness regarding similarity or complementarity between members, as well as interesting behaviour of cognitively central or peripheral members. The detected patterns serve as an input to algorithms which generate intelligent community-tailored support. The mechanism for extracting community knowledge sharing patterns is presented in Section 5.

Adaptive Notifications Generation: The purpose of detecting knowledge sharing behaviour patterns is to assist with providing personalised support where and when needed. Support in this work is designed as personalised notification messages that target individuals or groups of members who are detected in a specific pattern and will benefit from a specific message. The personalisation aspect lays in the content of the notification messages, containing information tailored to the detected behaviour of each member and target four important community participation and awareness aspects, such as participation of central and peripheral members, improving the community TM, and developing SMM. The mechanism for automatic generation of community-tailored notifications is presented in Section 6.

4 Community Model Extraction

This section will provide detail of the stages followed for extracting a community model. Firstly we will formalise the input data used for extracting the community model. Then the use of an ontology and the WordNet similarity mechanism will be described. At the end, algorithms for extracting a community model will be presented and discussed in detail.

4.1 Input Formalisation

Input formalisation is the first step towards the implementation of the community model. A conventional structure of log data stored by knowledge sharing applications is considered. A community environment contains elements related to the functioning of a knowledge sharing community, and includes a list of members M, set of resources R and set of folders F organised in a hierarchical structure. The community environment E is defined as $E{:}\langle M,R,F \rangle$. E is changing over time as a result of actions performed by community members, including a member registering to the community, a member leaving the community, creating a new folder, uploading a new resource to the environment, downloading a resource from the environment, adding a new description to a resource. These actions can cause the environment to evolve, e.g. topics to change or members to move into the periphery or the centre of the community. The actions are recorded in log data which can be analysed periodically to extract a community model and detect changes in E. A description of the tracking data is presented in Table 1. The log data also includes information about members, resources, and folders.

When a member m $(m \in M)$ joins the community, information about their name, email address and date of joining is recorded. Thus, members are represented as $m : \langle mName, mEmail, mDateJoin \rangle$. The community knowledge sharing space includes resources uploaded and downloaded by community members. A resource r $(r \in R)$ will be represented as tuple $r{:}\langle rCreatedData, rMetadata \rangle$, where $rCreatedData$ is information created by the member who uploads the resource, while $rMetadata$ is metadata associated with this resource. The metadata provides the first semantic layer used for community modelling. Resource keywords are used for semantically comparing resources.

When a member uploads a resource r, he/she creates some metadata about this resource. We denote this data with $rCreatedData{:}\ \langle rFolder, rName, rDescription, rRating, rCreator, rDate, rAssessors, rReaders \rangle$, where $rFolder$ is the folder storing the resource; $rName$ is the name of the resource (as given by the creator, and may be different from the original title of the resource), $rDescription$ denotes a set of resource descriptions provided by different members $rDescription : \{\langle rd_1, m_1 \rangle, \langle rd_2, m_2 \rangle, \ldots, \langle rd_n, m_n \rangle\}$; $rRating$ is a number which is the average rating given to that resource by community members, $rAssessors : \{\langle ra_1, m_1 \rangle, \langle ra_2, m_2 \rangle, \ldots, \langle ra_n, m_n \rangle\}$ represents the ratings given by members to this resource; $rCreator$ is a member $(rCreator \in M)$ who is the creator of the specific resource (usually the member who uploads the resource in the community knowledge sharing space); $rDate$ is the date the resource was uploaded; $rReaders$ indicates which members have read (downloaded) this resource.

In addition to the meta-data provided by the resource creator, there is existing metadata included in the resource itself. This is indicated wirh $rMetadata$ and represent the reseource metadata following the Dublin Core[4] schema (the basic and most conventional standard for online resources). The following elements have been selected from the Dublin Core metadata schema $rMetadata : \langle rTitle, rAuthor, rSource, rKeywords, rDatePublish \rangle$. $rMetadata$ can be either extracted from the uploaded resources or provided by the $rCreator$ when he/she uploads the resource.

Table 1 Summary of tracking data and resource metadata formalisation

[4] The Dublin Core Metadata Initiative (http://dublincore.org/), is an open organization engaged in the development of interoperable metadata standards for sharing resources on the web.

Description	**Formalisation**
Data about a member m	$m:\langle mName, mEmail, mDateJoin\rangle$
The creator of a resource r	$rCreator \in M$
Data about a resource r created by the member who uploaded the resource ($rCreator$)	*rCreatedData: <rFolder, rName, rDescription, rRating, rCreator, rDate, rAssessors, rReaders>*
Dublin Core based metadata for a resource r	*rMetadata :< rTitle, rAuthor, rSource, rKeywords, rDatePublish>*
Keywords of a resource r	$rKeywords:\langle k_1, k_2, \ldots, k_n\rangle$
Descriptions for r provided by members $m_1 \ldots m_n$	$rDescription:\{\langle rd_1, m_1\rangle, \langle rd_2, m_2\rangle, \ldots, \langle rd_n, m_n\rangle\}$
Assessments (ratings) for r provided by members $m_1 \ldots m_n$.	$rAssessors:\{\langle ra_1, m_1\rangle, \langle ra_2, m_2\rangle, \ldots, \langle ra_n, m_n\rangle\}$
Members who have read (downloaded) resources r	$rReaders: \{ m_1 \ldots m_n \}$

The most important metadata for our community modelling mechanism is $rKeywords:\langle k_1, k_2, \ldots, k_n\rangle$, which is used as a source for comparing resources and users. To compare a set of key words, we link them to an ontology which represents the community context. This enables positioning a resource within the community context, and allows enriching the keywords list with relevant concepts, as well as measuring how important a resource is for the community. The next section will provide detail about the use of the ontology for community model extraction.

4.2 Use of an Ontology

The domain in which the knowledge sharing community operates is represented as an ontology. The vocabulary composes the classes of the ontology and represents the domain topics of interest to the VC. The ontology Ω is selected according to the subject of focus of the community. The ontology is used as a source to semantically empower the algorithms for extracting the community model. For example in the case of the VC which is used in the experimental study in Section 7, an ontology that represents the Personalisation and Knowledge Management domain has been employed. The ontology has been built using concepts extracted from the folder hierarchy of the community space. The hierarchy is modified to represent logical relations (subClassOf), between concepts; the ontology consists of 159 classes. All concepts of the ontology were transformed into nouns so they can be used as a direct input in the WordNet similarity measure algorithm (described in Section 4.3).

The relevance of an uploaded or downloaded resource to the community is checked against the domain of the particular VC by using Ω. This is used to determine the value a resource has for the community, to identify similarity between resources, and to detect semantic similarities between members. Ω is used to determine the value a resource uploaded by a member and read by another member has to the VC. Consequently, the value of a resource r_i for the VC is defined as $V_{r_i} = Sim(rKeywords_i, \Omega)$ where $Sim(rKeywords_i, \Omega)$ is the similarity of the list of keywords $rKeywords_i$ of a resource r_i to the ontology Ω. The similarity is obtained using WordNet (see Section 4.3).

The second way the ontology is used is to define the similarity between community members based on the resources they are reading, uploading, and their interests. For each member m $(m \in M)$, a list of key words for that member ($mKeywords$) is composed by aggregating the key words for all resources m has created (uploaded). Using the ontology Ω, the list of keywords for each member m ($mKeywords$) is semantically enriched. This enables us to define member similarity based on the community context, where similarities and relations between members which are relevant to the community are rewarded and the less relevant relations are still counted but do not have strong influence on the strength of the relationship between members.

To semantically augment the keywords for a member m $(m \in M)$ ($mKeywords$) using the ontology Ω, we do the following. For each keyword $k \in mKeywords$

1. Check that k is a label for a concept in the ontology.
2. If $k \in \Omega$, pull from Ω all direct super classes, and all direct sub classes: $Super(k,\Omega) \cup Sub(k,\Omega)$ and use them to expand the key word list $mKeywords$.
3. If $k \notin \Omega$, perform a similarity check $Sim(k, \Omega)$ by using a WordNet similarity algorithm (e.g. in the implementation we used the algorithm presented in Section 4.3), in order to get the most similar concept C_{Sim} to k from the ontology. Once a similar concept C_{Sim} is selected, go to step 2 and perform the corresponding enrichment.

This is illustrated with the following example. Consider a member a and assume that the list of keywords for this member is $aKeywords = \{semantic\ web,\ knowledge\ sharing,\ context,\ collaboration\}$. Take $k = collaboration$, which is linked to a label of a concept in the ontology, i.,e. $k \in \Omega$. Then, we use the algorithm $Super(collaboration,\Omega) \cup Sub(collaboration,\Omega)$ to extract the direct super and sub classes of $collaboration$ from the ontology. In this example - for $Collaboration$ based on the ontology used the classes $Knowledge\ Management$ (as a super class) and $Information\ Sharing$ (as a sub class) will be returned. $Knowledge\ Management$ and $Information\ Sharing$ will then be added into the list $aKeywords$ to semantically enhance the list of keywords of member a. If $k = net$ and $k \notin \Omega$, then we perform a similarity check $Sim(net, \Omega)$ by using the WordNet similarity algorithm and the most similar concept $C_{Sim} = web$ will be returned. Then, the super and sub classes of web will be added to $aKeywords$.

The use of the ontology in this work is inspired by approaches followed in linked data where an existing, stable ontology is used to semantically augment and link content. In our case, this also allowed to measure similarity between resources and community members based on the concepts in the ontology. Instead of using a pre-defined ontology, the emerging community ontology could be integrated. Although the existing framework could be applied in this case (by simply taking a snapshot of the ontology at the time when the community model is being obtained as opposed to using the same ontology throughout), further research is needed to examine what similarity metrics would be appropriate when the community context changes dynamically. For instance, when the ontology changes, the importance of a resource would change, as it would be linked to a different sent of concepts at any different time. This will also affect the strength of relationships between members.

4.3 Measuring Semantic Similarity based on WordNet

Semantic similarity relates to measuring the similarity between concepts which are not lexicographically similar but have similar meaning (Varelas et al., 2005). In Natural Language Processing it is commonly argued that language semantics are mostly captured by nouns so it is common to built retrieval methods based on noun representations extracted from documents and queries (Varelas et al., 2005). WordNet is the most popular method for implementing semantic similarity (Liu et al., 2004; Seco et al., 2004; Tagalakis and Keane, 2005; Varelas et al., 2005).

The community modelling algorithms (presented in Section 4.4) utilise a mechanism for measuring the similarity between two lists of terms. If L_1 and L_2 are two lists of terms, we define a similarity function $Sim(L_1,L_2)$ which returns a number that indicates how close semantically the terms in both lists are. For this, we adapt the algorithm presented in (Seco et al., 2004), which calculates the semantic similarity between two words based on the WordNet taxonomic structure. The algorithm by Seco, Veale and Hayes (2004) accepts nouns as input and returns a decimal number (0, no similarity – 1, the same meaning) as an output, which represents the semantic similarity between two words. Following the work of Tagalakis and Keane (2005), the original formula has been modified as follows. Having two compounds cc_1 and cc_2 composed of $\{c_1^1,...,c_1^i\}$ and $\{c_2^1,...,c_2^j\}$ terms respectively, for each term of cc_1 we perform a similarity check with every term of cc_2 and store the highest value returned for each term c^i_1 of cc_1 in an array. The same is done for every term of cc_2, c^i_2 with all the terms of cc_1 and the highest value returned for each term c^i_2 is stored. All highest values for the terms of cc_1 and cc_2 are added up and divided by the sum of the total number of terms appearing in cc_1 and cc_2, $i + j$ (see Formula 1).

$$sim'(cc_1, cc_2) = \frac{\max\left(sim(c_1^1,\{c_2^1,c_2^2,...,c_2^j\})\right) + ... + \max\left(sim(c_1^i,\{c_2^1,c_2^2,...,c_2^j\})\right) +}{i+j}$$
$$\frac{+\max\left(sim(c_2^1,\{c_1^1,c_1^2,...,c_1^i\})\right) + ... + \max\left(sim(c_2^j,\{c_1^1,c_1^2,...,c_1^i\})\right)}{i+j} \qquad (1)$$

Formula 1 is used in the algorithms presented in the next section to extract the VC model.

4.4 Community Model Extraction Mechanism

The community model represents the whole community, including individual members, relationships between members, topics of interest and the cognitively central members. Hence, it consists of individual user models, a relationships model, the community domain (ontology), lists of popular and peripheral topics, and a list of cognitively central members. The main components of the community model are illustrated in Figure 2.

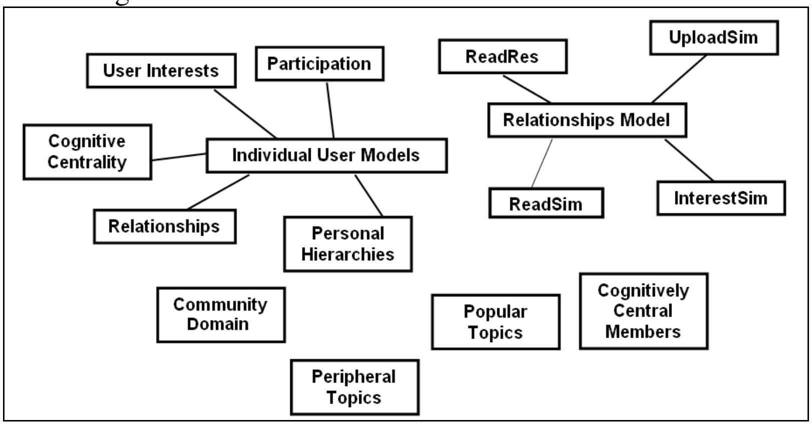

Figure 2 The components of the community model

The benefits of having this community model are to: (i) extract and store information about individual community members, in order to design and generate personalised messages; (ii) develop a relationships model of interactions and identify semantic links between community members that can be used in personalising the notification messages; (iii) represent the community domain, as an ontology, to semantically enhance the algorithms and measure how relevant the extracted relationships are to the rest of the community; and (iv) maintain a list of the most cognitively central members based on the semantic relevance their activities have to the community, and use this list to provide more personalised support to specific members. The rest of this section will provide a detailed description for each component of the community model, together with corresponding algorithms used for its extraction.

4.4.1 Relationship Model

We consider four types of semantic relationships between community members: *ReadRes* relationship indicates links based on reading resources uploaded by others, *ReadSim* and *UploadSim* relationships are based on similarity of read or uploaded resources, respectively, and *InterestSim* indicates relationship based on similarity of members' interests. The above relationships exist between community members and indicate semantic connections that can be represented in a graph (more detail of graph definitions used is given in Section 5). In this work, we assume that a downloaded resource has been read by the member who downloaded it. Thus downloading and reading is used interchangeably in the following sections.

ReadRes Relationship

ReadRes can be used to identify complementarity between community members, which can help to improve the community's transactive memory.

$ReadRes(a,b)$ relationship indicates that resources uploaded by member b are read by member a; the relationship strength corresponds to the number of resources and their relevance to the community domain. In other words, the closeness of the members depends not only on the quantity of the resources uploaded by one member and read by the other member, but also, and more importantly, on the relevance of these resources to the VC context. In this way, quality interactions in the community are rewarded. On the other hand, even if a relation is not very related to the community's topic, it will still be detected but the value of this relation will be lower.

Consider a resource r_i uploaded by member b and read by member a. We will denote the keywords for resource r_i with $rKeywords_i$. Considering the community domain which is represented by the ontology Ω, we define the value of r_i for the community as $V_{r_i} = Sim(rKeywords_i, \Omega)$, where the similarity is calculated based on the modified WordNet algorithm (Section 4.3). Let us denote $Z_r^{a \leftarrow b}$ 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 domain, i.e.:

$$ReadRes(a,b) = \sum_{i=1}^{Z_r^{a \leftarrow b}} V_{r_i} \qquad (2)$$

ReadSim and UploadSim Relationships

$ReadSim$ and $UploadSim$ relationships can also be important for discovering similarities between members. Making people aware of who else is holding knowledge similar to theirs can improve the community's transactive memory system. This can also improve the understanding of what is happening in the community which can be related to the development of shared mental models.

$ReadSim(a,b)$ indicates that members a and b have read semantically similar resources, while $UploadSim(a,b)$ indicates that a and b have uploaded similar resources. Here the value of a resource with respect to the ontology is not directly measured as in ReadRes, since what we are interested in the similarity of the resources that two members have uploaded or downloaded. The relevance of their relationship to the community will be reflected in the corresponding similarity calculation since the list of keywords they uploaded or downloaded is expanded based on the domain ontology.

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 and enriching with additional keywords extracted from Ω, as described in Section 4.2. Having the additional keywords extracted from the ontology, we then construct the extended list of keywords for each member. Let us denote these extended keyword lists as $aKeywords$ and $bKeywords$.

These lists are compared to find the similarity between the two members by using the extended WordNet similarity algorithm presented in section 4.3. Consequently, $ReadSim(a,b)$, is calculated as:

$$ReadSim(a,b) = Sim(aKeywords, bKeywords) \qquad (3)$$

$UploadSim(a,b)$ is calculated in the same way by using the resources uploaded by a and b.

$$UploadSim(a,b) = Sim(aKeywords, bKeywords) \qquad (4)$$

InterestSim Relationship

$InterestSim$ relationship can identify interest similarities and complementarities. 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.

$InterestSim(a,b)$ relationship represents the similarity of interests between members a and b.

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 member a and extending those with the concepts extracted from the ontology, a's personal list of interests

$\mathrm{I}_{ext(a)}$ is extracted. The extended lists of personal interests of member a ($\mathrm{I}_{ext(a)}$) and member b ($\mathrm{I}_{ext(b)}$) are compared using the extended WordNet similarity algorithm (Section 4.3) to calculate the interest similarity between a and b:

$$InterestSim(a,b) = Sim(\mathrm{I}_{ext(a)},\mathrm{I}_{ext(b)}) \quad (5)$$

The use of the ontology allows us to identify broader similar and complementary topics among community members. Although the interests of a member I_a are extended with terms from the ontology Ω to calculate similarity between members based on their interests (i.e. $InterestSim$), a member's interest in his/her individual user model will not be changed. The next section will give more detail of how the individual user models are extracted.

4.4.2 Individual User Models

Cognitive Centrality

The cognitive centrality is used to identify the importance of a community model to the VC. 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 (Lave and Wenger, 1991), knowing where unique knowledge is located can facilitate the transition from one subject area to another (Wegner, 1986). Awareness of who the central and peripheral members of the VC are can also help the improvement of the community shared mental models and transactive memory.

There are different approaches to measure centrality used in the social network area, mostly inherited from graph centrality (Nieminen, 1974; Freeman, 1979; Freeman et al., 1991; Borgatti and Everett, 2006; Latora and Marchiori, 2007). Freeman (1979) describes in a general review three types of centrality as developed in social network research: degree, betweenness and closeness centrality. CCen deals with a member who holds the most valuable knowledge in the community. In our approach the importance of the knowledge a member holds depends on the relationships a member has (semantic connections) with other members. Consequently, in this research we are following and adapting the degree centrality as introduced by Nieminen (1974) where the degree centrality of a point in a graph is measured according to how many points that given point is connected to in the graph.

Here we adapt the original formula as follows: $CCen(a)$ of member a is calculated as the number of all relationships member a is having with any member b (adding the value of the edge connecting the two members in the graph) considering the four relationship types (ReadRes, ReadSim, UploadSim and InterestSim).

$$CCen(a) = \sum_{b=1}^{n} \left(ReadRes(a,b)+ReadSim(a,b)+UploadSim(a,b)+InterestSim(a,b)\right) \quad (6)$$

where n represents the number of members in the community

User Interests

The interests of each user are stored in the individual user models. Interests are extracted based on the resources a member has uploaded and/or downloaded in the VC, or information users provide explicitly about their interests (if such a feature is available). The keywords (tags) of each of the resources member a is uploading or downloading are aggregated in a's individual model. Using $rKeywords$ for each resource uploaded or downloaded by a user, his/her interests are represented as a list of keywords with weights. For example, all keywords 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. The use of a weight $w(t,T_a)$ allows the mechanism to be used as an input in a system and display the keywords of a member as a tag cloud or folksonomy. If $w(t,T_a) \geq \Theta$ (Θ is a threshold), t is added to the interests of a denoted by I_a. I_a is presented as member a's personal list of interests. Threshold Θ can be adjusted according to the size of list T_a in such a way that will allow a list of interests I_a to be created for every community member. For example if T_a is small, then Θ will be small so keywords will be allowed to be added to the list I_a. Having Θ allows the approach to be flexible and accommodate close-knit VC of smaller or larger sizes. In the study presented in this paper (Section 7) the value of Θ was set to be 2. Not having a large number of resources, thus keywords shared within the VC, means that a value of Θ >2 was resulting in interests not being extracted for most members.

Participation, Relationships and Personal Hierarchies

In addition, the individual user model includes participation measures, relationships developed and hierarchies of folders built in the community by an individual member.

Participation: The frequency of knowledge sharing activities of a member (uploading or downloading) is stored in his/her individual user model - $uRate$ is the number of resources uploaded, and $dRate$ also is the number of resources downloaded.

Relationships: Each participating member in the VC is developing relationships with other members of the VC. These relationships $ReadRes, ReadSim, UploadSim, InterestSim$ are stored for each member in his/her personal profile.

Personal Hierarchies: Folders F and resources R created by each member in the VC compose the personal hierarchies that a member is creating. The personal hierarchies can be used in extracting the resources and folders a member has created, providing information for the notification messages generated for individual members, see Section 6.

4.4.3 Popular/Peripheral Topics and Cognitively Central Members

List of Popular and Peripheral Topics

In order for the VC to be able to adapt to changes (for example the main topic of interest of the VC is shifting or a new project comes and community members need to identify what resources in the community are relevant), a list of the most popular and peripheral topics has to be maintained. This will allow exploitation of knowledge that both cognitively central members (CCenM) and cognitively peripheral members (CPerM) have to offer in the VC. These lists are extracted based on the resources people are sharing in the VC and on the assumption that shared resources correspond to topics of interest of the VC members.

Based on the keywords of the resources members are sharing in the VC, two lists of topics are extracted: L_{Pop} represent the lists of popular topics and L_{Per} represents the list of peripheral topics. Assuming the keywords $rKeywords$ for each resource, a list $allKeywords$ is constructed which consists of all the keywords of all the resources in the VC. Each keywords is assigned a weight, $w(k, allKeywords)$ which represents the frequency of a keyword k in $allKeywords$. If $w(k, allKeywords) > \sigma_{Pop}$ (σ_{Pop} is a threshold), k is added to the list of popular topics L_{Pop}. If $w(k, allKeywords) < \sigma_{Per}$ (σ_{Per} is a threshold), k is added in the list of peripheral topics L_{Per}. L_{Pop} and L_{Per} are updated each time a new community model is extracted for the VC. Thresholds σ_{Pop} and σ_{Per} can be adjusted according to the number of keywords appearing in $allKeywords$. Thus, if the list of $allKeywords$ is small, then σ_{Pop} and σ_{Per} can be adjusted to let more keywords enter L_{Pop} and L_{Per} accordingly. This allows the approach to be applicable to close-knit VCs of different sizes. In the study presented in Section 7, after experimenting with different values, the values of thresholds were set to be $\sigma_{Pop} \geq 4$ and $\sigma_{Per} \leq 2$. Having these specific values for σ_{Pop} and σ_{Per}, allowed the topics added into the corresponding lists to represent the interests of cognitively central and peripheral members accordingly.

List of Cognitively Central Members

Having extracted the $CCen$ for every member in the VC, a list of members with the highest $CCen$ is composed. The purpose is to have a list of the members who are sharing the most valuable information to the VC at hand. This information can also be used in triggering the intelligent support described in Section 6. $CCen(a)$ represents the cognitive centrality for member a and $avg(CCen)$ defines the average cognitive centrality for the specific VC. If the $CCen(a) > avg(CCen)$, then member a is added in the list of cognitively central members for the VC under study.

5 Knowledge Sharing Patterns Extraction

This section presents graph based algorithms that automatically extract pre-defined static patterns from the community model defined earlier. The initial validation of the following graph based pattern detection algorithm has been presented in (Kleanthous and Dimitrova, 2010).

5.1.1 Definition and Detection of Relationships in Graphs

Each relationship in Section 4.4.1 defines a graph representing the corresponding links between community members: $G_{RS}(V_{RS}, E_{RS})$ is the graph derived for $ReadSim$, $G_{US}(V_{US}, E_{US})$ is the $UploadSim$ graph, $G_{IS}(V_{IS}, E_{IS})$ is the $InterestSim$ graph, and $G_{RR}(V_{RR}, E_{RR})$ is the graph for $ReadRes$. $G_{RS}(V_{RS}, E_{RS})$, $G_{US}(V_{US}, E_{US})$ and $G_{IS}(V_{IS}, E_{IS})$ are non-directed graphs of type $G(V, E)$ where V is the set of nodes representing community members and E is the set of edges representing the existence of the corresponding relation between both members in the nodes. The weight of the edge corresponds to similarity between the members, using the algorithms presented in Section 4.4.1. An edge is present in a relationship graph only if the weight of that edge is greater than a pre-set threshold value. The value of the threshold can be adjusted according to the density of connections appearing in a graph, in such a way that the resultant graph is meaningful for the purpose for which it is constructed. For example, if only strong relationships need to be extracted the threshold value will be high so only highly weighted edges will be extracted. A neighbourhood of a node v, denoted as $\mathrm{N}_G(v)$, represents the ego network (Degenne and Forse, 1999) of v and indicates members similar to v. Figure 3 shows example extracted graphs for InterestSim and ReadSim during the evaluation study (Section 7).

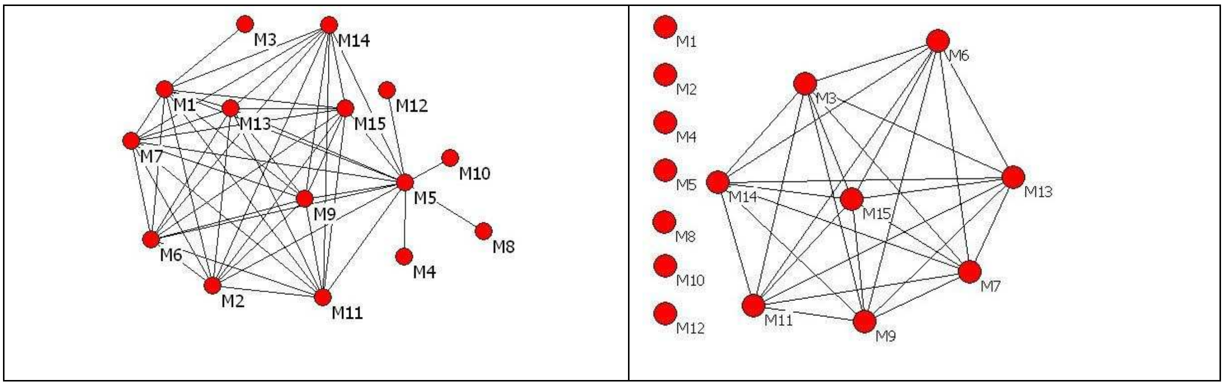

Figure 3 InterestSim relation on the left shows the similarities in terms of interests between community members at the end of the third period of the evaluation study (Section 7). In this graph, M10, M8, M4 and M12 only have an InterestSim with M5 but not among themselves. ReadSim relation on the right shows the similarities in terms of reading between the community members. On the left of that figure the disconnected dots represent the members who have not had a ReadSim with any other member in the community.

$G_{RR}(V_{RR}, E_{RR})$ is a directed graph (Gross and Yellen, 1999), see Figure 4, where the direction of each edge represents that a member (head) has read a resource uploaded by another member (tail). Each node v has an out-neighbourhood $\mathrm{N}_G^+(v): \{x \in V(G): v \rightarrow x\}$ representing community members who have downloaded resources uploaded by v, and in-neighbourhood, $\mathrm{N}_G^-(v): \{x \in V(G): x \rightarrow v\}$ representing members whose resources v has downloaded.

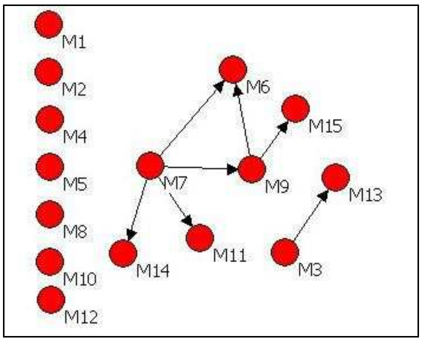

Figure 4 ReadRes is represented in this Figure during the second period of the evaluation study (Section 7). The arrows show that members have read resources by other members. For example members M14, M11, M9 and M6 have read resources uploaded by M7. The isolated dots on the left show the members who do not have any ReadRes relation with others in the community.

Let us denote member a and member b as 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 E_{RS}. $UploadSim$ and $InterestSim$ can be detected in the same way using the respective graphs for each relationship type. $ReadRes(a,b)$ exists if there is an edge $e_{ab} = (v_a, v_b)$, $e_{ab} \in E_{RR}$, i.e. b has read resources uploaded by a .

By analyzing the community relationships model and the individual user profiles, we can identify patterns of knowledge sharing behaviour related to TM, SMM, and CCen. The corresponding algorithms are presented in the next section.

5.1.2 Detection of Knowledge Sharing Patterns

A pattern is important if it can be detected and used in order to provide support to community members. We define seven patterns; for each pattern, we specify its relevance to TM, SMM and CCen, and describe how the pattern can be detected.

P1. Unexplored similarity between community members
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, as pointed out in (Harper et al., 2007). In addition, helping members understand that they hold complementary knowledge, (they might not be working in similar but related areas), improves the community TM system (Wegner, 1986) and can promote collaboration within the community (Ilgen et al., 2005).

Detection: To detect unexploited similarity between a and b, we extract the neighbourhoods of both members from $G_{RS}(V_{RS}, E_{RS})$. If one of the members does not belong to the other's neighbourhood, pattern P1 is discovered:

$$(\mathrm{N}_{RS}(v_a) \cap \mathrm{N}_{RS}(v_b) \neq \varnothing) \wedge (v_a \notin \mathrm{N}_{RS}(v_b)) \qquad (7)$$

In the same way, P1 is defined for the $UploadSim$ and $InterestSim$ relationships.

P2. Community members may not be aware of their similarity
Two members have $ReadSim$ with the same members and among themselves.

Importance: Community members may not be 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 the development of SMM (Mohammed and Dumville, 2001) (members will become aware of what others are working on), and enhance TM (Wegner, 1986) (members will know who they relate to in the community and how similar they are to others).

Detection: This pattern is detected by extracting the neighbourhoods of both members from $G_{RS}(V_{RS}, E_{RS})$. If one of the members belongs to the other's neighbourhood, pattern P2 is identified:

$$(\mathrm{N}_{RS}(v_a) \cap \mathrm{N}_{RS}(v_b) \neq \varnothing) \wedge (v_a \in \mathrm{N}_{RS}(v_b)) \qquad (8)$$

In a similar way, P2 is defined for the $UploadSim$ and $InterestSim$.

P3. Members not benefiting

A member uploads resources but does not download.

Importance: This pattern can be useful to identify members who are not downloading from the community. These members can be prompted to make the most of their time in the community.

Detection: Detection of P3 is done by using the upload and download rates of a member:

$$(uRate(a) > 0) \wedge (dRate(a) = 0) \tag{9}$$

P4. Members not contributing

A member who appears to *download but not upload* resources to the community can be detected similarly to P3, which can be denoted as

$$(uRate(a) = 0) \wedge (dRate(a) > 0) \tag{10}$$

P5. Important peripheral members not downloading

Members who do not download and occasionally upload resources that other members have downloaded.

Importance: We can use this pattern to motivate peripheral members (who have sporadically uploaded an important resource in the community but have not effectively engaged in other activities) to benefit from the community. Notifying them that others are interested in what they upload can motivate those members to start reading resources uploaded by the members they are similar to. This pattern may help to promote collaboration.

Detection: P5 is calculated using the upload and download rates for a and the out-neighbourhood in $G_{RR}(V_{RR}, E_{RR})$ to check that a uploads relevant resources:

$$(uRate(a) > 0) \wedge (dRate(a) = 0) \wedge \left(\mathrm{N}^{+}_{RR}(v_a) \neq \varnothing\right) \tag{11}$$

This pattern is a subset of P3. If both patterns P3 and P5 are detected, then only P5 will be considered.

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 relevant resources to start uploading, by showing them how similar their interests are to other members. This can improve the TM system of the community since members will be aware of others' interests (Wegner, 1986). Motivating them to upload to the community may help the community to sustain.

Detection: To detect P6, we check a member's upload and download rates and his/her neighbourhood in $G_{IS}(V_{IS}, E_{IS})$:

$$(uRate(a) = 0) \wedge (dRate(a) > 0) \wedge (\mathrm{N}_{IS}(v_a) \neq \varnothing) \tag{12}$$

Pattern P6 is a subset of P4. If both patterns P4 and P6 are detected, then only P6 will be considered.

P7. Unexplored complimentary similarity between members

Two members have $UploadSim$ but do not have $ReadSim$.

Importance: Members who upload similar resources in the community but are not reading similar resources, have similar and complimentary interests but are unaware of this. Making these people aware of their similarities and differences may improve the TM system since members will be able to identify where important knowledge, for them, is located (Ilgen et al., 2005). At the same time, this may improve the building of SMM (Mohammed and Dumville, 2001), since members can appreciate how everybody contributes to the community. Awareness of where complimentary knowledge is located may encourage collaboration.

Detection: P7 is identified using $G_{US}(V_{US}, E_{US})$ and $G_{RS}(V_{RS}, E_{RS})$, and checking that one of the members belongs to the other member's neighbourhood in $G_{US}(V_{US}, E_{US})$ but does not belong to the neighbourhood of that member in $G_{RS}(V_{RS}, E_{RS})$:

$$(v_a \in \mathrm{N}_{US}(v_b)) \wedge (v_a \notin \mathrm{N}_{RS}(v_b)) \qquad 13$$

Table 2 summarises the patterns, indicating when and how they can be used to generate adaptive notifications and how this relate to the key social processes – TM, SMM and CCen.

The above algorithms can capture patterns of behaviour in the VC at certain time points. All measures including CCen, are set to zero before the next time point is considered and the above pattern algorithms are applied. Consequently, if a member is detected as important at time point 1 and at time point 2 dropped his/her activity, then this change will be picked up and the relevant pattern will be triggered. A different approach has also been considered where temporal patterns have been defined and algorithms implemented (Kleanthous and Dimitrova, 2009).

6 Adaptive Notification Mechanism

This section will illustrate how the community model and detected patterns can be used to generate community-tailored support. There are four notification categories considering community participation and awareness aspects:

- *participation of Cognitively Peripheral Members (CPerM);*
- *participation of Cognitively Central Members (CCenM);*
- *improving the community TM system;*
- *developing SMM.*

We will provide here rationale for using each notification category.

Rationale for CPerM notifications: Studies have shown that acknowledging the uniqueness of peripheral members' expertise may increase their confidence, and thus improve their level of participation and contribution (Phillips, 2003; Thomas-Hunt et al., 2003). In addition, CPerM can be motivated to participate more by becoming aware of the importance of their unique expertise for the rest of the community (Thomas-Hunt et al., 2003).

Rationale for CCenM notifications: CCen members influence other VC members due to their status and knowledge. Research showed that less central members are influenced and usually follow the CCen members (Kameda et al., 1997). Hence, notifications for CCenM should aim at helping members from the periphery gain confidence and become influential. Participation of CCen members may be motivated by acknowledging their importance to the VC (Thomas-Hunt et al., 2003).

Rationale for notifications to improve TM: When a TM system is developed in a VC, members are able to locate knowledge important to them and identify who the experts in specific areas are (Wegner, 1986). By providing notification messages that include personalised information, we can help individuals in the VC become aware of what others are working on, who they are similar to and what resources might be of interest to them.

Rationale for notifications to improve SMM: Understanding what processes are happening in a community, what the VC purpose is, and being aware of the activities that relate members, creates an awareness and develops SMM (Mohammed and Dumville, 2001).

A notification may serve more than one of the purposes listed above. The notification categories will be targeted for both newcomers and existing members of the VC. The role of the VC members (e.g. student, supervisor, project coordinator), will not be considered since we are catering for equal membership as defined in the VC characteristics in Section 1. Similarly, as far as existing members are concerned, the period that a user has been a member of the VC is also not considered when notifications are generated. The reason for using newcomers and existing members is to examine the impact that the selected processes and notification messages can have in both categories of members.

Notification messages are triggered based on the detection of a pattern or a change in the VC. Consequently, different notification messages need to be generated according to the detection. Two types of notifications are defined: (i) Type1 - Notifications based on detected knowledge sharing patterns; (ii) Type2 – Notifications that combine data from the patterns detected and the community model. Table 2 shows which patterns will generate a Type 1 and Type 2 notification messages and how these are linked to the relevant processes.

This paper does not provide an exhaustive list of notification messages that can be generated, as notifications can vary according to the VC type, subject area, and size. The notifications provided here are just a sample of what can be generated, and illustrate how patterns detected using algorithms from Section 5 can be used to generate support for a close-knit VC.

Table 2 Summary of the main patterns and corresponding Type 1 and Type 2 notifications that will be generated for each pattern together with the expected affect on the community social processes.

Notification Type	Pattern	Affects
N1-1: Inform members of their unexplored similarity	P1: unexploited similarity between members	Collaboration, TM System, SMM
N1-2: Inform members of their similarity	P2: members unaware of their similarity	SMM, TM System
N1-3: Facilitate a member's integration by showing similar members	P3: members participating but not benefiting	Improve participation, Sustainability, TM
N1-4: Facilitate a member's integration by showing similar members	P4: members not contributing	
N1-5: Facilitate a CPerM who is downloading only to integrate	P5: peripheral members not downloading	Awareness, SMM
N1-6: Facilitate a CPerM who is uploading only to integrate	P6: peripheral members not uploading	
N1-7: Inform members of their complementary similarity	P7: unexplored member complementarities	SMM, TM System, Collaboration
N2-1: Exploit an important CCenM	P1 and CCen information	Motivate CCenM, develop TM and SMM
N2-2: Pair a CCenM with a CPerM	Information from Individual User Model (CCen) and Relationship Model	Help a CPerM integrate and Motivate a CCenM. Develop TM and SMM
N2-3: Welcome message to newcomers	Member must be detected as newcomer	Develop TM in newcomers

6.1 Notifications based on Detected Knowledge Sharing Patterns – Type 1

Notifications based on the knowledge sharing behaviour of members will be used to inform members of their status in the VC, how their behaviour affects themselves and other VC members, and to provide suggestions on how to exploit the material and knowledge available in the VC. We consider seven Type 1 notifications, as described below:

N1-1 (Inform members of their unexplored similarity) Members will be informed of the read, interest or upload similarity they appear to have with the same set of other members but not between themselves. The message can encourage members to read resources that others are reading and uploading. Links are provided to relevant resources along with the detected members. **Aim:** Develop TM and SMM since members will be informed of what others are working on.

N1-2 (Inform members of their similarity), will inform a group of members of the similarity they appear to have in terms of reading, uploading, or interests. Suggest other types of relationships that they might want to develop with these members by providing links to resources these members are uploading or downloading. In this message the members' ID will be mentioned and the relevant relationship type will be pulled from the Relationships Model. **Aim:** Improve TM and SMM by informing members of their similarity with others.

N1-3 (Facilitate a member's integration by showing similar members) A member who is only downloading will receive this notification which will develop awareness of how the member is related to others in the VC to help him/her integrate. A list of similar members will be provided in the message. **Aim:** Develop TM as a member will become aware of how he/she relates to others.

N1-4(Facilitate a member's integration by showing similar members) Similarly to N1-3, this notification targets members who only upload, and encourages them to start benefiting from the resources available in the VC. **Aim:** Similar to N1-3.

N1-5 (**Facilitate a CPerM who is downloading only to integrate)** A message will be sent to a CPerM, or a newcomer, who only downloads and appears to have similar interests to other members. The content of the message will include information on similar members and encourage the member to start contributing so others can benefit from his/her knowledge. **Aim:** Provide support to a CPerM and develop TM and SMM by providing awareness.

N1-6 (Facilitate a CPerM who is uploading only to integrate) message will be sent to a CPerM, or a newcomer, who only uploads and others appear to be interested in what he/she is uploading. The content of the message will include information on similar members and encourage the member to start downloading so he/she can benefit from the knowledge available in the VC. **Aim:** Similarly to N1-6, supports a CPerM to integrate and develops TM and SMM.

N1-7 (Inform members of their complementary similarity) Inform a group of members of their complementary similarity and show them resources uploaded by each other. Links to resources and members' IDs will be included in the message. **Aim:** Improve TM and SMM by providing information on the similarity between members.

6.2 Notifications based on Combining the Community Model and Detected Patterns –Type 2

Further information stored in the community model will also be used in order for notifications to be more effective. Type 2 notifications combine data from both the community model and the patterns presented above.

N2-1 (Exploit an important CCenM) it is used to inform a CCen member of how important he/she is among the VC members and, at the same time, to encourage that member to continue contributing and collaborating with less active members with whom he/she relates. The message will include the CCen rank of that member, member's ID and the IDs of less active members.
Aim: Motivate a CCenM and help develop TM and SMM in the VC.

N2-2 (Pair a CCenM with a CPerM) This notification will be sent to both the CCenM and the CPerM who are detected to be similar and encourage them to pair in order for the CPerM to benefit from the CCenM and integrate in the VC. The content of the message will differ for each member and can be seen in Table 3.
Aim: Motivate a CCenM and support CPerM to integrate. Develop TM and SMM.

N2-3 (Welcome message to newcomers) This message will be generated when a new member joins and will include information about other members and how they relate to the newcomer.
Aim: Support the integration and development of TM of newcomers. It is important for them to know how they relate to other members to realise the benefits of their membership in the VC.

6.3 Generating Notification Messages

Each notification has a goal, which relates to one of the categories defined, CPerM, CCenM, TM and SMM. The content of each notification message is designed to promote the goal of the notification. Two different formats of personalised notification messages have been generated varying the level of detail (see Table 3 for illustration).

Table 3 Examples of the two personalised message formats: notifications sent to members M7 and M2 during the evaluation study reported in Section 7. For more examples see (Kleanthous Loizou 2010, Appendix D).

Member Id	Example of first personalised message	Example of second personalised message
M7	You appear to upload similar resources with M9 and M3. You may find it helpful to see the resources these members are uploading and downloading. Use the links provided below to navigate through the resources: General link to the VC space is provided.	Your influence to this VC is dropping due to stop uploading valuable resources. The resources you previously uploaded have been valued in this VC. Start sharing your knowledge again and keep your centrality up. You also appear to have reduced your download activity in this VC. Use the links below to navigate through resources that might be of your interest Resources by M9: Link to resource provided Resources by M5: Links to resources provided
M2	You appear to have similar interests with M6, M7 and M9. You may find it helpful to see the resources these members are uploading and downloading. General link to the VC space is provided.	You appear to have reduced your download activity in this VC. Use the links below to navigate through resources that might be of your interest. Resources by M5: Link to resources uploaded by M5 Resources by M11: Link to resources uploaded by M11

Having two different message forms, allowed us to examine whether a more personalised approach rather than a less personalised approach was preferred by VC members (see Section 7). In the *first set of notifications*, members received information according to the patterns detected including similar members (no links to profiles), and only links to the general read history of the VC were sent. In the *second set of notifications*, more personalised information relevant to a given member (e.g. links to specific resources that might be of interest, links to profiles of members who have similar interests with that member, links to specific relevant folders) are provided.

Next we give a description of the formalisation of the adaptive notification mechanism and introduce the notations used.

6.3.1 Formalisation of Adaptive Notifications Mechanism

The derived patterns in Section 5 are used as an input for generating notification messages. For every notification message a standard structure is followed:

- *Detection* – the situation that triggers the notification - the knowledge sharing pattern detected.
- *Target Users* - the list of community members to whom the notification will be sent.
- *Goal* – defines the aim of the notification in relation to TM, SMM, and CCen.
- *Content Template* - the textual template, together with corresponding parameters, used to generate the notification.

Selected representative examples of notification definitions are given in

Table **4**.

The following notations are used. The VC members are represented as a set of members $\mathcal{M} = \{M_1, M_2, ..., M_n\}$ where n, is the total number of community members. A subset of $\mathcal{M}$ is derived, $\mathcal{M}' = \{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ which represents the members extracted in a detection - such that $\mathcal{M}' \subseteq \mathcal{M}$. Additionally, three more sets have been derived: $CCenM$ (a set of the cognitively central members), $CPerM$ (a set of the cognitively peripheral members) and $Newcomers$ (the set of all the new members of the VC); $CCenM \subseteq \mathcal{M}$, $CPerM \subseteq \mathcal{M}$, and $Newcomers \subseteq \mathcal{M}$. If a notification can be generated for more than one detection, the $< RelationshipType >$ is used to indicate the type of relationship from the community model.

A threshold value is used in the algorithms when constructing the list of target members $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ to whom a notification should be sent. In the study in Section 7, after experimentation, the threshold value was set to 3, i.e. the three most similar members were added to the target members list $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ that was used in sending the notification messages. Numbers greater than three (e.g. five most similar members) have not selected, since due to the closeness of the VC most members would have received the same messages and this eliminates the personalised aspect of the approach. On the other hand, if too few members (e.g. one or two) are selected to be notified about a similarity or a relationship, then this makes the impact of the message to the VC very insignificant and does not help in promoting TM and SMM within the VC.

Table 4 Selected illustrative examples of the definition of Notification Messages – For a more comprehensive list of the notifications defined please refer to (Kleanthous Loizou, 2010)

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Notifications Based on Knowledge Sharing Behaviour Patterns				
N1-1 (Inform members of their unexplored similarity)	P1: Members $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ have $< RelationshipType >$ with the same members but not among themselves.	$\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$	Inform members of their similarity and encourage them to read resources the others' are reading. Develop TM and SMM	For every M_{i_j} : *"Did you know you have a* $< RelationshipType >$ *similarity with* $T \setminus \{M_{i_j}\}$*. You may find it helpful to check the resources these members are reading and uploading. Follow the links below:"*
N1-4 (Guide member's integration by showing similar members)	P3: Member $\{M_{i_j}\} \in \mathcal{M}$ is downloading only $\{M_{i_j}\}$ has a $< RelationshipType >$ with $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$	$\{M_{i_j}\}$	Develop awareness of the member relates to others and provide information on where resources important to him are located. Develop TM.	*"Share your knowledge with the rest of the community by start uploading resources.* $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ *have* $< RelationshipType >$ *with you and will benefit from what you share with them."*
Combination of Patterns and Information from community model				
N2-1 (Exploit an important CCenM)	$P_1 \wedge \{M_{i_j}\} \in CCenM$	$\{M_{i_j}\}$	Let a CCenM know of his/her importance in the VC, encourage him to continue and suggest he/she pairs with less active members to help them integrate.	*"You are an important member connecting* $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\} \setminus \{M_{i_j}\}$ *Keep up the good work and upload more interesting resources. Can you suggest resources that these members may read? You may wish to contact each member"* Also generate *N1-1*
N2-2 (Pair a CCenM with a CPerM)	$\{M_{i_1}\} \in CPerM \wedge \{M_{i_2}\} \in CCenM$ $RelationshipType(M_{i_1}, M_{i_2})$	$\{M_{i_1}\}$	Let a CPerM know of his/her relationship with a CCenM and suggest pairing with the CCenM to help him integrate.	*"You have a* $< RelationshipType >$ *with* $\{M_{i_2}\}$ *who is an important member in this VC. Check what* $\{M_{i_2}\}$ *is uploading and downloading using the links below. You can also contact* $\{M_{i_2}\}$ *if any help is needed."*
N2-3 (Welcome message to newcomers)	$\{M_{i_j}\} \in Newcomers$ $\{M_{i_j}\}$ has an $InterestSim$ with $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$	$\{M_{i_j}\}$	Inform a newcomer of people with similar interests in order to help that member start benefiting from the VC. Help him integrate and develop TM.	*"Welcome to the community! Based on the information you have provided, the following members* $\{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ *might have uploaded resources that could be of interest to you. Links:"*

7 Evaluation Study with a Knowledge Sharing Community

An evaluation study is conducted to examine the effect of community-adapted notifications on individual members and on the VC as a whole (knowledge sharing). The overall framework was employed in a VC to derive a community model, extract knowledge sharing patterns and generate adaptive notifications sent to individual members via email.

The aim of the evaluation is to identify the effects of the notification mechanism so it can be employed in supporting close-knit knowledge sharing communities. Specifically, it examines what influence intelligent notification support, designed based on TM, SMM and CCen, can have on individual members and the functioning of a VC as a whole. The following questions are addressed:

Effect on the community as a whole: is CCen shifting between members; do peripheral members become more central; do members develop links and follow resources from others?

Effect of notifications on oldtimers: have oldtimers followed the notifications, and, if not, why; in what ways (if any) can the notifications be useful for oldtimers; do notifications motivate oldtimers to engage in the community; do oldtimers become more confident to contribute; is there any effect on the TM and SMM of oldtimers; does oldtimers' behaviour change following the notifications?

Effect of notifications on newcomers: have newcomers followed the notifications, and, if not, why; in what ways (if any) can the notifications be useful for newcomers; do notifications motivate newcomers to integrate in the community; do newcomers become more confident to contribute; is there any effect on the TM and SMM of newcomers; does newcomers' behaviour change as a result of the notifications they receive?

7.1 Study Design

This section provides information about the community and outlines the experimental study.

7.1.1 Outline

Platform: In this study the close-knit VC was created in BSCW system[5]. BSCW is a general tool for cooperation over the web which supports the main knowledge sharing activities, such as upload, download, search for resources, synchronous and asynchronous communication, and version control (Wolfgang et al., 2004). The convenience of using BSCW is that the system keeps tracking data about every activity/modification in the shared collaboration space, and every member has access to this data. The data tracked by the system was necessary in order for us to derive the community model. In addition BSCW provides generic functionality that is reflected in its tracking data. For example, resource keywords and tags can be added similarly to the way used in CiteULike[6] to describe resources; members can collaboratively rate or add/edit the description of a resource as it can be done in shared wiki spaces. In this work, we have exploited only generic tracking data from BSCW (resource name, formal and informal keywords, uploading and downloading activity, members join/leave), that can be tracked in other collaborative Web2.0 and Semantic Web applications. The only difference is that in this case we had full access and control[7] of the full tracking data including time-stamps. During the study, members have only received messages sent by us as part of the evaluation and have not used the notifications provided by BSCW.

Community: The close-knit community included *15 members* (researchers and doctoral students) from different research groups working on similar research topics around *Personalisation and Intelligent Knowledge Management*. The community members were working on different projects and some of them participated in joint seminars. The members were based in two countries (UK and USA), some people knew each other and belonged to a physical community (attended weekly seminars together) but others were working remotely. Eight members were *oldtimers* (existing members), and seven were *newcomers* (new members invited to join the VC during the study). Eleven

[5] http://public.bscw.de/

[6] CiteULike is a web environment for sharing academic citations: http://www.citeulike.org/

[7] For example in the case of CiteULike, you can have a private group of people sharing citations, but members do not have full access or control of the activity tracking data.

members in this community were research students working on separate projects, two members (M2 and M6) acted as supervisors for the research students participating in the community, and two members (M4 and M5) were active researchers in their fields but were not directly engaged with supervising students from this community (both members worked at remote geographic locations and had not met most of the other members). A virtual space for reference sharing (hereafter referred to as the VC) was created in the BSCW system. Gradually members embraced the idea of having a VC, and used the BSCW to create folders, upload and download shared references.

The community was created by the first author who acted as the experimenter and did the initial seeding with relevant content. Although she was a member of the community, her activity and contribution to the VC was not considered in the data collected, and she has not replied to any of the questionnaires circulated. The second author was also a member of the community (M2), she was one of the oldtimers. She was not involved in the study design, data collection, and analysis. We have excluded the data submitted by M2 to the questionnaires in the results reported in the paper. However, M2 was referred to in several answers by community members, and we have preserved these references in the reported analysis.

The most popular activity in the VC before the experimental study with notification generation was uploading papers. During the pre-study period, which lasted 21 months, there were several phases of high activity and times when there was no activity in the VC (although activity in the physical community continued). The high activity periods relate to collaboration work on different research projects. Although the VC was created for members to share resources with each other, most of the members shared resources in small teams of two–three people since they were collaborating among themselves but not with everyone (e.g. working on joint projects or organising workshops). Thus, although interesting resources were uploaded by some members, other members had not looked at those (no one had downloaded any of the resources M5 uploaded in the VC since many were not even aware of the existence of that member - M5 was not part of the physical community involving most of the other members.

The VC had some duplicate resources. In five out of the six occurrences, M7 re-uploaded a resource that was already in the VC space. M7 was working on a joint project with M2 and M6 and was uploading resources relevant to that project. This shows that M7 did not know that the resources were already there. It is important to note that M7 was the most CCenM of this VC before the experimental study began. The above occurrences indicate a lack of TM and SMM in both the physical and virtual community and show the need for some intervention to provide better awareness of links with VC members who are remotely collaborating. Even when people were involved in the same physical community (11 members were working in the same lab) they had developed SMM and TM with their supervisors and close friends but not with people who might be similar or share the same interests, which may help develop useful connections.

Method outline: This work aims at providing intelligent support to knowledge sharing virtual communities. Consequently, it was vital to assess the effect the notification messages had on the knowledge sharing in the VC by comparing collected data *before and after the notifications were generated*. Two groups were considered - the existing members of the community (oldtimers), and the newly joining members (newcomers). Comparing the findings from both groups gives us a better idea of the effect notifications could have on the main user categories. All members were asked to complete online questionnaires conducted with the help of a web survey tool (prior and during the study) to examine issues relevant to TM, SMM and CCen and also members' opinions about the notifications they had received. During the study, five notification types were sent to all members (three of type N1-1, four of type N1-2, one of type N1-4, one of type N1-6 and seven of type N2-3). The data analysis combined data collected from the questionnaires with data extracted in the community model (e.g. participation of members, CCen, relationships).

Notification messages were sent during the middle of the day each time. This was not planned but ensured that members were active at that time of the day and messages would be noticed.

Data: *Objective data* was the log data collected over the duration of the study (2 months) using the BSCW activity tracking features. In order to keep the input as generic as possible, we collected data concerning only the basic functionality of the system, such as uploading/downloading, naming a resource, and providing keywords/tags. The tracking data was pre-processed and transferred into database tables in line with the input format to the algorithms developed. The Java algorithms for community modelling, pattern detection, and notification generation were run.

In addition to the tracking data, *subjective data* was collected using questionnaires that combined open ended questions, choice questions (multiple and single answers), and alternative selection questions. Two types of questionnaires were used. The first one was sent to both oldtimers and newcomers only one time and aimed at extracting initial interests of members and to assess issues relevant on TM, SMM and CCen. The second questionnaire was sent two times to oldtimers (after each set of notifications generated) and one time to newcomers. The purpose of this questionnaire was to evaluate the effect of notifications and assess aspects relevant to TM, SMM and CCen in the VC. **Extracting interests:** In the first questionnaire the first question asked members to provide 5-10 keywords/phrases that described their research interests. Based on this data combined with what was extracted in the community model the first set of notifications was generated. **Assessing TM, SMM and CCen of members:** In both questionnaires, eight questions provided members with the list of all members in the community and asked them to select three members that considered as similar to them in terms of reading/uploading, who they might contact for information, who they believe they can benefit from in terms of the knowledge that member holds and also who the three most CCen members of the community are. These questions allowed us to collect evidence of TM, SMM and the perception members had for CCen. **Participation/Activity:** The first questionnaire contained multiple selection questions and was asking members about their participation and activity in the community (uploading/downloading) and reasons for that, allowing them to select the option "other" where free text could be added. **Relevance of Notifications:** In the second questionnaire, there was a binary question asking members if the information they received with the notifications was relevant to them and for the negative answer a mandatory text box was provided for additional comments. A multiple selection question followed a positive answer asking members to select how the information they received helped them. **Actions following the notifications:** The second questionnaire provided a binary question asking members if they followed the links provided with the notifications. A positive answer was leading to a multiple selection question where members could select the activity resulted (uploading/downloading/non of the two). If members selected "none of the two", they were directed to a text box for providing more information for taking no action. **Motivation and Confidence:** In the second questionnaire members were asked if the notifications received motivated them to remain active and if they were feeling more confident to contribute.

The questionnaire data was transformed into a suitable format and analysed using spreadsheets (MS Excel) and a statistical package (SPSS for Windows). The replies to open ended questions were analysed, coded and quantified by the first author (Section 7.1.3).

7.1.2 Stages

Pre-Study Period: This period acted as a seeding period during which eight members were invited to join the VC space, and encouraged to upload or download resources according to their interests and needs. Immediately at the start of the study, the *first questionnaire* was given to the existing members (oldtimers) to collect their interest and compose an Individual User Model for each member. Additionally, this questionnaire assessed issues relevant to TM, SMM and CCen of this community before any interventions have been done. The community model acquisition mechanism was employed to extract an initial community model, based on which algorithms for discovering knowledge sharing patterns were applied.

During the **Second Period**, the knowledge sharing patterns were used to decide what notification messages should be triggered (following the first format of messages, as described in Section 6.3) to provide members with relevant information based on their individual user models and the community relationship models. Individualised notifications were sent to each oldtimer. They included general messages pointing at relevant users – according to their derived relationships and patterns. A week after the email notifications were sent, a *second questionnaire* was sent to all oldtimers. This allowed examining the effect of the first format of notifications to oldtimers. The data extracted from the questionnaire along with log data was used to assess the effect and benefits of the first set of notifications. During this time, seven new members were also invited to join the VC (i.e. the new members joined two weeks after the study started). The newcomers had to reply to an *initial* *questionnaire* which was used to extract their individual interests and to assess issues related to TM, SMM and CCen of newcomers prior to receiving any notifications.

During the **Third Period,** based on data extracted from the newcomers' initial questionnaire, the data in the derived community model, and the application of the static knowledge sharing patterns, welcome email messages were sent to the newcomers (two weeks after their registration) with information relevant to the interests of each member. At the same time, pattern detection and notification generation algorithms were applied to the VC interaction data, which was used for generating the *second set of notifications* sent to every member (each member received tailored messages, as described in Section 6.3). The form of these notifications was different from the first round – in addition to pointing at relevant members (as in the first round), we included a list of relevant papers – extracted based on the relationships members had with other VC members - providing the links to these papers in the email message. Thus, each member could go straight to the BSCW system from the notification message he/she received.

In the **Fourth Period**, members were asked to complete a *final questionnaire*, which was sent to them a week after the second round of notifications was sent. Data from the generated community model (e.g. participation, behaviour and patterns) along with a comparison between the second and final questionnaires was used to assess the effects of the notifications on the VC as a whole.

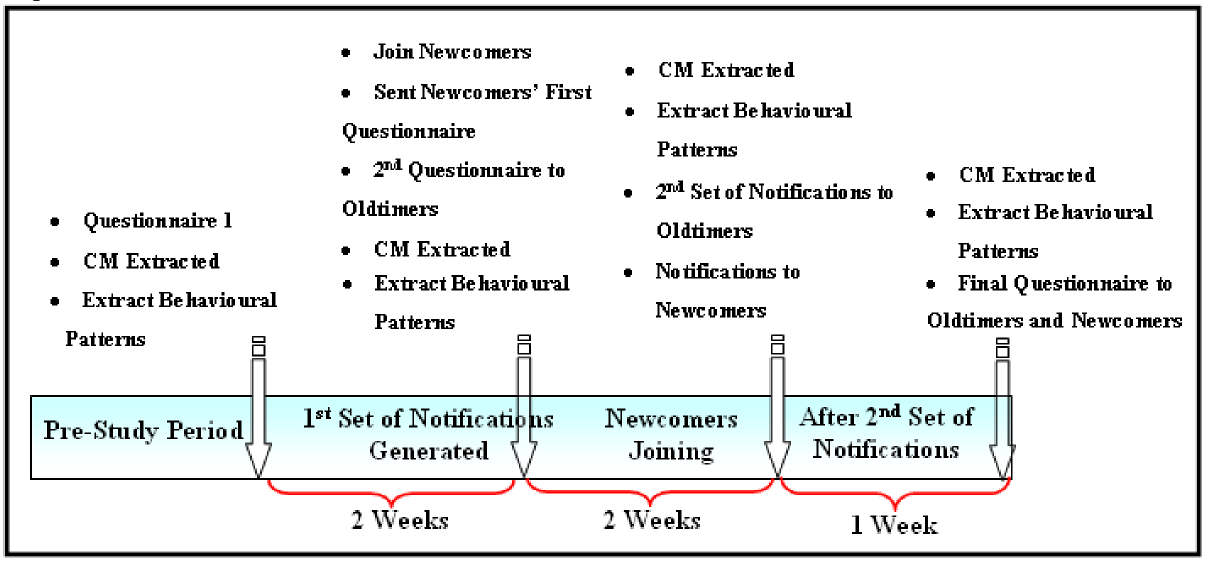

Figure 5 Evaluation timeline including the periods and methods used during each period.

7.1.3 Data Analysis

The results extracted from the questionnaires were statistically analysed in order to understand the effect of notifications. The first eight questions in all questionnaires asked members to list three related community members, such as people whom they would read papers from, people who may benefit from papers they upload, etc. In order to statistically analyse the replies, we needed to quantify them in a uniform way. The reply of each member was a set of members (three). Similarly, the community model indicated a set of members found to be related to each member. Let us define two sets, $\mathcal{A}$ denotes the top most similar members to a given member as derived in the community model. $\mathcal{C}$ denotes the set of members selected by a given member as a reply to a questionnaire. $\mathcal{B} = \mathcal{A} \cap \mathcal{C}$ represents the set of members who appear in both the selection of a member (as indicated in his/her questionnaire replies) and in the community model that was generated at the time the questionnaire was issued. Adapting precision, recall and F1 metrics (Herlocker et al., 2004; Lo and Lin, 2006; Olson and Delen, 2008), we can quantify the replies of the questionnaires as follows.

Following *precision* metrics (i.e. the ratio of relevant items selected to number of items selected), we consider the ratio P of the overlap between selections made by a given member and the set of members extracted in the community model for this member over the number of selections suggested in the community model: i.e. $P = \frac{|B|}{|C|}$ [8]. Following *recall* metrics (the ratio of relevant items selected to total number of relevant items available), we consider the ratio *R* of the overlap between selections made by a given member and the set of members extracted in the community model for this member over the selections made by the member, i.e. $R = \frac{|B|}{|A|}$. To combine both metrics into one number, we will adapt the standard F1 metric (which combines precision and recall into a single number): $F1 = 2 \times \frac{P \times R}{P + R}$.

*F*1 is computed for the replies to every question in each pair of the questionnaires (Questionnaire 1 – Questionnaire 2, Questionnaire 2 – Questionnaire 3 and Questionnaire 1 – Questionnaire 3). The statistical *Wilcoxon non-parametric test* is applied to compare the mean *F*1 scores before and after the notifications.

7.2 Findings

We will discuss the findings from the experimental studies following the main objectives and considering the effect of notifications on the VC as a whole, as well as on each of the two groups – oldtimers and newcomers. For the purpose of consistency as well as anonymity 'he' will refer to a community member male or female.

7.2.1 Effect of Notifications on the Community

In all four periods, the activity in the community included uploading and downloading resources, 237 resources in total. One member was only uploading and one was only downloading. Five members (all newcomers) were isolates and never uploaded or downloaded resources. Eight members uploaded and downloaded from the VC. During the second period (after the first set of notifications generated for oldtimers), there was no uploading. With the second set of notifications, we noticed resource uploading from a newcomer (M15) - third period. This minimal uploading activity, and given that the VC was at the forming stage, indicates that after the second round of notifications some activity in the VC started.

Downloading took place during all periods except the second period. During the third period, after the second set of notifications was sent, the downloading resumed and continued until the end of the experimental study. It is encouraging to see that with the triggering of notification messages oldtimers (e.g. M9), as well as newcomers (e.g. M14, M15) had downloaded resources from the VC.

CCen was shifting between members during the experimental study (see Figure 6).

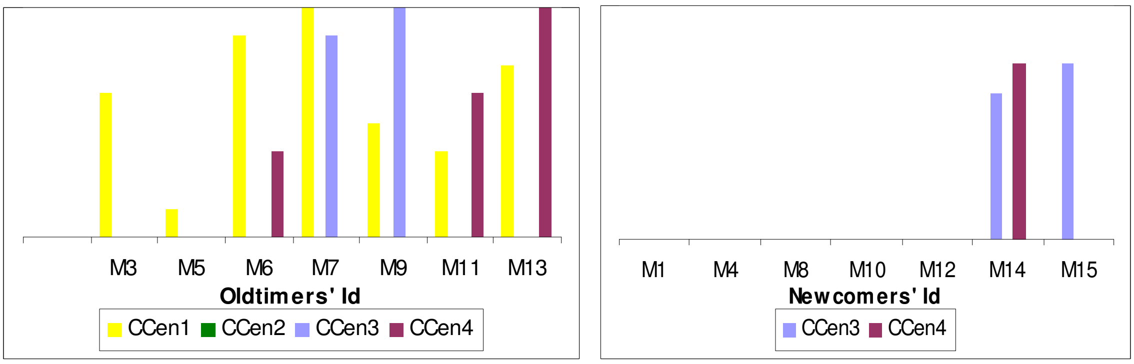

Figure 6 CCen variations during the experimental study[9]. The bars show the CCen rank of the centrality of each member during each period. The left figure shows the CCen of oldtimers and the right figure shows the CCen of

[8] $|S|$ denotes the cardinality of a set *S*

[9] M2 excluded from this graph to ensure objectivity of the approach

newcomers. In both groups, members' centrality starts to improve after the first set of notifications was sent during the third and fourth periods (CCen3 and CCen4, respectively).

Since the activity of all oldtimers dropped, and resumed during the experimental study, CCen also dropped. The important fact is that after notifications were generated members began to gain CCen. For example M11, who had CCen=0 during the second period, started uploading and downloading in the fourth period and had CCen = 0.789. M14 and M15 (both newcomers) also gained CCen during the third period. M9 had CCen=0 during the first two periods, and after the generation of notifications M9 became more active and was detected as the most cognitively central member. An interesting fact was noted – during the first questionnaire M9 could not indicate related members or cognitively central members. In the second questionnaire, M9 identified correctly that M7 was one of the CCenM. This demonstrates that after the notifications were generated, M9 became more active and aware of what was happening in the VC.

During the study, the member with the highest CCen changed three times (after each set of notification sent) demonstrates that the notification had some positive effect on the VC functioning as a whole. At the end of the first period, the two most CCen members were M7 and M6. At the end of the third period, CCen shifted to M9 and M7. During the fourth period, CCen shifted to M13 (see Figure 6). The shifting of CCen to different members in the VC shows a fairly dynamic VC where different members engage at different times. M14 who was a newcomer was peripheral at the beginning of the third period but gained centrality at the end of this period. There were four cases when a CPerM read resources uploaded by CCenM. Two newcomers M14 and M15 read resources uploaded by two different CCenM - M7 and M9, respectively. In two other occasions oldtimers who used to be in the periphery of the VC before the second period and completely inactive, downloaded resources uploaded by central members. The above detections provide some evidence that notifications acted as a trigger to promote the resources uploaded by central members, and hence helped CPerM to identify relevant resources they were previously unaware of.

7.2.2 Findings from the Oldtimers' Pre-Test Questionnaire

At the end of the first period, the first questionnaire was given to oldtimers to assess issues relevant to the community's TM, SMM and CCen.

One member uploaded but did not download from the VC and according to his reply, this member (i) had difficulties identifying in which folder resources relevant to his interests were stored, and (ii) was only interested in resources uploaded by specific members.

Members were asked to identify who they believed were the two most central members in the community. It is important to note that from the answers we received on the first questionnaires members were influenced mostly by what they knew about the physical community and not by what was happening in the VC. For example, most members regarded M2 (who led several of the projects community members participated in) as one of the CCenM, although M2 had not made valuable contributions to the VC. Furthermore, this shows that members had made the assumption that CCenM in the VC were the senior researchers in the overall community, which in fact was not the case.

Members were asked why this specific VC was created in order to evaluate their SMM before the notifications were generated. The responses are documented in Figure 7. The three most popular answers are "*To Share resources*", "*To keep important papers in one place*" and "*So others in the group can see what we are reading*". SMM in a VC requires all members to have a shared understanding of what the purpose of the creation of a specific VC is. In this case the results show a common understanding since they have all picked the "*To share resources*" and "*To keep important papers in one place*" options. One member selected the first option "*To socialise*" which does not represent the purpose of the community under study. A positive observation is that no members have selected the "*I don't know*", which shows that all members had an opinion on the purpose of this VC.

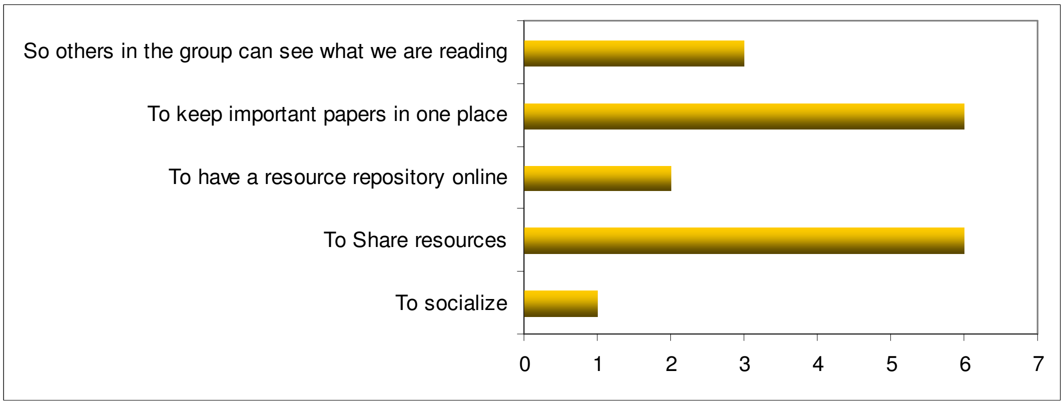

Figure 7 Replies of seven oldtimers (M2 is excluded from this graph) to the question "*Can you please state in your own understanding, why the Personalisation & Intelligent Knowledge Management VC has been created?*"

7.2.3 Effect of Notifications on Oldtimers

Two sets of notifications were generated and sent to members with different formats according to the description in Section 6.3. The first format of notifications which included the general links to the VC has been evaluated using the second questionnaire (at the end of period 2), and the second format of notifications has been evaluated using the third and last questionnaire (at the end of period 4).

From the second questionnaire (after the first set of notifications was sent) we can see that the opinion of members changed regarding CCen with the exception of M3, M6 and M7. For example, M5 selected at the first questionnaire M2 and M7 as the CCen members of the VC while at the second questionnaire M5 selected M2 and M3 who were the members mentioned as similar to M5 in the first set of notifications sent to that member (note that M5 had never met in person or collaborated in any way with M3). Similarly, M11 selected M2 and M6 as the CCenM replying in the first questionnaire (these two members are supervisors in the physical community). In the second questionnaire, M11 indicated M13 and M7 as central (both members appearing as similar to M11 in the first notification message). These observations show that the notification messages make people aware of other members in the community. In contrast, there were members, such as M3, M6 and M7, who did not change their opinion about community centrality across the three questionnaires. Their answers represent who they considered as CCenM in the physical community rather than the VC.

Have oldtimers followed the notifications sent? Two members reported they had followed the links in the first notifications and downloaded from the VC. Two members followed the links provided in the notifications but no actions were taken after that. The reason, as reported by the members, was lack of time:

"I am planning to do so, it was just a busy month for me."(M5, Questionnaire 2)

Four members did not follow the links in the notifications for the following reasons. One member stated that the information was not relevant to him, one member stated that he had not noticed the links in the notification and two members mentioned lack of time.

The situation was different after the second set of notifications. The results show that all oldtimers followed the links in the notifications. One member uploaded resources due to the notifications, one member downloaded resources, and seven members, although they had followed the links to the VC, had not uploaded or downloaded from the VC. Members stated that lack of time was the reason for not taking any action after they followed the links. *"I was busy at that time. I've just checked the message."(M3, Questionnaire 2).* The second set of notifications included more personalised information providing links to resources uploaded by similar members. We can infer here that compared to the general links provided in the first set of notifications, the second set was more appealing to oldtimers to follow and explore the links provided.

In what ways were the notification messages useful for oldtimers? Oldtimers rated the information they received in both message formats as relevant to them.

The information received through the first set of notifications (the beginning of the second period) helped VC members identify people with similar interests. Six members agree that the messages helped identify people they could contact for information, as well as to identify who was uploading similar resources. In addition, five (out of seven oldtimers) replied that the messages helped them identify people and resources that could be useful:

"I have discovered one connection which I didn't think of before."(M7, Questionnaire 2)

"Have not been using it (the VC) for a while...a message like this may be enough to remind me that there is a pool of information there for me to visit/revisit."(M6, Questionnaire 2)

Similar results were obtained from the third questionnaire which looked at the effect of the second set of notifications that followed a different format of personalised information (see Figure 8). Three members suggested they were motivated to upload resources and identify who the CCenM were because of the notification they received. As one of the members commented:

"The papers that were recommended to me sounded very interesting and of high quality. Until now I haven't been active in the community, but I have come across some papers that could be of interest to others. It would be nice to contribute to the community and give something back." (M11, Q 3)

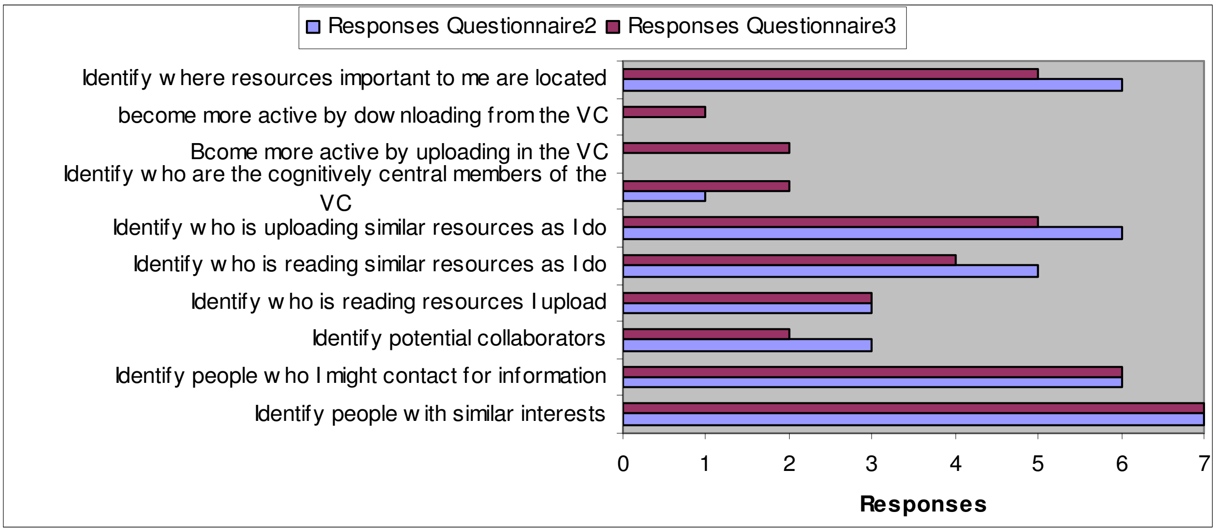

Figure 8 Replies of the seven oldtimers (M2 is excluded from the data presented in this graph) to the question how information received with the notifications helped members

From people's replies to the questionnaires after the notifications were sent, we can see that oldtimers at first tend to believe that the notifications provided a way of creating awareness (develop TM and SMM) in the VC, while after the second round of notifications members believed that the notifications were not only providing awareness but helped members to remain/become active.

Have the notifications motivated oldtimers? Although, there is not much evidence that messages motivated members, four members stated that the first set of notifications helped them remain active by visiting the VC space (uploading/downloading). As a VC member noted: *"When I read the notification e-mail, it motivated me to look at who have the same interests and read similar resource with me. That means, I can download the interesting resources from them or might take advice from them. If I did not receive the e-mail, I would forget to contribute to the community."(M9, Questionnaire 2)*

After the second set of notifications, one member believed the notifications motivated him to download and three members believed that the messages would motivate members to upload resources. Six oldtimers indicated that the information they received in the second set of notifications helped them in their everyday practice and motivated them to become/remain active: *"As a result of the email I uploaded and downloaded some resources" (M11, Questionnaire 3)*

"The notifications show a list of people who are interested in similar topics as me. It is useful to look at those papers uploaded from them" (M3, Questionnaire 3)

Have oldtimers become more confident to contribute Four oldtimers agree that receiving notifications in both formats would boost their confidence in contributing to the VC: *"I now feel more confident to share/download information from the community" (M9, Questionnaire 2)*

Four members selected the "Neither agree nor disagree" option and commented that:

"I don't think a message is enough to change my behaviour. But, if I get this regularly over a period of time which showed some activities it may rekindle my use of BSCW."(M6, Questionnaire 2)

I would not say I feel more confident, I may feel more engaged to this. However, unfortunately other activities retained me to contribute to the community. (M7, Questionnaire 3)

Based on the above findings, we cannot conclude that the notifications had an impact on the confidence of oldtimers to contribute in the VC, even though four members stated that they felt more self-assured in uploading and being active in the VC.

Have notifications had an effect on TM and SMM of oldtimers? One of the purposes of generating the notification messages was to develop TM and SMM. In all three questionnaires to oldtimers, members were asked to identify three other members from the VC who**:** may have similar research interests (Q8), may read similar resources (Q9), may upload similar resources (Q10). Following the method described in Section 7.1.3, the data was statistically analysed and compared among the three questionnaires. With regard to Q8 and Q10, the Wilcoxon non-parametric signed test shows no significant changes between the three questionnaires. This outcome was expected since the experiment ran for a relatively short period of time between the generated notifications sets. On the other hand, the results for Q9 show a marginal statistical difference between the questionnaire answers (Table 5), i.e. before notifications and after the two rounds of notifications were generated.

Table 5 Wilcoxon signed non-parametric test results for Q9 from oldtimers questionnaire. Results extracted from the questionnaires compared to the data in the community model for all members. The results show marginal statistical significant difference between the replies of questionnaires 1 and 3.

Wilcoxon Signed Non-Parametric Test for Q9		
F1 metric for Q9 Replies	**Z**	***p* (2-tailed)**
FQ1 – FQ3	-2.236	0.025

Although statistical results do not show much difference, looking at the data we can identify some interesting cases that show the influence of notifications on oldtimers. For example, in the first questionnaire M5 selected M2, M7 and M13 for Q8 (members with similar interests), while in the second questionnaire M5 changed his opinion and selected M6 instead of M13 (which actually was closest to that member's interests). Furthermore, M3 selected M2, M9 and M7 in the first questionnaire and changed his opinion to M2, M1 and M6. It is interesting to note that in M3's selection, a newcomer - M1 - was added. This shows that M3 acknowledged the addition of a new member and identified the similarity in interests they had.

With respect to Q9 (members who read similar resources), five (out of 7 oldtimers) changed their opinions after the first set of notifications. Changes can be seen also after the second set of notifications (between the second and third questionnaires). An interesting example is M3 who selected in the second questionnaire M2, M1 and M6, while after the second set of notifications M3 changed his opinion and thought that M8 was reading similar resources to him (note that M1 and M8 were newcomers to the community). This selection reflects what was happening in the physical community, and indeed M1, M8 and M3 realised their similarities and engaged in research activities outside the VC. Members M6 and M9 did not have an opinion about relevant members in the first questionnaire but after the notifications they made selections. In the case of M9, he selected two of the members who appeared also in his community model. In general the overlap between members' selections and what was extracted from the community model increased between the first and third questionnaires ($p = 0.025$, Table 5).

The selections of members for Q10 (members who upload similar resources) have changed also between the three questionnaires. Members' selections show they became aware of the newcomers, for example M7 selected M8 and M1 as having uploaded resources similar to his. Furthermore M5 was becoming aware of his similarity with M8 (in questionnaire 2 after the first notifications were sent) and with M7 (in questionnaire 3 after the second notifications were sent); although M5 had not worked with either of these two members.

There were noticeable changes following the two sets of notifications. Although members' selections initially reflected what was happening in the physical community, their opinions after the notifications (questionnaires 2 and 3) changed. Members became aware of people who joined the VC recently or people whom they had not met physically. This can be considered a positive effect attributed, to some extent, to the notifications. However there are clear cases where the members' opinion did not change and were far from the community model. In summary, although notifications did not have considerable impact on the actual behaviour of members, there was evidence that in some cases notifications helped members to develop a better awareness of what was happening in the VC (TM and SMM improved). The slow development of TM and SMM in the VC was also confirmed in some user comments.

"I have discovered one connection I didn't think of before"(M6, Questionnaire 2).

"It (notification message) *shows the list of people who are interested to similar topics with me, so it's useful to look at those papers from them." (M9, Questionnaire 3)*

7.2.4 Findings from the Newcomers' First Questionnaire

The first questionnaire was sent to the seven newcomers a few days after they joined the VC (i.e. at the end of the second period). The primary purpose of this questionnaire was to extract an initial list of interests for each joining member in order to generate their individual user models. In addition, this questionnaire helped in assessing issues related to the newcomers' awareness (TM, SMM).

Out of the seven newcomers, five did not participate at all in the VC prior to the generation of notifications. One member uploaded and downloaded and one member only uploaded to the VC. M1 and M12 commented that they did not participate in the VC due to lack of time. M12 reported that in his opinion *"I don't really work in similar areas with any of the other members"(M12, Newcomers Questionnaire 1)*. However, this member is in the same research group with four other members of the VC and, based on his individual user model, the community model indicated similarities between M12 and members M15, M6 and M9 (these members are in M12's research group). Another member, M14 only downloaded and noted, *"I am a new member and don't know what others are interested in"(M14, Newcomers' Questionnaire 1).* This indicates the initial lack of confidence, which can be attributed to a lack of awareness of how the newcomer could contribute to the community – lack of TM and SMM. To properly analyse the effect of notifications on newcomers' confidence to participate, longer term studies would be required.

With respect to CCen, the newcomers' replies to the first questionnaire show that they were unaware of who the cognitively central members were. Newcomers selected predominantly their supervisors (five out of seven) as the central members.

With regard to SMM, newcomers were asked to state why the specific VC had been created (see Figure 9). Several options were given from which the members could select several. Six (out of seven) newcomers chose the option *"To Share resources"* and three (out of seven) selected *"to keep important papers in one place"* and *"to have a resource repository online"*. Two members selected, *"to socialise"*, which does not represent the purpose of the VC. One member commented: *"To improve my awareness of the field & get up to date information about the field"(M12, Newcomers' Questionnaire 1).* The replies indicate that the newcomers'expectations joining the community was to share papers, which is in line with the purpose of this VC.

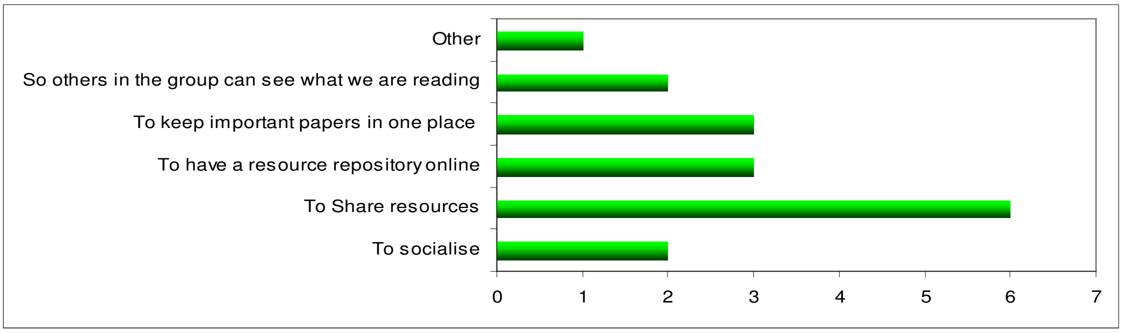

Figure 9 Seven newcomers replied to the question "Can you please state in your own understanding why the Personalisation & Intelligent Knowledge Managament VC has been created?" One member replied "Other " and he specified: *"To improve my awareness of the field & get up to date information about the field"*.

7.2.5 Effect of Notifications on Newcomers

The notification to newcomers was delivered as a welcome message providing relevant information and links (members and resources) according to individual interests of each member extracted from the newcomers' questionnaire.

After notifications were sent to newcomers their opinions about the CCen members changed. For example, M4 selected M5 as a cognitively central member in the final questionnaire, despite the fact that M4 never met M5. In the first newcomers' questionnaire, M14 selected his supervisor - M2 – and M13 to be the CCenM. After the notifications were sent, M14 thought that M13 and M7 were the central members. Similarly, M15 initially considered M2 and M6 (the two supervisors) as the two CCenM in the VC, but after the notifications were sent, M15 thought that M6 and M9 were central. The observed changes in the opinions about cognitive centrality indicate that the notifications may help build newcomers' awareness of who the influential members in the VC are.

Have newcomers followed the notifications sent? Three (out of seven) newcomers followed the notifications and two downloaded resources from the VC. One newcomer (M12) did not upload nor download any resources, commenting: *"My main research interest is in a different field." (M12, Questionnaire 3).* As discussed earlier, M12 did have InterestSim with others in the VC and although the notifications did bring this similarity to M12's attention, this member did not consider such information valuable. Four newcomers did not follow the notifications. One of them mentioned he had not noticed the links provided in the message and the others pointed at time restrictions.

For newcomers, it is harder to follow the notifications and upload/download resources from the VC than for oldtimers (newcomers were introduced to both a new community and a new software environment). Nevertheless, three newcomers (out of seven) followed the links to the VC, which is a positive indication that some newcomers may benefit from the notification approach. However, this conclusion should be taken with caution and can be validated in future longer term experimental studies.

In what ways were the notification messages useful for newcomers? 6 newcomers rated the information received with the notification messages as relevant to them. One member suggested that the information he received was not directly relevant to his research.

According to the information collected through questionnaire 3, notifications helped newcomers identify where resources important to them were located, identify people with similar interests, become more active by downloading and identify people they might contact for information. Two members stated that the information they received allowed them to become more active by uploading, to identify who is uploading similar resources as they do, to identify who is reading similar resources as they do and identify potential collaborators. One member mentioned that he got help in identifying who the central members of the VC were. Members provided further comments on how the notifications could help them integrate into the VC:

"Notifications reminded me that some of the resources in the VC could be useful for my current work!" (M4, Questionnaire 3)

"The notifications were a useful approach in sharing/reading resources and communication with others."(M10, Questionnaire 3)

Although only three members have followed the notifications and only two of them had an activity in the VC, five members agreed that the information they received would motivate them to be active.

"Having the notifications will make me aware about the community" (M15, Questionnaire 3)

Newcomers see the notifications as an awareness feature (to help them develop TM and SMM) that helps them identify their similarities with others in the VC, and the cognitively central members. In addition, notifications can serve as a motivational tool encouraging newcomers to visit the VC space. However, it is very difficult to motivate new members to contribute to a VC (Brazelton and Gorry, 2003), and it is even harder if their research interests do not directly fit with the VC (e.g. M12). The fact that five (out of seven) newcomers in the VC found the notification messages motivational, gives encouraging support that notifications could be a way of motivating and keeping newcomers active in a VC. Further experimental studies with larger user numbers would be needed to systematically examine the extent to which newcomers can be motivated by adaptive notifications.

Have newcomers become more confident to contribute? Five out of seven members agree that the information they received helped them build confidence in uploading/downloading from the VC.

“I’ will keep working/collaborating in VC and spend more time on navigation.”

“Based on what other people have contributed I'll find it easier to evaluate whether an article may be of interest to others in the community.” (M4, Questionnaire 3)

According to the newcomers’ comments, we can infer that the information they received with the notifications could have helped some members develop confidence in remaining active in the VC and consequently which could facilitate newcomers’ integration. However, the data from the study is insufficient to make a general conclusion about the possible connection between notifications and newcomers’ confidence.

Have the notifications had an effect on the TM or SMM of newcomers? The TM and SMM of newcomers were examined using the first and second newcomers’ questionnaires. Similarly to oldtimers, newcomers had to identify three other members from the VC who: may have similar research interests to them (Q8), may read similar resources to them (Q9), and may upload similar resources to them (Q10). Following the method described in Section 7.1.3, the data from both questionnaires was compared and statistically analysed. Wilcoxon non-parametric test was applied on the data collected for Q8, the results show a small difference with $p = 0.024$ (Table 6) between the two questionnaires. This can be due to newcomers’ interests extracted according to what they had provided as interests and thus TM or SMM was easier to be captured through this data. In terms of Q9 nothing could be extracted from the community model for the newcomers except M14 and M15 who were reading resources from the VC. The data for these two members showed that they had changed their opinions with respect to their selections after receiving notifications. M14 selects M1, M13 and M2 in the first questionnaire. After he received the notifications, M14 selects M1, M13 and M7. For Q10, the community model extracted information only for M15 who was the only newcomer uploading resources to the VC. The selection M15 made at the first questionnaire did not change after he received notifications. It is important to note that the selection of M15 represents what was happening in the physical community and not in the VC since M10 and M12 (selected by M15 as similar in uploading) were members from his research group and supervised by the same supervisor.

There are interesting observations with respect to Q8. Five out of seven members had changes in their opinion on who had similar interests to them. In the first community model extracted for newcomers the interests of each member were derived based on the keywords provided in the first questionnaire. Based on what members replied, the algorithms extracted the three most similar members in terms of interests to every newcomer. Although this model was based on the keywords they have provided there is overlap only on two occasions. After notifications were generated and the second community model extracted, there was a greater overlap between the members selected by newcomers in the second questionnaire and what has been extracted in the community model (Table 6). For example, M1 selected M2, M13 and M8 as the most similar members in the first questionnaire. In the second questionnaire, he selected M2, M13 and M3, which was exactly what was indicated in the community model. M4 selected only M2 in the first questionnaire, but in the second questionnaire M4 also selects M3 (and similarities were also detected in the community model). M8 had selected M2, M7 and M3 at the beginning but after the notifications M8 added M1, which is a link present in the community model as well. These examples show that new members joined the VC with limited (or no) TM but after the notifications they became more aware of who had similar interests to them. This is also confirmed in the newcomers’ comments, e.g.:

“VC helps me identify/discover broader details on members’ interests and locate additional resources I am not aware of” (M1, Questionnaire 3)

Table 6 Wilcoxon non-parametric test results for Q8 for newcomers. Results extracted from the questionnaires compared to the data in the community model for all members. The results show marginal statistical significant difference between the two questionnaires.

Wilcoxon Signed Non-Parametric Test for Q8		
F1 metric for Q8 Replies	**Z**	***p* (2-tailed)**
FQ1 – FQ2	-2.264	0.024

Have newcomers had an increase in their CCen? CCen for newcomers is calculated in the same way as for oldtimers (see Section 4.4). Figure 6 summarises the CCen variations, 2 newcomers had an increase in their centrality. During the first and second periods of the experiment, the centrality is 0 for all newcomers since these members were not members of the VC yet. Period 3 is the time newcomers were invited to join the VC and received the notification messages. M14 and M15 increased their CCen after the first and second set of notifications sent.

8 Discussion and Conclusions

This section will discuss the evaluation results focusing on key lessons learnt regarding community-adapted support, which can be beneficial for researchers in personalisation and user-adaptive systems. We will then conclude revisiting the research questions stated in Section 1.

8.1 Lessons Learnt

8.1.1 Supporting Community Social Processes

This research started with the assumption that support for communities should be given in a holistic way, considering the community as an entity. We looked at three key processes - TM, SMM and CCen – identified as important for the VC to grow and sustain. The evaluation results give confidence in the selection of these three processes as the basis for community support. It has been shown that problematic patterns relevant to TM, SMM, and CCen can be defined and identified by analysing community log data. By providing notification messages we aimed to create awareness among members with respect to who the CCenM are, how VC members relate to each other, and the purpose of the VC (TM, SMM). Based on the results extracted from the questionnaires and the community model, we can conclude that although the notification messages have not had an actual effect on the uploading/downloading activity, they had an effect on members' *awareness* and perception of how they related to other members in the VC. Furthermore, in some cases this awareness was transferred to the physical community, where in two occasions members engaged in discussions after discovering they had common interests in the VC.

The study confirms that community-adapted support based on TM, SMM, and CCen can lead to improved community awareness. Nevertheless, long term studies are necessary for any affirmative claims about the positive effect of using these processes for intelligent community support (e.g. the study showed that members needed a long time to conceptualise what was happening in the VC and this would affect their behaviour in the VC). Our focus has been on specific community processes which could be monitored by analysing log data. More comprehensive approaches would be needed to consider other processes, e.g. trust, motivation, and group efficacy.

8.1.2 Taking into Account Community Stages and Characteristics

Following the main **stages** of a VC as presented in (McDermott, 2000), we can conclude that the adaptive notifications approach presented in this paper is more suitable for the Grow and Sustain stages of a community where members are trying to make connections and keep the community active. There was small evidence from this study that the notification messages helped some of the oldtimers to be more active and also that they were beneficial for newcomers and helped some of them to integrate.

The study showed a strong backing for considering different community **actors** when providing community-adapted support. Our experimental design distinguished between the newcomers and oldtimers, which helped identifying different effects. We expected the newcomers' participation in the VC to be more influenced by notifications but only two newcomers engaged into an activity in the VC. According to the newcomers' comments, time was a problem since many of them were working against deadlines during the period when the study took place. One of the newcomers was leave at some point during the study. On the other hand, one of the inactive newcomers became active after the end of the study. Newcomers only received one round of notifications and this has not provided enough time for them to integrate properly within the VC. Nevertheless, some of them integrated and became aware of the similarities they had with others. The effect on oldtimers was different – we noted that they became more active following notifications sent to them, some of which led to helping newcomers integrate. Notably, the importance of a 'disconnected' oldtimer (outside the physical community) was 'discovered' by other members, and they started to benefit from the resources this member had uploaded.

An important finding about the behaviour of oldtimers is that participation in the VC is influenced by the **physical community**. In most of the cases what is happening online, in a VC, is part of a larger environment, usually within particular organisational settings. Knowledge sharing happens in the physical space, as well as in the virtual. In this research we have not considered any physical interactions of the community members when modelling or supporting the VC. When the community has only virtual interactions and members do not have physical contact, the approach proposed here would be a feasible way for modelling a VC. The evaluation shows that members might consider an influential person from their physical community to be a CCenM in the VC (although this may not represent the virtual activity of that member). This will not be captured in the tracking data, and hence will be missed in the community model. There is a need for further research to develop extended approaches that exploit the convergence of physical and virtual spaces when providing community support. For example, an open community model can be used to allow members to inspect and modify the community model to reflect what is happening in the physical space. The discrepancy between the community model generated by the system and the open model modified by members can be used to better target the notification messages.

The evaluation study showed that the **virtual space** used would impact the functioning of the community. The study used the BSCW system as a virtual space, which had advantages and disadvantages. BSCW is a robust system and the functionality is stable and well-tested. It also allowed us to keep the tracking data used as generic as possible. However, BSCW has a specific style of interaction, which some of the members were not familiar with. Most members had not used BSCW before and also, during the study, were not using BSCW for any other activities in their practice. Several members commented that they were less motivated to participate because they disliked the BSCW interaction style. We can assume that some of the negative results obtained may be attributed to the BSCW platform following studies that people tend to perform best when the tools are similar to what they are used to and also what appeals to their working style (Uruchrutu et al., 2005). This highlights the importance of taking into account a broad range of human factors, which we feel would be a feasible direction for future work in community-adapted support.

8.1.3 Using Tracking Data

The approach followed in this research is based on analysis of tracking data from the VC. Obviously, there are elements that can be captured by this kind of data and others that cannot. The advantage of extracting a community model based on tracking data is that it represents the actual interaction of members with the resources available in the VC. Furthermore, if members are working explicitly online, then a community model is a good source to represent a VC and use it to provide support. In the case when the VC is an extension of the physical community, tracking data is insufficient to capture all community interactions. Hence, no matter how robust the algorithms developed are, one cannot do much if there is no enough input. However, using the tracking data we can discover connections between members that they were unaware of. For example most of the research students involved regarded their supervisors as the only members they were connected to at the beginning of the study, but after the notifications they discovered that they were also connected to other members in the VC that they did not know about before. Although the notifications did not influence radically the behaviour of members, there was indication of positive effects on the awareness of people in the virtual space which enhanced their view of the physical space.

A disadvantage of using the tracking data for extracting a model of a VC is that members' interests change, and if the VC does not represent this change (e.g. when a member is not using the VC regularly) the extracted community model will not represent the current interests of the members. Opening the community model to the VC members and allowing them to modify their individual user models can be a possible way to address this problem. A second and well known problem of relying on tracking data to extract a community model is the cold-start problem. When people did not use the VC to upload or download resources, the algorithms were not able to extract connections among members. To overcome the cold start problem, we used the first questionnaires to gather information about members' interests. However, when members did not download/upload papers, the extracted community model included only interest similarity graphs. This is especially an issue when the VC is voluntary, like in our case, and there is no explicit incentive for people to participate.

8.1.4 Interference and Motivation

Our study looked only at notifications, sent as emails to community members. Some participants saw the approach of receiving notification messages as an interruption of their practice. On the other hand, some members found the targeted notification messages to be useful reminders. Members commented in favour of notifications and there was some evidence that notifications could motivate people to engage in the community. In other cases, notifications acted as a reminder that there was a pool of information that could be exploited. Similarly, some members mentioned the information in the notification messages influenced their confidence in contributing to the VC. We can conclude that the study found that notifications would be a useful approach to influence the awareness of the community as a whole (though there is a caution that the approach would not be uniformly accepted). Although there were members who benefited from the notifications, there were members who were not influenced by the notifications. This can be attributed to the volunteering principle for participating in the VC and the lack of any explicit incentives (see below).

Regarding the different message formats, members' responses show a clear preference for the more targeted (personalised) messages rather than the general ones sent with the first notifications. During the experimental study we have employed two formats of notifications. One was containing general links that pointed to the VC space, and the second included more personalised links pointing at relevant members and specific resources and folders containing information relevant to that specific member. Although there was more activity after the second (more personalised) notifications, this might also be a result of the members adapting to the methods in the study. Other approaches for designing notification messages can be considered, implemented and validated using the current framework. Social theories can inform the design of persuasive, motivational, and incentive driven messages that can influence members to contribute to the community and help them see the added value of their participation (Cialdini, 1993; Kollock, 1999; Preece et al., 2004; Rafaeli et al., 2004; Preece, 2009). It is more straightforward to facilitate participation in a community when some reward mechanism linked to participation can be given, e.g. students' participation in an online learning community can be encouraged with a small reward to their course mark (Cheng and Vassileva, 2005). When the VC is voluntary, further incentives and motivation strategies are needed. Consequently, an interesting continuation of our approach will be to investigate theories and design message content that will facilitate community participation and examine their impact on a close-knit VC.

8.1.5 Wider Applicability of the Proposed Approach

An appealing continuation of this work is the further application to different types of social web groups. Communities of Practice are an attractive possible application of our framework. It will be interesting to examine whether our approach will have a benefit to community of practice. Especially, when people are located in geographically dispersed areas, it will be interesting to examine what effects (if any) the approach can have on the knowledge sharing behaviour of members in such communities. Actors involved in communities of practice usually have different roles in the community: CCenM, CPerM and facilitators. A future extension would be to examine how to exploit these roles in providing support for knowledge sharing in communities of practice. A community of learners can be another possibility. Extracting semantic relationships among students who are sharing resources as part of a module and providing support through notifications to these members will allow us to examine what effects will this have on their sharing behaviour given that no incentive is given. For example, will students be self motivated and socially influenced by their CCen peers to contribute to the VC, how will lurkers react to the notifications? A more recent trend concerns the support of Collaborative Innovation Networks, which might exist in organisations or in the educational sector and involve actors from geographically dispersed areas and of diverse knowledge and skills. In these networks people are brought together to work towards the generation of innovation. It is believed that members of these networks need to be supported in identifying their complementarities more importantly than their similarities. It will be interesting to investigate what connections need to be modelled in this kind of social groups, and how the our framework should be adapted.

8.1.6 Study Design Choices

A limitation of the evaluation study can be the formulation of the open questions in all questionnaires. Members were asked to select three members from a list of all VC members as their reply to each question. This allowed us to extract the conceptualisation of members on who the CCenM of the VC were, who was uploading/reading resources similar to them, etc. What would have been a better approach was to provide a checkbox next to a person's name and allow members to select as many VC members as they want to define their similarities in the community. Furthermore this would have made the use of Precision, Recall and F1 metrics more meaningful. Since we had equal number of members selected as replies in the questionnaires (three members selected), and we had also three members (closest to a member), extracted in the community model the value for Precision, Recall and F1 was the same. In a more general approach where the selection number is not fixed, the precision, recall, and F1 metrics would be more informative.

Another aspect of the study design is the choice made that the experimenter was a member of the community. The questionnaires were given before the community model was extracted in each period in order to mitigate the influence on the experimenters' behaviour. When participants know the experimenters, they might reply in the questionnaire in a biased way. Having this issue in mind the experimenter tried to mitigate any noise in the data. In addition, the questionnaires have been structured in such a way that members' replies could not be adapted in order to please the experimenters, since they did not know what could be a correct answer. In fact, there was no way to be positively biased, as it was not known what the community model was at the time the questionnaires were conducted. Moreover, despite knowing the experimenter (the first author), some members did not contribute to the community, which shows that there was no bias to try to please the experimenter. The participation of the second author did not impact in any way the experimental study, as this author was not involved in the study design, conducting, and analysis. The subjective data from this community member have been excluded from the reporting in the paper.

On the other hand, being members of the VC, the authors could gain an insight of what was actually happening in the community, which enabled the interpretation of the results in a more meaningful way. For example, the important issue about the link between virtual and physical community or the links formed between members outside the VC could not be picked had the authors not been members of the physical and virtual community.

8.2 Conclusions

We will conclude the paper by revisiting the three research questions stated in Section 1.

How to extract a computational model to represent the functioning of a community as a whole by using semantically enhanced system log data?

We have formalised the input data to capture essential information about members, including information about users (member Id, email, date joined the community), activity data (uploading/downloading), resources (name, keywords (tags), description, rating) and an ontology representing the VC domain. Appropriate algorithms have been developed to extract a community model based on tracking data and semantically enriching this data using an ontology. We have described a general model for VCs that consists of individual user models of the community members, several relationship graphs, a list of popular and peripheral topics, and a list of the cognitively central members. Generic community tracking data have been used to extract this model, together with an ontology used to extract semantic relationship graphs. The algorithms for extracting relationship graphs have been kept flexible and can be adjusted according to the input data at hand. A study with archival data from an existing VC was conducted. Patterns of community behaviour were detected, and provided as the basis for community-tailored support. The proposed community modelling mechanism can be used in other knowledge sharing applications where key words or tags are associated with shared resources. The existence of an ontology is beneficial but not necessary (in the absence of a domain ontology, WordNet alone can be used as the source for discovering semantic relationships). The key limitation of the community modelling approach is its sole reliance on system log data. Although this ensures generality and wider applicability, we should point out that further extension is needed to take into account other input about the user/community behaviour (e.g. participation in external communities, characteristics of the physical community, user roles). Further work is also needed to improve the efficiency of the proposed community modelling algorithms.

How can user modelling, adaptation and personalisation techniques be utilised to support processes which are important for the functioning of close-knit virtual communities?

We have developed graph-based algorithms to analyse the extracted community model and identify knowledge sharing patterns. Knowledge sharing behaviour patterns in a VC have been defined, following selected processes (TM, SMM and CCen) important for the effective functioning of close-knit communities. Section 5 demonstrated how these patterns can be detected and used to provide community-tailored support in the form of personalised notifications. Although the interventions we consider - email notifications - are sent to individual members, the content of the notifications is based on what is known about the overall community behaviour regarding the specific community processes chosen. The adaptive notification mechanism defines why and how a notification is generated, according to detected knowledge sharing patterns. This allows deploying the approach in different close-knit VCs. In this paper, we have demonstrated how tracking data extracted from a widely used knowledge sharing system - BSCW - can be used in designing and extracting a community model and providing community-tailored support. The evaluation study has confirmed the feasibility of the approach and has pointed at additional aspects to be taken into account in selecting content and shaping the form of notifications: members' status (oldtimers versus newcomers), member's roles and relationships in the physical community (e.g. group leader, seminars moderator, supervisor, collaborator), frequency and timing of notifications (e.g. based on certain time points or when significant events are detected).

Can adaptive support, driven by community processes, affect the functioning of the community?

An experimental study was conducted with a knowledge sharing VC. Results provide evidence that supporting the development of TM, SMM and CCen in a VC can be beneficial for community knowledge sharing. The results of the evaluation show that notification messages can have a positive effect on members (both newcomers and oldtimers). Two formats of notification messages (general and personalised) have been sent to VC members. The second message format (personalised information for each member pointing at relevant members and providing links to resources in the VC) was preferred by members. In both cases, members rated the notification messages as relevant to them. In general, notification messages can be used for motivating members to keep active in the VC and, in the case of newcomers, to upload and download resources. The confidence of members slightly increased after receiving notifications and a slow development of TM and SMM was shown in members' comments. Members reported that they were becoming aware of the resources and people available in the VC. Some newcomers and oldtimers increased their activity after receiving notification messages. Finally, the results show evidence that monitoring the CCen of members can be used to support the knowledge sharing in a close-knit VC. The evaluation study has allowed us to draw wider implications for a newly forming research direction in personalisation and user-adaptive systems which considers holistic, community-adapted personalised support.

Acknowledgements

The work presents the PhD studies of the first author, conducted at the University of Leeds and funded by the School of Computing. The work is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no ICT 257184 (DICODE project). Special thanks go to the participants in the experimental study. The authors are grateful to the reviewers whose constructive comments helped improve the quality of the paper.

References

Ardissono, L. and Bosio, G. (2012): Context-dependent awareness support in open collaboration environments, *UMUAI,* vol. 22, no. 3, pp.223-254.

Ardissono, L., Bosio, G. and Segnan, M. (2011): *An Activity Awareness Visualization Approach Supporting Context Resumption in Collaboration Environments*, Proceedings of the International Workshop on Adaptive Support for Team Collaboration 2011 held in conjunction with the 19th International Conference on User Modeling, Adaptation and Personalization, UMAP 2011, CEUR Workshop Proceedings Girona, Spain.

Baghaei, N. and Mitrovic, T. (2007): *From modelling domain knowledge to metacognitive skills: extending a constraint-based tutoring system to support collaboration*, Int. Conf. on User Modeling 2007, Springer Corfu, Greece.

Barley, S., Dutton, W., Kiesler, S., Resnick, P., Kraut, R. and Yates, J. (2004): Does CSCW need organization theory?, *CSCW '04: Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work*, ACM Chicago, Illinois, USA.

Borgatti, S. P. and Everett, M. G. (2006): A Graph-theoretic perspective on centrality, *Social Networks,* vol. 28, no. 4, pp.466-484.

Brazelton, J. and Gorry, A. (2003): Creating a knowledge-sharing community: if you build it, will they come?, *Communications of the ACM,* vol. 46, no. 2, pp.23-25.

Bretzke, H. and Vassileva, J. (2003): *Motivating Cooperation on Peer to Peer Networks*, 9th Int. Conf. on User Modelling 2003, Springer Berlin / Heidelberg USA.

Chakrabarti, D. and Faloutsos, C. (2006): Graph mining: Laws, generators, and algorithms, *ACM Computing Surveys,* vol. 38, no. 1, pp.2.

Cheng, R. and Vassileva, J. (2005): *User Motivation and Persuasion Strategy for Peer-to-Peer Communities*, 38th Hawaii International Conference on System Sciences Hawaii, USA.

Cheng, R. and Vassileva, J. (2006): Design and evaluation of an adaptive incentive mechanism for sustained educational online communities, *Journal of User Modeling and User Adaptive Interaction,* vol. V16, no. 3, pp.321 - 348

Cialdini, R. B. (1993): *Influence: Science and practice*. New York, NY, US: HarperCollins College Publishers.

De Choudhury, M., Sundaram, H., John, A. and Seligmann, D. (2007): *Contextual Prediction of Communication Flow in Social Networks*, WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Silicon Valley, CA, USA.

Degenne, A. and Forse, M. (1999): *Introducing Social Networks*. London: Sage Publications Ltd.

Ester, M., Kriegel, H.-P., Xu, X. and Clustering, K. (1996): *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, AAAI Press Portland.

Falkowski, T., Barth, A. and Spiliopoulou, M. (2007): *DENGRAPH: A Density-based Community Detection Algorithm*, IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Silicon Valley, CA, USA.

Falkowski, T. and Spiliopoulou, M. (2007): *Users in Volatile Communities: Studying Active Participation and Community Evolution*, Proc. of the International Conf. on User Modelling, LNCS Vol. 4511/2007, Springer Berlin / Heidelberg

Corfu, Greece.

Farzan, R., DiMicco, J. and Brownholtz, B. (2009): *Spreading the honey: a system for maintaining an online community*, Proceedings of the ACM GROUP 2009 conference, ACM Florida, USA.

Fischer, G. and Ostwald, J. (2001): Knowledge Management: Problems, Promises, Realities, and Challenges, *IEEE Intelligent Systems,* vol. 16, no. 1, pp.60 - 72.

Freeman, L. (1979): Centrality in social networks conceptual clarification, *Social Networks,* vol. 1, no. 3, pp.239.

Freeman, L. C., Borgatti, S. P. and White, D. R. (1991): Centrality in valued graphs: A measure of betweenness based on network flow, *Social Networks,* vol. 13, no. 2, pp.141-154.

Fu, Y., Xiang, R., Liu, Y., Zhang, M. and Ma, S. (2007): *Finding Experts Using Social Network Analysis*, WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Silicon Valley, CA, USA.

Gross, J. and Yellen, J. (1999): *Graph theory and its applications*. London: CRC Press.

Harper, M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., Kraut, R. and Riedl, J. (2007): *Talk amongst yourselves: inviting users to participate in online conversations*, Proceedings of the 12th int. conf. on Intelligent user interfaces (IUI'07), ACM Honolulu, Hawaii, USA

Herlocker, J., Konstan, J., Terveen, L. and Riedl, J. (2004): Evaluating collaborative filtering recommender systems, *ACM Trans. Inf. Syst.,* vol. 22, no. 1, pp.5-53.

Hubscher, R. and Puntambekar, S. (2004). Modeling Learners as Individuals and as Groups. *Adaptive Hypermedia and Adaptive Web-Based Systems*: pp. 300-303.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M. and Jundt, D. (2005): Teams in Organizations: From Input - Process - Output Models to IMOI Models, *Annual Review of Psychology* vol. February 2005, no. 56, pp.517 - 543.

Kameda, T., Ohtsubo, Y. and Takezawa, M. (1997): Centrality in Sociocognitive Networks and Social Influence: An Illustration in a Group Decision-Making Context, *Journal of Personality and Social Psychology,* vol. 73, no. 2, pp.309.

Kay, J., Maisonneuve, N., Yacef, K. and Reimann, P. (2006): *The Big Five and Visualisations of Team Work Activity*, Intelligent Tutoring Systems 2006, Lecture Notes in Computer Science Volume 4053/2006, Springer Berlin / Heidelberg Taiwan.

Khan, J. and Shaikh, S. (2006): *Relationship Algebra for Computing in Social Networks and Social Network Based Applications*, WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Hong Kong.

Kleanthous Loizou, S. (2010). Intelligent Support for Knowledge Sharing in Virtual Communities. School of Computing. Leeds, University of Leeds. **PhD**.

Kleanthous, S. and Dimitrova, V. (2009): Detecting Changes over Time in a Knowledge Sharing Community, *Proc. of the 2009 IEEE/WIC/ACM Int. Joint Conf. on WI and IAT*, IEEE Computer Society Washington, DC, USA Milan, Italy.

Kleanthous, S. and Dimitrova, V. (2010): Analyzing Community Knowledge Sharing Behavior, *User Modeling, Adaptation, and Personalization*, Springer Berlin / Heidelberg Big Island, Hawai.

Kollock, P. (1999). The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace. *Communities in Cyberspace*. Smith, M. and Kollock, P., London: Routledge: pp. 220–239.

Kunegis, J., Lommatzsch, A. and Bauckhage, C. (2009): The slashdot zoo: mining a social network with negative edges, *WWW '09: Proceedings of the 18th international conference on World wide web*, ACM Madrid, Spain.

Latora, V. and Marchiori, M. (2007): A measure of centrality based on network efficiency, *New Journal of Physics,* vol. 9, no. 6, pp.188-188.

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

Lin, Y.-R., Chi, Y., Zhu, S., Sundaram, H. and Tseng, B. (2008): *Facetnet: a framework for analyzing communities and their evolutions in dynamic networks*, WWW '08: Proceeding of the 17th international conference on World Wide Web, ACM Beijing, China.

Lin, Y.-R., Sundaram, H., Chi, Y., Tatemura, J. and Tseng, B. (2007): *Blog Community Discovery and Evolution Based on Mutual Awareness Expansion*, WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Silicon Valley, CA, USA.

Liu, S., Liu, F., Yu, C. and Meng, W. (2004): *An effective approach to document retrieval via utilizing WordNet and recognizing phrases*, SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, ACM Sheffield, United Kingdom.

Lo, S. and Lin, C. (2006): *WMR--A Graph-Based Algorithm for Friend Recommendation*, WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society Hong Kong.

Masthoff, J. (2004): Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers, *User Modeling and User-Adapted Interaction,* vol. 14, no. 1, pp.37-85.

McDermott, R. (2000): Community Development as a NAtural Step: Five Stages of Community Development, *Knowledge Management Review,* vol. 3, no. 5 (November/December 2000).

Mohammed, S. and Dumville, B. C. (2001): Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries, *Journal of Organizational Behavior,* vol. 22, no. 2, pp.89 - 106.

Nieminen, J. (1974): On the centrality in a graph, *Scandinavian Journal of Psychology,* vol. 15, no. 1, pp.332-336.

Nonaka, I., Toyama, R. and Konno, N. (2000): SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation, *Long Range Planning,* vol. 33, no. 1, pp.5-34.

Olson, D. L. and Delen, D. (2008): *Advanced Data Mining Techniques*. Berlin: Springer.

Pal, A., Farzan, R., Konstan, J. A. and Kraut, R. E. (2011): *Early Detection of Potential Experts in Question Answering Communities*, User Modeling Adaption and (2011) Springer-Verlag Girona, Spain.

Paletz, S. and Schunn, C. (2010): A Social-Cognitive Framework of Multidisciplinary Team Innovation, *Topics in Cognitive Science,* vol. 2, no. 1, pp.73-95.

Paramythis, A., Lau, L., Demetriadis, S., Tzagarakis, M. and Kleanthous, S. (2011): International Workshop on Adaptive Support for Team Collaboration (ASTC-2011), *International Conference on User Modeling, Adaptation, and Personalization (UMAP2011)*, CEUR Workshop Proceedings.

Phillips, K. (2003): The Effects of Categorically Based Expectations on Minority Influence: The Importance of Congruence, *Pers Soc Psychol Bull,* vol. 29, no. 1, pp.3-13.

Pierrakos, D. and Paliouras, G. (2010): Personalizing Web Directories with the Aid of Web Usage Data, *IEEE Trans. Knowl. Data Eng,* vol. 22, no. 9, pp.1331-1344.

Preece, J. (2009). An event-driven community in Washington, DC: Forces that influence participation. *Handbook of research on urban informatics: The practice and promise of the real-time city*. Foth, M. H., IGI Global, PA, USA**:** pp. 87-96.

Preece, J., Maloney-Krichmar, D. and Abras, C. (2003). History and Emergence of Online Communities. *Encyclopedia of Community*. Wellman, B., Berkshire Publishing Group**:** pp.

Preece, J., Nonnecke, B. and Andrews, D. (2004): The Top 5 Reasons For Lurking: Improving Community Experience for Everyone, *Computers in Human Behaviour,* vol. 2, no. 1.

Puntambekar, S. (2006): Analyzing collaborative interactions: divergence, shared understanding and construction of knowledge, *Computers & Education,* vol. 47, no. 3, pp.332 - 351.

Rafaeli, S., Barak, M., Dan-Gur, Y. and Toch, E. (2004): QSIA - a Web-based environment for learning, assessing and knowledge sharing in communities, *Computers & Education,* vol. 43, no. 3, pp.273-289.

Sankaranarayanan, K. and Vassileva, J. (2009): *Visualizing reciprocal and non-reciprocal relationships in an online community*, n Proceedings of International Workshop on Adaptation and Personalization for Web 2.0 (AP-Web 2.0 2009) at UMAP'09, CEUR Workshop Proceedings Trento, Italy.

Schmidt, K. (2002): The Problem with `Awareness': Introductory Remarks on `Awareness in CSCW', *Computer Supported Cooperative Work (CSCW),* vol. 11, no. 3 - 4, pp.285-298.

Seco, N., Veale, T. and Hayes, J. (2004): *An Intrinsic Information Content Metric for Semantic Similarity in WordNet*, ECAI'2004, the 16th European Conference on Artificial Intelligence, IOS Press Valencia, Spain.

Shami, S., Yuan, C., Cosley, D., Xia, L. and Gay, G. (2007): That's what friends are for: facilitating 'who knows what' across group boundaries, *Proceedings of the ACM 2007 GROUP conference*, ACM Florida, USA.

Song, X., Tseng, B., Lin, C.-Y. and Sun, M.-T. (2005): *ExpertiseNet: Relational and Evolutionary Expert Modeling*, Lecture Notes in Computer Science 3538, Springer, Verlag Heidelberg.

Tagalakis, G. and Keane, M. (2005): *How Understanding Novel Compounds is Facilitated by Priming from Similar, Known Compounds*, Proceedings of the Cognitive Science Society Stresa, Italy.

Thomas-Hunt, M., Ogden, T. and Neale, M. (2003): Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange Within Groups, *MANAGEMENT SCIENCE,* vol. 49, no. 4, pp.464-477.

Upton, K. and Kay, J. (2009): Narcissus: group and individual models to support small group work, *Proceedings of the 17th International Conference on UMAP*, Springer Trento, Italy.

Uruchrutu, E., MacKinnon, L. and Rist, R. (2005): *User Cognitive Style and Interface Design for Personal, Adaptive Learning. What to Model?*, International Conference on User Modeling 2005 Edinbourgh, Scotland.

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. and Milios, E. (2005): *Semantic similarity methods in WordNet and their application to information retrieval on the web*, WIDM '05: Proceedings of the 7th annual ACM international workshop on Web information and data management, ACM Bremen, Germany.

Viermetz, M. and Skubacz, M. (2007): *Using Topic Discovery to Segment Large Communication Graphs for Social Network Analysis*, Web Intelligence, IEEE/WIC/ACM International Conference on, IEEE Computer Society Silicon Valley, USA.

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

Wellman, B. (2001). The Rise of Networked Individualism. *Community Networks Online*. Keeble, L., Taylor & Francis, London**:** pp.

Wenger, E. (2000): Communities of Practice and Social Learning Systems, *Organization,* vol. 7, no. 2, pp.246.

Wolfgang, P., Uta, P.-B., Wolfgang, G., Tom, G., Sabine, K. and Scafer, L. (2004). Presenting Activity Information in an Inhabited Information Spaces. *Inhabited Information Spaces Living with your Data*. Snowdon, N. D., Churchill, F. E. and Frecon, E., Springer, UK**:** pp.

Zhang, J., Ackerman, M. and Adamic, L. (2007): Expertise networks in online communities: structure and algorithms, *Int. Conf on WWW 2007*, ACM Alberta, Canada.