Intelligent Support for Knowledge Sharing in Virtual Communities

Styliani Kleanthous Loizou

Submitted in accordance with the requirements for the degree of
Doctorate of Philosophy

UNIVERSITY OF LEEDS
The University of Leeds
School of Computing

May, 2010

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Publications

Some of the work in this thesis has been presented prior to thesis submission:

Static knowledge sharing patterns that defined in **Chapter 5** and parts of the adaptive notifications from **Chapter 7** discussed in:

Kleanthous, S. and Dimitrova, V. (2010) Analyzing Community Knowledge Sharing Behavior, in P. De Bra, A. Kobsa, and D. Chin (Eds.): UMAP 2010, LNCS 6075, pp. 231–242, Springer – Verlag, Berlin Heidelberg.

The detection of knowledge sharing patterns over time has been discussed in **Chapter 6** and in:

Kleanthous, S. and Dimitrova, V. (2009) Detecting Changes over Time in a Knowledge Sharing Community, Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology – WI/IAT'09, Volume 1, pp. 100-107, IEEE Computer Society, Washington, DC, USA.

The advantages of the ontology integration in this approach have been discussed in:

Kleanthous, S. and Dimitrova, V. (2009) Semantically Enhanced Community Model for Identifying Relationships and Centrality in Virtual Communities, TEL-CoPs'09: 3rd International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice at the 4th European Conference on Technology Enhanced Learning, Nice, France.

Chapter 4 and **Chapter 5** have been discussed at:

Kleanthous, S. and Dimitrova, V. (2008) Modelling Semantic Relationships and Centrality to Facilitate Community Knowledge Sharing, W.Nejdl et al.(Eds.): AH'08, LNCS 5149, pp. 123-132, Springer – Verlag, Berlin Heidelberg.

An overview of the approach has been discussed in:

Kleanthous, S. (2007) Semantic-Enhanced Personalised Support for Knowledge Sharing in Virtual Communities, C.Conati, K. McCoy and G. Paliouras (Eds.): UM 2007 LNAI 4511, pp. 465 – 469, Springer – Verlag, Berlin Heidelberg .

An early version of community model definition and input formalisation presented in **Chapter 4** and at:

Kleanthous, S. and Dimitrova, V. (2007) A Semantic-Enhanced Approach for Modelling Relationships and Centrality in Virtual Communities, SociUM: 1st Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities at 11th International Conference on User Modeling, Corfu, Greece.

Theoretical issues in VCs discussed in **Chapter 2**, and the proposed framework, **Chapter 3**, discussed also in:

Kleanthous, S. and Dimitrova, V. (2006) Towards a Holistic Personalised Support for Knowledge Sharing in Virtual Learning Communities, 1st Workshop on TEL Communities of Practice, at 1st European Conference on Technology Enhanced Learning, Crete, Greece.

Acknowledgements

Without the encouragement and support of many people the completion of this thesis would have not been possible. I would like to extend my sincere appreciation to them, even though this gesture does not adequately express my appreciation.

I am indebted to my supervisor *Dr. Vania Dimitrova* for her unique insight and warm guidance. She has always been encouraging and enthusiastic and it has been a genuine pleasure working with her. I owe a great debt of appreciation to her for the time, effort, support and valuable guidance she offered.

I would like to express my deep gratitude to the people in the *Intelligence Augmentation forum*, at the School of Computing of the University of Leeds, for the many stimulating discussions and useful comments we shared during this journey. I wish to offer special thanks to the people who helped completing the evaluation of this research project. I would also like to thank the *School of Computing, University of Leeds* for funding my studies.

I would like to express my special thanks to my examiners, *Dr. Daniel Kudenko*, University of York and *Dr. Lydia Lau*, University of Leeds. Their invaluable guidance and insight during the viva enhanced the presentation and quality of this thesis to the outmost extent.

Special and deep thanks to *Demetra Loizou*, my sister in-law who has proof read part of this thesis and to my aunt *Athena Charalambous* who has acted as a childminder for my son during the difficult time of writing the thesis. I could never thank you enough.

I would also like to take this opportunity to let my mother *Olvia*, my father *Philppos*, and my siblings, *Kleanthis, Nikolas and Vicky,* know how much their constant support and encouragement has meant to me, especially during the final stages of this journey. I would not have been able to finish without your support.

Last and by no means least, I would like to express my deepest gratitude to my husband *Andreas* and our son *Nikolas* for all their commitment and patience as well for their help, care, support, and listening ears. All my love goes to you and our little *Nikolas*. Thank you for being a great family and for your endless love.

To My Family

Abstract

Virtual communities where people with common interests and goals communicate, share resources, and construct knowledge, are currently one of the fastest growing web environments. A common misconception is to believe that a virtual community will be effective when people and technology are present. Appropriate support for the effective functioning of online communities is paramount. In this line, personalisation and adaptation can play a crucial role, as illustrated by recent user modelling approaches that support social web-groups. However, personalisation research has mainly focused on adapting to the needs of individual members, as opposed to supporting communities to function as a whole.

In this research, we argue that effective support tailored to virtual communities requires considering the wholeness of the community and facilitating the processes that influence the success of knowledge sharing and collaboration. We are focusing on closely knit communities that operate in the boundaries of organisations or in the educational sector. Following research in organisational psychology, we have identified several processes important for effective team functioning which can be applied to virtual communities and can be examined or facilitated by analysing community log data. Based on the above processes we defined a computational framework that consists of two major parts. The first deals with the extraction of a community model that represents the whole community and the second deals with the application of the model in order to identify what adaptive support is needed and when.

The validation of this framework has been done using real virtual community data and the advantages of the adaptive support have been examined based on the changes happened after the interventions in the community combined with user feedback.

With this thesis we contribute to the user modelling and adaptive systems research communities with: (a) a novel framework for holistic adaptive support in virtual communities, (b) a mechanism for extracting and maintaining a semantic community model based on the processes identified, and (c) deployment of the community model to identify problems and provide holistic support to a virtual community. We also contribute to the CSCW community with a novel approach in providing semantically enriched community awareness and to the area of social networks with a semantically enriched approach for modeling change patterns in a closely-knit VC.

Contents

Appendices

Figures

Tables

Abbreviations

BSCW	Be Smart Cooperate Worldwide
CCen	Cognitive Centrality
CCenM	Cognitively Central Member
CM	Community Model
CoI	Community of Interest
CoIs	Communities of Interest
CoP	Community of Practice
CoPs	Communities of Practice
CPerM	Cognitively Peripheral Member
CSCW	Computer Supported Cooperative Work
IUM	Individual User Model
RM	Relationships Model
SMM	Shared Mental Models
TM	Transactive Memory
VC	Virtual Community
VCs	Virtual Communities

Conventions

The meaning of *member* and *user* in this thesis is assumed equal. They refer to a person joined a virtual community and has interacted with the virtual community's space. In the same line of thought, *user model* refers also to a virtual community's *member model.* They can be used in the same context to name the model build inside a computer system to present the interactions and knowledge of a person joined the virtual community.

A *resource* refers to the *academic papers, links to papers or other academic material* people are sharing in the virtual community's space. The term *community* is used to mean *virtual community* as defined later in the thesis, unless otherwise is specified.

Throughout this thesis we will use male gender for the member, which is only for convenience. To clarify, he shall be taken to mean he or she and his shall be taken to mean his or her.

Note that throughout the thesis, *we* refers to the author and *our* refers to the author's.

Chapter 1

Introduction

The process of knowledge sharing among individuals has been investigated for many years in the fields of psychology, social psychology and education. It has been proven that individuals are constructing knowledge though interaction in their social context (Vygotsky, 1978). Different environmental factors can affect communication between people and thus affect the sharing of knowledge. Although sharing of knowledge happens even through everyday conversations, it has been shown that certain processes can improve and empower knowledge sharing (Wenger, 1998; McDermott, 2000; Mohammed and Dumville, 2001; Ilgen et al., 2005). Knowledge can be tacit and explicit (Lewis and Allan, 2005). People are not often aware of the tacit knowledge they possess or how it can be valuable to others, while explicit knowledge is articulated, codified, and stored in certain media and can be easily transmitted to others. Tacit knowledge develops in long term personal interaction with the person possessing the knowledge, and can be transferred from one person to another. On the other hand, explicit knowledge can be articulated and transferred from person to person through some form of interaction.

Communities of people that are connected with interpersonal ties, where people socialise, have a sense of belonging and support each other through interaction, are the primary medium where humans share knowledge (Wellman, 2001). Recently, with the advancement of the internet, communities are being built online and constitute an important part of most peoples' lives (Preece et al., 2003). Online or virtual communities (VCs) are one of the most effective mediums for sharing knowledge formally and informally (Wenger, 2000). Organisations spend a fair amount of money to build and maintain closely-knit communities where employees share experiences, information and knowledge even if they are not located in the same geographic area. This is considered to be one of the most inexpensive and successful ways of managing knowledge in an organisation (Nonaka et al., 2000). Educational institutions and academia are another sector where closely-knit VCs have been very successful (Puntambekar, 2006). Researchers across the globe are coming together in a VC to collaborate on large or small projects, organise academic conferences or just share resources and develop collective knowledge.

However, for closely-knit VCs to be successful more than just people and technology is needed (Fischer and Ostwald, 2001). Researchers in this area need to have a good understanding of what the problems in virtual communities are and what processes have to be supported to enable effective knowledge sharing to be effective and to add value to all members (McDermott, 2000). Organisational psychology has identified processes that have an impact on the collective knowledge sharing and are important for the effective functioning of teams which share common characteristics with closely-knit communities (Mohammed and Dumville, 2001; Ilgen et al., 2005).

This thesis examines how processes which are important for the functioning of communities, can be supported with intelligent techniques, and what effect this could have on the knowledge sharing and functioning of closely knit VCs. We focus in three key processes: *Transactive Memory* (TM) (members are aware how their knowledge relates to the knowledge of the others) (Wegner, 1986), *Shared Mental Models* (SMM) (members develop a shared understanding of the key processes and the relationships that occur between them) (Mohammed and Dumville, 2001) and *Cognitive Centrality* (CCen) (members who hold strong relevant expertise can be influential; it has been shown that members of effective communities gradually move from being peripheral to becoming more central and engaged in the community) (Ilgen et al., 2005). Following these processes, the thesis will present a novel approach to supporting closely-knit VCs as a whole, as opposed to existing approaches that focus solely on supporting individual members.

The hypotheses that drive this PhD are that*: Providing intelligent support tailored to the community as a whole and supporting TM, SMM, and monitoring CCen can be beneficial for knowledge sharing and community functioning. In addition, monitoring static and time depended patterns of knowledge sharing behaviour of members can enable a more targeted support to be generated and help the community to share knowledge more efficiently, in a more sustainable manner, for a longer period of time.*

Following the above hypotheses this research will address three research questions:

- How to extract a computational model to represent the functioning and evolution of the community as a whole by using semantically enhanced tracking data?
- By using that model how can intelligent functionality be provided to support the development of TM, building of SMM and monitoring of CCen?
- How can intelligent support of the above processes affect the functioning of the community?

We will present a computational framework that consists of a mechanism for extracting a community model (CM), algorithms for analysing the CM to derive knowledge sharing patterns and detect changes in the VC, and a mechanism to generate adaptive notifications sent to individual members but aimed at improving the functioning of the community as a whole. Two assumptions have been made at the outset of this research:

- TM, SMM and monitoring CCen within a community are important for the functioning of a VC (Ilgen et al., 2005).
- Resources shared by community members represent the topics of interest of the VC members, and correspond to the knowledge a member holds (Song et al., 2005; Cheng and Vassileva, 2006).

This PhD thesis combines methods from the areas of user modelling and adaptive systems, social networks and graph theory. As a result, a number of original contributions are made to the following research communities:

- *User modelling and adaptive systems* - the thesis presents (a) a novel framework for holistic adaptive support in virtual communities, (b) a mechanism for extracting and maintaining a semantic community model based on the processes identified, and (c) deployment of the community model to identify knowledge sharing patterns and to provide holistic support to a virtual community.
- *Computer supported cooperative work* - the thesis contributes to this research area with a novel approach for providing semantically enriched community awareness.
- *Social Networks* - the thesis presents a semantically enriched approach for modeling change patterns in a closely-knit VC, which can be applied to a broad range of social networks and web-based communities.

This PhD thesis is organised in nine chapters. Chapter 2 will set the scope of this PhD and outline the foundations for the remaining chapters. We will justify the need for providing intelligent support for knowledge sharing in closely-knit VCs. Community in the broad sense will be defined and the distinction between physical and virtual community will be discussed. Furthermore, processes relevant to the functioning of VCs will be presented and the three processes selected will be discussed in detail. Problems relevant to knowledge sharing in VCs will be presented explaining their importance.

Community modelling approaches by other authors will be discussed in Chapter 3, where the computational framework proposed in this research will be outlined. The two main parts of the framework – acquisition of a CM and application of the CM for community-tailored support - will be briefly introduced and the components of each part will be explained. The procedure for the acquisition of the community model will be presented at this point followed by a description of a mechanism for maintaining the CM. The chapter will also give a brief summary of the use cases conducted where the CM has been applied to an existing VC. A description of this VC that was based on sharing of resources in Semantic Web by researchers will be given in Chapter 3. The community will then be used for the validation of the framework components presented in Chapters 4, 5 and 6.

Chapter 4 will provide a detailed definition of the CM. First, we will give details of the input used for extracting the CM and how this input is formalised. Then, the CM mechanism components, namely Individual User Model (IUM), Relationships Model (RM), Popular/Peripheral Topics, List of Cognitively Central and Peripheral Members (CCenM and CPerM), will be presented. Chapter 4 will end with the description of an initial study which applied the CM mechanism over tracking data from the Semantic Web VC. This provided a way to validate the CM algorithms, and to identify problematic cases in the VC by analysing the CM with a visualisation tool. The problems identified at this stage informed the development of algorithms for the automatic detection of these patterns.

Chapter 5 will extend Chapter 4 defining an approach for automatic detection of problematic patterns in a VC. Relevant literature on extracting graph patterns in social networks and communities will be discussed at the beginning to position our approach in the related work. Since in Chapter 4, relationships among community members are extracted as graphs, we will develop graph-based algorithms to automatically extract patterns of members' knowledge sharing behaviour. Similarly to Chapter 4, the algorithms will be validated with a study using tracking data and the CM from the Semantic Web VC. The results of the study will check the feasibility of automatically detecting knowledge sharing patterns related to TM, SMM, and CCen.

Chapter 6 will examine the modelling of community evolution. It will start with presenting relevant approaches that employ graph based techniques to model community evolution, which will position the approach followed in this PhD in the related work. We will then explain in detail our approach for extracting community changes over time. Graph-based algorithms will be defined and implemented based on the CM described in Chapter 4. A study will be presented where the evolution algorithms have been validated by applying them on the CM extracted from the tracking data of the Semantic Web VC. The results will enable us to test the feasibility of the proposed approach for modelling evolution patterns in closely knit VCs.

The graph-based algorithms defined in chapters 5 and 6, will inform the generation of intelligent support tailored to the VC as a whole. This thesis will consider intelligent support in the form of individualised notification messages. The mechanism for defining notification messages based on the CM defined in Chapter 4 and the knowledge sharing patterns defined in Chapters 5 and 6, will be presented in Chapter 7. A distinctive characteristic of the notifications is that although they are sent to individual members, their objective is related to the TM, SMM and CCen in the community as a whole. The rationale behind the notifications will be given and corresponding association will be made between each notification pattern and the processes followed in this research. A detailed formalisation of the notifications will be provided in the form of rules, which will be illustrated with examples.

The effect of notifications to a VC will be examined with an experimental study presented in Chapter 8, which will give the summative evaluation of our framework. The evaluation will focus on the validation of the notification messages and the effect they may have on the VC as a whole, as well as on specific categories of members, including oldtimers (exiting members) and newcomers (newly joining members). The results will be discussed following specific questions addressed in the study: *Effect on the community as a whole*: questions concern the types of change patterns that can be recognised after notifications are sent; whether CCen shifted between members; do peripheral members become more central; whether members develop links and follow resources from others. *Effect of notifications on oldtimers:* questions will examine if oldtimers followed the notifications, and if not why; in what ways (if any) can the notifications be useful for oldtimers; were oldtimers motivated to engage in the community due to notifications; have oldtimers become more confident to contribute; any effect on the TM and SMM of oldtimers; any changes in the activity of oldtimers as a result the adaptive notifications they receive? *Effect of notifications on newcomers:* have newcomers followed the notifications, and if not why; what is the usefulness of the notifications for newcomers; were newcomers motivated to integrate in the community; do newcomers become more confident to contribute; any effect on the TM and SMM of newcomers; do newcomers' activity change as a result of the adaptive notifications they receive? The evaluation of the notifications will be done using a different VC which was active during the evaluation period, which enabled us to examine the impact of notifications.

Finally, Chapter 9 will summarise the main aspects of this research, discuss the applicability and generality of our approach, outline the main contributions, point at limitations of the framework, and discuss future directions in terms of application and extension of our approach.

Chapter 2

Knowledge Sharing in Virtual Communities

2.1 Introduction

Virtual communities appear to offer a unique approach for people to communicate online and, accordingly, share knowledge through interaction. There is a broad literature on virtual communities that will be narrowed here in terms of the type of community in question, the actors involved and the technology employed regarding any specific virtual community under study. Although communities constitute an effective way to share knowledge, there are also problems that act as obstacles to knowledge sharing. The aim of this chapter is to delineate the scope of this PhD by defining the area of supporting virtual communities.

The next sections will set the theoretical foundations behind this research and define the processes that are imperative in supporting the effective sharing of knowledge among virtual community members. Furthermore, this chapter will stress the need to adopt a holistic intelligent support which will be driven by selected processes underlying the functioning and sustainability of virtual communities.

2.2 Virtual Communities

During the last decade, academics and practitioners have been searching for techniques to support knowledge extension and sharing (Puntambekar, 2006). Online communities appear to be an exceptional approach bringing together people from diverse backgrounds (Preece et al., 2003), providing support for collaboration, and – through this collective knowledge sharing – facilitating the creation of shared understanding (Lewis and Allan, 2005; Puntambekar, 2006).

Community as a term may be defined in various ways. Different disciplines such as sociology, information technology, business, psychology and education (Preece et al., 2003), vary in their perception of what constitutes a ‘community’. From a social sciences perspective, a general definition for a physical community of people is:

“Communities are networks of interpersonal ties that provide sociability, support, information, sense of belonging, and social identity” (Wellman, 2001).

Furthermore, McMillan and Chavis define the sense of community as:

“a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members’ needs will be met through their commitment to be together”(McMillan and Chavis, 1986).

In general, in a physical community, people may share common interests, belong to the same department, live in the same village, work on the same project, or attend the same course.

With communities existing in almost every area of our lives, the emergence of the internet brought about the development of online communities (Preece, 2001). The broad literature offers different definitions of the terms online community and virtual community and the two are usually used interchangeably (Preece et al., 2003). A widely used definition considers virtual communities as:

“a set of users who communicate using computer-mediated communication, and have common interests, shared goals, and shared resources” (Lazar and Preece, 2002).

Similarly, virtual communities have been defined as:

“people with shared interests or goals for whom electronic communication is a primary form of interaction” (Dennis et al., 1998).

The above definitions of virtual communities do not exclude the physical interaction of their members. Often a physical community is complemented with virtual interaction and shared resource repository and, to this effect, its members may share knowledge in a more effective manner. An example of this kind involves a group of nurses working at the same hospital and who are using a community system in order to share medical articles and experiences online. Although, the definitions do not coincide, physical and virtual communities are not mutually exclusive. Irrespective of the nature of the community in question, the key element is the presence of the human factor: groups of people that interact with each other to either achieve a goal, to share knowledge, socialise or even entertain each other (Lewis and Allan, 2005).

2.2.1 Types of Virtual Communities

A virtual community is a social network of individuals who interact through specific media, potentially crossing over geographical and political boundaries in order to pursue mutual interests or goals (Rheingold, 1993), engage in discussions and eventually share their knowledge. In this area, knowledge sharing is broadly used to represent the sharing of resources, information and eventually knowledge. Thus, the notions of knowledge and information are used interchangeably in the literature (Kim, 2000; Fischer, 2001; Bieber et al., 2002; Brazelton and Gorry, 2003; Cheng and Vassileva, 2005; Shen et al., 2006). In terms of participation, members who communicate with the use of the VC can either be collocated or in geographically dispersed areas.

There is no consensus among researchers regarding the different types of VCs available on the web (Eunyoung and JoongHo, 2009). In general, the term virtual community is an umbrella term encompassing many different kinds of groups of people who choose to interact using the internet. The online interaction may involve the sharing of recourses, the use of a chat room or a wiki, a weblog where common interests are discussed, or an asynchronous or synchronous discussion forum (or a range of the above facilities) (Bell and Heinze, 2004; Lewis and Allan, 2005; Pan and Leidner, 2003). Virtual communities embrace different authority levels. For instance, some members may have access to more actions than others in a forum (Preece et al., 2003; Klein et al., 2005). They may differ as regards to their size, the number of participants, how long a community exists and for how long it will continue to operate. They may have a different purpose and operate either at a local (e.g. people working for the same organisation), national, or an international community level. Although there are communities build specifically for sharing knowledge (knowledge sharing communities) what happens in the vast majority of cases involving virtual communities is knowledge sharing through interaction (knowledge sharing in community). Popular sub-definitions of virtual communities are: communities of practice and communities of interests. Such definitions not only constitute different definitions but they also represent different types of communities.

Communities of practice (CoPs) as defined by Wenger, McDermott and Snyder (2004) are:

“groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis”.

Members of CoPs are bound by the task they undertake together in their respective community. Through their common practice sharing knowledge is effectuated and the members eventually learn according to legitimate peripheral participation – newcomers of the community learning through participating in minor activities of the community and eventually share knowledge, learn and become core members of the community (Lave and Wenger, 1991; Engestrom, 2001). Examples of this type of community are: architects, urban planers, research groups, software developers, and end-users (Fischer, 2001). Participation in this type of community is voluntary where the members are engaging in a common practice. It is not a prerequisite for the community to operate online in order to be considered as CoP. A newcomer becomes a full member by gradually learning the community's practice (experiences of past members transferred to new members) (Chanal and Kimble, 2010). The common practice members' engagement is what distinguishes this type of community from communities of interests (Wenger, 1998; Fischer, 2001).

In contrast to CoPs, **communities of interest (CoIs)** are composed of members brought together in order to solve a particular problem or alternatively have a common passion and interest. They usually operate online, are more temporary than CoPs, and as a result of their heterogeneity (that is, due to the lack of common practice) they are potentially more innovative and easier to transform than other community structures (Fischer, 2001). An example of CoIs is a team interested in software development which is composed of software designers, marketing specialists, psychologists, and programmers, who jointly undertake to solve a problem together. The mode of learning for CoI members is different from that of CoPs members. In a simultaneous process, they must be taught to communicate with and learn from others (Engestrom, 2001), bearing in mind that the members involved can have different views and perhaps employ different vocabulary to describe things.

Communities are usually distinguished between loosely-structured and closely-knit ones. In a **loosely-structured community**, membership is freely available to anyone, and despite the fact that members have to log-in to the system to upload resources or participate in discussions, there are no restrictions in terms of interacting with material on the community's space. Examples of this kind of community, which is outside the scope of this research, are large social network sites like: CiteULike[1] – members use the community to create personal and/or group libraries of academic resources online. Online is their only means of interaction; Del.icio.us[2] **-** members share bookmarks online. Membership is freely available but non-members have access to the material as well. People who belong to this community live in geographically dispersed areas and hence they use the internet to communicate and share bookmarks; MetaFilter[3] is a community weblog where its members can post comments which can be viewed and/or discussed by the community. Most of the content is also available to non-members.

---

[1] http://www.citeulike.org/

[2] del.icio.us is a social bookmarking website -- the primary use of del.icio.us is to store bookmarks online, which allows people to access the same bookmarks from any computer and add bookmarks from anywhere: http://del.icio.us/

[3] MetaFilter is a weblog that anyone can contribute a link or a comment to: http://www.metafilter.com/

Closely-knit VCs exist in small organisations, the educational sector, or are composed of researchers who share similar interests and are working together on large or small projects and to this effect they are sharing knowledge. In this type of community there is closed membership, members have to be invited in order to join. There is a common purpose for the creation of the community which is identifiable by its participants. Membership may be equal or there might be a facilitator that leads the VC to the ultimate goal. There is a common repository and/or discussion forum within which the members are sharing resources and are engaging in high level dialogues and discussions. In this way, such repositories and/or forums act as a medium to share knowledge effectively. Although most of the activity is taking place online, closely-knit communities might also involve physical interaction. This is illustrated by communities operating in BSCW (Wolfgang et al., 2004) or Wiki based VC like AWESOME (Bajanki et al., 2009) or any controlled-membership VC system. CoPs and CoIs are also representative examples of closely-knit communities.

In this PhD we follow the definition provided by Lazar and Preece (2002) since it is broad and includes the elements that characterize a virtual community for the purposes of this research. The scope of a VC is narrowed in order to focus our attention on and consequently to consider closely-knit communities that may exist either in the 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*. Importantly, this definition does not exclude the physical interaction of members along with the virtual. We assume that members are sharing resources through the VC space and that such resources represent the knowledge and interests any given community member has on an area. Two VCs have been employed in the formative and summative evaluations of this PhD: both VCs are operating on BSCW system. The first one (Semantic Web VC) is used in the formative evaluation phase in Chapter 4, Chapter 5 and Chapter 6. This VC is composed of researchers at different stages of their professional life and operating completely online in view of the fact that these researches are located in different countries and have been collaborating in projects by exchanging resources through the VC. The second VC used for the summative evaluation (Chapter 8) also consists of researchers located in different professional stages. Notably, this VC is partially operating in a physical context where most of the members know each other but it also includes members (three) located in other institutions who have never met some of the other members. This VC initially undertook its activity in the physical context but later on its members proceeded to move their sharing resources activity online.

2.2.2 Community Stages and Actors

Virtual communities are voluntary communities and have a lifecycle during which they created, developed and sustained (Wenger, 1998; McDermott, 2000; Wenger, 2001). Different proposals have been put forward regarding the stages a community goes through during the course of its life. Wenger (1998) introduced the following five stages where: a) members identify commonalities (Potential); b) community identifies its potential (Coalescing); c) members come together and engage in developing a practice (Active); d) members' activity drops (Dispersed); e) community is no longer active but members remember it (Memorable). Five steps have also been utilized by McDermott (2000) to describe the lifecycle of a community: Plan, Start-up, Grow, Sustain and Close (Figure 2.1). At a later stage Wenger, McDermott and Snyder (2002) combined their ideas and created a third lifecycle for communities that included Potential, Coalesce, Mature, Sustain and Transform (Wenger et al., 2002). Recently a different lifecycle has been proposed which is very similar to the above and consists of: inception, creation, growth, maturity, and death (Alicia and Gondy, 2009). A similar approach to the community's lifecycle is the one introduced by Ilgen et al. (2005) concerning small teams in organisations. Teams share similar characteristics to closely-knit VC and CoPs (Wenger et al., 2002) and to this effect the approach is considered as relevant. In their approach three stages are defined - Forming, Functioning and Finishing – where each stage is divided into sub-stages: Forming involves the stages of trusting, planning and structuring. In this stage the members learn to work together, trust each other and identify where knowledge is located in the community. Functioning involves bonding, adapting and learning. It is at this stage that members create bonds with each other and with the team as a whole. They learn who holds unique knowledge that can be exploited when the team needs to adapt to changes and it is through this interaction that eventually learning takes place. Finally the team enters into the finishing stage (Ilgen et al., 2005). In all of the above lifecycles the community appears to be planned and generated, begins to expand and be very active, then appears to be losing breadth and finally stops functioning. The most general approach is the one presented originally by McDermott (2000) and is discussed in more details below.

Plan: The need of people to share or find knowledge brings about the emergence of a VC. Such communities must have the potential to grow. Thus, the people who are to be invited in the VC have to be carefully chosen. It is imperative to invite key people to join the community in order for the community to initiate its course and eventually grow.

Start-Up: Communities often start with a spike of activity but when members realise the level of effort required to share knowledge the excitement quickly falls. At the beginning members are not familiarized with the other members' interests. Thus, they cannot tell what is valuable to share. Time is needed for relationships to build up among members and realise where the value of the community lies in order for it to grow. Consequently, relationship building should be encouraged in order for members to help each other and promote the true value of the community towards its members.

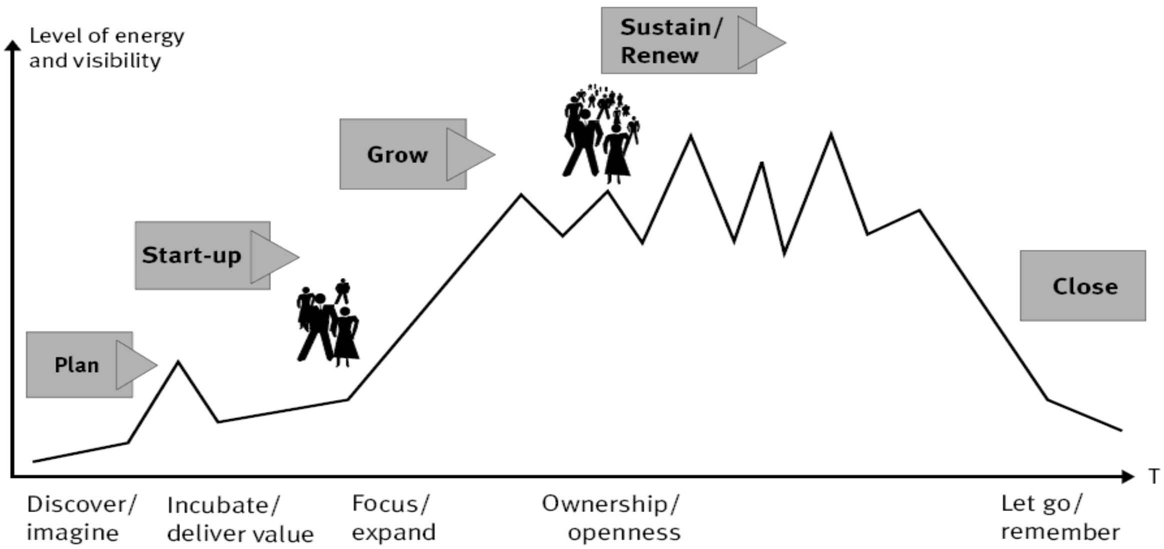


Figure 2.1 The stages of community development. Taken from McDermott (2000).

Grow: A community expands when knowledge is effectively shared among the members and there is value from which people may gain. Newcomers bring new ideas and perspective to the communities. A community thrives from new people and new knowledge. Communities should have a focus but at the same time new members should be invited to the core. A mentorship programme can facilitate the growth of the community.

Sustain: Communities change and expand during their maturity period. New people join the community with new interests and knowledge, new relationships are created and activity has its lows and ups. The main problem faced by communities that grow older is how they are going to sustain and adapt to changes. During a long period of operation the people involved in the community create a sense of belonging and may seem cold or uninterested towards new people or ideas. In order to remain active communities must be open to new ideas and new people and be prepared to shift their focus.

Close: Communities close in two ways: they either lose members over time until there is no activity in the community or some core members just keep what has been shared in the community. The key aspect during this stage is for the community to close before it stops functioning itself and identify who are the key members who will articulate the core knowledge generated in the community so as to carry it on.

As discussed in the above stages, core members (oldtimers) are equally important as newly joining members (newcomers) in the functioning of a VC. Most communities are composed of individuals with different degree of expertise. Experts in an area, novices, or individuals with an average level of knowledge usually coexist in the same community (Klein et al., 2005). Members who are located in the heart of the community (cognitively central), or alternatively members found in the periphery of the community (peripheral), all play their part in contributing towards the sustainability of a community (Lave and Wenger, 1991).

Oldtimers or core members are considered to be those members who have been invited to join the community during the planning stage (Figure 2.1) of the community and have been members of the community since then. Oldtimers usually create and develop ties and relationships among themselves through community interaction (Lave and Wenger, 1991; Fischer, 2001). They usually consider themselves as the owners of the collective knowledge of the community and may sometimes be reluctant to other members joining (McDermott, 2000). For example, oldtimers might reduce their sharing due to the lack of developed trust among themselves and towards the newcomers of the community (Ilgen et al., 2005).

Newcomers are also important for the community since they are in possession of new knowledge that needs to be incorporated in to the community so as to renew it and help in sustaining it (Lave and Wenger, 1991; McDermott, 2000). The importance of newcomers is stressed by Lave and Wenger (1991) in the development of legitimate peripheral participation. Based on this theory, newcomers eventually learn and move to the centre of a community thus becoming themselves oldtimers (Lave and Wenger, 1991).They have to be supported through different approaches in order to identify their position in the VC and comprehend what is the added value of their membership (Vygotsky, 1978). For example, apprenticeships programs are considered as effective tools to help newcomers integrate in a community (physical or virtual).

In CoPs, CoIs and knowledge sharing VCs **experts** and **novices** are considered of equal importance. This way knowledge moves in both directions experts provide support and answer the questions of less experienced members, whilst at the same time the experts have the chance to view things through the eyes of novice members in a manner perhaps not considered by them before (Puntambekar, 2006). Dina Lewis and Barbara Allan (2005), theorists who support the importance of human interaction, support the view that individuals construct their knowledge as a result of interactions with others within a particular social context (Lewis and Allan, 2005). Along this line, knowledge sharing can be considered as an important process which needs to be supported in a way which would enable both experts and novices to benefit from this interaction. Experts can include newly joining members in the community and novices can be members who belong to the core of the community. The maturity of a member in the community does not coincide with the level of expertise held by that member on given subject.

The presence of central as well as of peripheral members is imperative to the survival of the community (Lave and Wenger, 1991). **Central members** represent the life of the community (McDermott, 2000). These are the members who share the most valuable information with the rest of the community (Ilgen et al., 2005), keep the community active and help peripheral members to learn and integrate in the community (Lave and Wenger, 1991; Leskovec et al., 2005; Lewis and Allan, 2005). At the same time **peripheral members** are of equal importance to the community since their interaction within the community provides new insight and perspective to the other members (Wenger, 1998). Peripheral members help the community to sustain for a longer period of time and to adapt to changes of the environment (Ilgen et al., 2005). For example, as interests change and areas of focus shift in the community, peripheral members hold unique knowledge that can be exploited so as to relocate the community's focus. These are the members who will attain central positions and take over the growth of the community (McDermott, 2000). Furthermore, they may found boundaries of centrality and peripherality of any given community. The value of these people cannot be overestimated. On the contrary, it is necessary to be equally supported and acknowledged since they are the ones who are closest to centrality (compared to peripheral members) and may compose the future core of the VC (Lave and Wenger, 1991; Fischer, 2001).

In this PhD, we aim to provide a holistic support tailored to the needs of the community as whole. The community lifecycle provided from McDermott (2000) is followed. Accordingly, we focus on the three stages experienced by the community: Start-up, Grow, and Sustain. Oldtimers need to be supported in order to engage in the community and to create relationships among each other. Newcomers have to be supported in order to identify the benefit of sharing knowledge in a community and create relationships with other community members. All members, newcomers and oldtimers alike, must be aware of what is happening in the community and what similarities exist between them and the other members of the community. Members who share knowledge which is the closest one to the community's central topic are considered to be central members and while members who are sharing resources of less importance to the community are regarded as peripheral members of the VC.

In every type of community (physical or virtual), there are problems that cause the community not to function properly and not be able to work in a sustainable manner. The next section will discuss problems that impact knowledge sharing in a VC.

2.2.3 Problems in Communities

A common misconception is to believe that VC will be effective when people and technology are present. Fischer (2001) recognises the challenges in the process of knowledge sharing in communities. Challenges in VCs can cause the community to have a very short “Sustain” period and arrive in the “Close” stage very early. Lack of trust is a big problem between community members especially when they are not involved in a physical community and virtual interaction is their only means of communication (Hartley, 1999; Kreijns et al., 2002). In an environment of uncertainty, community members do not believe in each other and thus the community cannot function as a whole (Hoadley and Kilner, 2005; Ilgen et al., 2005). Incompatibility of the systems used by the people in the community can cause confusion, chaos and the community is inhibited from benefiting from what is actually available (Laurillard, 2002; Koper et al., 2004; Lewis and Allan, 2005). For example, in a resource sharing community, where a member is using a different system for managing his resources, it is very difficult for that member to change his routine. Poor resource sharing, can also affect the development of shared understanding between members of the community (Mulder et al., 2002; Hoadley and Kilner, 2005; Ilgen et al., 2005). For example, duplication of resources in the VC space demonstrates that members are not aware of what is already available in the VC.

In addition, newcomers in a community usually face difficulties of integration into the community as well as identifying the added value of sharing knowledge within the specific community (Lave and Wenger, 1991).At the same time they get to know how they can relate to the oldtimers of the community. Newcomers need to feel welcomed by the community and accordingly they must be equipped with the necessary support so as to discover what is relevant to them. Also, the expertise of newly joining members must be acknowledged. In effect, it is essential to make those members feel important to the community (McDermott, 2000). Similarly to newcomers, the new perspective peripheral members bring to the community, needs to be recognize so that these members do not lose interest and disengage prematurely (Fischer, 2001).

Furthermore, oldtimers face problems in establishing relationships and identify where knowledge important to them is located in the community (McDermott, 2000). The members may lose interest and consequently the community stops functioning. Lack of participation is an important problem that communities come to face (Lave and Wenger, 1991; Ardichvili et al., 2003; He, 2004; Gouvea et al., 2005; Rashid et al., 2006; Shen et al., 2006). Where members choose not to participate (previously active members becoming inactive) an indication is provided as regards the lack of fresh resources or knowledge entering the VC. Thus, less active members are gradually becoming disinterested and the community quickly enters the “Close” stage of its lifecycle.

In this PhD we hypothesise that by employing technology, identifying and supporting key processes in the VC and monitoring static and time depended patterns of knowledge sharing behaviour in the community, we can overcome the above problems and help the community to share knowledge more efficiently in a more sustainable manner for a longer period of time.

2.3 Community Processes to be Supported

Fischer and Ostwald (2001) stress the need for appropriate support for the effective functioning of virtual communities (Fischer and Ostwald, 2001). This requires a good understanding of what is happening within a community, and what processes influence the success of knowledge sharing. Researchers in the field of social sciences have examined the functioning of VCs and the processes that affect the knowledge sharing between its members (McMillan and Chavis, 1986; Romm et al., 1997; Hung and Der-Thanq, 2001; Kerr and Tindale, 2004; Ilgen et al., 2005). Information sharing theory suggests that people are sharing the greatest amount of information with others when they feel familiar within the social context, that is sharing is perceived as a constituent part of the environment (meaning people and technology) (Constant et al., 1994). Consequently, the processes that influence the environment of VC members have to be supported.

Potency deals with the collective belief that members can be effective together. Having members to believe in the effectiveness of their group helps the VC to achieve their highest performance (Ilgen et al., 2005). Trust towards other members of the community and their competence as well as psychological safety affect knowledge sharing in a community (Edmondson, 1999; Ilgen et al., 2005). Trust in general is a very important cognitive process that needs to be properly supported and developed among community members. For example, if members do not trust each other then they will not share what they are holding within the community. Self efficacy is also vital given that members need to believe that their community is performing well in terms of knowledge sharing and generation (Kerr and Tindale, 2004). Bonding among community members is also regarded as important as members feel supported and that they can count on each other for new information and knowledge when such need arises (Fischer, 2001; Ilgen et al., 2005). Furthermore, bonding enables the members to develop relationships and become aware of each others' contribution. The ability of the community to adapt to its changing environment and to have the potential to shift its focus has also been scrutinized by social sciences researchers.

In addition, research in organisational psychology has identified that effective teams and groups which are operating in the boundaries of an organisation build transactive memory (TM), develop shared mental models (SMM), and become aware which members are their cognitively central (CCen) and which members are regarded as peripheral ones (Wegner, 1986; Hollingshead, 2000; Mohammed and Dumville, 2001; Hollingshead and Brandon, 2003; Kerr and Tindale, 2004; Ilgen et al., 2005). Since we are dealing with closely-knit communities with characteristics similar to those of groups and teams (Wenger, 2001), the above processes can also be applied to a broader context to ascertain what kind of support should be provided to a VC. Three processes (TM, SMM and CCen) have been selected. These are considered as important in this research in view of their direct impact on knowledge sharing. Furthermore, they can be investigated by employing the means of analysing the tracking data of a closely-knit VC. Below we discuss in more detail the three processes identified as important to be supported in this research.

2.3.1 Transactive Memory

Transactive Memory (TM) deals with the relationship between the memory system of individuals and the communication that occurs between them (Wegner, 1986; Hollingshead, 1998). The focus is on encoding, storage and retrieval of information. Therefore, a TM system can provide the option to recall previously visited areas and subjects, and to identify relevant knowledge (Wegner, 1986; Mohammed and Dumville, 2001). The notion of TM and the development of TM system has proven to be very promising for the functioning of teams and groups (Wegner, 1986; Hollingshead, 2000; Mohammed and Dumville, 2001; Hollingshead and Brandon, 2003). Wegner (Wegner, 1986) points out that TM is concerned with:

"the prediction of group and individual behaviour through an understanding of the manner in which group processes and structures information".

TM helps group members to be aware of one another's expertise and to divide responsibilities with reference to different knowledge areas. The key element behind the ability of a TM system to function is for the divergent information held in members' head to be known by the other members. To illustrate this we assume that member A's memory can act as an extension of member B's memory. If B is aware of what A knows, he/she should be able to get access to A's knowledge and the information possessed by A. Virtual communities can benefit from a TM system since members will become aware of the knowledge held by other members. Furthermore, the promotion of TM creates awareness on who is interested in what and facilitates the identification of complementary knowledge. To this effect, the opportunities for collaboration among community members are potentially enhanced.

2.3.2 Shared Mental Models

Shared Mental Models (SMM) are defined as the

"team members' shared, organised understanding and mental representation of knowledge about key elements of the team's relevant environment" (Mohammed and Dumville, 2001).

Studies confirm that collaborative knowledge exploitation can be improved if community members have a shared understanding of the environment, situation and task at hand (He, 2004). One of the main objectives of community formation is the development, through knowledge sharing and communication, of a shared understanding of the context in which community members act, and consequently the creation of a shared understanding of the world(Merali and Davies, 2001; Puntambekar, 2006).

The development of SMM among VC members will promote awareness regarding the relations/similarities members have with each other in the community. Furthermore, this will allow oldtimers to identify members with similar interests and recognise what is considered as important (for the community as a whole) that accordingly is worth sharing. In this context, the integration of newcomers will be easier since they will realise what constitutes the purpose of the community and understand what similarities they have with oldtimers. Building this kind of awareness among members will make resource sharing more effective given that members will know what is important for other members and they will be able to evaluate what they share.

2.3.3 Cognitive Centrality

Cognitive Centrality (CCen) considers the importance of the contribution of individual members having regard to the wider community context (Ilgen et al., 2005). Specifically CCen is defined as:

"The greater the degree of overlap between the information a member holds and information other members hold on average, the greater the degree of centrality for that member"

Members who share a significant amount information which is valuable for the whole community become cognitively central (CCenM) and play a vital role in the smooth functioning of a community. These members are considered as key ones in view of the fact that the sustainability of the community depends on them. Conversely, if a CCenM holds a piece of information relevant to the community and fails to share it with the others then the community will go to great efforts to find out about it, otherwise it might never know. CCenM are important for the maintenance of activity in the VC and the addition of new material. They can also guide new members during their integration in the community. It is therefore essential to detect whether influential members continue to be active and to help them understand their importance for the effective functioning of the VC. In addition, monitoring the activity of CCenM can be used as a motivational mechanism for CPerM to become more active.

On the other hand, peripheral members can sometimes hold unique knowledge, and can also be important for effective knowledge sharing. Having many members on the periphery of the VC causes the community to become inactive, thus shortening the “Sustain” stage of the lifecycle. Lave and Wenger (1999), give a description of how members initially found in the periphery gradually strengthen their integration in the community so as to become central. Consequently, these are the members who by undertaking a proper integration, they will become at a future stage the heart of the community., Accordingly, they help the community to adapt in relation to possible changes which would keep the community active for a longer period of time.

2.4 Support Needed

The above processes can affect the functioning of VC, and can act as indicators of the kind of support that may be necessitated in a given situation. This will be illustrated below with the aid of several scenarios. We will show that the support to be provided to a VC has to be tailored to the community’s needs and serve the needs of both newcomers and oldtimers (Wenger, 2000). Furthermore, intelligent support should add value to the creation and sharing of knowledge between members in addition to facilitating the functioning of the community as a whole.

2.4.1 Support to Newcomers

Newcomers are newly joining members who need to identify their role in the community and what their gain will be from participating in that community. Support is needed in order to effectuate the quick integration of these members in the community’s knowledge processes. This would improve their sharing experiences and can bring about a positive effect on the overall functioning of the community. For example, consider a person named Chris who is interested in social tagging for e-learning and is joining a VC where members share information about technology-enhanced learning. Chris has no background about what was previously happening in the community, does not know about the interests and knowledge of other members, is unsure whether there are any relevant resources on the topic he is interested in, and does not know what he contribution can be towards the community. Help should be provided to Chris enabling him to identify people or knowledge important to him in the community under consideration. Furthermore, Chris should be given assistance in order to identify the kind of knowledge held by him while at the same time the other members become aware that Chris holds valuable knowledge, referred to as TM and SMM. Such assistance facilitates his introduction in the community. Moreover, given that social tagging is identified as a peripheral topic for this community, Chris may be encouraged to elaborate on its relation with personalised learning, which is the main focus, i.e. CCen, of this community. This will be beneficial for him (he may discover relationships he was previously unaware of and may become a more central member of this community) and for the community (new topic will be connected to the community's domain which can improve the processes of knowledge sharing and construction).

2.4.2 Support to Oldtimers

Existing members (oldtimers) should be given help in order to integrate and become active participants in the community's knowledge processes. For example, consider Jane who is an existing member of this community and is interested in intelligent tutoring systems. She is regularly uploading and downloading resources and is actively engaged in discussions with other members. Jane is one of the cognitively central members of this community. We assume that another member – Mark – is interested in student modelling something which Jane is familiarized with (as she has participated in discussions dealing with this topic and has uploaded relevant resources). Mark and Jane should be given help and support in order to discover that they have joint interests. In this way, not only themselves but the other members of the community as well can benefit by combining their knowledge and consequently extending the community's TM.

Jane is now working on a new project and needs to find information on ontologies - a topic she is not very familiar with. Assistance can be granted to her in order to allocate relevant resources within the community and establish contacts with members knowledgeable in the area, all these being related to the community's TM system. Jane may also be encouraged to upload more resources on ontologies and examine the link of this topic with technology enhanced learning. If this new topic is of interest to many members, it will move close to the community's CCen.

The community has to adapt to changes in its environment. This may lead to a shift of the central area of interest as well as to transformation of participation (Wenger, 2000). Consequently, active contributors may become passive members, while others who used to be peripheral participants may become CCenM (Kerr and Tindale, 2004; Ilgen et al., 2005). For example, Jane may gradually reduce her participation or even stop contributing to the community. If such changes are detected, CCenM like Jane who are moving towards the periphery can be encouraged to participate more actively in the community's knowledge processes.

2.4.3 Support to VC as a Whole

People categorise and organise their resources differently according to specific characteristics, diverse conceptualisations, searching habits, etc.(Berlin et al., 1993; Indratmo and Vassileva, 2005). Confusions may occur and disagreements are inevitable (Indratmo and Vassileva, 2005) and these can have an impact on the effective functioning of a VC (Berlin et al., 1993; Wu and Gordon, 2004; Golder and Huberman, 2005). Consider for example several members of the community who are interested in the use of context in systems of technology-enhanced learning. Each member uploads resources perceived as important to them and relevant to the projects they are engaged in. Jane considers context from an Artificial Intelligence perspective and links it to encoding different viewpoints in an ontology.

Chris associates context with the conditions in a learning environment, while Mark is engaged in a mobile learning project where context is used to represent location-based information. Appropriate support for effective knowledge sharing would encourage members to develop awareness regarding these issues, which can form part of SMM.

People participating in VCs share an information space and may be engaged in active communication. These are preconditions for collaboration often associated with effective VCs, where members either work together on a joint project or share a common desire to produce better services (Dillenbourg, 1999). Collaboration among community members can be encouraged in two ways. Firstly, support should be provided to help members build a common understanding as to what constitutes the purpose of the community, who is involved and what their interests are, what tasks people are undertaking, what is happening in the community and how it progresses over time. These issues relate to building SMM and developing a good TM system. Secondly, interaction between community members can be encouraged in order to create more opportunities for collaboration. Situations where members will possibly benefit from communication with others can be identified.

To sum up, TM, SMM and CCen relate to the effective functioning of a community and are critical in defining intelligent support tailored to the needs of the community as a whole. TM is effective in promoting the quick integration of newcomers to the community, improving the benefits of oldtimers so as to motivate their participation, and encourage further collaboration amongst all members. SMM is a pre-requisite for effective knowledge sharing and is directly linked with awareness and information localisation; it is also a key factor in facilitating collaboration between community members. CCen can be helpful in putting into perceptive and relating to the community's domain the knowledge of newcomers and existing members, as well as monitoring the changes which occur within the community over time.

2.5 Existing Approaches to Support Knowledge Sharing in VC

2.5.1 Example Systems

This section will review what computational methods have been developed to address TM, SMM, and CCen by employing several representative systems.

Answer Garden (Ackerman, 1998) supports organisational memory where recorded knowledge is made retrievable and by knowledgeable individuals are made accessible. The emphasis given is in helping people find answers to their questions thus saving the frequently asked questions. The functionality of the system pursues the following procedure: the user types a question in the system, the system searches and matches the question according to a natural language processing approach. If the answer is not in the system or the user is not satisfied with the answer provided, then he can relay the question to an expert. The expert will then reply with an answer. If the question is frequently asked and/or important then the answer will be posted to the system. This way the database expands and the organisational memory along with it.

BSCW (Basic Support for Cooperative Work) (Wolfgang et al., 2004) system has been developed by the Institute for Applied Technology at Fraunhofer. Its main goal is the transformation of the Web from a primarily passive information repository to an active cooperation medium. In BSCW, users have to register and fill in a profile before they are able to access the workspace. After registration, they have the choice to belong to more than one community, and to have shared as well as private workspaces. With regard to shared workspaces, users share their resources with other members but anything uploaded on the private workspace is private to the specific user. The main activities undertaken in BSCW are uploading, downloading and rating of resources, synchronous and asynchronous communication, version control and search facilities. Resources are organised in a nested folder hierarchy and the creator is free to name them. Besides, given that users have certain access rights one user may not be able to edit the folder or the resource created by another user. When a user uploads a resource, he has to rate it according as he thinks fit. Additionally, an event awareness facility is available, which keeps the members of a community informed as to when and where changes have been made in the community's space.

Comtella (Cheng and Vassileva, 2005) is a small-scale peer-to-peer application implemented at the University of Saskatchewan, Canada. It allows the development of online community and enables the sharing of resources among its members. People may share files (e.g. research papers) and services (e.g. help each other). In community supporting systems high participation of members is imperative to the functioning of the community as such. Along this line, motivation and reward mechanisms have been implemented. The techniques that are mostly used by the system consist of visualisations. Firstly, there are three levels of memberships available which are determined based on the member's contribution to the community. The more new the resources a member is sharing with others are, the more stars this member is receiving. According to the number of stars held by a member, membership is identified as gold, silver and bronze. This classification is publicly available to the community and consequently members are motivated to enhance their participation so as to upgrade their membership. Additionally, peers are able to rank uploaded resources thus making higher quality contribution possible. Members with higher membership levels receive more points to give out. This means that they are more influential to the community. Notably, this is a very simple mechanism. Another technique used by Comtella is the clustering of peers based on their interests whereby clusters are presented as galaxies. A peer with more than one interest area might appear in more than one galaxy. When the mouse is hovered over the galaxies, the full name of the interest category (category's topic) appears, along with the number of members in that category. When the galaxy icon is clicked it explodes to show the cluster of peers with their full details and their stars.

In the more recent implementations of Comtella, a reward mechanism, based on the contribution of each individual member to the community has been introduced. A simple community model is built based on the number of total contributions to the community's current topic (QC) and the community reward factor (FC). Each time a new topic is introduced in the community a new QC is set up by the administrator. This is done in order to reward people and hence motivate them to participate.

GIMMe (Lindstaedt, 1996) is a web based group memory system which helps groups to capture, store, organise, structure, share, and retrieve electronic mail conversations. The system serves as a central repository of all mail which has been distributed within the group. A daemon running in the background periodically collects all mail messages which are sent to the group memory. Members themselves can also send important emails directly to the group memory. Additionally, messages are categorised against the group's category hierarchy which is based on the message's subject line. A message can be linked to different categories or subcategories so it can cover the wider information space. Users can navigate along the category hierarchy to find a mail they are interested in. Furthermore, can also create, copy, move or delete categories as well as the mail messages belonging to them. This way they can manually categorise messages already stored in the memory. GIMMe supports free text queries over the group memory using Latent Semantic Indexing algorithm. Thus, GIMMe helps users to overcome the vocabulary problem since Latent Semantic Indexing is based on co-occurrence of words and can retrieve documents even if these documents do not contain the exact search words. GIMMe is a system that deals only with organisational memory and does not promote knowledge sharing directly. In fact, it supports the group and not the individual. Categorising email conversations in an organized structure can help members to find previously discussed topics.

KSE (Merali and Davies, 2001) is a system of information agents in charge of organising, summarising and sharing knowledge from a number of sources, including the World Wide Web, an organisation's intranet or other users. Users in KSE are clustered into communities of interest with related or overlapping interests. An agent corresponds to each user. This agent holds the user's profile which is updated according to the usage of the system by the user.

Jasper II (Merali and Davies, 2001) is an extension of KSE. In Jasper II users can store to their own agents only the relevant meta-information (created by the user), about information, which considered as important. The system then matches the information stored in the users' profiles. Along with this, phrases extracted from this information are added to or removed from their profiles, if this is what the user wants to. Users have the option to search for other users in the system. Alternatively, the system can also automatically suggest users with similar interests. In addition to that, the system promotes the development of shared understanding by being able to capture the individual perspective of the user. Nevertheless, given that a user who is posting a document is assigning his meta-information to that document, the development of cognitive consensus is not supported by the system. A member's personal view and understanding of a concept, does not match someone else's. Thus, a second user who might be interested in that piece of information might not be able to reach it since the meta-information assigned to it is not comparable to what his agent is holding in his profile.

MILK (Agostini et al., 2003) is a system aimed at a) supporting communities of interest along with the official organisational structure and b) cluing together knowledge associated with people, communities and informal knowledge. The main constituent part of MILK is the Metadata Management System (MMS). MMS is responsible for capturing and organising knowledge on profiles which includes metadata describing various aspects of involved elements. The system is designed to be independent of any archiving systems and it can be coupled with any existing database or document management system. MILK is using ontology to manage services such as insert and delete terms, unify synonyms and multi-language entries, categorise elements by keywords, compute similarities, support categorisation of new keywords within a domain and compute statistics on term usage.

As the user interacts with the system his profile is updated accordingly. The metadata associated to the documents the user is interacting with are linked to other users' profiles. This way, people, communities and documents involved are kept in association. When a community is created in the system, the user must position this community within a specific node of the ontology tree. Additionally, documents may be associated with terms selected from the ontology. The system is also selecting, grouping, organising and presenting any kind of information related to what the user is doing as well as to the content of his actions. This information is surrounded by contextual information regarding each document or person presented by the system. Personalised services are provided to the user on the basis of their profiles and context of use.

NuggetMine (Goecks and Cosley, 2002) is an intelligent groupware application that facilitates opportunistic sharing of information nuggets among a group. Information nuggets are small self-contained information pieces, such as, an interesting URL, a book title or an article, or the time, date and place of an important event. The system mainly supports three tasks. Firstly, users can quickly submit nuggets while working on their desktop; secondly, captured nuggets are displayed at appropriate times without disturbing the user; and finally, NuggetMine automatically develops content-based and attribute-based associative networks to help users navigate and find nuggets of interest. Once the user submits a nugget, it is added by the system in a repository. Thus, an association is created between this nugget and the existing ones. Following the submission, NuggetMine contacts the original source (server where the URL is stored) and receives meta-information about the nugget, all of which are saved along with the nugget. This meta-information is updated regularly by the system. Consequently, an interested user can view the nugget with the meta-information available along with any other nuggets which are related to the one he has chosen.

OntoShare (Davies and Duke, 2004) is an ontology based knowledge sharing environment which encompasses a community of practice modelled according to the interests of the user, as they stored in any given user profile. OntoShare is a WWW environment and has the capability to extract key words from WWW pages and other sources of information shared by a user. Subsequently, it shares this information with other users in the community whose profiles predict interest regarding the information under consideration. It supports a degree of ontology evolution based on the kinds of information users are sharing and the concepts to which they assign this information. OntoShare uses RDFS to specify the classes included in the ontology as well as their properties. Accordingly, RDF populates this ontology with instances of shared information. The main objective of this system is to provide personalised support to the individual members by retrieving explicit knowledge available in the system's storage space.

When a file is uploaded, the system checks the profiles of all members and sends notification to those users whose profiles match relatively strongly with the keywords of the shared file. Being able to locate specific information, which has just been made available, may facilitate the development of transactive memory as well as the promotion of knowledge sharing in the community. However this support is in some way implicit due to the lack of an explicit representation of a community model in the system. Additionally, the facility "Documents for me", display to the user the most recently stored information which matches the user's profile. At the same time, the system compares the concepts associated with each document to the existing ontology of the community, and proceeds to add or remove concept(s) from the user's profile after consulting the user to this effect. A user's profile can also be manually edited by the user himself/herself so that it is up to date with the user's preferences. Moreover, the user can search for other members in the community who share the same interests as himself/herself. One of the key features of OntoShare is the creation of evolving ontologies. Every time a file is shared, the system suggests to the member who is sharing the files, a set of concepts (from the existing ontology) to which the information could be assigned. The user can accept some of the concepts and/or suggest the addition of other concepts to the system. By adding a user's own conceptualisations to the ontology, the ontology itself is expanding.

AWESOME Dissertation Environment (Lau et al., 2009) is being developed as an institutional demonstrator at the University of Leeds and it is based on MediaWiki[4]. It follows Web 2.0 approaches for collective content creation and information sharing in an informal way. The semantic wiki environment is customised for the sharing of practice in an academic writing community and includes several features. Content based annotations based on pre-defined categories and following user-defined properties are used in order to find information, as well as to help with the community's moderation. Semantic forms for enhancing ease of use and structure are employed so as to ask questions and make content related comments. Dynamic queries are used to give a semantic-enhanced view about the practice and functioning of a community. Tag cloud and folksonomy have been employed to provide a general outlook of the environment and to indicate popular topics. Furthermore, embedding of text, audio, and video enables the sharing of dissertation experiences in a flexible way. Semantics is added via a special markup in the main text of a wiki page. Categories, which correspond to core ontology concepts, are also included. The categories are organised in a hierarchy which is interpreted as an OWL ontology. Properties express binary relationships between one semantic entity (a wiki page or a category) and another data entity or data value. Users have full control over the definitions of properties and the values they assign. Each category or property can be connected to a wiki page along with a description which can help the users to get more information as well as to use them appropriately.

The systems presented above have been selected for the reason that they address, to a certain degree, the concepts presented in Section 2.3. Notwithstanding, none of the above systems is providing support or functionality to community members in relation to all three processes. The following sections will discuss what functionality is provided by the above systems in supporting the three processes selected as having central importance in this PhD.

2.5.2 Support for Transactive Memory

The building of transactive memory is supported, to a certain degree, by all systems. The most common technique used to facilitate the development of TM is *a search facility* which helps users allocate relevant knowledge and people. BSCW (Wolfgang et al., 2004) provides a standard search function through resource titles. On the other hand, MILK (Agostini et al., 2003) prescribes for the search of experts or information in the community based on the information stored in people's profiles and on the metadata associated to resources. However, this approach is prone to inaccuracy:

---

[4] http://www.mediawiki.org/wiki/MediaWiki

metadata is defined by the members who upload the resource and the profiles are solely determined by the users' interactions with the system. These problems are addressed in KSE/Jasper and OntoShare which provide enhanced search facilities based on keyword extraction from entire documents (Merali and Davies, 2001; Davies and Duke, 2004). Moreover, KSE/Jasper and OntoShare enable users to search for other members with similar interests based on dynamically maintained user profiles open for inspection and change by the users. Answer Garden and GIMMe also illustrate the use of natural language processing techniques in providing support to the development of transactive memory (Lindstaedt, 1996; Ackerman, 1998). AWESOME allows semantic searching of documents or words which appear in postings. Answer Garden uses a text retrieval engine to allocate "expert" answers to a user's question, employing simple dialogue to clarify the question. Although identifying expertise can be related to TM, Answer Garden maintains anonymity of user contributions not allowing allocation of community members in possession of such expertise. GIMMe makes use of latent semantic indexing to facilitate the search through a vast repository of email conversations. Moreover, it extracts group categories on the basis of previously visited issues, which are potentially important for TM.

While search relies on the users' pulling for information, *notifications and recommendations* are push techniques. BSCW notifies users every time changes are made to the community space (who uploaded what, who read what etc). This may implicitly promote the development of awareness regarding who knows what. Nonetheless, users may not notice important information because the notifications are not tailored to reflect the user's current interests; as this is done in OntoShare according to a simple content-based filtering mechanism. While recommendations have been found to constitute useful personalisation techniques, their current application in VC focuses solely in supporting the individual.

Semantic-enhanced technologies have also been applied in order to support the development of TM. NuggetMine and MILK use metadata with reference to resources so as to associate newly added pieces of information with old ones (Goecks and Cosley, 2002; Agostini et al., 2003). However, this approach relies only on metadata and does not take into account information about the people who shared/read the resources, which is a crucial factor for the construction of TM. GIMMe and BSCW maintain a hierarchal structure of categories that can facilitate knowledge allocation. Nevertheless, the categories are freely constructed by the users themselves and can become messy. This may hinder resource allocation and expertise finding, thus impeding the TM's development. On the other hand, OntoShare and AWESOME use an ontology of domain categories in order to identify knowledge and similarities between users and resources respectively.

2.5.3 Support for Shared Mental Models

Making members aware of what is happening in the community is considered as important and is supported by the majority of the systems in different ways and in varying degrees. *Visualisation* techniques allow users to become aware of what is happening in the community in general and they have been utilized by two systems for the development of SMM. The development of SMM is promoted in Comtella (Bretzke and Vassileva, 2003) where galaxies visualisations illustrate the convergence of topics. BSCW also uses visualisation techniques to support the development of SMM. Users can explore an information space map which shows each folder and the activities in it, indicated with small rectangles. Another visualisation, presents as towers in a city, the number of papers which are included in a folder. Visualisation techniques form a useful overview of what is happening in the community but appear to be inadequate in providing a deep understanding of the conceptual processes which are taking place within the community.

Semantic – aware techniques have been explored in furthering the development of SMM in Jasper II and MILK. Jasper II supports the creation of shared understanding by capturing the individual perspective in the form of annotations typed in by the users (Merali and Davies, 2001). Similarly, MILK supports contextual awareness in the community based on the meta-information users are typing (Agostini et al., 2003). However, meta-data provided solely by users may be inaccurate, incomplete or even contradictory. A shared ontology is used by MILK which allows users to associate uploaded documents to the terms on the ontology tree. To this effect, users have to agree upon a specific point of view which would be represented in the ontology, even if it may not always be shared by all community members. On the basis of an ontology, AWESOME makes associations with reference to postings in addition to relating documents together. This facilitates the promotion of awareness in relation to other documents which are similar to the one posted by any given member. Furthermore, interesting topics can be bookmarked and revisited later on. Thus, members can easily remind themselves of previously visited postings. The appearance of notifications in the environment on login, are representative of the latest activity that is taking place in the community which also provides the foundations for the building of SMM.

2.5.4 Monitoring Cognitive Centrality

Cognitive centrality is addressed partly in Comtella through a *reward mechanism* whose aim is to encourage participation in online communities. Each member earns points on the basis of the ratings accorded to him or her with reference to the resources uploaded by the said member (Cheng and Vassileva, 2005). Comtella uses *visualisation techniques* to represent cognitive centrality. In a recent version of the system, stars of different size and brightness have been employed in order to indicate which of the members are contributing to the community valuable resources (judged by the ratings). Similarly, in an earlier version of the system, galaxies represented topics of potential interest to the community. The closer to the centre of the galaxy a member is, the more central (judged by the number of papers uploaded) he/she is considered to be (Bretzke and Vassileva, 2003). Quantitative mechanisms are used to calculate cognitive centrality in Comtella. Notably, such mechanisms do not take into account neither the cognitive influence of members nor the relevance of their contribution to the community's context.

2.5.5 Open Issues

To summarise the above discussion, support is provided to partially facilitate the development of TM and SMM and in order to assist the monitoring of CCen. Basic and semantic search techniques are employed by most of the systems to promote TM development in the community. In addition, the engagement of notifications and recommendations in the systems described above provides members with awareness in relation to the community's activity. The most common support is granted to TM. Semantic enhanced technologies have also been used as an alternative approach in supporting the development of TM and SMM in the community. Furthermore, visualisations have been employed so as to encourage awareness of what is happening in the community, promote SMM and monitor CCen. CCen has been granted further support by providing reward mechanisms whereby members get rewarded for being central in the community.

Despite the fact that systems attempt to support TM, SMM and monitoring CCen, the absence of a versatile community model in conjunction with the adaptation to the individual rather than the community, constitute the main obstacles to their holistic success. All the systems discussed above provide support to one or two of the processes identified. None of these systems is supporting all the three processes and the support they are providing is not based on a community model that represents the community. This research purports to develop a framework for a holistic community support based on a community model to be subsequently used to support the building of TM, SMM, and monitoring of CCen. The computational framework consists of two major parts. The first part deals with the development of a community model, which represents the whole community and focuses on the processes discussed above. The second part considers the intelligent support offered in order to improve the functioning of the community. Chapter 3, discusses in detail the modelling approach and the support given to VCs within the context of in this PhD. Figure 3.1 illustrates the architecture of our framework which follows the general architecture of user-adaptive systems defined in (Jameson, 2003).

2.6 Summary

VCs represents a relatively new approach which brings together people sharing common interests who need to collaborate with others or who merely enjoy sharing knowledge in itself. Closely-knit VCs are very popular and exist in most organisations and universities around the world. Although an effective way for sharing knowledge, support is necessitated in order to overcome the problems acting as a barrier to this process. A number of processes which have an impact in the functioning of the community are available. Nevertheless, not all of them have an impact on knowledge sharing. Processes from the field of organisational psychology have been identified as being important for the functioning of VCs, namely TM, SMM and CCen.

The aim of this project is to provide holistic intelligent support to closely-knit VCs based on TM, SMM and CCen. Consequently, knowledge sharing systems have been reviewed in order to determine what support has been provided by the existing approaches as regards the development of TM, the establishment of SMM and the monitoring of CCen. The results of this review demonstrate that none of the systems reviewed supports all the processes in a VC as a whole. To this effect, this PhD comes to fill in a gap and develops a computational framework which consists of a community model extracted from the tracking data of real VCs, which have been used to provide intelligent support to VC on the basis of the three processes identified. The next chapter will elaborate on the defined approach.

Chapter 3

Modelling Virtual Communities

3.1 Introduction

The main objective of our work is to find out how intelligent techniques can be used to provide support for knowledge sharing in VC. Chapter 2 discussed processes considered as important for the functioning and knowledge sharing in VC, and showed that these processes were not supported in a holistic way in existing systems. In order to provide a holistic intelligent support to facilitate the knowledge sharing in VCs, problematic cases need to be identified and used as input in generating personalised support. This requires a model of the community. Thus, the first step in our framework is to incorporate appropriate techniques to extract a community model, which can be analysed to identify problems within the VC that can be used in defining what intelligent support may be needed and when.

The aim of this chapter is to present the overall framework for intelligent support in a knowledge sharing VC. We will provide a brief overview of existing community modelling approaches and differentiate the approach followed in this PhD. The next section will describe related work on modelling communities in user modelling and social network research areas. We will then present in Section 3.3 the proposed computational framework that is followed for extracting, maintaining and applying the community model, which is used in following chapters for identifying problems in knowledge sharing that can be addressed in a VC. The components of the community model will be defined and briefly described. To inform the design and to validate the framework proposed in this PhD thesis, we will use archival data from an existing VC operated in the BSCW system. Section 3.4 provides a description of the BSCW VC that is used in the studies described in Chapter 4, Chapter 5 and Chapter 6.

3.2 Community Modelling Approaches

Modelling virtual communities has recently become very popular in different research areas. In user modelling, modelling group of members provides the grounds for generating group recommendations (Masthoff, 2004). In social networks, community modelling aids the discovery of relationships between people and among communities (Lin et al., 2008). Modelling user interests, relationships, or a group/community, in general, can provide useful insights to inform what support can be offered to community members. **Interests** can be derived based on items users are sharing, tags/keywords associated with a user, a description provided by the user on his interests, or keywords extracted based on discussions. Similarly, **relationships** between members are extracted based on user activities in the community (or the group), discussions members engage in, sharing habits within the community/group, or the spread of expertise and knowledge. For the purpose of this PhD approaches from both user modelling and social network areas are considered as important and discussed in this section.

A fairly simple and easy to implement **community model** is presented in (Cheng and Vassileva, 2006). It is based on a list of topics derived based on the resources members of the VC are sharing. In addition, a reward factor is considered to measure the relevance of each contributed resource to the current topic that the VC is working on. The **individual user model** consists of the reputation measure of a member in the VC and the data describing the user's current membership level in order to calculate the reward factor (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 but there is a crucial difference. Users' **interests** are modelled in (Bretzke and Vassileva, 2003) based on how frequently and how recently users have searched for a specific area from the ACM taxonomy, and user **relationships** are derived based on any successful download or service that took place between two users. In contrast, our approach employs the metadata of the resources shared in the community along with the ontology and derives a semantically relevant list of interests for every user.

A different approach is followed by Tian et.al. (2001) where the **community model** represents the **interaction activities** that happen in the VC (Tian et al., 2001). All interactions are associated to 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 the social interaction activities of users and allied with the lexicon developed. The approach presented in this PhD 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 the ontology. Consequently, the data used to extract the interest similarity relation ($InterestSim$) are semantically enriched. Moreover, the community model is semantically richer, since it contains more than the interactions between community members having also the personal hierarchies created by each member, relationship model, cognitive centrality and the ontology which represents the community domain.

User interests have been extensively studied. For example, (Davies et al., 2003) present an approach where user **interests** are extracted as keywords from the user profiles and other web content shared by a user with the community. 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) is extracting interests based on the tags users are creating for items posted online. **Relationships/associations** are derived between users based on their tags even if they are not directly connected by a social graph. The latter approach is similar to the one followed in this project - both approaches consider that members can be connected in interest similarity even if they have not read any resources uploaded by each other.

Furthermore, **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** on a discussion structured community based on the posting/replying threads. According to the discussion topics a member of the community is contributing to, his interests and expertise are extracted, based on which user 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 recorder on 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 previous publications and two people are considered related if they have publications in the same research area. This PhD adds to the above approaches. Our approach does not aim at identifying expertise alone, but also derives a person’s influence in the VC based on the relationships he/she has developed with others, which benefits the VC as a whole.

Recent 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** is holding the conceptual understanding of a user and a **graph based 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 on a graph. Kay et al. (2006) uses the notion of interaction network to represent **relationships** between users in a learning community. Two members are related if they have modified the same resource and hence they appear connected in the interaction graph. Falkowski et al. (2007) consider the exchange of messages as interaction between two users, represented in a graph. A **relationship** exist between two users if they have engaged in 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 in this PhD to graph community modelling is the considering 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.

Section 3.3 will outline the computational framework for providing support to knowledge sharing VC and will present the proposed community modelling approach.

3.3 Computational Framework to Support Knowledge Sharing in VC

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 closely-knit VC (Chapter 1). Based on this, and following the general architecture of user-adaptive systems presented in (Jameson, 2003), a framework is outlined for providing intelligent support to VCs which includes two parts (see Figure 3.1):

- acquisition of a community model that represents the whole community and focuses on aspects related to TM, SMM, and CCen;
- application of the community model to offer intelligent support and improve the functioning of the community.

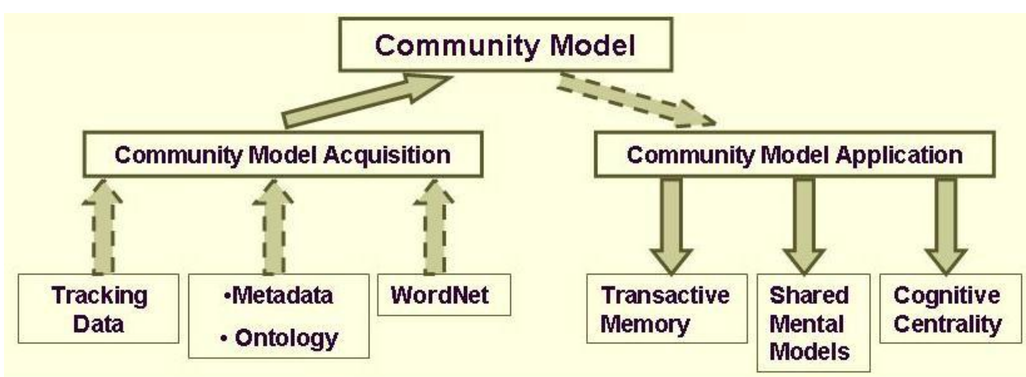


Figure 3.1 Architecture of our computational framework to provide holistic support to closely-knit VC.

The purpose of the following sections is to provide an overview of, the mechanism for deriving the community model (Section 3.3.1), the components of the proposed community model (Section 3.3.2) and how this model can be applied for providing support to VCs (Section 3.3.3).

3.3.1 Community Model Acquisition

Extracting a community model involves (i) identifying the input data, (ii) formalising the input data and (iii) defining the community model components. Our framework uses as input for a community model **tracking data** from the knowledge management system the community uses. To identify the key input attributes, we examined tracking data extracted from a knowledge sharing system, BSCW (see Section 3.4), that supports collaboration over the web and offers basic functionality for resource sharing. The tracking data extracted (including information about members, reading/uploading resources, folder creation/deletion, rating resources) is generic and common to most knowledge sharing systems. Thus, the approach described in this PhD can be contemplated as generic, as the algorithms developed for extracting the community model can be adjusted to suit input data from different social systems.

In addition to the tracking data, we consider **metadata** which can be extracted from the resources people are sharing. Dublin Core metadata standard has been followed, including information about resource title, publication place, authors, keywords, and date of publication ($rMetadata : \langle rTitle, rAuthor, rSource, rKeywords, rDatePublish \rangle$). 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 CM acquisition algorithms, an **ontology** Ω representing the main topics of the community is employed. The ontology is used as an input to the CM acquisition algorithms to represent the *community domain*. The domain of the VC is represented by the vocabulary relevant to the main area of interest of the community under study. Section 0 gives a detailed description of the ontology and how it is used in this PhD. Furthermore, **WordNet** is used as a source for measuring the semantic importance of a resource to the VC and the semantic similarity between members and resources.

A detailed description and formalisation of the input data is presented in Section 4.2.

3.3.2 Community Model

The community model represents the whole community, including members $\{M\}$, relationships $\langle RelationshipType \rangle$, topics of interest L and the cognitively central members $CCen$. Hence, it consists of individual user models IUM, a relationships model RM, the community domain (ontology) Ω, lists of popular and peripheral topics L_{Pop}, L_{Per} respectively, and a list of cognitively central members $CCen$. The main components of the community model are illustrated in Figure 3.2.

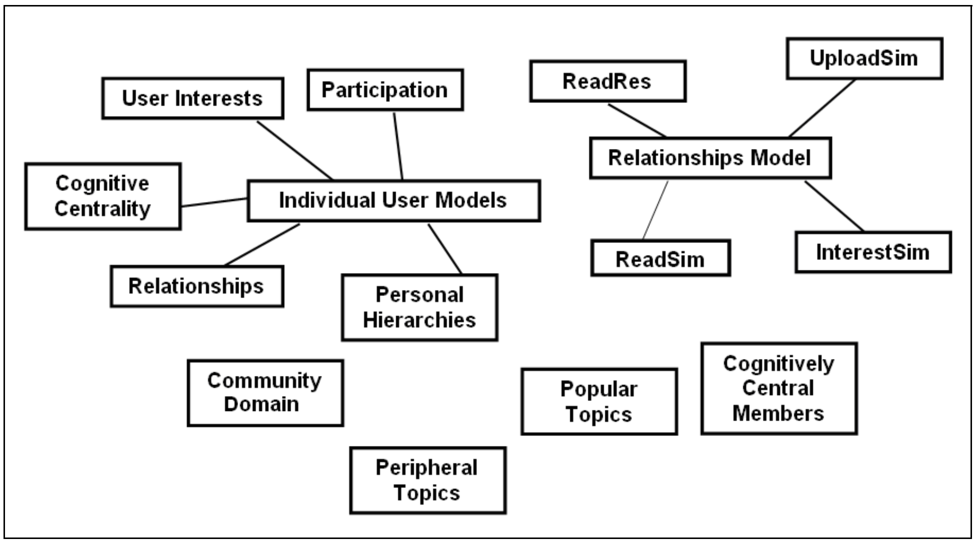


Figure 3.2 The components of the community model as defined in this PhD

Individual User Models maintained for every member of the community and include user interests I, type of participation (uploading $uRate$, downloading $dRate$), how cognitively central that member is $CCen$, what relationships $\langle RelationshipType \rangle$ he has with other members in the community, and personal hierarchies of folders F and resources R created by that member.

Community Domain is important in order to judge the cognitive centrality and influence of community members. Domain as a term has been used to serve different purposes. With regard to VC, domain is most often considered as the general area of interest for the whole community (Zhdanova, 2005). This can be represented as a list of topics (Bretzke and Vassileva, 2003) or a more complex structure linking topics to an ontology. In this PhD an ontology Ω is used to represent the community domain and is further discussed in section 0.

Popular and Peripheral Topics are identified according to the keywords of the resources uploaded. The popular topics list L_{Pop} holds the most popular topics within the community. The peripheral topics list L_{Per} holds the least popular topics of the community. Having these two lists will help us in better exploiting the cognitively central members of the community along with defining the cognitive centrality of a particular member.

Cognitively Central Members $CCen$ are influential to the rest of the community. When problems with knowledge sharing are detected, these members can be prompted to take corresponding actions. The list of the most cognitively central members is used as an input in the algorithms for generating intelligent support.

The above components of the community model are discussed in more detail in section 4.3.

3.3.3 Community Model Application

Application of the community model provides us with an insight of what support is needed in the VC and when. According to the information extracted in the CM, patterns (static and dynamic), of problematic cases can be extracted and used as an input for providing intelligent notifications to VC members. We illustrated the application of the community model with three case studies. Firstly, we conducted a case study (Chapter 4) to examine the application of the community modelling algorithms to analyse the log data in an existing community. This helped us identify problems that could have been spotted earlier and addressed properly to help the community sustain. A CM extracted for the SW VC described below (Section 3.4) was used to identify problematic cases *manually* with the help of the visualisation tool NetDraw[5]. These patterns pointed out possible problems that may exist in a VC, were used as a guideline in developing graph algorithms that automatically pick problems with community knowledge sharing and relationship formation.

The second case study extracted community models of the same VC over three months (Chapter 5). Automatic pattern detection algorithms were developed and applied to the CM. This revealed cases that could be used as an input in providing intelligent support to community members. This study revealed the need to identify changes occurring over time in the VC.

Consequently, a third study was conducted where algorithms for detecting community evolution were implemented and applied to the data used in the second study (Chapter 6). This detected changes in knowledge sharing and relationships of members in the VC. The results were used in defining support that could have been provided to community members.

[5] NetDraw is a free program written by Steve Borgatti of Analytic Technologies, for visualizing social network data. The program reads UCINET system files, UCINET DL files, Pajek files, and VNA format, which allows saving network and attributes data together, along with layout information. NetDraw can be downloaded from http://www.analytictech.com/netdraw/netdraw.htm.

The fourth study conducted in this thesis looked at the generation of intelligent support based on CM application, which is defined in Chapter 7. Support is considered in the form of personalised notification messages sent to VC members to inform them of connections they have with other members and resources or people that might be of their interest. This has been applied to a community at its starting stage, which allowed us to examine benefits of notifications for community forming and bonding.

The next section will present the VC from which archival tracking data was used in the first three case studies.

3.4 A Semantic Web Virtual Community in BSCW

To validate the algorithms developed in Chapter 4, Chapter 5 and Chapter 6 , they were employed to extract patterns from archival tracking data of a real community which both authors belonged to. The VC in all three studies included 34 members (researchers and doctoral students) from two research groups working on similar research topics in the area of Semantic Web, sharing documents and research papers (referred here as resources) with the BSCW[6] system that provides general support for collaboration over the web (Applet, 1999). BSCW allows people to create folders in a hierarchical manner, upload, download papers, rate the papers, provide tags and descriptions, and engage in discussions. The system provides logs of actions performed by the community members e.g. who and when has joined/left the community, who and when has uploaded/downloaded papers, any descriptions added or changed about a resource or a folder, any ratings added changed and if a folder or resource deleted from the community space. The system requires a user name and password to enter the VC a user is registered to. Once a member is logged in to the VC space he can see a description of the VC (Figure 3.3) and two awareness icons: footprint icon – provides general information on who has joined/left the community, folders created dropped or changed; and glasses icon – provides details on who as read a resource and when. The awareness icons are also provided for each folder individually and describe the activity in that specific folder (Figure 3.4).

---

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

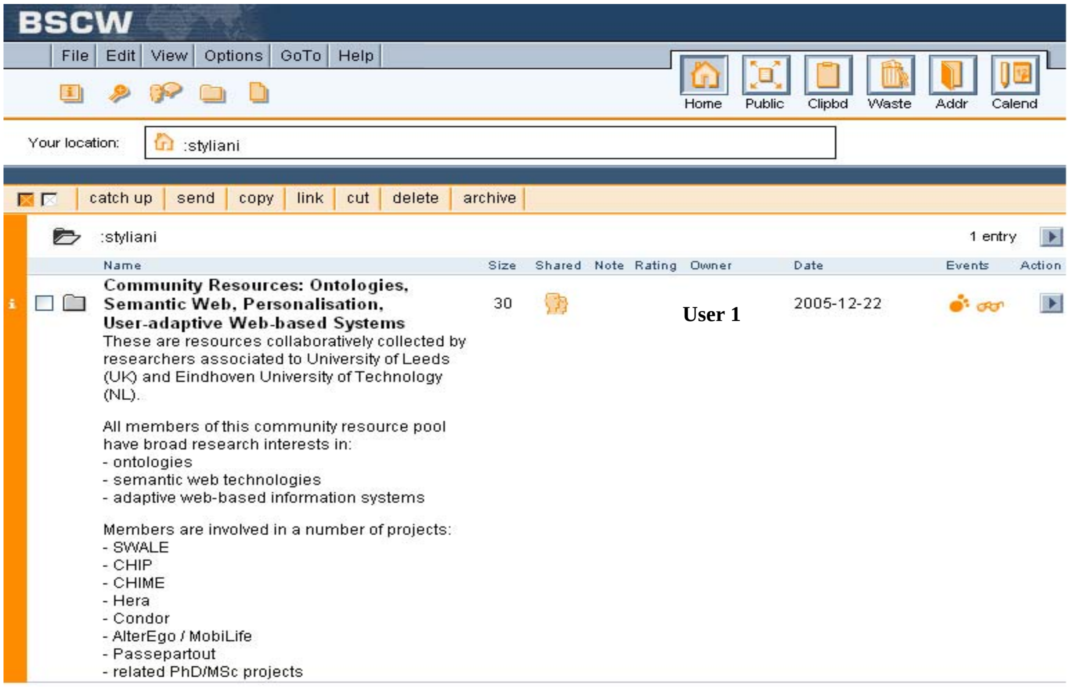
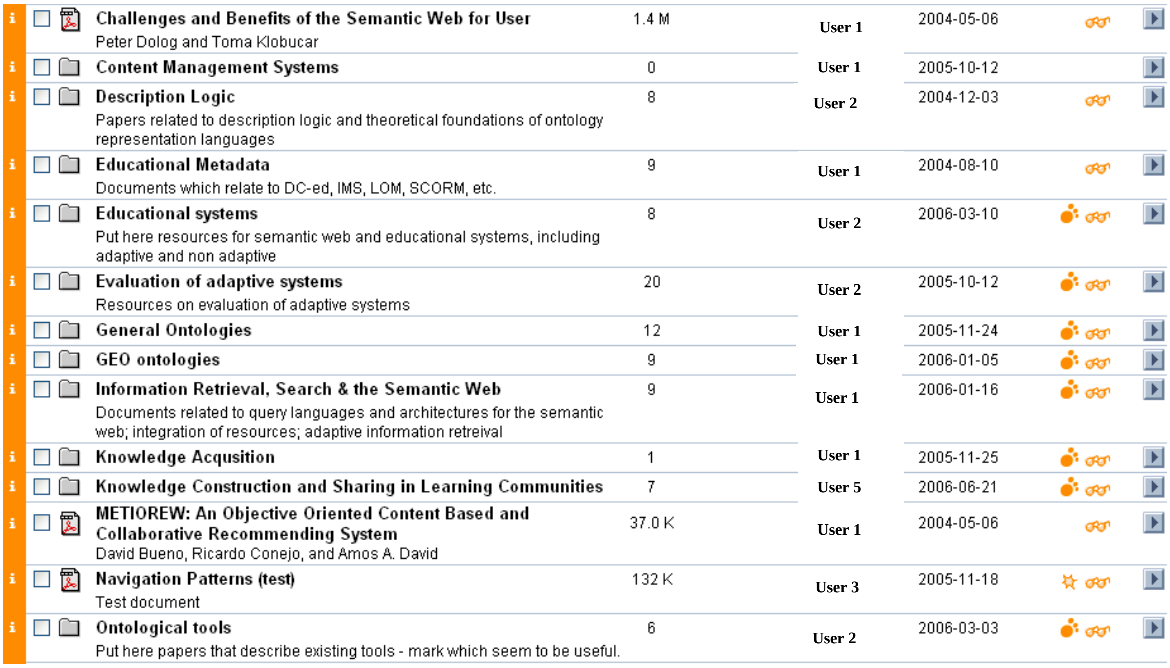


Figure 3.3 The VC environment after a member logs in to BSCW. User names have been substituted with codes to comply with data protection regulations

In the SW VC, people were working on large projects and were using BSCW to create folders and upload and download resources (Figure 3.4) from the folders created as part of their every day research practice. It is important to mention here that this VC has not been built by the members for the purpose of forming ties. Thus, it is not a VC focused on social relationships. Instead it reflects VC members' shared interests. The groups were based in two European countries, thus no physical interaction was taking place among members. Some members knew each other but many had never met. Two members were leading researchers in the field and the others were mostly research students and staff. Log data was collected over 15 months using the BSCW activity tracking features. Data was collected about members using only the basic functionality of the system, such as uploading/downloading and naming a resource, which is provided in any virtual community for knowledge sharing.



Challenges and Benefits of the Semantic Web for User Peter Dolog and Toma Klobucar	1.4 M	User 1	2004-05-06
Content Management Systems	0	User 1	2005-10-12
Description Logic Papers related to description logic and theoretical foundations of ontology representation languages	8	User 2	2004-12-03
Educational Metadata Documents which relate to DC-ed, IMS, LOM, SCORM, etc.	9	User 1	2004-08-10
Educational systems Put here resources for semantic web and educational systems, including adaptive and non adaptive	8	User 2	2006-03-10
Evaluation of adaptive systems Resources on evaluation of adaptive systems	20	User 2	2005-10-12
General Ontologies	12	User 1	2005-11-24
GEO ontologies	9	User 1	2006-01-05
Information Retrieval, Search & the Semantic Web Documents related to query languages and architectures for the semantic web; integration of resources; adaptive information retreival	9	User 1	2006-01-16
Knowledge Acqusition	1	User 1	2005-11-25
Knowledge Construction and Sharing in Learning Communities	7	User 5	2006-06-21
METIOREW: An Objective Oriented Content Based and Collaborative Recommending System David Bueno, Ricardo Conejo, and Amos A. David	37.0 K	User 1	2004-05-06
Navigation Patterns (test) Test document	132 K	User 3	2005-11-18
Ontological tools Put here papers that describe existing tools - mark which seem to be useful.	6	User 2	2006-03-03

Figure 3.4 Folder and resource organisation in BSCW. User names have been substituted with codes to comply with data protection regulations

The CM algorithms extracted relationships between all 1122 pairs of members. The activities monitored included uploading and downloading of resources, 244 resources in total. Four members were *only uploading* while thirteen were *only downloading*. Eight members were *isolates* and never uploaded or downloaded resources. There was a gradual decline in the uploading and downloading of resources in the observed period. During the beginning of the monitored period (Month1 – Month8) members were uploading and downloading papers. After that the activities minimised for all members, and during the last few months of the monitored period (Month13 – Month15) there was no uploading and very little downloading. The community was gradually declining and has stopped its activity at the moment. The use of data from a community that did not sustain ensures that problems did exist, so the researcher could see if the algorithms could spot those problems. Note that there were no interventions or any experimental conditions while the community was active.

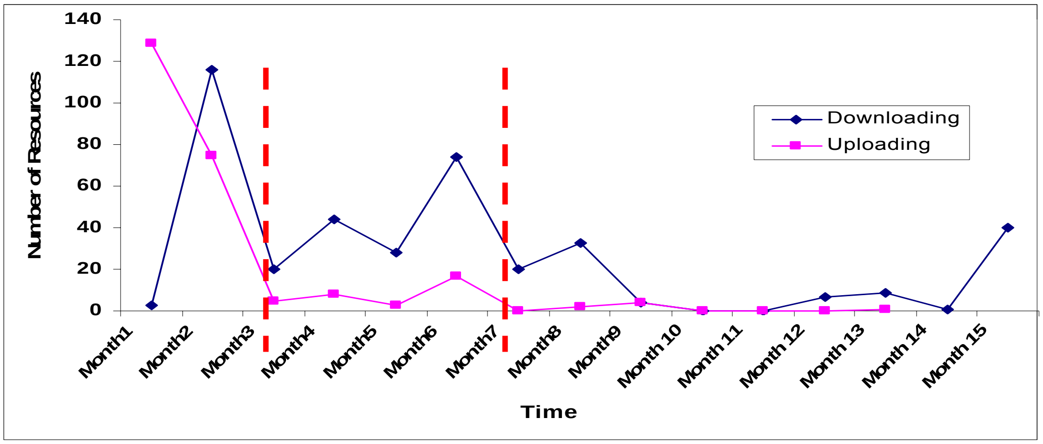


Figure 3.5 Uploading (blue) and downloading (pink) activity in the BSCW virtual community.The red dotted lines show the drop of activity in two different occasions

3.5 Summary

User models are employed in adaptive and intelligent systems to inform a more accurate adaptive functionality to a system targeting the individual needs of a specific user. In our research community modelling is engaged in order to capture the properties of the VC in a CM, analyse and apply the CM to provide intelligent support to VC members individually but aimed to support the functioning of the VC as a whole. Information relevant to knowledge sharing and to TM, SMM and CCen stored in the CM, can be used to generate appropriate support to targeted individuals but aimed at improving the knowledge sharing in the VC.

The CM has been outlined in this chapter based on the information that could be extracted from the tracking data of a real VC. Related work shows that there is no agreement on what should be in a CM, as long as it serves the purpose of the intended support or the purpose of the research at hand. For example, as shown in Section 3.2, different CM have been built in every work but served the purpose of the relevant project equally well. Moreover, in some cases a systematic approach has followed to extract a CM and implement the required support whether in some other cases a more ad-hoc approach has been followed. In our work a computational framework has been built to aid at a step-by-step approach the extracting, maintaining and applying of a CM.

This chapter outlined the computational framework followed in this PhD, and introduced the main components of the CM. Following the framework, tracking data from a VC has been extracted and enriched with semantics (ontology and metadata) in order for the CM to represent relevant semantic information and relationships. We have kept the data used from the SW VC as general as possible in order for the approach to be generic and applicable to VCs operating with different knowledge sharing systems.

In order to validate the CM extracted, real data from a VC has been employed and the results are presented in the three subsequent chapters. Chapter 4 will discuss the components and present the algorithms developed in deriving the CM in more detail. A case study performed in validating the CM will also be discussed.

Chapter 4

Definition of the Community Model

4.1 Introduction

Defining the community model is the first step towards the implementation of the framework presented in Chapter 3. The first step is to formalise input data - tracking data from a knowledge sharing application, an ontology and an existing semantic similarity algorithm will be used. This is described in Section 4.2. The definition of the VC model consists of a *relationships model*, *individual user models*, lists of *popular and peripheral topics* and a list of *cognitively central members* of the VC. The model is presented in Section 4.3. The model defined in this study is a general model that can be adjusted according to the VC under study. This may include defining more relationships by following the approach presented in this chapter.

A study with the SW VC, which was introduced in Section 3.4, is conducted to validate the community modelling algorithms. Section 4.4 presents the study and gives examples of knowledge sharing patterns identified. Visualisations are provided using NetDraw in order to better illustrate the cases discovered. We discuss how the discovered patterns relate to *transactive memory, shared mental models* and *cognitive centrality* in the VC.

4.2 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. In addition, semantic features, such as metadata and an existing ontology are exploited.

4.2.1 Tracking Data Formalisation

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 H_F. The community environment E is defined as $E{:}\langle M,R,F,H_F \rangle$.

E is changing over time as a result of actions performed by community members, including:

`join_community` – a member is registering to the community;

`leave_community` – a member is leaving the community;

`create_folder` – a new folder is created by a member;

`upload_resource` – a new resource is uploaded in the environment;

`rate_resource` – a member is assessing a resource;

`download_resource` – a resource is downloaded from the environment;

`add_resource_description` – a new description is added to a resource;

`add_folder_description` – a new description is added to a folder.

The above 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.

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 $m : \langle mName, mEmail, mDateJoin \rangle$. 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 to this resource.

A user creates $rCreatedData: \langle rFolder, rName, rDescription, rRating, rCreator, rDate, rAssessor, rReader \rangle$ 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), $rDecription$ denotes a set of resource descriptions $rDescription : \{\langle rd_1, m_1 \rangle, \langle rd_2, m_2 \rangle, \ldots, \langle rd_n, m_n \rangle\}$, where $\langle rd_q, m_a \rangle$ is the description q given by member a; $rRating$ is a number which is the average rating given to that resource by community members, $rAssessor : \{\langle ra_1, m_1 \rangle, \langle ra_2, m_2 \rangle, \ldots, \langle ra_n, m_n \rangle\}$ where $\langle ra_q, m_a \rangle$ represents the assessment q given by member a; $rCreator$ is a member $(rCreator \in M)$, who is the creator of the specific resource; $rDate$ is the date the resource was uploaded; $rReader$ records the access to the resource by community members, $rReaders: \{\langle m_1, r_1 \rangle, \langle m_2, r_1 \rangle, \ldots, \langle m_n, r_1 \rangle\}$ where $\langle m_a, r_q \rangle$ indicates that member a has read resource q.

rMetadata represents formal metadata following the Dublin Core[7] schema, which is the basic and most conventional standard for online resources. The following elements have been selected from the Dublin Core metadata schema $rMetadata : \langle rTitle,\ rAuthor,\ rSource,\ rKeywords,\ rDatePublish \rangle$, $rKeywords : \langle k_1, k_2, \ldots, k_n \rangle$ is used for comparing resources, as described in Section 4.3.1 below. *rMetadata* have been either manually extracted from the uploaded resources or provided by the *rCreator* when he uploaded a resource. The metadata provides the first semantic layer used for community modelling. A summary of member and resource tracking data formalisation is presented in Table 4.1 below.

Table 4.1 Summary of member and resource tracking data formalisation

Description	Formalisation
Data about a member m	$m : \langle mName, mEmail, mDateJoin \rangle$
A member who created a resource	$rCreator \in M$
Data about a resource r created by the member who uploaded the resource (*rCreator*)	*rCreatedData: <rFolder, rName, rDescription, rRating, rCreator, rDate, rAssessor, rReader>*
Resource elements from Dublin Core Metadata Schema	$rMetadata :< rTitle,\ rAuthor,\ rSource,\ rKeywords,\ rDatePublish >$
Keywords of a resource	$rKeywords : \langle k_1, k_2, \ldots, k_n \rangle$
Description of a resource r	$rDescription : \{\langle rd_1, m_1 \rangle, \langle rd_2, m_2 \rangle, \ldots, \langle rd_n, m_n \rangle\}$
A member m who assessed a resource	$rAssessor : \{\langle ra_1, m_1 \rangle, \langle ra_2, m_2 \rangle, \ldots, \langle ra_n, m_n \rangle\}$
A member m who has read resources	$rReaders: \{\langle m_1, r_1 \rangle, \langle m_2, r_1 \rangle, \ldots, \langle m_n, r_1 \rangle\}$

In addition, the next section will describe the use of an ontology for semantic enhancement of community tracking data.

4.2.2 Use of an Ontology

There are different definitions and interpretations of the term “ontology” (Noy and McGuinness, 2001). This PhD follows the classical definition by Gruber (1993) that an ontology is a description of the concepts and relationships that can exist for an agent or a community of agents. For the

[7] The Dublin Core Metadata Initiative, is an open organization engaged in the development of interoperable metadata standards that support a broad range of purposes and business models: http://dublincore.org/

purposes of this PhD an ontology Ω is a formal explicit description of a vocabulary in a domain of discourse where entities are associated by definitions (classes, properties and restrictions)(Gruber, 1993; Noy and McGuinness, 2001). The ontology is considered here as the community domain and represents the domain in which the knowledge sharing community operates.

The vocabulary composes the classes of the ontology and represents the topics of interest of the VC in use. The classes are structured in a tree like hierarchy and have a sub-class, super-class relation. The VC environment E is dynamically changing according to members' actions. The ontology Ω can also change according to the subject of focus of the community but the algorithms consider the current snapshot of the ontology and do not exploit the ontology evolution. For example in the case of the SW VC, which is used in the studies in sections 4.4, 5.5 and 6.4, an ontology that represents the Semantic Web domain has been employed. In this study the ontology is passed as an input to the algorithms for extracting the CM in order to semantically empower those and extract a more accurate CM (see section 4.3). A sample ontology Ω (Appendix A), has been created to serve the purpose of the VC considered in this research. The ontology has been build using concepts extracted from the folder hierarchy of the BSCW SW community space. The hierarchy modified accordingly to represent a logical relation (subClassOf), between concepts. All concepts of the ontology were transformed into nouns so they can be used as a direct input in the WordNet similarity measure algorithm (more details in Section 4.2.3). This ontology consists of 159 classes all connected by the subClassOf relations. A sample of the ontology used is represented in Figure 4.1. For example the concept $Ontology$ is represented in the ontology indicating that $Concept$ is a sub-class of the class $Ontology$ and is a super-class of the class $Concept\_Modelling$.

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 defined using appropriate WordNet similarity measure, e.g. the current implementation of the community modelling algorithm uses the similarity measure library described in (Seco et al., 2004) (see Section 4.2.3). The second way the ontology is used is to define the similarity between VC members based on the resources they are reading, uploading and their interests (Section 4.3.1). Having the list $rKeywords$ for a member, we do the following:

(i) We first check that a keyword exists in the ontology: $k \in \Omega$,

(ii) If yes, we 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 k for the member.

(iii) If $k \notin \Omega$ we perform a similarity check $Sim(k, \Omega)$ by using a WordNet similarity algorithm (e.g. in the implementation we used the algorithm presented in (Seco et al., 2004)[8]),in order to get the most similar concept C_{Sim} to k from the ontology. When a concept is selected, we repeat steps (ii) and (iii) above.

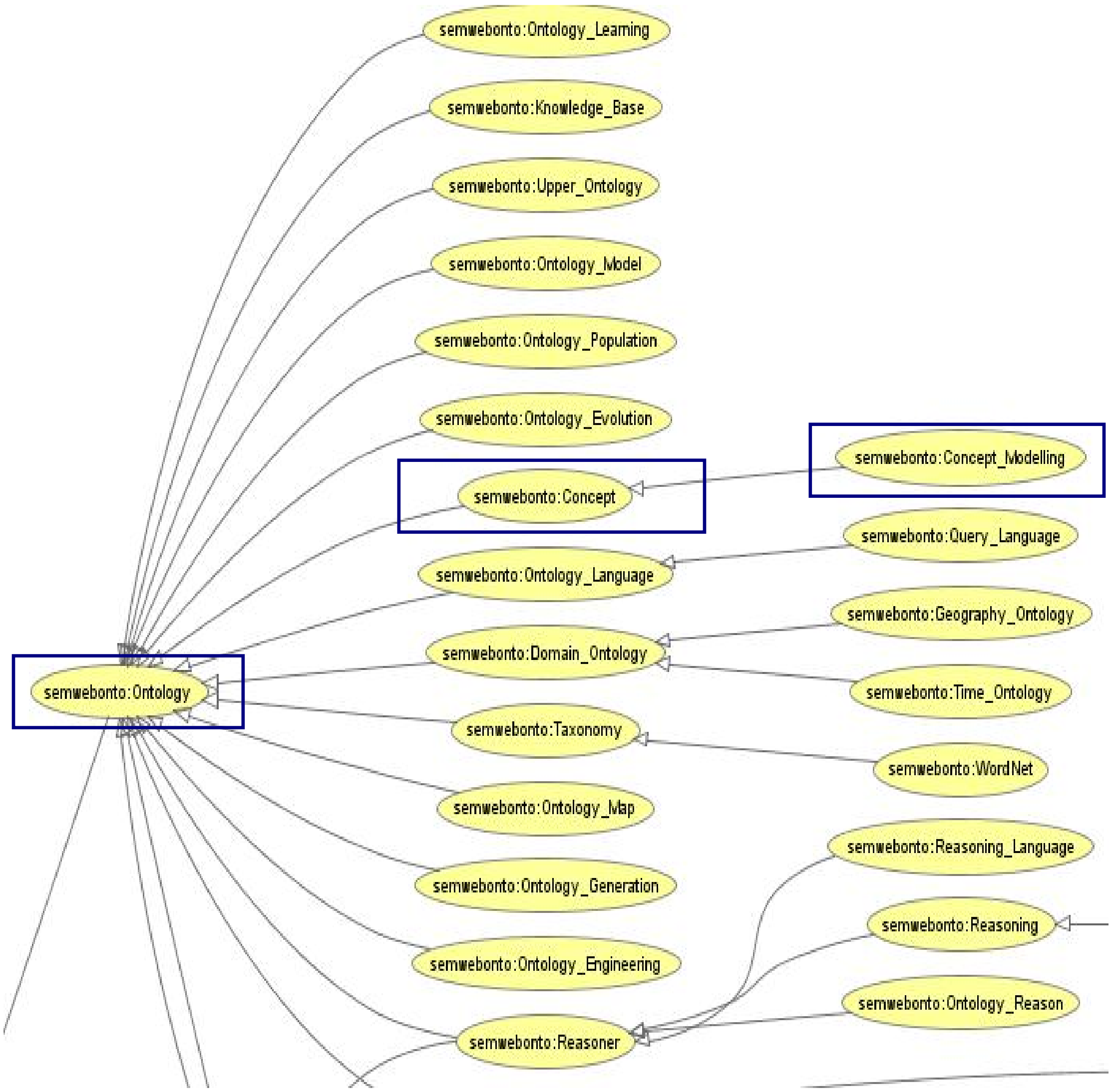


Figure 4.1 Example from the ontology developed

---

[8] Other possible Wordnet based similarity APIs can be used, e.g. Ted Pedersen's library: http://wn-similarity.sourceforge.net/ or the API developed in the RESULT project: http://nlp.shef.ac.uk/result/software.html

This is illustrated with the following example. Assume that $aKeywords$ is the list of keywords extracted for member a, and $aKeywords:\{semantic\ web, knowledge\ sharing, context, collaboration\}$. Assume that $k = collaboration$ and $k \in \Omega$. 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. This is done using an appropriate ontology reasoner, e.g. in the implementation we used the Jena reasoner[9] with Java API. A flowchart represents the algorithm used is presented in Figure 4.2. In the above 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.

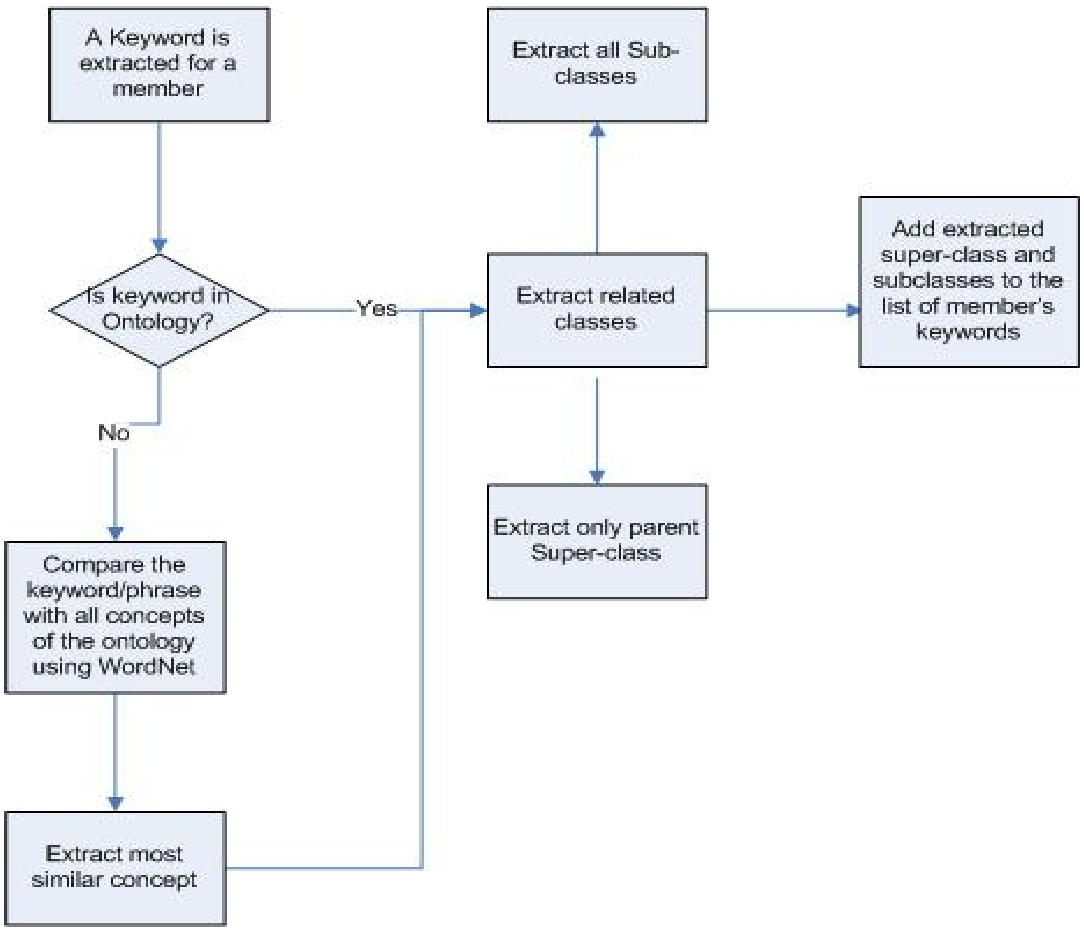


Figure 4.2 Flowchart representing the extraction of super and sub classes from the ontology

---

[9] http://jena.sourceforge.net/inference/

4.2.3 Extended WordNet Semantic Similarity Measure

Semantic similarity relates to measuring the similarity between concepts which are not lexicographically similar but have similar meaning (Varelas et al., 2005). WordNet is an on-line lexical reference system developed at Princeton University (Miller, 1995) that attempts to model the lexical knowledge of a native speaker of English. It can be seen as an ontology for natural language terms and contains nouns, verbs, adjectives and adverbs grouped into synonym sets (synsets). 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 and evaluating semantic similarity algorithms (Liu et al., 2004; Seco et al., 2004; Tagalakis and Keane, 2005; Varelas et al., 2005).

The algorithms in section 4.3 utilise a mechanism for measuring similarity between two lists of terms. If L_1 and L_2 are two lists of terms, we define a similarity procedure $Sim(L_1, L_2)$ 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's 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 (formula 4-1).

$$sim(c1, c2) = 1 - \frac{ic_{wn}(c1) + ic_{wn}(c2) - 2 * sim_{res}(c1, c2)}{2} \qquad \text{4-1}$$

In formula 4-1 $c1$ and $c2$ are the two concepts compared, $ic_{wn}(c1)$ and $ic_{wn}(c2)$ are the information content values expressed as functions of the hyponyms each of them has, and sim_{res} corresponds to Resnik's (1995) similarity function that accommodates the information content values (Resnik, 1995). The formula is then linearly normalised to constrain the output to values between 0 and 1.

The Wordnet similarity algorithm does not support compound similarity (e.g. "knowledge management" and "knowledge capture"), thus a modification of this algorithm was necessary for the purpose of this project. Following the work of Tagalakis and Keane (2005), formula 4-1 is 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_1 of cc_1 in an array. The same is done for every term of cc_2, c_2 with all the terms of cc_1 and the highest value returned for each term c_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 appear in cc_1 and cc_2, $i+j$. Below is the modified version of the formula 4-1:

$$sim'(cc_1,cc_2)=\frac{\max\left(sim(c_1^1,\{c_2^1,c_2^2,...,c_2^j\})\right)+...+\max\left(sim(c_1^i,\{c_2^1,c_2^2,...,c_2^j\})\right)+}{i+j}$$
$$\frac{+\max\left(sim(c_2^1,\{c_1^1,c_1^2,...,c_1^i\})\right)+...+\max\left(sim(c_2^j,\{c_1^1,c_1^2,...,c_1^i\})\right)}{i+j} \quad \text{4-2}$$

Formula 4-2 is used in the algorithms presented in the next section to extract the VC model.

4.3 Community Modelling Mechanism

This section will outline algorithms to generate a model of a VC, comprising a Relationships Model, Individual User Models, Lists of Popular and Peripheral topics of the VC, and a List of CCen members. As input for the community modelling algorithms we consider the *tracking data* and *metadata* (described in Section 4.2.1), the *community domain / ontology* Ω (outlined in Section 0), and the modified *WordNet algorithm* (presented in Section 4.3.2)

4.3.1 Community Relationships Model

We consider the four types of semantic relationships between users: $ReadRes$ 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 details on graph definitions will be given in Chapter 5).

ReadRes Relationship

$ReadRes(a,b)$ relationship indicates that resources uploaded by member b are read by member a; the relationship strength corresponds to the relevance of the resources to the community domain. In other words, the closeness of the members depends not only on the quantity of the common resources they read but also, and more importantly, on the importance of these resources for the VC. If two members read many materials that are not very connected to the community, their relationship will be indicated but it's strength will be low as the connection is not valuable for the community. On the other hand, users may read fewer common resources but they can be highly relevant to the community domain; such relationship will have a higher value for the VC.

$ReadRes$ can be used to identify complementary knowledge among people, and this helps to improve the community's *transactive memory* (Section 2.3).

Consider a resource r_i uploaded by b and read by a. We will denote its keywords 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 (Seco et al., 2004) algorithm (section 4.2.3).

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.:

$$ReadRes(a,b) = \sum_{i=1}^{Z_r^{a \leftarrow b}} V_{r_i} \qquad 4\text{-}3$$

ReadSim and UploadSim Relationships

$ReadSim(a,b)$ indicates that members a and b have read semantically similar resources, while $UploadSim(a,b)$ indicates that a and b have uploaded similar resources. These relationships can be important for discovering similarities that members may not know of. 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* (Section 2.3).

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 the additional keywords extracted from Ω, as described in section 0. 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 4.2.3. Consequently, $ReadSim(a,b)$, is calculated as:

$$ReadSim(a,b) = Sim(aKeywords, bKeywords) \qquad 4\text{-}4$$

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

$$UploadSim(a,b) = Sim(aKeywords, bKeywords) \qquad 4\text{-}5$$

InterestSim Relationship

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

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, a's personal list of interests I_a is extracted, following the algorithm presented below (Section 4.3.2). The personal list of interests of a member is further extended by extracting the neighbourhood concepts from the ontology Ω, in the same way this was done in the $ReadSim(a,b)$ algorithm (section 0). The extended lists of personal interests of members a and b - I_a and I_b - are compared using the extended Wordnet similarity algorithm (Section 4.2.3) to calculate the interest similarity between a and b:

$$InterestSim(a,b) = Sim(\mathrm{I}_a, \mathrm{I}_b) \qquad 4\text{-}6$$

The next section will give details of how the individual user models are extracted.

4.3.2 Individual User Models

Cognitive Centrality

Cognitive centrality measure is used to locate knowledge inside the community that is important to the community members. This can be helpful to identify the central members and how they contribute to the community. It can also be useful in identifying unique knowledge held by peripheral members. This is important for the community's sustainability and flexibility - interests might shift in time (Lave and Wenger, 1991), knowing where unique knowledge is located can facilitate the transition from one subject area to another (Wegner, 1986). Being aware of the central and peripheral members of the community can also help the improvement of *shared mental models* and *transactive memory.*

There are different approaches on centrality as used in 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 (Freeman, 1979) describes in a general review three types of centrality as developed in social network research. *Degree centrality* considers that a point is central according to how many adjacent points it is connected to. *Betweeness centrality* is based on the frequency with which a point falls between two other points. The importance of this type of centrality is that a point with a high betweeness centrality is controlling the communication inside the social network. *Closeness centrality* represents how depended a member is on other members if he/she needs to pass a message to other members in the network. In other words, it deals with the distance of a given point from all other points in the network.

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. Formula 4-7 illustrates the degree centrality approach introduced by (Nieminen, 1974). Nieminen's measure is the count of the degree or number of adjacencies for a point p_k, where $a(p_i, p_k) = 1$ if and only if p_i and p_k are connected by a line, and 0 otherwise.

$$C_D(p_k) = \sum_{i=1}^{n} a\left(p_i, p_k\right) \qquad 4\text{-}7$$

Here we adapt Formula 4-7 as follows: $CCen(a)$ of member a is calculated as the number of all members b to whom a is connected considering the four relationship types defined in section 4.3.1:

$$CCen(a) = \sum_{b=1}^{n} ReadRes(a,b)+ReadSim(a,b)+UploadSim(a,b)+InterestSim(a,b) \qquad 4\text{-}8$$

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 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, and extending those with the concepts extracted from the ontology (see Section 0), his interests are represented as a list of terms with weights. For example, all terms that member a has shown any interest in are aggregated in the list T_a, where every term $t \in T_a$ has weight $w(t,T_a)$ that indicates the frequency of t in T_a. If $w(t,T_a) \geq \theta$ (θ is a threshold), t is added to the interests of a denoted with I_a. I_a is presented as the member a's personal list of interests. 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 a given member. For example if T_a is small then θ will be small so terms will allowed to be added in the list I_a. Having θ allows the approach to be flexible and accommodate closely-knit VC of smaller or larger sizes.

Participation, Relationships and Personal Hierarchies

Participation: The frequency of knowledge sharing activities of a member (uploading or downloading) is stored in his individual user model - $uRate$ is the number of resources downloaded, and $dRate$ also is the number of resources downloaded The participation rates are used in algorithms that detect change patterns in the community, described in Chapter 6.

Relationships: Each participating member in the VC is developing relationships with other members of the VC (section 4.3.1). These relationships $ReadRes$, $ReadSim$, $UploadSim$, $InterestSim$ are stored for each member in his personal profile. This information is used in algorithms for generating personal notifications provided to individual members but aimed at benefiting the community as a whole, see Chapter 7.

Personal Hierarchies: Folders F and resources R created by each member in the VC are composing the personal hierarchies that a member is creating. The personal hierarchies can be used in extracting the resources and folders a member has created and be used as content information in notification messages generated to members, see Chapter 7. For example when two members a and b are detected with a $ReadSim(a,b)$ and a notification is generated to member b the resources R_a stored in the personal hierarchies for member a are included in the notification message N generated as suggested reading for member b.

4.3.3 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 the knowledge CCenM and 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} and L_{Per} represent the lists of popular and peripheral topics respectively. 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 represent the frequency of a keyword k in $allKeywords$. If $w(k, allkeywords) > \sigma_{Pop}$ (σ_{Pop} is a threshold), k is added in 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 CM 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} will be small as well so terms can be added in L_{Pop} and L_{Per} accordingly. This allows the approach to be applicable to closely-knit VC of different sizes.

List of Cognitively Central Members

Having extracted the $CCen$ for every member in the VC (section 4.3.2) a list of the most $CCen$ members is extracted. 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 Chapter 7. $CCen(a)$ 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.

4.4 Identifying Knowledge Sharing Patterns – Study with a Virtual Community

To validate the community modelling algorithms, we have applied them to extract a model of the SW VC (Chapter 3). The study was conducted to examine whether the application of the community modelling algorithms to analyse the log data could identify problems that could have been spotted earlier and addressed properly to help this community sustain.

The community log data was stored in a text file, fully anonymised, and then converted to database tables. The tables were used as input for the community modelling algorithms described in section 4.3, and implemented in Java. All keywords converted into nouns in order to be used as an input to the WordNet similarity algorithm. Here representative examples of phenomena discovered about the community are presented, and discussion is offered on how this can be used for adaptive support. In the illustrations below, excerpts from the community model are rendered with NetDraw.

4.4.1 Relationships Model

The relationship model indicated strong semantic links between members which were often not explored in the community. According to the community model, the members who never uploaded resources in the community had in fact *ReadRes* similarity (see Figure 4.3). There are links with two groups – the group including members M31, M29, M15, M3, M13, M22, M23, M29, and M17 and with the group of members M33, M20, and M12. The situation in Figure 4.3 indicates that the community's transactive memory system is not well-developed, which points at the need for appropriate support. For example, automatic messages can be generated to point out to member M29 (who is actively engaged in the community) that he has a relation with member M19 (who is not uploading). Providing such awareness can improve the transactive memory, develop members' understanding of what the others are doing and facilitate collaboration.

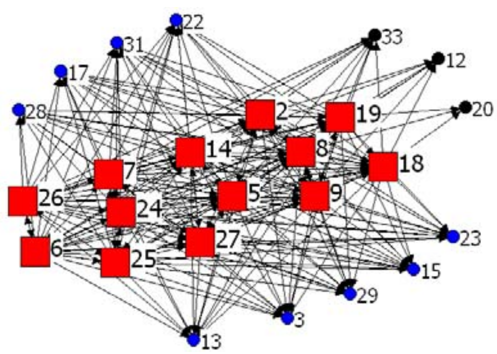


Figure 4.3 The members not uploading to the community, in rectangles, had *ReadRes* similarity with the same people. These links were unexplored in the community. Members should be made aware of their similarity in reading resources uploaded by the same members.

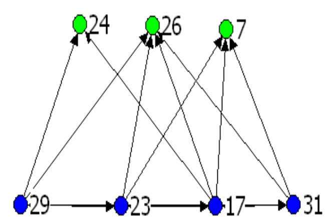


Figure 4.4 *ReadSim* between members M24, M26 and M7, who are reading resources uploaded by the same people but are not aware of this similarity.

Another interesting case concerns members M7, M24 and M26 who have ReadSim relation with almost the same people but have no connection among themselves (Figure 4.4). Interestingly, these people are coming from the same research group and, as indicated in the community model, they have not explored (and are perhaps unaware of) their connections via the community. Making these members aware of their similarities with others may motivate them to better participate in the community, see (Harper et al., 2007). It can also facilitate knowledge sharing between these people, who appear to be interested in the same topics (Fischer and Ostwald, 2001), and may promote collaboration.

Initial results showed most nodes of the graphs representing *ReadSim*, *UploadSim*, and *InterestSim* relationships to appear strongly connected. This confirmed our expectations for the community model (when people are working in similar areas their interests and the resources shared tend to be semantically similar).

4.4.2 Cognitive Centrality

The centrality of each member (Figure 4.5) was calculated based on formula 4-8. Members M31, M29 and M17 are indicated as the three most central members of this community. This closely corresponds to the real world - members M17 and M31 are the facilitators of the two research groups involved in this community, while member M29 is a researcher who actively contributed resources relevant to the community domain. Centrality can be influenced by different circumstances. For example, members M6 (newcomer) and M25 (oldtimer) gained some centrality due to actively downloading from the community. Such members might be aware of the cognitive processes in the community and can provide valuable information to the others. Member M13 on the other hand, was an old-timer actively engaged by both reading and uploading resources to the community. This member was indeed involved in most projects and could be quite influential to the community.

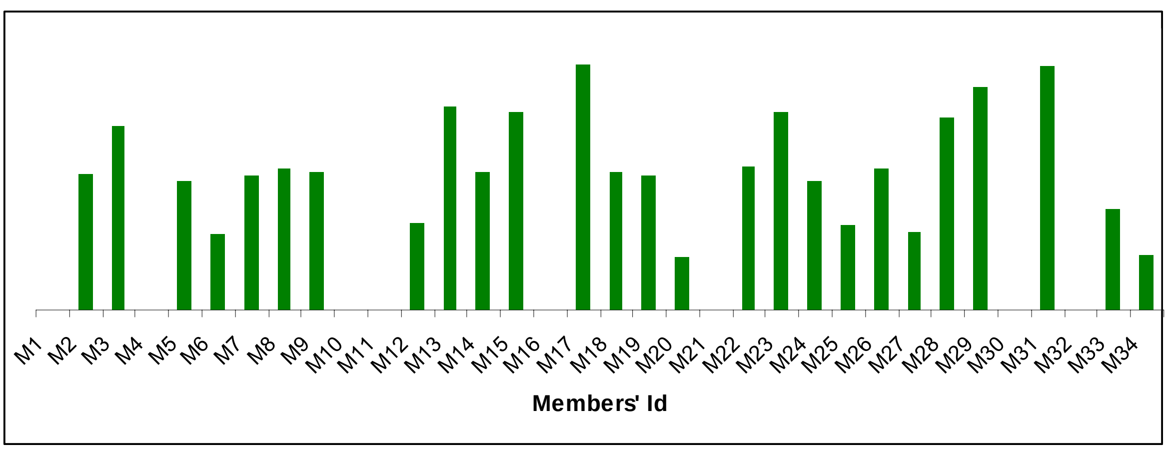


Figure 4.5 Community members' cognitive centrality. The bigger the bar the higher the CCen of a member. Numbers represent community members and column height represent members' cognitive centrality.

It is interesting to compare the centrality of two members M5 (newcomer) and M26 (old-timer) – who have not uploaded resources. Member M26 appears to be more central to the community than member M5 although M26 has read fewer resources (twenty-one in total) than M5 (who read fifty resources in total). This indicates that M26 has read resources that are closer to the community's interests and illustrates the effect of the community's domain on deriving relationship values (section 0).

The centrality measure can be a way to motivate people to contribute and remain active, e.g. in (Cheng and Vassileva, 2006) centrality is visualised to encourage participation. We consider *push mechanisms* where tailored notification messages (Chapter 7) can be sent to members based on their centrality. For example, members M5, M6 and M25 can be encouraged to contribute to the community since they already have similarities with the rest of the community. Indicating the most central members can be beneficial for the community. They can be asked to point others at valuable resources, e.g. when member M28 (peripheral) is searching for a topic which member M31 (central) seems to have information about, we can display a message to direct M28 to M31 for further help. Also, a newcomer like M6 can be integrated faster if they are mentored by a cognitively central member with similar interests.

4.4.3 Individual Cases

Information about individual users' engagement can be combined with the relationships model to identify cases where individuals can be given support in order to improve the functioning of the community as a whole. For instance, member M12, who was actively involved in projects with community members, has not downloaded anything, and has uploaded only one resource read by many members (Figure 4.6). M12 was identified as a fairly central member, as what he shared was important to the community. This member can be informed that people are interested in his resources and that there are other members uploading similar resources, which can motivate M12 to engage and can improve the knowledge sharing.

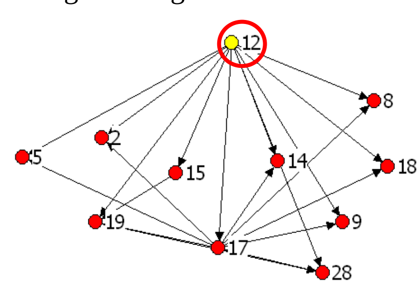


Figure 4.6 *ReadRes* relationships of member M12.
The graph shows the members who have read the resource uploaded by M12.

A typical problem for the effective functioning of communities is the *integration of newcomers* (newly joining members). There were several newcomers who did not integrate in the community during the analysed period. For example, member M14 was very active during the first two months after his joining but then became fully disengaged. The relationships model indicates that M14 read resources similar to those read by others and had similar interests to other members (Figure 4.7). The community model helped recognise similar behaviour followed by other members (e.g. M25 and M19) - downloading actively for some time and becoming disengaged afterwards. This might be an indication that these members were struggling to find their way in the community's knowledge space and were uncertain about their role in the community. Such members could be helped to become aware of their cognitive relationships with others, so they may be motivated to remain actively involved.

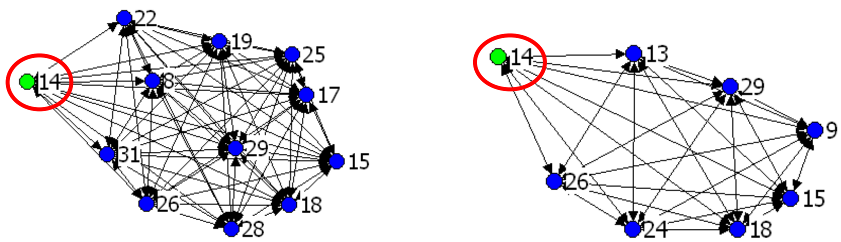


Figure 4.7 *ReadSim* (left) and *InterestSim* (right) ego networks for member M14. The above networks show that the resources member M14 was reading were similar to the resources several members on the community were reading too. Also the derived interests of member M14 are similar to the interests of other community members. These links were unexplored, and member M14 became disengaged from the community.

Another interesting newcomer case is member M33 who was inactive at the beginning but then started contributing to the community. She uploaded a total of eleven resources all of which were highly relevant to the community, but only one resource was read by one other member (Figure 4.8). Member M33 was a visitor for a year at one of the research groups whose leader was member M31. The relationship model indicated that many members uploaded similar resources to M33. Unfortunately, these links were never exploited and the VC as a whole did not benefit from the knowledge "shared" by M33. M33 was isolated throughout and did not benefit from participating in the community. The example shows how the community model helped detecting an isolated niche which hinders the effective knowledge sharing. Based on *ReadSim* or *InterestSim* relationships, oldtimers that have similar interests or are reading similar resources and are actively engaged in the community can be approached.

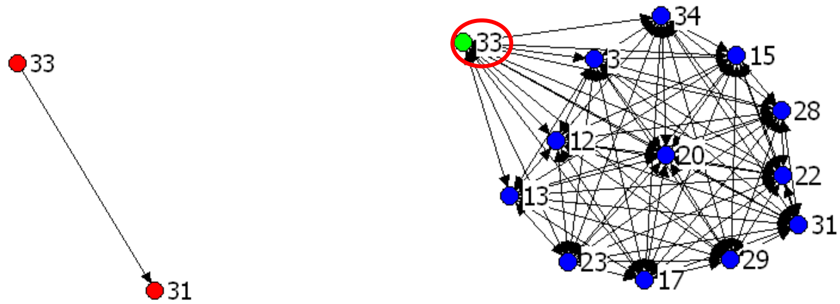


Figure 4.8 *ReadRes* for member M33 (left) and ego network for *UploadSim*(M33) (right). Despite the similarity with other member, M33 did not integrate in the community.

For instance, a message can be sent to member M31 to help the newcomer M33 to integrate in the VC. Member M33 could also been notified that others had similar interest and were uploading relevant resources. At the same time, oldtimers could be encouraged to look into interesting resources uploaded by newcomers. In general, such support aims at improving the community's transactive memory and can motivate members to remain engaged.

In short, the study showed that the community model represented cases which actually happened in the reality. If the community model was done by a human (the community moderator for instance) this would have taken significant effort for processing the log data. Furthermore, the study pointed out that the community model could be analysed to indicate problematic cases with both existing or newly joining members, which could be related to TM, SMM, and CCen. The CM analysis was done manually here with the help of visualisation software. This analysis informed our next step aimed at automatically analysing the CM and detecting community knowledge sharing patterns.

4.5 Summary

This chapter presented the definition of the community model followed in this PhD. The formalisation of input data has been discussed and the algorithms developed for extracting the community model presented. The input data is kept in a generic format so the approach can be generic and used in other knowledge sharing applications. The relationship model algorithms can be adjusted according to the input data at hand. A study with log data from an existing VC was conducted. It has enabled us to identify patterns of community behaviour detected with the community model, and provides the basis for automatic detection of community patterns and dynamic community-tailored support, which will be presented in the following chapters.

Chapter 5

Analyzing Community Knowledge Sharing Behaviour

5.1 Introduction

Based on theories on TM, SMM, CCen, and collaboration support, this research aims to support members of a closely-knit community to answer questions like “Who knows more about subject A?”, “Do others in this community know what I know?”, “Who shares the most valuable resources in this community?”, “Whose knowledge is important to me?”, “To whom is my knowledge important?”, and “What others are doing in this community?”. Studies in CSCW community also looked at the above awareness issues, and stress that the outcome of a member’s action in the community, can influence others’ actions (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 a good TM system, establishment of SMM, and exploiting CCen, can facilitate the knowledge sharing in VC (Ilgen et al., 2005).

Chapter 4 presented algorithms for deriving a community model which were applied to generate a model of a real VC based on archival tracking data. The extracted community model was analysed using visualisation techniques, based on which interesting cases were identified and discussed according to their relevance to the main processes followed in this PhD – TMM, SMM, and CCen. In order to provide intelligent support to VC members, the relevant situations/patterns have to be discovered *automatically* in order to provide an input to algorithms which generate intelligent community-tailored support. The automatic detection of community knowledge sharing patterns is the main goal of this chapter.

We will present graph based algorithms that will automatically extract pre-defined static patterns from the community model defined earlier. The chapter starts with a brief discussion of related work to position our approach in the relevant literature. We then introduce main definitions and notations for representing and mining a graph that represents semantic relationships between community members. Following these notations, algorithms that detect community knowledge sharing patterns by identifying graph features and combining this with the individual user models are presented. The algorithms are applied over the SW VC community model derived in Chapter 4 to derive patterns about the knowledge sharing behaviour in that community. The study validates the algorithms and indicates that the knowledge sharing patterns can be used as an input for generating intelligent support to VC members.

5.2 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). A graph is defined as a set of nodes, where pairs of nodes might be connected by edges (Gross and Yellen, 1999). These connections between edges and nodes compose the structure of a graph (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 PhD, semantic relationships based on knowledge sharing. Structural patterns in social networks refer to the many mathematical attributes of graphs that can be recognised in a network (e.g. cliques, degree centrality, structural equivalence, and structural holes)[10]. These are quantitative approaches that do not consider any semantics and thus have yet to be applied for the investigation of social “roles” (e.g. newcomers, oldtimers) or social “power” (e.g. CCen, CPer) (Chakrabarti and Faloutsos, 2006). This section gives a brief overview of relevant graph mining approaches in social networks on identifying relationships and/or interactions in peoples’ graphs.

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). The common characteristic of these approaches is their focus on structural attributes of communities and in extend structural 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).

Viermetz and Skubacz (2007) apply text mining techniques upon email conversations to extract patterns/networks of people and their relationships. Keywords are extracted from email

[10] More details on this approaches can be obtained from: Chakrabarti, D., et al. (2006): 'Graph mining: Laws, generators, and algorithms,' *ACM Computing Surveys*, vol. 38, no. 1, 2.

communication to form vectors that each one represents a single message. A network of those messages is built. Then clusters of similar messages can be 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. This way the centrality of actors is measured according how central the topics of the messages exchanged by a given actor are. Compared to our work, the main difference is that the text analysis does not consider the relevance of each email message to the overall community. Thus, extracted centrality measure considers only the keywords associated with a user and does not take into consideration the importance of these keywords to the community as a whole. Following the algorithms presented in Chapter 4, 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.

A publication network is used in Khan and Shaikh (2006) to extract predefined algebraic functions that represent social relationships in a network (Khan and Shaikh, 2006). The patterns developed have been applied to an existing publication network to extract reviewers for a specific paper. Although the algebraic functions developed are called relationships, they are concerned with structural functions of graphs and have no added semantics. For example, in the extracted network a binary 1 represents an edge exists between two nodes and a 0 denotes no edge between the two nodes. Thus, the edge 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, the algebraic functions proposed in Khan and Shaikh (2006), are applied to the network of that person. The main difference to our approach is the absence of semantics from the patterns developed. Khan and Shaikh (2006) are only concern with the structural characteristics of a social network graph (matrices and sets operations), and the algorithms developed are based on those structural characteristics.

A recommendation tool developed by Lo and Lin suggests friends to community members based on exchange of messages (Lo and Lin, 2006). Sending messages to each other is the connection that exists between two community members in their network. The content of the message is not considered and there is no use of any semantic data. Compared to our work we are considering the keywords of the resource shared by a member and in addition the type of the connection between two members (e.g. if two members read, upload similar resources or resources shared by each other).

All the above approaches have developed static pattern algorithms and employed them to analyse people networks. Our approach focuses on the modelling of *semantic relationships* via graphs, i.e. an edge connecting two members represents their semantic similarity to each other, and the relevance of this link to the community's domain. This combined with the theoretical underpinning, enables us to contribute to the area by developing a graph-based approach for *qualitative* analysis that automatically detects relevant interaction patterns.

5.3 Definition and Detection of Relationships in Graphs

This section will introduce the main notations needed for defining community knowledge sharing patterns. Let a and b denote two community members. As specified in Chapter 4, we consider the following relationships: $ReadRes(a,b)$ indicates that resources uploaded by b are read by a, and its strength corresponds to the relevance of the resources to the community domain; $ReadSim(a,b)$ indicates that a and b have read semantically similar resources; $UploadSim(a,b)$ indicates that a and b have uploaded semantically similar resources; and $InterestSim(a,b)$ represents the similarity of interests between a and b.

Each relationship 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 two members (nodes), the strength is calculated by applying the algorithms presented in Chapter 4. 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 way that the resulted graph is meaningful for the purpose is constructed. For example, if only strong relationships need to be extracted the threshold value will be high so only high weighted edges will be extracted. A neighbourhood of a node v, denoted as $N_G(v)$, represents the ego network (Degenne and Forse, 1999), of v and indicates the members that v has corresponding similarity with.

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

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

In addition, the automatic detection of patterns exploits information from the individual user profiles, including: (i) $uRate(a)$ and $dRate(a)$ which denote the upload and download rate of a (Chapter 4); and (ii) $CCen(a)$ which indicates how important the knowledge a holds is for the VC, calculated as the sum of all the relationships a has with any member b (described in Chapter 4). 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.

5.4 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 types of patterns. For each pattern we point out its relevance to TM, SMM, and CCen, and define 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 complimentary knowledge 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:

$$\left(\mathrm{N}_{RS}(v_a) \cap \mathrm{N}_{RS}(v_b) \neq \varnothing\right) \wedge \left(v_a \notin \mathrm{N}_{RS}(v_b)\right) \qquad 5\text{-}1$$

In the same way, P1 is defined for $UploadSim$ 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 are not aware of how similar they are in terms of uploading, reading or interests with other members of the community. Detection of this pattern can be used to promote 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:

$$\left(\mathrm{N}_{RS}(v_a) \cap \mathrm{N}_{RS}(v_b) \neq \varnothing\right) \wedge \left(v_a \in \mathrm{N}_{RS}(v_b)\right) \qquad 5\text{-}2$$

In a similar way, P2 is defined for $UploadSim$ 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. Support can be provided to those members in order to benefit and make the most of their time in the community.

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

$$\left(uRate(a) > 0\right) \wedge \left(dRate(a) = 0\right) \qquad 5\text{-}3$$

P4. Members not contributing

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

$$\left(uRate(a) = 0\right) \wedge \left(dRate(a) > 0\right) \qquad 5\text{-}4$$

P5. Important peripheral members not downloading

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

Importance: We can use this pattern to motivate peripheral members to benefit from the community. Notifying him/her that others are interested in what he/she uploads can motivate that member to start reading resources uploaded by the members he/she has similarity with. 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:

$$\left(uRate(a) > 0\right) \wedge \left(dRate(a) = 0\right) \wedge \left(\mathrm{N}_{RR}^{+}(v_a) \neq \varnothing\right) \qquad \text{5-5}$$

P6. Important peripheral members not uploading

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

Importance: This pattern can be used to motivate people who are only downloading from the community to start uploading, by showing them how similar 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 sustain.

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

$$\left(uRate(a) = 0\right) \wedge \left(dRate(a) > 0\right) \wedge \left(\mathrm{N}_{IS}(v_a) \neq \varnothing\right) \qquad \text{5-6}$$

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 similarity and difference 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 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 $G_{RS}(V_{RS}, E_{RS})$:

$$\left(v_a \in \mathrm{N}_{US}(v_b)\right) \wedge \left(v_a \notin \mathrm{N}_{RS}(v_b)\right) \qquad \text{5-7}$$

Table 5.1 summarizes the importance of each pattern to the community processes. The patterns indicate when and what interference can be made, as discussed next.

Table 5.1 Summary of how the detection of a pattern can affect the relevant processes

Pattern	Affects
P1: unexploited similarity between members	Collaboration, TM System, SMM
P2: members unaware of their similarity	SMM, TM System
P3: members participating but not benefiting	Improve participation, Sustainability
P4: members not contributing	Improve participation, Sustainability
P5: peripheral members not downloading	SMM, Collaboration
P6: peripheral members not uploading	TM System, Collaboration
P7: unexplored member complementarities	SMM, TM System, Collaboration

5.5 Detection of Knowledge Sharing Patterns – Study with a Virtual Community

To validate the algorithms for detecting knowledge sharing behaviour patterns, we conducted a study with the archival data from the SW VC presented in Chapter 3. This approach was chosen because the log data gives an inside of what has happened in a *real knowledge sharing community* during a fairly *long period from active functioning to standstill*. The evaluation approach followed is somewhat similar to evaluation using simulated data, applicable when large amount of data is needed, data is too expensive to collect, or when people have to be involved and there is no available sample. The major advantage of our evaluation approach is the use of longer term authentic data. Since the author actively participated in the community and few other members were still available for clarification, we were able to check the appropriateness of patterns recognised involving these members and the suitability of the notifications that could have been generated to community members.

We validated 60% of some 90 detected pattern occurrences (from 7 pattern types). A pattern was validated when:

- the detected relationships between members were appropriate (which was checked by looking at the resources members shared);
- the log data of the follow up behaviour confirmed the pattern, e.g. when it was found that members might have not been aware of their similarity, in their subsequent interactions they indeed did not read papers from each other; when members were available, it was confirmed that they were indeed unaware of the detected similarities;

- one of the community moderators confirmed that the detected useful patterns indicated situations when intervention could have been made.

5.5.1 Data Analysis

We needed to indicate *when* detected patterns would have been useful and *what* interventions could have been triggered then. A quantitative summary of the community participation identified that the activity minimized rapidly in Month3 and in Month7 (see Figure 3.5). Thus, we applied the algorithms on the data collected in Month4, Month5 and Month6 - the months between the two activity drops. The community model presented in Chapter 4 was used. Relationships were extracted between 1122 member pairs (considering the relationship between any two members apart from themselves). The community model was used as input to the pattern detection algorithms, implemented in Java following the definitions in Section 5.4.

The detected patterns involving the author who was an active community member were checked and the resources downloaded/uploaded examined. The patterns were also discussed with the supervisor (one of the cognitively central members of the VC) and clarified with former VC members available for contact. The results are presented below.

The application of the patterns on the data collected uncovered that the community had in general poor TM system and SMM, and collaboration between community members was difficult to achieve. We will focus on the crucial period (months 4-6) when some intervention could have been beneficial.

5.5.2 Results

The results in Month4, Month5, and Month6 show that each month the VC was coming closer to a halt. The analysis regarded the discovered patterns as relevant. The analysers (the author and her supervisor) appeared aware of some patterns (e.g. an inactive member whom they had similarity with and could have helped him to better integrate, people working on a similar project), but other patterns showed links that the researchers were unaware of and had to validate by examining the resources read/uploaded. We present below examples which illustrate how patterns could have been used to generate notifications to community members. It is important to note that the visualisation tool is used to illustrate pattern which were *automatically* detected. We will point at possible notifications that could be sent, which will be defined in Chapter 7.

P1. Unexplored similarity between community members.

In Month4 P1 was detected for 10 pairs, while in Month6 this pattern was detected for 22 pairs of members. This situation creates a problem to the VC as it shows that a TM system is not in place, people are not aware of whom they have similarity with and there is a lack of SMM as members do not know what others are working on. Figure 5.1 illustrates pattern P1.

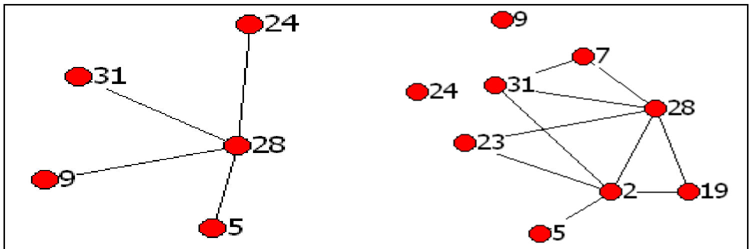


Figure 5.1. Left**:** ReadSim in Month4 – M9, M24, M31 have ReadSim with the same members as M5 but not with M5. Right: ReadSim in Month5 showing members M9 and M24 to be disengaged.

In Month4, members M9, M24, M31, and M5 had *ReadSim* with M28 but did not have *ReadSim* among themselves. M28 was found to be the connecting node between these four members (Figure 5.1 Left). In Month 5, M5 and M31 continue in the same situation but members M9 and M24 have stopped contributing or downloading from the community (Figure 5.1 Right).

Furthermore, in Month4, members M5 and M9 were detected to have *InterestSim* with M24 and M28 but not between themselves; M23 has *InterestSim* with M28 (Figure 5.2). Members M5, M9, M23, M24, and M28 have closely related interests but might have not been aware of their similarity with each other.

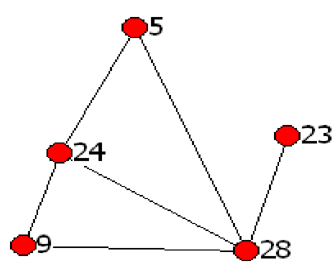


Figure 5.2 In Month4 M5 and M9 have InterestSim with M24 and M28 but there is no InterestSim between M5 and M9.

Consequently, in Month5 and Month6, members M9, M23, and M24 became inactive. Notification *N1-3* (Table 7.1, Chapter 7) could have been sent in Month 4 to M28, M5, M9, M24, and M31. Notification *N3-1* (Table 7.1, Chapter 7) could have been sent in Month 5 to M28 who is the connection to these members and to M31 detected as cognitively influential member.

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

Examples of this pattern are members M9, M24 and M31. In Month4, M9 and M24 had *ReadSim* among themselves and with M31 (Figure 5.3 left), while in Month5 M9 and M24 disengage from the community. Where these members aware of the similarity they had with each other and with member M31 (CCenM)? Additionally, the pattern detection algorithms found that in Month5 members M2, M5, M7 and M19 had *InterestSim* with each other (Figure 5.3 right). In Month6, members M2 and M7 were detected as disengaged from the community and the activity of members M5 and M19 was very low. Notification *N3-2* (Table 7.1, Chapter 7) should be sent to M31 making this member aware of his influence in the VC and the decline in participation of relevant peripheral members might have been helpful to keep M9 and M24 engaged. Making M9, M24, M2, M7, M19, and M5 aware of their similarity with notification *N1-2* (Table 7.1, Chapter 7) could have helped them engage back in the community.

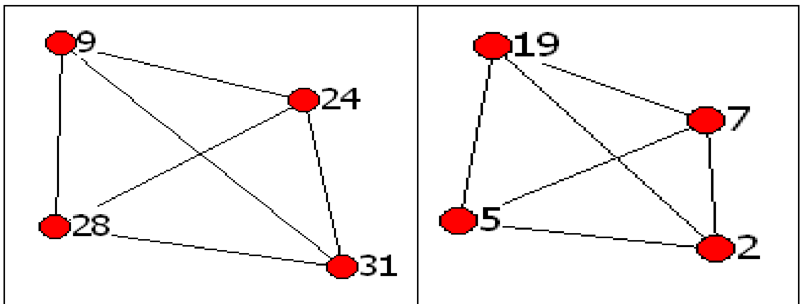


Figure 5.3 Left: ReadSim between members in Month4.
Right: InterestSim between members in Month5.In both cases members appear to have similarity among themselves but are they aware of the similarity they have?

P3 & P4. Members not benefiting or not contributing.

In Month4, one member was only uploading and four members were only downloading. By Month6, two members were only uploading and nine members were only downloading. The patterns detected in Month6 showed that members began to disengage from the community either because they lost interest or because they could not find information useful for them. Downloading only excessively is a behaviour that newcomers develop when they struggle to locate information important for them. For example, in Month6, member M19 downloaded 33 resources, without uploading anything from Month4 to Month6. Members like M19 can be supported by sending *N1-5* or/and *N2-5* (Table 7.1, Chapter 7) providing them with information of members with similar interests or members who are reading similar resources.

P5. Important peripheral members not downloading.

Example for this pattern is member M33 who was detected uploading but not downloading from the community, while other members were interested in what M33 was uploading. In Month 6, this member disengaged from the VC feeling that she had not benefited from this community. The detection of this pattern could have been used to generate appropriate notifications. *N1-3* (Table 7.1, Chapter 7) could have been sent to M33 in Month4 and Month 5, making her aware of her importance to the VC. In addition, in Month5, Member M31, who was a cognitively central member, has read resources uploaded by member M33 could have been sent notification *N3-1* (Table 7.1, Chapter 7). This might have helped form a link between M33 and M31 to motivate and channel M31’s contribution.

P6. Important peripheral members not uploading.

An example for this pattern is member M5. In Month4, M5 joined the community and extensively downloaded resources. Based on this, he had *InterestSim* with members M24 and M28. In Month5, M5 was still only downloading and had *InterestSim* with M2, M7 and M19. In Month6, M5 stopped participating. Notification *N1-5* (Table 7.1, Chapter 7) could have been sent to M5 in Month4 helping him identify people in the community or resources of his interest. Notification *N1-6* (Table 7.1, Chapter 7) could have been sent in Month5 pointing to M5 that his knowledge is relevant to the community, helping him discover relevant resources, and encouraging him to contribute to the VC.

P7. Unexplored complimentary similarity between members.

Member M13 had *UploadSim* with member M33 but not *ReadSim* with M33. The detection of this pattern could have been used to trigger *N1-7* (Table 7.1, Chapter 7), in order to keep M33 and M13 aware of their similarity and motivate them to read resources uploaded by each other.

In summary, the study indicates that the approach can be beneficial when an active VC starts to experience problems. The detected patterns can provide a better understanding of these problems and suggest possible interventions. The analysis of what was *automatically detected* in the study corresponds to what was *manually detected* in an earlier study in Chapter 4. Moreover, the automatic analysis discovered patterns *missed by the human analyzer* who was exploring only visualization tools in the study discussed in Chapter 4. A careful look into the human-missed problems confirmed their importance to the functioning of the community. Hence, the advantage of the approach presented here is the ability to discover patterns when large log data is collected, and the suggestion of corresponding community-tailored interventions.

5.6 Summary

This chapter has described a new approach to identify static knowledge sharing behaviour patterns in a VC driven by processes important for the effective functioning of closely-knit communities. We have shown how these patterns can be detected and used to provide community-tailored support using examples of notifications that can be generated (full list of the notification messages defined can be found in Chapter 7). The examples used are representative of what patterns can be discovered, how they can be automatically detected, and how the detection can be used. This work does not aim at defining an exhaustive list of patterns that can be discovered in a VC. Indeed, patterns can vary from community to community depending on the topic, people and the VC purpose. New patterns can be included as long as they can be defined with appropriate graph characteristics and the underlying CM is in place.

Although static knowledge sharing patterns can be useful in identifying problematic cases especially during the starting phase of a VC, changes over time are crucial when identifying how a community is functioning and what support would be needed. The static patterns defined in this chapter are the foundation to discover how a community evolves over time. Chapter 6 will define algorithms for detecting change patterns in a knowledge sharing community. They will be applied on the SW VC data to extract changes in the VC to indicate further cases when intelligent support to VC members could have been generated.

Chapter 6

Detecting Changes over Time

6.1 Introduction

To identify what support can be offered to a virtual community, a deep understanding of the functioning of the community is needed. For this, it is crucial to detect evolution patterns of the community knowledge sharing behaviour captured in the community model. The focus of this chapter is on the analysis of the community model to detect changes over different time periods in order to identify when and how the community's TM system can be improved and SMM developed, as well as who is part of the CCen and how they can influence the VC.

Monitoring how relationships and activities are changing in a closely-knit VC is vital for knowledge sharing, collaboration, and community sustainability (Palla et al., 2007). Explicitly making people aware of how their similarities, in addition to their activity, are changing can influence their actions, and thus, help them engage in the community (Schmidt, 2002). Along this vein, we are looking at three types of knowledge sharing change patterns which can indicate that some intervention to the community may be beneficial: *(a) members are moving to the periphery (changes in CCen and neighbourhood), (b) members are not exploiting relationships between them (TM can be improved), (c) members are not integrating effectively (better SMM need to be developed).*

This chapter will start with positioning our research on detecting changes over time in a knowledge sharing VC in the relevant body of research by comparing with similar approaches. Section 6.3 will describe the algorithms for detecting community change patterns and their importance with regard to TM, SMM, and CCen, as well as for analysing changes after an intervention to the VC. Section 6.4 will show how community change pattern algorithms have been applied in a study with tracking data from the BSCW community (presented in Section 3.4) to identify what community-tailored support may be provided. Section 6.5 will summarise and conclude the chapter.

6.2 Relevant Approaches for Analysing Community Evolution

Analysis of community evolution refers to different approaches for detecting changes over time in large or small people networks represented as graphs. Existing approaches are examining mainly structural changes of social networks (e.g. density, degree distribution, average distance, clustering coefficient) by comparing the characteristics of graph instances at given time points (Leskovec et al., 2005; Falkowski et al., 2006; Asur et al., 2007; Falkowski and Spiliopoulou, 2007; Lin et al., 2007; Palla et al., 2007; Lin et al., 2008).

Several studies have looked into properties indicating time evolution of people networks. In (Palla et al., 2007), two social networks are monitored through a series of timestamps. Edges represent existence of a co-authorship or a telephone conversation, in a co-authorship network and mobile phone network, respectively. Leskovec et al (Leskovec et al., 2005) propose evolution models based on an analysis of temporal evolution of several networks based on densification power laws and shrinking diameters in real graphs. Falkowski and colleagues (Falkowski et al., 2006; Falkowski et al., 2007; Falkowski and Spiliopoulou, 2007) detect community evolution by finding structural overlapping between sub-graphs in a large network extracted in different time periods. A user's involvement in the community is calculated as the number of interactions of that member with other members, and is represented as edge weights in the graph. Research in online social network dynamics analyzes the evolution of links in social networks. For example, Kumar et al (2006) study the evolution of links in Flickr and Yahoo!360 - a model of the evolution of online social networks is proposed based on analysis of the network density (Kumar et al., 2006).

Recently, evolution based on topic similarity has been investigated. For example, communities have been identified based on topic similarities between users who are mutually aware of each other's presence in the network (Lin et al., 2007). Five possible patterns of community evolution have been considered (one to one derivation, merge, split, extinct, and emerge), and corresponding graph mining algorithms developed to identify the resulting communities. Edge weight is measured according to the degree of interest in interaction between two people in the network. In (Song et al., 2005), a graph model is used to describe the relationship and temporal evolution of a person's topic expertise in a network based on authorship of publications. Each expert is presented with a graph where edge weights denote the strength of connection of one topic to another. Time is divided in temporal sliding intervals, and the strength of the nodes as well as the structure of the network is considered in the evolution graph.

All approaches mentioned above have a similar purpose: they monitor how the network/graph under investigation is evolving over time in order to get an insight of the community. The indicated changes are not related to any particular processes of the community but rather at what can be derived from analysing two graph instances. In this project, the purpose of detecting changes is to exploit the extracted information in order to provide intelligent support to the community as a whole. A principle difference from the existing work is that this work aims to detect change patterns connected to specific processes related to effective functioning and sustainability of a VC. To the best of the author's knowledge, there is no other approach which examines community evolution with regard to TM, SMM, CCen.

Another notable difference with related community evolution approaches is that this research combines several graphs and other community parameters (e.g. user profiles) when deriving change patterns. The algorithms presented in Section 6.3 combine different relationship graphs and analyse their structural changes with regard to members' neighbourhood. Additional parameters based on members' activities represented in the user profiles are also considered.

Finally, our work adds to existing research in modelling communities as graphs. We exploit graphs based on semantic relationships between community members which are derived by analysing the members' knowledge sharing activities and interaction with each other. In contrast with the existing methods, which consider simple indicators for a relationship (e.g. direct connection), here semantic techniques are exploited to derive possible relationships between members (Chapter 4). The relationships are derived, and the members' behaviour is analyzed to identify when members should be informed of beneficial relationships that they may be unaware of.

6.3 Detection of Community Changes through Time

In the context of this research, we will monitor two different types of changes in the community. The first deals with the *detection of changes over time in order to identify when interventions may be needed*. This assists us to discover issues with the community that need to be resolved in order for the community to better function as an entity. Secondly, after interventions are triggered, we need to detect whether any *changes have occurred in the VC as a result of the interventions*. We extract graphs for each relationship defined in Section 4.3, and compare graph parameters at sequential time points $t-1$, t, $t+1$. Time point t represents the present time, and $t-1$, $t+1$ represent the time points before and after t respectively. Section 6.3.1 will provide the algorithmic patterns used to monitor change patterns to aid intervention along with examples. Section 6.3.2

gives the definitions of the algorithms developed to capture changes occurred as a result of the interventions.

6.3.1 Detecting Community Changes to Aid Intervention

Analysing the community model, we can detect change patterns in the relationships between members. For example: (a) two members have similar interests and read similar resources at time $t\text{-}1$, while at time t they continue to have similar interests but no longer read similar resources; (b) at time $t\text{-}1$, two members read similar resources, as well resources uploaded by each other, however, at time t they continue to read similar resources from others but no longer each other's resources.

As discussed in Section 5.3, four graphs have been extracted for the four different relationship types. $G_{RS}(V_{RS}, E_{RS})$ represents the graph derived for $ReadSim$, $G_{US}(V_{US}, E_{US})$ represents the graph $UploadSim$, $G_{IS}(V_{IS}, E_{IS})$ the graph for $InterestSim$ and $G_{RR}(V_{RR}, E_{RR})$ corresponds to the graph extracted for $ReadRes$. Let us denote members a and b to be members of the same VC, thus nodes in the above graphs. $CCen(a)$ indicates the cognitive centrality for a member a of the specific VC. In this section the graph notation will be used as described in Chapter 5. We will define below several change patterns, justifying their importance and providing examples from the real SW VC presented in Section 3.4.

Change Pattern 1: Members moving to periphery.

Members moving to the periphery can be encouraged to remain active and contribute. Such members can be detected in three cases. *(a) A member who is moving to the periphery due to his cognitive centrality as time passes, (b) A former influential member is moving to the periphery due to his cognitive centrality falling compared to previous time point and (c) A member's neighbourhood starts shrinking compared to the previous time point.*

Importance: Members who become inactive or share items, that are not relevant to the VC's key areas, are losing $CCen$ and are moving to the periphery of the VC. Having many people in the periphery is causing the VC to have reduced activity, other members to lose interest, (since they don't have new material to read), and is breaking the knowledge sharing chain increasing the number of members who are not participating actively.

Similarly, since for every relationship we can extract the neighbourhood of a member (members with whom a specific member has a semantic relationship with), then we can detect when a member's neighbourhood is shrinking. This allows tracking members who are about to start moving to the periphery of the VC, or are becoming inactive (stop reading or contributing), and help them to remain active and keep an interest to this VC. In this way the active lifetime of the VC is prolonged and the community members are encouraged to remain active and benefit from their time in the VC.

A special category of members are the cognitively influential members who have high $CCen$. They are important for the maintenance of activity in the VC and the addition of new material, and can also guide new members how to integrate in the community. It is therefore essential to detect whether influential members stay active and help them understand their importance for the effective functioning of the VC.

Early detection of members moving to the periphery will help us intervene by sending them encouraging messages to remain active and share resources valuable for the community. TM (Wegner, 1986) and SMM (Mohammed and Dumville, 2001) can be promoted by informing members of what is happening in the community and what relations/similarities exist between them and others. This may also promote collaboration as members can identify potential collaborators who are working on relevant areas.

Detection Case (a): Member a is moving to the periphery if $CCen(a)_{t-1} - CCen(a)_t > thr_{CCen}$ where thr_{CCen} indicates the chosen threshold value for considering a decrease in members' centrality. If the above situation is detected, we have to examine what causes the centrality to fall. Consequently, if $uRate(a)_{t-1} > uRate(a)_t$ the member has to resume uploading important resources to the community. If $dRate(a)_{t-1} > dRate(a)_t$, a has reduced the reading of resource uploaded by others. If neither the $uRate(a)_t$ nor the $dRate(a)_t$ have been reduced, the decrease in $CCen$ is caused by reducing the quality of the resources uploaded and/or downloaded by a.

Example: Let us consider member M15 from the SW VC (Section 3.4). At time $t-1$, $CCen(M15)_{t-1} = 0.78$, $uRate(M15)_{t-1} = 2$, $dRate(M15)_{t-1} = 0$ $avg(CCen)_{t-1} = 0.70$. At time t, $CCen(M15)_t = 0$, $uRate(M15)_t = 0$ and $dRate(M15)_{t-1} = 0$. If we assume a threshold $thr_{CCen} = 0.02$ then M15 would be considered as moving to periphery due to a drop in his uploading.

Detection Case (b): If a is an influential member, we have $CCen(a)_{t-1} > avg(CCen)$, $avg(CCen)$ being the average cognitive centrality of this community. To check if an influential member is moving to the periphery we use:

$$(CCen(a)_{t-1} - CCen(a)_t > thr_{CCen}) \wedge (uRate(a)_{t-1} > uRate(a)_t)$$

Example: To illustrate let us consider member M31 from the SW VC. At time $t-1$, $CCen(M31)_{t-1} = 6.39$, $uRate(M31)_{t-1} = 4$, $avg(CCen)_{t-1} = 1.11$. At time t, $CCen(M31)_t = 5.99$ and $uRate(M31)_t = 0$. If we use a threshold $thr_{CCen} = 0.5$, the influential member M31 is detected as moving to periphery.

Detection Case (c): For each relationship $UploadSim$, $ReadSim$ or $InterestSim$ we are taking the number of vertices that appear in a given member's neighbourhood, and compare those for time points $t\text{-}1$ and t. Here is a demonstration using $ReadSim$. If for a member v_a his neighbourhood for $ReadSim$ presents the following behaviour for times $t\text{-}1$ and t $|\mathrm{N}_{RS}(v_a)_{t-1}| > |\mathrm{N}_{RS}(v_a)_t|$, then the neighbourhood of member v_a for $ReadSim$ is shrinking. The same can be done for $UploadSim$ and $InterestSim$. For $ReadRes$ we have to check for both, the out-neighbourhood $\mathrm{N}^+_{RR}(v_b)$, thus the members who read what v_a is uploading and also the, in-neighbourhood $\mathrm{N}^-_{RR}(v_a)$ thus the members v_a reads from. If $|\mathrm{N}^+_{RR}(v_a)_{t-1}| > |\mathrm{N}^+_{RR}(v_a)_t|$ it means that less people are reading what that member is uploading. Similarly if $|\mathrm{N}^-_{RR}(v_a)_{t-1}| > |\mathrm{N}^-_{RR}(v_a)_t|$ then that member is reading from less people than previously.

Example: To illustrate let us consider member M31's $ReadRes$ relationships. At time $t-1$, the in-neighbourhood of M31 was larger than the neighbourhood at time t: $|\mathrm{N}^-_{RR}(v_{31})_{t-1}| > |\mathrm{N}^-_{RR}(v_{31})_t|$. Then we detected that M31has reduced/stop reading resources from other members. At the same time period ($t-1$), the out-neighbourhood of M31 was larger than at time t, $|\mathrm{N}^+_{RR}(v_{31})_{t-1}| > |\mathrm{N}^+_{RR}(v_{31})_t|$ then we detected that others reduced/stopped reading resources uploaded by M31.

Change Pattern 2: Changes in the behaviour of members indicate unexplored relationships.

Analyzing the community model, we can detect changes in the relationships between members. For example: *(a) two members have similar interests and read similar resources at time t-1, while at time t they continue to have similar interests but no longer read similar resources; (b) at time t-1, two members read similar resources, as well resources uploaded by each other, however, at time t they continue to read similar resources from others but no longer each other's resources; (c) at time t-1 two members are reading resources uploaded by each other and have interest similarity,* *however at time t they continue to have interest similarity but they are not reading resources uploaded by each other anymore.*

Importance: Discovering such cases is an indication that members are missing important information available to them in this VC. Detecting that after a period of time these people do not read similar resources, while they still have similar interests, or that they are reading similar resources but they stopped reading resources uploaded from each other, can be an indication that they are not aware of the relationships they have. This detection can be employed to make these people aware of the similarity they have and show them how they can benefit from that. For example, notification messages can be sent to such members to encourage them to continue reading from each other and to make them aware of their similarity. This can promote awareness helping the VC to develop better TM (Wegner, 1986), SMM and collaboration (Ilgen et al., 2005).

Detection: This pattern is detected by comparing the neighbourhood of a member in the corresponding relationship graphs. To illustrate, we will consider case (b). Having $G_{RS}(V_{RS},E_{RS})_{t-1}$, $G_{RR}(V_{RR},E_{RR})_{t-1}$ and $G_{RR}(V_{RR},E_{RR})_t$, $G_{RS}(V_{RS},E_{RS})_t$, the vertices v_a and v_b will appear in all four graphs. We need to detect at time $t-1$ one of the members (vertex), in the neighbourhood of the other, in both graphs $G_{RS}(V_{RS},E_{RS})_{t-1}$ and $G_{RR}(V_{RR},E_{RR})_{t-1}$. If this case exists, we need to also check that at time t one of the members does not appear in the others neighbourhood for $G_{RR}(V_{RR},E_{RR})_t$ but still appears for $G_{RS}(V_{RS},E_{RS})_t$ graph. Using the graph notations from Section 3, the pattern is detected with the following calculations:

$$\left[(v_a \in \mathrm{N}^{+}_{RR}(v_b)_{t-1}) \wedge (v_a \in \mathrm{N}_{RS}(v_b)_{t-1})\right] \wedge \left[(v_a \notin \mathrm{N}^{+}_{RR}(v_b)_t) \wedge (v_a \in \mathrm{N}_{RS}(v_b)_t)\right] \qquad \text{Case (b)}$$

Similar to case (b), cases (a) and (c) can be derived using the formulas below.

$$\left[(v_a \in \mathrm{N}_{IS}(v_b)_{t-1}) \wedge (v_a \in \mathrm{N}_{RS}(v_b)_{t-1})\right] \wedge \left[(v_a \in \mathrm{N}_{IS}(v_b)_t) \wedge (v_a \notin \mathrm{N}_{RS}(v_b)_t)\right] \qquad \text{Case (a)}$$

$$\left[(v_a \in \mathrm{N}^{+}_{RR}(v_b)_{t-1}) \wedge (v_a \in \mathrm{N}_{IS}(v_b)_{t-1})\right] \wedge \left[(v_a \notin \mathrm{N}^{+}_{RR}(v_b)_t) \wedge (v_a \in \mathrm{N}_{IS}(v_b)_t)\right] \qquad \text{Case (c)}$$

Example: To illustrate let us consider M28 and M31 for Case (b). M28 belonged to the out-neighbourhood of member M31 in $G_{RR}(V_{RR},E_{RR})_{t-1}$, thus ($M28 \in \mathrm{N}^{+}_{RR}(M31)_{t-1}$), and M28 belongs to the neighbourhood of M31 in $G_{RS}(V_{RS},E_{RS})_{t-1}$, ($M28 \in \mathrm{N}_{RS}(M31)_{t-1}$), and at time t, M28 does not belong to M31's out-neighbourhood for $G_{RR}(V_{RR},E_{RR})_t$, thus $M28 \notin \mathrm{N}^{+}_{RR}(M31)_t$ but still belongs to the neighbourhood of M31 at $G_{RS}(V_{RS},E_{RS})_t$ thus $M28 \in \mathrm{N}_{RS}(M31)_t$ then we have the case on change pattern 2.

Change Pattern 3: Members are not integrating effectively.

Members are considered to have integration problems if they either do not establish a reading relationship $ReadRes$ or an upload relationship $UploadSim$ with other members. Gradually, such members are becoming isolated from the community. For example: *(a) A newcomer has* $UploadSim$ *but no* $ReadRes$ *with other members for consecutive time points; (b) A newcomer has* $ReadRes$ *but not* $UploadSim$ *for consecutive time points; (c) A member has* $UploadSim$ *but no* $ReadRes$ *with other members for two consecutive periods; (d) A member has* $UploadSim$ *with others, however no one is reading what that member is uploading for consecutive time periods.*

Importance: Detection that a member is not uploading any resources, or that he is not uploading relevant resources in the VC, or that he is not reading relevant resources, shows that this member is not integrating effectively. In such cases, the members can be suggested resources that may be of their interest or directed to people with similar interests. In addition, cognitively central members with interests similar to an inactive member can be asked to contact the isolating member and help him/her find out how to benefit from knowledge sharing activities in the VC. In this way, SMM (Mohammed and Dumville, 2001) and TM (Wegner, 1986) can be promoted.

Detection: Similarly to the previous pattern, we consider a member's neighbourhood in corresponding relationship graphs. To illustrate, we will use case (b). If a is a newcomer at time t, v_a should not be present in the relationship graphs at time $t-1$, i.e. $v_a \notin V_{all_{t-1}}$. We then check that for time t, the specific member has a $ReadRes$ relation with other members but he does not have $UploadSim$. Consequently, for $ReadRes$ we check that the in-neighbourhood of v_a is not empty, $\mathrm{N}^{-}_{RR}(v_a)_t \neq \emptyset$, and for $UploadSim$ the neighbourhood is an empty set, $\mathrm{N}_{US}(v_a)_t = \emptyset$:

$$\left[(v_a \notin V_{all_{t-1}}) \wedge (\mathrm{N}^{-}_{RR}(v_a)_t \neq \emptyset) \wedge (\mathrm{N}_{US}(v_a)_t = \emptyset)\right] \wedge \left[(\mathrm{N}^{-}_{RR}(v_a)_{t+1} \neq \emptyset) \wedge (\mathrm{N}_{US}(v_a)_{t+1} = \emptyset)\right] \qquad \text{Case (b)}$$

To satisfy change pattern 3, the above should also be checked at time $t+1$. Cases (a), (c) and (d) can be detected following a similar approach. The formulas are presented below:

$$\left[(v_a \notin V_{all_{t-1}}) \wedge (\mathrm{N}^{-}_{RR}(v_a)_t = \emptyset) \wedge (\mathrm{N}_{US}(v_a)_t \neq \emptyset)\right] \wedge \left[(\mathrm{N}^{-}_{RR}(v_a)_{t+1} = \emptyset) \wedge (\mathrm{N}_{US}(v_a)_{t+1} \neq \emptyset)\right] \qquad \text{Case (a)}$$

$$\left[\left(\mathrm{N}_{US}(v_a)_{t-1} \neq \emptyset\right) \wedge \left(\mathrm{N}^{-}_{RR}(v_a)_{t-1} = \emptyset\right)\right] \wedge \left[\left(\mathrm{N}_{US}(v_a)_t \neq \emptyset\right) \wedge \left(\mathrm{N}^{-}_{RR}(v_a)_t = \emptyset\right)\right] \qquad \text{Case (c)}$$

$$\left[\left(\mathrm{N}_{US}(v_a)_{t-1} \neq \emptyset\right) \wedge \left(\mathrm{N}^{+}_{RR}(v_a)_{t-1} = \emptyset\right)\right] \wedge \left[\left(\mathrm{N}_{US}(v_a)_t \neq \emptyset\right) \wedge \left(\mathrm{N}^{+}_{RR}(v_a)_t = \emptyset\right)\right] \qquad \text{Case (d)}$$

Example: Let us consider the newcomer M5. If at time $t-1$, $v_{M5} \notin V_{all_{t-1}}$ and the in-neighbourhood of M5 for $ReadRes$ time t is a non-empty set, $N^{-}_{RR}(v_{M5})_t \neq \emptyset$, and the neighbourhood of M5 for

$UploadSim$ is an empty set $N_{US}(v_{M5})_t = \varnothing$, then we need to check the situation at time $t+1$ (two consecutive times of monitoring). If at time $t+1$, $N_{RR}^{-}(v_{M5})_{t+1} \neq \varnothing$ and $N_{US}(v_{M5})_{t+1} = \varnothing$ then M5 is satisfying Case (b).

The following section will present the algorithms developed to capture changes in the behaviour of members due to the interventions triggered.

6.3.2 Detecting Community Changes Occurred after Intervention

After the above cases have been discovered, notifications (the notification generation mechanism is presented in detail in Chapter 7) are sent to community members in order to help them benefit the most from the community and in extending the community to remain active. In this section the algorithms developed to capture any changes in the behaviour of members due to the interventions are presented. The algorithms described below correspond to the discoveries of each of the cases of the previous section. Thus for every change pattern on section 6.3.1 there is a change pattern in this section. In section 6.3.2 we follow a different structure from above since the importance of each pattern has already been discussed. The importance of monitoring this type of changes, (changes occurred after intervention), is a way of evaluating the intervention mechanism triggered due to the detection of change patterns presented in section 6.3.1.

Change Pattern 1: Members moving to periphery.

Cases: (a) A member who is moving to the periphery due to his cognitive centrality as time passes, (b) A former influential member is moving to the periphery due to his cognitive centrality falling compared to previous time point and (c) A member's neighbourhood starts shrinking compared to previous time point.

Detection Case (a): To check whether a given member a has recovered his cognitive centrality we check that $CCen(a)_t > CCen(a)_{t+1}$.

Detection Case (b): The centrality and upload rate have to be checked for member a. If $((CCen(a)_t < CCen(a)_{t+1}) \wedge (uRate(a)_t < uRate(a)_{t+1}) \wedge (CCen(a) > Avg(CCen)))$ then that member is becoming influential again.

Detection Case (c): In this case the neighbourhood of every member for each relationship is being checked whether it has remain at least the same in terms of size as at time t or grown. Let us illustrate using $ReadSim$: $|N_{RS}(v_a)_t| \le |N_{RS}(v_a)_{t+1}|$ similarly it can be done for $UploadSim$ and $InterestSim$. For $ReadRes$ we have to check both the in-neighbourhood and out-neighbourhood of a member. So, $|N^{-}_{RR}(v_a)_t| \le |N^{-}_{RR}(v_a)_{t+1}|$ and $|N^{+}_{RR}(v_a)_t| \le |N^{+}_{RR}(v_a)_{t+1}|$.

Change Pattern 2: Changes in the behaviour of members indicate unexplored relationships.

Cases: (a) two members have similar interests and read similar resources at time $t\text{-}1$, while at time t they continue to have similar interests but no longer read similar resources; (b) at time $t\text{-}1$, two members read similar resources, as well resources uploaded by each other, however, at time t they continue to read similar resources from others but no longer each other's resources; (c) at time $t\text{-}1$ two members are reading resources uploaded by each other and have interest similarity, however at time t they continue to have interest similarity but they are not reading resources uploaded by each other anymore.

Detection Case (a): The intervention triggered based on change pattern 2 detection, aims at encouraging the two members to build $ReadSim$. The algorithm in this case has to check if one of the members is at the neighbourhood of the other's for both $ReadSim$ and $InterestSim$. Thus, $(v_a \in N_{IS}(v_b)_{t+1}) \wedge (v_a \in N_{RS}(v_b)_{t+1})$.

Detection Case (b): In order to check whether the interventions triggered influenced the behaviour of targeted members we check if one member appears in the in-neighbourhood or the out-neighbourhood of the other member for $ReadRes$, thus if $v_a \in N^{+}_{RR}(v_b)_{t+1}$ or $v_a \in N^{-}_{RR}(v_b)_{t+1}$.

Detection Case (c): In this case the in-neighbourhood or the out-neighbourhood of one of the members is checked to include the other member. Thus, $v_a \in N^{+}_{RR}(v_b)_{t+1}$ or $v_a \in N^{-}_{RR}(v_b)_{t+1}$.

Change Pattern 3: Members are not integrating effectively.

Cases: (a) A newcomer has $UploadSim$ but no $ReadRes$ with other members for consecutive time points; (b) A newcomer has $ReadRes$ but not $UploadSim$ for consecutive time points; (c) A member has $UploadSim$ and not $ReadRes$ for two consecutive periods; (d) A member has $UploadSim$ with others, however no one is reading what that member is uploading for consecutive time periods.

Detection Case (a): The intervention triggered for this pattern should encourage this member to start reading resources. Consequently, the algorithm checks if the member in question is a newcomer at time $t-1$ and if the in-neighbourhood of $ReadRes$ for that member is a non-empty set: $\left(v_a \in V_{all_{t-1}}\right) \wedge \left(\mathrm{N}^{-}_{RR}(v_a)_{t+1} \neq \varnothing\right)$.

Detection Case (b): First the algorithm checks if the person was a newcomer at time $t-1$ and then if his neighbourhood for $UploadSim$ relation contains other members, thus he has an uploading similarity with others in the VC. Thus, $(v_a \notin V_{all_{t-1}}) \wedge (\mathrm{N}_{US}(v_a)_{t+1}) \neq \varnothing$.

Detection Case (c): We need to check here if the in-neighbourhood of that member for $ReadRes$ is a non-empty set. This will show that member started reading resources uploaded by other members $\mathrm{N}^{-}_{RR}(v_a)_{t+1} \neq \varnothing$.

Detection Case (d): The algorithm in this case checks that the out-neighbourhood for $ReadRes$ of that member is a non-empty set, $\mathrm{N}^{+}_{RR}(v_a)_{t+1} \neq \varnothing$.

The next section will present the application of the change patterns discussed in section 6.3.1 and gives examples of results extracted. The algorithms have been also applied as part of an evaluation study presented in Chapter 8.

6.4 Application of Algorithms Detecting Community Changes – Study with a Virtual Community

Section 6.4 will present the application of the algorithms presented in section 6.3.1 on the community model obtained from tracking data from BSCW SW VC (see Section 3.4). Section 6.4.1 will present the time points selected. The detection of change patterns will be discussed in Section 6.4.2. Examples from the real data used will be provided along with visual illustrations extracted from NetDraw, where possible.

6.4.1 Data Analysis

As described in Section 3.4, the BSCW VC in this study included 34 members (researchers and doctoral students) from two research groups working on similar research areas, sharing documents and research papers (referred here as resources) with the BSCW system that provides general support for collaboration over the web (Applet, 1999). People were using BSCW to create folders and upload and download resources from the folders created.

We collected log data of 15 months using the BSCW activity tracking features. To ensure the generality of the approach we used only data collected about members using the basic functionality of the system, such as uploading/downloading and naming a resource, which is provided in any virtual community for knowledge sharing.

The algorithms extracted relationships between all 1122 pairs of members. There was a gradual decline in the uploading and downloading of resources in the observed period, and the community has now stopped functioning. The main objective of the study was to find out whether any problems with the community could have been discovered earlier by detecting any of the change patterns defined in Section 6.3.1, what support could have helped this community sustain. A summary of the community activity from the community model (see Figure 3.5) indicated that the community was fairly active before Month3, the activity then had a dip in Month3 but again improved until another dip in Month7, after which the community activity never recovered and the community died.

Hence, we consider that any intervention that could have been helpful to sustain the community should have been done between Month3 and Month7. We therefore applied the community evolution algorithms on the community model derived in Month4, Month5 and Month6, and tried to indicate what interventions could have been done in these months.

6.4.2 Change Patterns Detected

For each change pattern detected, illustrative examples will be provided pointing out how the detected pattern could be used to inform possible support that could have been provided.

Change Pattern 1: Members moving to periphery.

Detection: During the time period Month4 – Month5 three members, M9, M15 and M24, were detected to be shifting to the community periphery. M9 and M24 stopped downloading resources from the community, and thus their centrality dropped (Figure 6.1). On the other hand, M15 used to be one of the influential members who uploaded resources to the VC but completely stopped contributing or downloading. The detection between Month5 and Month6, showed that two other influential members, M2 and M31, shifted to periphery and stopped contributing resources during Month5 and Month6.

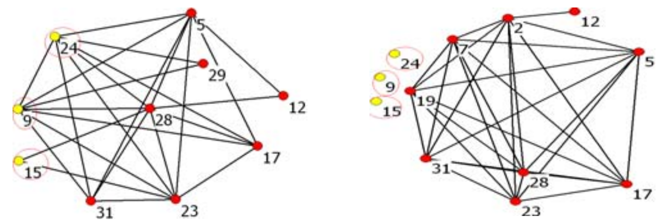


Figure 6.1 M9, M15 and M24 appear to have relationships with others in the community in Month4 (left). During Month5 all members appear inactive with no connections with others.

Furthermore, among others, M9 and M24 appear to have their neighbourhoods shrinking for *ReadSim*, *InterestSim* and the in-neighbourhood of *ReadRes* between Month4 and Month5. Similarly, M28 who also appears to be one of the cognitively central members is detected to have a neighbourhood shrinking for *UploadSim* the in-neighbourhood of *ReadRes*. This means that M28 reduced his relevant uploading and downloading to/from the VC. In the period between Month5 and Month6, member M31 who is one of the CCen members of the VC is detected to have reduced neighbourhood for *InterestSim*, and both in- and out-neighbourhoods of *ReadRes*. This denotes that this member has reduced his reading from the VC and also the resources he is uploading are not of interest to the others. M17 who is the CCen member of this VC appears to have a reduced neighbourhood for *UploadSim* during the same period of time.

Support & Benefits: The detected movement to the periphery indicates that people who could influence the VC could have been encouraged to continue contributing, which could have helped the VC remain active. Having influential members actively engaged often motivates others to engage in the VC too. In this way, the cycle of knowledge sharing could have been kept active to benefit all community members.

Support can be provided to the detected members moving to the periphery in the form of notifications, letting them know about the drop of their *CCen* and trying to motivate them by pointing out how popular the resources they previously uploaded were. This is a way to make members understand how important the knowledge they hold is to the rest of the VC. This can improve the TM of the community since members will become aware of who is interested in what they are interested, and may also promote collaboration.

Additionally shrinking of the neighbourhood of members is either due to their drop of activity (reading/uploading), or for *ReadRes* out-neighbourhood due to others not reading from them. This situation can cause the VC to stop functioning after a short period of time. There is a need to support the detected members in order for the VC to sustain and for them to continue benefiting from their membership.

Interventions can be used based on the data collected from this pattern in order to encourage members of the community to continue being active and contribute/benefit from the VC. By showing to them people with similar interests, or how popular a resource they uploaded is, may encourage them to engage with the VC in a more effective, for them and for the VC, way.

Change Pattern 2: Changes in the behaviour of members indicate unexplored relationships.

Detection: M5 and M23 were detected to have $ReadRes$ and $ReadSim$ in Month4 but appeared to have only $ReadSim$ in Month5. This situation can be seen in Figure 6.2Figure 6.3Figure 6.4.

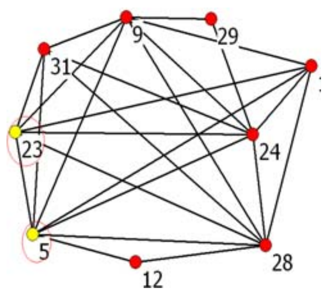


Figure 6.2 $ReadRes_{Month4}$ and $ReadSim_{Month4}$ for both members

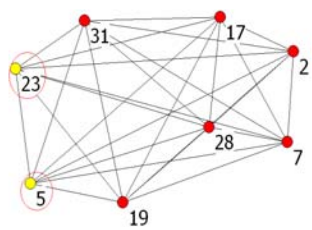


Figure 6.3 Represents the graph for $ReadSim_{Month5}$ whereM5 & M23 appear to have a connection among them

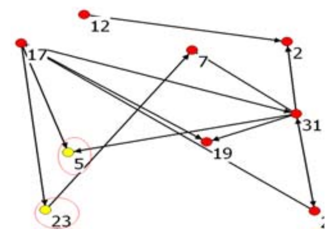


Figure 6.4 M5 and M23 are not connected in the $ReadRes_{Month5}$ graph

In the period from Month5 to Month6, five pairs of members, (M7 and M23, M7 and M31, M23 and M17, M28 and M31, M31 and M17), have been identified to satisfy change pattern 2. These are illustrated in Figure 6.5 and Figure 6.6. Figure 6.6 show that the detected members indeed were reading similar resources and they were reading also resources uploaded from each other during Month5.

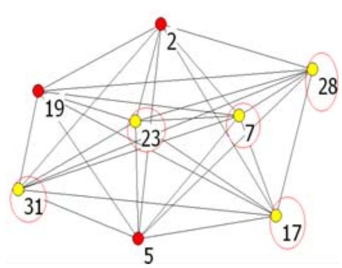


Figure 6.5 The five members and the $ReadSim_{Month5}$ relation they have with others in the community

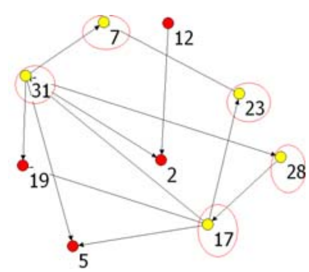


Figure 6.6 $ReadRes_{Month5}$ relation between the detected members

Figure 6.7 represents the $ReadSim_{Month6}$ where members detected above continue to read similar resources. However, they appear not to have $ReadRes_{Month6}$ (Figure 6.8), thus they are not reading resource uploaded from each other. M5 and M28 appear to read similar resources during Month4 and they have also similar interests during that time. In Month5, they stopped reading similar resources but still have similar interests. In the period Month5-Month6, M5 is detected again along with M19. The latter detection is important since both members are newcomers to the community and they need to be supported accordingly.

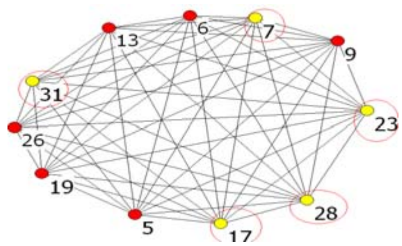


Figure 6.7 $ReadSim_{Month6}$: members detected continue to read similar resources in Month6

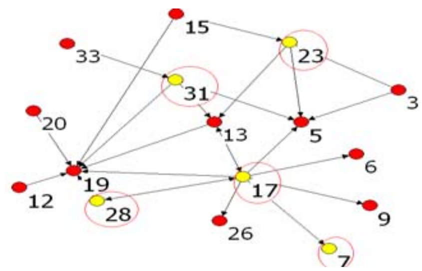


Figure 6.8 $ReadRes_{Month6}$: Members appear not to read resources uploaded from each other.

Support & Benefits: Members detected in this pattern have to be supported and encouraged to continue reading from each other and also to read resources the others are reading. Ignoring what others have uploaded to the VC can lead to missing important resources. The members may not be aware what is happening in the community and may not be aware how their interests/expertise is related to the VC.

The members detected with this change pattern could have been informed of their similarity in terms of interests and reading, and could have been offered recommendations what similar members are reading. This can develop awareness and promote TM (Wegner, 1986). Since members can become aware of who is working on similar knowledge areas, they can discover further opportunities for collaboration.

Change Pattern 3: Members are not integrating effectively.

Detection: Applying this change pattern, M5 was detected as a newcomer for Month4 and appeared to have some *ReadRes* relationships but not any *UploadSim* relationships during Month4 and Month5. This member downloaded 21 resources when he first joined the community in Month 4 and 6 resources during Month5. He appeared not to have uploaded anything during these two months. M5 is also detected in this pattern between Month4 and Month6 when he downloaded 14 more resources without uploading anything to the VC. During Month5 and Month6, M19 was also detected as a newcomer with the above behaviour. M19 downloaded 4 resources in Month5 and 33 resources in Month6, but did not upload any resources in Month6.

Support & Benefits: The behaviour of excessive downloading that both members developed shows that they struggled to find their way to the resources that were interesting and relevant to them. Consequently, support could have been provided to these newcomers by notifying them who works in similar areas as they do, and recommending them important resources. This may help develop SMM in the VC along with possible collaboration between members

Detection: M28 appears to be one of the cognitively central members is detected to have a neighbourhood shrinking for *UploadSim* the in-neighbourhood of *ReadRes*. This means that M28 reduced his relevant uploading and downloading to/from the VC. In the period between February and March 2006, member M31 who is one of the CCen members of the VC is detected to have reduced neighbourhood for *InterestSim*, and both in- and out-neighbourhoods of *ReadRes*. This denotes that this member has reduced his reading from the VC and also the resources he is uploading are not of interest to the others. M17 who is the CCen member of this VC appears to have a reduced neighbourhood for *UploadSim* during the same period of time.

Support & Benefits: Interventions can be used based on the data collected from this pattern in order to encourage members of the community to continue being active and contribute/ benefit from the VC. By showing to them people with similar interests, or how popular a resource they uploaded is, may encourage them to engage with the VC in a more effective way, both for the individual members and for the whole VC.

6.5 Summary

This chapter explored the potential of defining and utilizing community change patterns to identify when intelligent support is needed to support a community to function better as an entity. Change patterns have been described in two categories: detecting changes to aid support to be triggered and detecting changes in members' behaviour due to the support provided. Section 6.3.1 discussed three types of change patterns to aid interventions and provided rational for considering each pattern showing how it can be related to three main processes (TM, SMM, CCen) which are crucial for effective and sustainable VCs. The results extracted from the study with the BSCW VC (section 6.4) show how change detection can be used to identify interventions that may help a VC function better and sustain.

In the study presented here, we have used data from a closely-knit VC operating on the BSCW system. The approach though is generic and is applicable broadly to any closely-knit community for knowledge sharing - a relationship model suitable for the specific community has to be built in the form of graphs and the evolution algorithms can be adapted accordingly. It is important to note that this PhD does not aim to provide an exhaustive list of change patterns that can be discovered in a VC. Change patterns can vary from one community to another according to the community's topic, purpose, and members. Using the basic principle presented in this research, further patterns can be defined.

Chapter 7 will provide detailed description of the support that can be provided due to the changes discovered in this chapter.

Chapter 7

Community Adapted Notifications

7.1 Introduction

This PhD aims to provide intelligent support to VC driven by key processes discussed in Chapter 2 - Transactive Memory, Shared Mental Models and Cognitive Centrality. Chapters 3-6 presented a mechanism for community modelling and algorithms to detect knowledge sharing patterns in a virtual community. This chapter will illustrate how the community model detected patterns can be used to generate community-tailored support. We will presents a mechanism for generating adaptive notifications aimed at supporting TM, SMM, and CCen. Notification messages will be generated to provide personalised information to members assisting them to integrate and benefit from their membership in the VC. Notifications will be generated by using patterns presented in Chapters 5 and 6 and also based on information pulled from a CM following Chapter 4.

The next section will discuss the rationale behind our approach and will give an outline of each notification message. Section 7.4 will provide the formalisation and detailed description of adaptive notifications. Examples using the VC data collected and presented in Chapter 5 and Chapter 6, will demonstrate how the notifications could be used to address community problems.

7.2 Relevant Techniques for Supporting Virtual Communities

There is a growing interest in providing intelligent support for teams, groups and communities. This section will provide an overview of relevant approaches on supporting small groups of people working together over the internet. Although there might be numerous different approaches on providing support to communities and networks online, in this section we have selected the closest and most relevant approaches to the mechanism presented in this PhD.

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), 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.

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). 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.

The closest to our approach is research on intelligent group/community interventions, e.g. notification (Ardissono et al., 2009), feedback (Baghaei and Mitrovic, 2007), or promotion of cognitively central members (Bretzke and Vassileva, 2003; Farzan et al., 2009). 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.

Section 7.3 will provide the rationale for providing community adaptive notifications and explain how these relate to the processes identified as important and followed through this PhD project (TM, SMM and CCen).

7.3 Community Adapted Notifications

The purpose of detecting knowledge sharing behaviour patterns is to assist with providing 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 designed notification messages fall under four categories which target: (a) participation of Cognitively Peripheral Members (CPerM); (b) participation of Cognitively Central Members (CCenM); (c) improving the community TM system; (d) developing SMM. We will provide here the rationale for using this kind of support.

<u>Rationale for CPerM notifications:</u> 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 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 are influential to 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 to gain confidence and become influential. The participation of CCen members may be motivated by acknowledging their importance to the community (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 important knowledge to them and identify who the experts in specific areas are (Wegner, 1986). By providing notification messages that include personalized information, we can help individuals in the VC to become aware of what others are working on, who they are similar to and what resources might be of their interest.

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 a awareness and develops SMM (Mohammed and Dumville, 2001).

A generated 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. Furthermore, the role of the VC members (e.g. student, supervisor, project coordinator), will not be considered since we are allowing for equal membership as defined in the VC characteristics in Chapter 2. 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.

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. Three types of notifications will be defined:

- Type1 - Notifications based on detected knowledge sharing patterns (Chapter 5)
- Type 2 - Notifications based on detected changes through time (Chapter 6)
- Type 3 – Notifications that combine data from the patterns detected and the CM.

This chapter will not provide an exhaustive list of notification messages that can be generated as these can vary according to the type, subject area and number of members of a VC. The notifications provided here are just a sample of what can be generated and how the detections on Chapter 5 and Chapter 6 can be used to generate support for a closely knit VC.

7.3.1 Notifications based on Detected Knowledge Sharing Patterns

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 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) Detected members will be informed of the read, interest or upload similarity they appear to have with the same members but not between themselves. The message will encourage members to read resources the others' are reading and uploading. Links will be 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, interests or uploading. Suggestion of other type of relationships that they might want to develop with these members by providing links to resources these members are uploading or downloading, will be provided. In this message the members' ID will be mentioned and 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 detected member relates to others in the VC and help him integrate. A list of similar members will be provided in the message.

Aim: Develop TM as a member will become aware of how he relates to others.

N1-4(Facilitate a member's integration by showing similar members) Similarly to N1-3 this message will target members who only upload and encourage 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)** message will be sent to a CPerM, or a newcomer, who only downloads and appear to have similar interests to other members. The content of the message will include information on members with a similarity and suggest to that member to start contributing so others can benefit from his 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 is uploading. The content of the message will include information on members with a similarity and suggest to that member to start contributing so others can benefit from his knowledge.

Aim: Similar to N1-6, supports a CPerM to integrate and develops TM and SMM.

N1-7 (Inform members of complementary similarity they have) Inform a group of members of their complementary similarity and show them resources uploaded by each other. Links to resources and members' id will be included in the message.

Aim: Improve TM and SMM by providing information on the similarity between members.

7.3.2 Notifications based on Detected Changes through Time

Notifications sent to VC members due to the changes detected in the members' status or relationships will inform members of these changes and provide suggestions in order for the whole VC to benefit. This section provides the outline for Type 2 notifications below.

N2-1 (Motivate a former CCenM who is moving to periphery)*:* Detecting a CCen member who is moving to the periphery will generate a message encouraging that member to start uploading and downloading again by explaining how important he used to be for the VC.

Aim: Support a CCenM and improve TM in the VC by informing a member of the value his resources have.

N2-2 (Motivate a member who stopped downloading): A member moving to periphery (CPerM) due to stop downloading will receive a message which will provide information on the impact the resources he uploaded previously had to the VC and encourage that member to remain active. The CCen drop of this member will be mentioned in the message as a mechanism for motivation. **Aim:** Support a CPerM to integrate and develop TM and SMM.

N2-3 (Motivate a member who stopped uploading) Similarly to N2-2 members detected in this pattern need to be motivated to start downloading, by showing to them members with similar interests and resources uploaded by those members.

Aim: Similar to N2-2 support will be provided to CPerM to help develop TM and SMM.

N2-4 (Notify members they might be missing important information) will target two or more members who used to have a relationship. The message will provide awareness of the similarity these members had and encourage them to keep this relation active. Information from the relationships model will be used and the IDs of all the detected members will be mentioned in the message.

Aim: SMM and TM can be developed due to the awareness this message will provide.

N2-5 (Motivate a newcomer to contribute): A newcomer who is detected to download only for a period of time will generate a message which will provide information on members with similar interests to him and who might be interested in the knowledge this newcomer holds.

Aim: Develop TM and help a CPerM integrate.

N2-6 (Inform members of resources they might be missing) This message will target a group of members who appeared to have read and interest similarity at some time point and only have interest similarity the next. It will attempt to provide awareness of the similarity the members detected have with each other and provide links to resources each other is reading. Detected members' id will be provided in this message along with links to resources.

Aim: Improve TM and SMM with the information that will be provided.

N2-7 (Inform members of resources uploaded by similar members) With this notification a group of members who used to read resources uploaded by each other and have similar interests will be targeted. It will inform members of the similarity they have in terms of interests and encourage them to read resources uploaded by each other. Links to resources will be provided along with the ids of the detected members.

Aim: Develop TM and improve SMM in the VC by providing awareness.

N2-8 (Encourage a member to benefit from the VC) will be sent to a member who is detected to only uploading for two consecutive periods. Suggestions of people with similar interests will be provided along with information on resources that might be of that person's interest.

Aim: TM can be developed since CPerM will be supported and become aware of members with similar interests and where interesting for them resources are stored.

N2-9 (Promote a member in the VC) The detection for this notification will be a member who uploads resources but other members are not reading them. The notification message will target the set of members who appear to have a similarity in uploading ($UploadSim$) with the detected member, inform them of the similarity they have with that member and suggest resources that member is uploading for them to read.

Aim: Improve TM since others will become aware of what that member is uploading.

7.3.3 Notifications based on Combined Data from the Community Model and Detected Patterns

Further information stored in the CM will also be used in order for notifications to be more effective. Type 3 notifications combine data from both the CM (Chapter 4) and the patterns presented at Chapter 5 and Chapter 6.

N3-1 (Exploit an important CCenM) it is used to inform a CCen member of how important he 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 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.

N3-2 (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 7.1.

Aim: Motivate a CCenM and support CPerM to integrate. Develop TM and SMM.

N3-3 (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.

7.4 Generating Notification Messages

In the CM application phase, identifying a behaviour or change pattern is the first step in generating notification messages to the VC members. The notifications target individual or group of members in the VC who were detected in one of the behaviour or change patterns presented in Chapter 5 and Chapter 6. Consequently, the input for the generation of a notification will be the output of a pattern. Each notification has a goal, which relates to one of the categories defined in Section 7.3, CPerM, CCenM, TM and SMM. The content of each notification message is designed to promote the goal of the notification and follow the definitions in Sections 7.3.1 and 7.3.2. A third category of notifications that includes combination of the detected patterns and information extracted on the CM is also considered. Two different formats of personalised notification messages have been generated. In the first set of notifications, only the links to the folder structure and the read history of the VC have been sent. The second set of notifications personalized information relevant to a given member (e.g. resources that might be of interest, members who have similar interests with that member), are provided in the notifications. The template of the message generated remains the same (see Table 7.1), only the links included in the message differ.

Section 7.4.1 will give a description of the formalisation of the notification messages generation mechanism and introduce the notation used.

7.4.1 Formalisation of Adaptive Notifications Mechanism

For every notification message a standard structure is followed:

- *Detection* – The situation that triggers the notifications e.g. knowledge pattern detected.
- *Target Users* - the list of community members detected at a given pattern to whom a notification should be sent.
- *Goal* – defining the aim of the notification, related to TM, SMM, and CCen
- *Content Template* - pattern of the text that will be sent.

Table 7.1 provides the definitions of the notifications that will be generated with the detection of a pattern. In Table 7.1 the following notation is 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, $\mathcal{M}' = \{M_{i_1}, M_{i_2}, ..., M_{i_n}\}$ such that $\mathcal{M}' \subseteq \mathcal{M}$ which represents the members extracted in a detection. In $\mathcal{M}'$ a member $\{M_{i_j}\} \in \mathcal{M}'$ for all j. Additionally, three more sets have been derived by applying the algorithms presented in Chapter 4 $CCenM$ (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), and are represented as $CCenM \subseteq \mathcal{M}$ $CPerM \subseteq \mathcal{M}$, and $Newcomers \subseteq \mathcal{M}$. Furthermore, if a notification can be generated for more than one detection, the $< RelationshipType >$ is used to indicate the type of relationship from the CM.

Table 7.1 Definitions of Notification Messages. Pages 100 -105

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 $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have $< \mathit{RelationshipType} >$ with the same members but not among themselves.	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	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* $< \mathit{RelationshipType} >$ *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:"*
N1-2 (Inform members of their similarity)	P2: Members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have a $< \mathit{RelationshipType} >$	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Provide awareness of similarity between detected members. Improve TM and SMM.	For every M_{i_j} : *"You appear to have a* $< \mathit{RelationshipType} >$ *similarity to* $T \setminus \{M_{i_j}\}$. *You may find it helpful to see the resources these members are uploading and downloading. Follow the links below to navigate through the resources."*
N1-3 (Guide member's integration by showing similar members)	P3: Member $\{M_{i_j}\} \in \mathcal{M}$ is downloading only $\{M_{i_j}\}$ has a $< \mathit{RelationshipType} >$ with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Develop awareness of how the member relates to others and help him integrate. Develop TM.	*"Share your knowledge with the rest of the community by start uploading resources.* $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ *have* $< \mathit{RelationshipType} >$ *with you and will benefit from what you share with them."*

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Notifications Based on Knowledge Sharing Behaviour Patterns				
N1-4 (Guide member's integration by showing similar members)	P4: Member $M_{i_j} \in \mathcal{M}$ is uploading only M_{i_j} has a $< RelationshipType >$ with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Develop awareness of the member relates to others and provide information on where resources important to him are located. Develop TM.	*"You have $< RelationshipType >$ similarity with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$. You may find what they are uploading interesting and useful. Follow the links to navigate through resources these members are uploading"*
N1-5 (Guide a CPerM who is downloading only, to integrate)	P5: $\{M_{i_j}\} \in CPerM$ is downloading only $\{M_{i_j}\}$ has a $< RelationshipType >$ with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Help a CPerM integrate by acknowledging their importance and referring to similar members in the VC.	*"You appear to have $< RelationshipType >$ similarity with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$. Share your knowledge with these members by start uploading resources. They will benefit from what you share with them as you are benefiting from what they share with you"*
N1-6 (Guide a CPerM who is uploading only, to integrate)	P6: $\{M_{i_j}\} \in CPerM$ is uploading only $\{M_{i_j}\}$ has a $< RelationshipType >$ with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$		*"$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ find what you are uploading very interesting. You may find what they are uploading interesting and useful. Follow the links to navigate through resources these members are uploading"*

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Notifications Based on Knowledge Sharing Behaviour Patterns				
N1-7 (Inform members of complementary similarity they have)	P7: Members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have ReadSim but not UploadSim	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Improve TM and SMM by informing members of the similarity they appear to have and the relationships they can develop among themselves	For every M_{i_j} : *"You are reading similar resources as* $T \setminus \{M_{i_j}\}$ *. You might be interested in what this member is uploading. Follow the links below to navigate through the resources uploaded by these members."*
Notifications based on Changes through Time				
N2-1 (Motivate a former CCenM who is moving to periphery)	CP1: $\{M_{i_j}\} \in CCenM$ is moving to periphery.	$\{M_{i_j}\}$	Let member know of their CCen drop and motivate them to become active again. This can influence TM system development and promote collaboration.	*"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."*
N2-2 (Motivate a member who stopped downloading)	CP2: Member $\{M_{i_j}\} \in \mathcal{M}$ is moving to periphery due to stop downloading.	$\{M_{i_j}\}$	Motivate members to continue benefiting by providing links to relevant to them resources available.	*"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. "*
N2-3 (Motivate a member who stopped uploading)	CP2: Member $\{M_{i_j}\} \in \mathcal{M}$ is moving to periphery due to stop uploading.	$\{M_{i_j}\}$	Motivate member to share knowledge with the rest of the VC.	*"Resources you have previously uploaded have been very useful to other members. They have read your resources. Continue sharing with others and keep your centrality up."*

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Notifications based on Changes through Time				
N2-4 (Notify members they might be missing important information)	CP3: Members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have $ReadRes$ and $ReadSim$ at time $t\text{-}1$ but at time t they do have only $ReadSim$ and not $ReadRes$	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Let members know of the similarity they appear to have and inform them of resources uploaded by each other. Develop TM and SMM.	For every M_{i_j} *"You appear to read similar resources as* $T \setminus \{M_{i_j}\}$. Check what you have missed using the links provided below.*"*
N2-5 (Motivate a newcomer to contribute)	CP4: A newcomer $\{M_{i_j}\} \in Newcomer$ has $ReadRes$ but not $UploadSim$ for two consecutive periods of monitoring with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Acknowledge the importance of the knowledge a newcomer holds and motivate him to contribute by showing members with similar interests. Develops TM and helps a CPerM to integrate.	*"You appear to have* $< RelationshipType >$ *similarity with* $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$. *Share your knowledge with these members by start uploading resources. They will benefit from what you share with them as you are benefiting from what they share with you "*
N2-6 (Inform members of resources they might be missing)	CP5: Members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have $ReadSim$ and $InterestSim$ at time $t\text{-}1$ but at time t they do have only $InterestSim$	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Create awareness of the similarity the detected members have and provide links to resources each other. Improve TM and SMM with this notification.	For every M_{i_j} *"You appear to have similar interests to* $T \setminus \{M_{i_j}\}$. *Keep reading what these people are reading following the links below."*
N2-7 (Inform members of resources uploaded by similar members)	CP6: Members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ have $ReadRes$ and $InterestSim$ at time $t\text{-}1$ but at time t they do have only $InterestSim$	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Inform members of the similarity they have with others in term of interests and encourage them to read resources uploaded by these members to redevelop $ReadRes$	For every M_{i_j} *: "You appear to have similar interests with* $T \setminus \{M_{i_j}\}$. *Use the links below to see what these people are uploading and benefit from their knowledge"*

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Notifications based on Changes through Time				
N2-8 (Encourage a member to benefit from the VC)	CP7: Member $\{M_{i_j}\} \in \mathcal{M}$ has $UploadSim$ but no $ReadRes$ with members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ for two consecutive periods $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Inform the detected member of the uploading similarity he appears to have with others and suggest resources to read. Help a CPerM to integrate and improve TM.	*"You appear to have an upload similarity with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$. Start benefiting from what these people are uploading. Follow the links below to read resources that might interest you"*
N2-9 (Promote a member in the VC)	CP8: Member $\{M_{i_j}\} \in \mathcal{M}$ has $UploadSim$ with members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ and no one is reading what that member is uploading after two consecutive periods $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	Let members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ know of the similarity they have with member M_{i_j} and suggest resources M_{i_j} is uploading for them to read.	*"You appear to have an uploading similarity with M_{i_j}. Follow the links below to read resources uploaded by that member."*
Combination of Patterns and Information from CM				
N3-1 (Exploit an important CCenM)	$P_1 \wedge \{M_{i_j}\} \in CCenM$	$\{M_{i_j}\}$	Let a CCenM of his importance in the VC, encourage him to continue and suggest he pairs with less active members to help them integrate.	*"You are an important member connecting $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\} \setminus \{M_{i_j}\}$ Keep 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*

Type	Detected Situation	Target Members (T)	Notification Goal	Content Template
Combination of Patterns and Information from CM				
N3-2 (Pair a CCenM with a CPerM)	$\{M_{i_1}\} \in CPerM \wedge \{M_{i_2}\} \in CCenM \wedge$ $RelationshipType(M_{i_1}, M_{i_2})$	$\{M_{i_1}\}$	Let a CPerM know of his 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."* Also generate: *N1-1 or N1-2*
		$\{M_{i_2}\}$	Let a CCenM know of how important he is in this VC and his relationship with a CPerM. Suggest to contact this member to help him integrate.	*"You are uploading very interesting resources for this VC. You are a very important member and you can help others through their journey in this VC. Use the links provided to contact $\{M_{i_1}\}$ who appears to have a $<RelationshipType>$ with you to help him."* Also generate: *N1-1 or N1-2*
N3-3 (Welcome message to newcomers)	$\{M_{i_j}\} \in Newcomers$ $\{M_{i_j}\}$ has an $InterestSim$ with $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ where $\{M_{i_j}\} \notin \{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$	$\{M_{i_j}\}$	Inform a newcomer of people with similar interests in order to help that member start benefiting from the VC. This helps the integration of newcomers and the development of TM.	*"Welcome to the community! Based on the information you have provided, the following members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ might have uploaded resources that can be of your interest. Use the links below to read their resources."*

Using the defined notifications this research aims at providing support to a knowledge sharing VC in three stages of its life: start-up, grow and sustain as defined in Chapter 2:

Start-up: This stage can be supported by providing information to members on how they relate with others in the VC to help them start benefiting from what is available (e.g. N1-1, N1-3, N2-5, N3-3).

Grow: Helping the VC to grow can also be achieved through the defined notifications, since members are developing a TM and are establishing SMM. Notifications that provide information on what other members are doing in the VC, how they can relate to others and where important information is located (e.g. N1-5, N1-7, N2-4, N3-1).

Sustain: Sustainability of the VC can be achieved by monitoring the activity of members and providing awareness and motivation for them to remain active as long as possible. Notifications can provide (e.g. N1-2, N1-4, N2-1, N2-8, N3-2).

The following section will provide examples of the above defined notifications.

7.5 Example Community Adapted Notifications

In order to set the above definition in context, this section is demonstrating how notifications will be generated to VC members (see Table 7.2). The examples are based on detected patterns from the SW VC studies in Chapter 5 and Chapter 6, and show how the notifications defined in Section 7.4 could have been used in order to address problems in that community[11].

The examples below illustrate what could have been done to prevent community from dying. However, we cannot check what the effect of these examples could have been. In order to examine the effect of notifications, we need to apply them to a real community, which will be presented in Chapter 8.

---

[11] Note that the Table 7.2 does not present an exhaustive list of possible notifications that can be generated.

Table 7.2 Examples of Adaptive Notifications Using Real Data

Detection	Target Members	Notification Goal	Notification Message
Knowledge Sharing Behaviour Patterns			
Members $\{M5, M9, M24, M31\}$ have *ReadSim* with $\{M28\}$ but they do not have *ReadSim* among themselves.	$\{M9\}$ The same message will be sent to $\{M24, M31, M5\}$	Inform the targeted users of the similarity they have and encourage them to read resources the others' are reading. Develop TM and SMM.	**N1-1***"Did you know you have a* $<RelationshipType>$ *similarity with* $M5, M24, M28, M31$. *You may find it helpful to check the resources these members are reading and uploading. Follow the links below:"*
	$\{M28\}$	Let a CCenM of his importance in the VC, encourage him to continue and suggest he pairs with less active members to help them integrate.	**N3-1***"You are an important member connecting* $M5, M9, M24, M31$ *Keep the good work and upload more interesting resources. Can you suggest resources that these members may read? You may wish to contact each member"*.
A CPerM, $\{M33\}$ is uploading only in the VC, while $\{M31\}$, a CCenM is interested in what $\{M33\}$ was uploading.	$\{M33\}$	Help a CPerM integrate by acknowledging the knowledge he holds and provide information of similar members in the VC	**N1-6** *"* $M31$ *find what you are uploading very interesting. You may find what they are uploading interesting and useful. Follow the links to navigate through resources these members are uploading"*
	$\{M31\}$	Let a CCenM know of how important he is in this VC and his relationship with a CPerM. Suggest to contact this member to help him integrate.	**N3-2***"You are uploading very interesting resources for this VC. You are a very important member and you can help others through their journey in this VC. Use the links provided to contact* $M33$ *who appears to have a* $<ReadRes>$ *with you to help him."*
Changes through Time			
$\{M5\}$ and $\{M23\}$ were detected to have *ReadRes* and *ReadSim* in Month4 but appeared to have only *ReadSim* in Month5	$\{M23\}$ Same message will be sent to $\{M5\}$	Let detected members know of the similarity they appear to have and inform them of resources uploaded by each other. Develop TM and SMM.	**N2-4** *"You appear to read similar resources as* $M5$. *Check what you have missed using the links provided below."*
$\{M19\}$ detected as a newcomer and downloading only for two consecutive time periods.	$\{M19\}$	Acknowledge the importance of the knowledge of a newcomer. Motivate him to contribute by showing members with similar interests	**N2-5***"You appear to have* $<ReadSim>$ *similarity with* $M13, M17, M28, M31$. *Share your knowledge with these members by start uploading resources. They will benefit from what you share with them as you are benefiting from what they share with you "*

7.6 Summary

In this chapter the rationale for adaptive notification generation has been given, namely notifications for CCenM, CPerM, for the development of TM and the establishment of SMM. The definition of each notification has been discussed in detail. The formalisation of adaptive notification mechanism provided an insight into how and why a notification is generated. Examples of how the notifications could have been generated to VC members have been given. An application of the defined notifications will be provided in Chapter 8 where the CM acquisition and application algorithms will be used to provide support to a functioning VC. This will enable us to evaluate the use of adaptive notifications, focusing on the four categories presented in Section 7.3.

Chapter 8

Evaluation of Adaptive Notifications

8.1 Introduction

In the previous chapters we presented algorithms for the extraction of a CM, the identification of static behavioural patterns, and patterns indicating how a knowledge sharing community changes through time. Each component has been validated via a corresponding case study using archival data from an existing VC. Chapter 7 presented an output of the adaptation framework - a mechanism for generating notifications to community members, and provided details of when and what messages will be sent according to the CM. In this chapter, we will validate the overall framework by evaluating the output – the notification mechanism presented in Chapter 7 - with an active knowledge sharing community. This will enable us to examine the effect notification messages may have on the behaviour of community members, and how this can affect the functioning of the community in general.

The chapter starts with a review of relevant evaluation approaches to justify the selected evaluation method. An experimental study is then presented. We outline the evaluation methodology referring to the overall objectives and research questions and making a brief reference to the formative evaluation done over archival data from the SW VC. The summative evaluation of the notification messages is then presented; the results obtained are reported and discussed.

8.2 Evaluation Approaches

Evaluation is an important part of the development of adaptive systems. This ranges from checking that the components in an adaptation framework work as intended to identifying possible benefits and drawbacks from the overall approach. We will review relevant evaluation approaches, and will outline the main method used in this PhD.

8.2.1 Relevant Evaluation Approaches

This section will present general approaches that have been used in evaluating adaptive social systems and awareness systems. Methods vary according to what is evaluated, the number of users involved, and the duration of the evaluation studies.

Common methods used for evaluating intelligent systems that integrate adaptive features include control group studies (comparing versions of the system “with” and “without” adaptivity) and simulations (evaluating the system with simulated users or archival data). Adaptation in social environments is a new trend, and there are just few examples of evaluation studies. These studies utilise either *control group* or *simulation-based* methods. For example, Cheng and Vassileva (Cheng and Vassileva, 2006) use a control group study to evaluate an adaptive reward mechanism aimed at encouraging more valuable contributions from users in an educational community. Similarly, in (Farzan et al., 2009) a control group study has been conducted to evaluate a reward mechanism that helps users to discover new content and promote social interactions in a VC. In both systems, two groups of users were employed - one using a system with adaptivity features and another using the same system but without the adaptive features. An alteration of the control group based method will be consider here – taking a system without adaptation features and analysing the user knowledge sharing behaviour with that system, we will gradually include adaptive notifications and will examine what effect this can have on the use of the system and the community as a whole.

Simulation of users is preferred as a method of evaluation when large amount of data is needed and data is too expensive to collect, or when people have to be involved and there is no available sample (Vanlehn et al., 1996). This method has been used, among others, in user modelling (Millan and Perez-De-La-Cruz, 2002; Shlomo et al., 2007) and social networks system (Menges et al., 2008; Stocker and Larkin, 2008) where modelling the individual but also the relations among people is a crucial part of the evaluation. A method inspired by simulation-based evaluation was utilised in this thesis – we used archival data from a real community to validate the algorithms proposed for community modelling and knowledge sharing patterns detection.

There is an emerging consensus among the user modelling community that evaluation of adaptive systems should be conducted in a layered manner, component by component (Paramythis and Weibelzahl, 2005), following the *layered evaluation* approach (Karagiannidis and Sampson, 2000). The first layer focuses on evaluating whether the information extracted about the user is correct and relevant and ensures that the user model is adequate for the adaptation. The second layer evaluates the adaptation mechanism – are the adaptation decision valid and accurate. Applications of the layered evaluation approach confirm its benefits over the traditional “as a whole” evaluation where adaptive systems are evaluated without separating the user model and adaptivity decision parts (Brusilovsky et al., 2004).

A similar approach is the two stage *formative* and *summative* evaluation technique that has been widely used in both intelligent tutoring environments and user-adaptive systems (Mark and Greer, 1993; Kosba et al., 2007). In the formative stage, the system is evaluated for its usability and effectiveness of each of its components. Drawbacks in the algorithms and system performance have to be pointed out and solved. In the summative stage, which usually takes place in real settings, the effectiveness of the overall system is evaluated. Both formative and summative evaluation stages will be conducted to evaluate the framework developed in this PhD.

8.2.2 Formative Evaluation

The formative evaluation phase provides useful information with respect to the validity of the developed algorithms and points out required modifications. This part of the evaluation has been done in the form of three case studies using the SW VC, as discussed in Chapter 4, Chapter 5, and Chapter 6. In the first case study (Chapter 4), the CM acquisition algorithms were validated. A model of the SW VC (including building relationship graphs and individual user models) was derived and examined with a visualisation tool. During this case study, important patterns, relevant to TM, SMM, and CCen were identified manually by the researcher (who was a member of the VC). In the second case study, the graph-based algorithms for deriving static knowledge sharing patterns in a VC were applied to the extracted community model (Chapter 5). The automatically derived patterns were validated comparing them to human derived patterns. In the third case study, the graph-based algorithms for extracting changes through time were applied to the community mode of the SW VC (Chapter 6). Similarly to the previous studies, the derived patterns were validated by looking at specific cases in the VC.

The key strength of our formative evaluation studies is the use of long term authentic data. We consider that using a community which did not sustain is an appropriate choice for formative validation of the framework - the problems did exist and we could see if the algorithms could detect them; there were no interventions or any experimental conditions while the community was active. The members used a popular knowledge sharing platform as their main way of sharing papers both on their projects and between themselves.

The formative evaluation via each case study with the SW VC was discussed in the corresponding chapter. The remaining part of this chapter presents our summative evaluation.

8.2.3 Summative Evaluation

The summative evaluation phase focused on the effect of the adaptive notifications to individual members (oldtimers and newcomers) and to the VC as a whole (knowledge sharing). The overall framework was employed in a real VC to derive a CM, extract knowledge sharing patterns and detect knowledge sharing changes through time, and generate adaptive notifications sent to individual members via email. It is important to note that the VC used in this phase is different from the semantic web VC used in the three formative evaluation case studies in chapters 4-6 (that community is no longer active). Hence, the notification mechanism was evaluated with an active community.

Aim and questions addressed in the summative evaluation

The aim of the evaluation is to identify the effects of the notification mechanism so it can be employed in supporting virtual knowledge sharing communities. The summative evaluation will *examine 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 will be addressed in this second part of evaluation:

Effect on the community as a whole: *what change patterns can be recognised after notifications are received; 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; do oldtimers' activity change as a result the adaptive notifications they receive?*

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; do newcomers' activity change as a result of the adaptive notifications they receive?*

8.3 Experimental Study Outline

To validate the notification mechanism, the algorithms were employed to extract patterns in a real community which the author belongs to. This section provides information about the community, outlines the experimental study, and presents the stages followed.

8.3.1 General Information about the Case Study

Community: The community included *15 members* (researchers and doctoral students) from different research groups working on similar research topics around *Personalisation and Intelligent Knowledge Management*. The members were working on different projects and some of them participated in joint seminars. The participants were asked to use BSCW share resources between themselves by creating folders and uploading/download resources. Members were based in two countries (UK and the 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). 11 out of 15 members in this community are research students working on separate projects, 2 members act as research supervisors (M2 and M6) for the research students participating in the community, and 2 members are active researchers in their fields (M4 and M5) but are not directly engaged with supervising students from this community (both members worked at remote geographic location and had not met with most of the other members). Furthermore, since there existed a physical community and most people were communicating and exchanging ideas, the BSCW virtual space (hereafter referred to as the VC) was created for them to have an online environment for sharing and storing resources. Gradually members began to embrace the idea and they were using BSCW workspace to share their resources and collaborating using the VC.

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 where several phases of high activity and times when there was no activity in the VC (although activity in the physical community continued). Figure 8.1 shows the activity in the VC prior the study. The dotted lines indicate four drop periods of uploading activity. 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 tend to 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 since M5 was not part of the physical community). This situation demonstrates lack of SMM and TM system within the VC and between VC members.

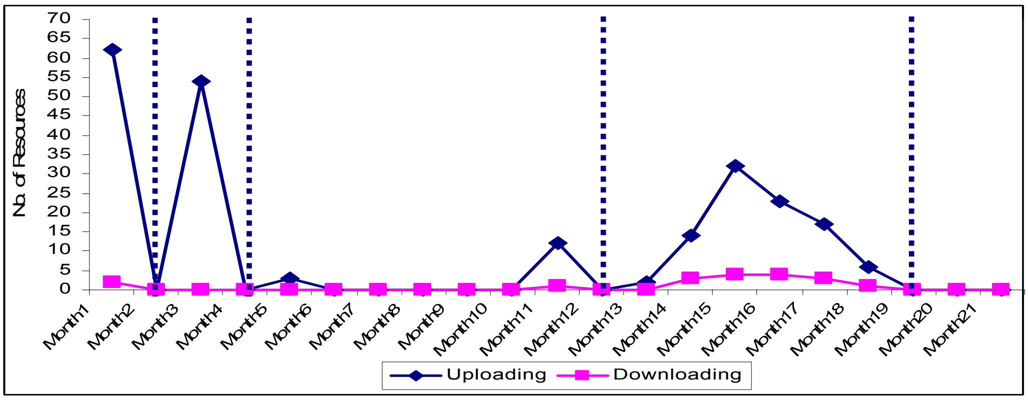


Figure 8.1 Uploading and downloading activity during the pre-study period. Note the dotted lines show drop of activity. Before the study began, there was no uploading/downloading activity in the VC.

The VC had some duplicating resources (Table 8.1). In five out of the six occurrences, M7 has 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 resource was already there, and also that M7 either did not search for the resource or could not find them in the VC. It is important to note that M7 was the most CCenM of this VC before the experimental study begun. M2 also uploaded a resource that was previously uploaded by M13. This is an interesting situation since M2 is the supervisor of M13 and neither of the two members knew that the other member had looked at the specific resource and uploaded it in the shared space. Also M2 uploaded the specific resource when working with M5 on a different project but other members were unaware of the activity of these two members. The above occurrences manifest lack of TM and SMM in both the physical and virtual community.

Table 8.1 Duplicated resources before the experimental study. First column shows the resource Id in database, second column is the Id of the member who first uploaded the resource, third column is the member id that re-uploaded the resource and the last column shows the title of the duplicated resource

RId	Member ID (1st submission)	Member ID (2nd submission)	Resource Title
37	M6	M7	Expert Recommender: Designing for a Network Organization
23	M2	M7	An ontology for supporting communities of practice
131	M13	M2	Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report
31	M13	M7	tagging, communities, vocabulary, evolution
54	M13	M7	Blog Community Discovery and Evolution Based on Mutual Awareness Expansion
50	M13	M7	Users in Volatile Communities: Studying Active Participation and Community Evolution

The above examples show the need for some intervention to provide better awareness of links with VC members who are remotely collaborating. Even when people are 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 PhD 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. The data analysis combined data collected from the questionnaires with data extracted in the CM (e.g. participation of members, CCen, relationships).

Data: *Objective data* was the log data (screenshots of the interface the data collected from are presented in Appendix C) collected over the duration of the study (2 months) using the BSCW activity tracking features. Similarly to the studies in chapters 4-6, 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 (see Chapter 4). The Java algorithms for community modelling, pattern detection (both static and change), and notification generation were run.

In addition to the tracking data, *subjective data* was collected using questionnaires (Appendix B) that combined open ended questions, choice questions (multiple and single answers), and alternative selection questions. 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 researcher (see Section 8.3.3).

8.3.2 Stages of the Experimental Study

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 CM acquisition mechanism was employed to extract an initial CM, based on which the behavioural pattern algorithms (Chapter 5) were applied. The change pattern algorithms have not been used at that point since there was no activity just before the study began, and it was not feasible to apply these algorithms.

During the **Second Period**, the knowledge sharing patterns were used to decide what notification messages should be triggered (follow the first format of messages as described in Section 7.4) 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. A week after the email notifications were sent, a *second questionnaire* was sent to all oldtimers. This enabled to examine the effect of the first format of notifications to oldtimers. The data extracted from the questionnaire along with log data (application of static patterns (Chapter 5) and change patterns (Chapter 6)) 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.

In the **Third Period,** based on data extracted from the newcomers' initial questionnaire, the data in the derived CM, and the application of the static knowledge sharing patterns (Chapter 5), 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 the algorithms for detecting change patterns (Chapter 6) 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 7.4). 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 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 CM (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 to the VC as a whole.

A summary of the experimental study timescale is presented in Figure 8.2.

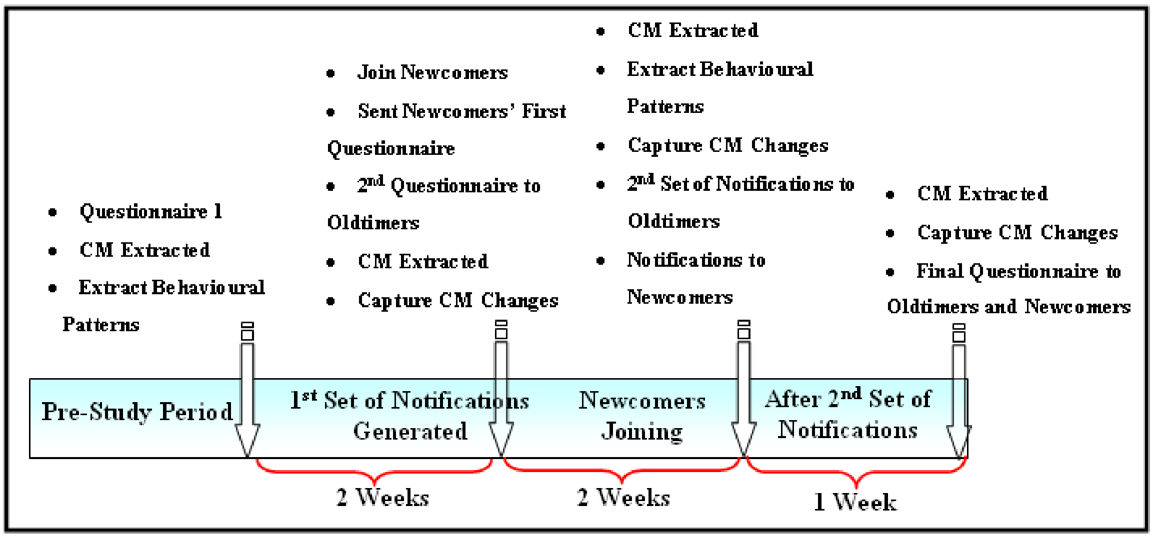


Figure 8.2 Experimental study timeline.

8.3.3 Data Analysis

The results extracted from the questionnaires were statistically analysed in order to understand the effect of notifications. The first 10 questions in all questionnaires asked members to list three related community members (two members in one question), such as people whom they would read papers from, people who may benefit from papers they upload, etc. (Appendix B). 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. Similarly, the CM indicated a set of members found to be related to each member. Figure 8.3 illustrates the combination of both sets. $\mathcal{A}$ denotes the top most similar members to a given member as derived in the CM. $\mathcal{C}$ 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 questionnaire replies) and in the CM 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 CM for this member over the number of selections suggested in the CM: i.e. $P = \frac{|B|}{|C|}$[12]. 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 CM 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): $F1 = 2 \times \frac{P \times R}{P + R}$.

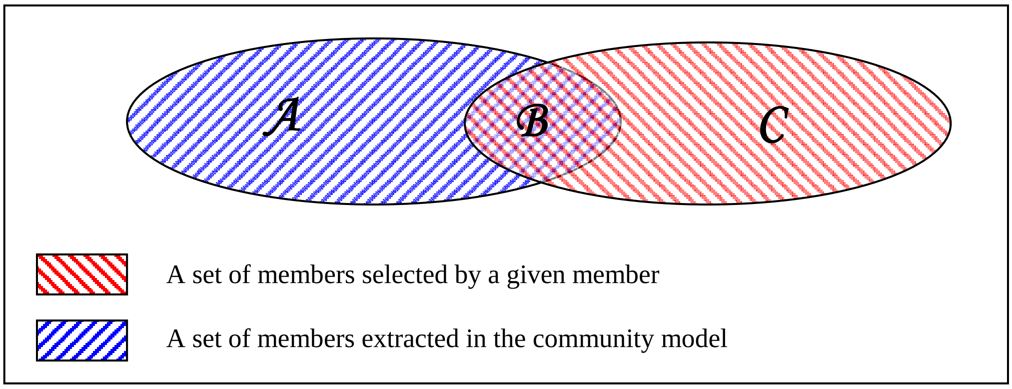


Figure 8.3 The three sets considered in our metrics.

$F1$ is computed for the replies to every question in each of the questionnaires. Statistical *Wilcoxon non-parametric test* is applied to compare the mean $F1$ scores before and after the notifications.

8.4 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.

8.4.1 Effect of Notifications on the Community

In this section general facts for the community are presented and the effect of the notifications to the community in general has been examined.

The CM algorithms extracted relationships between all pairs of members. The activity monitored 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. 237 resources have been uploaded over a period of 23 months. During the second period (after the first set of notifications generated for oldtimers), there was no uploading. With

[12] where $|S|$ denotes the cardinality of a set S

the second set of notifications, we indicated resource uploading from a newcomer (M15) in the third period and from an oldtimer (M2) in the fourth period (note that M2 was inactive till the end of the second period). This minimal uploading activity indicates that after the second round of notifications, the VC was “waking up”.

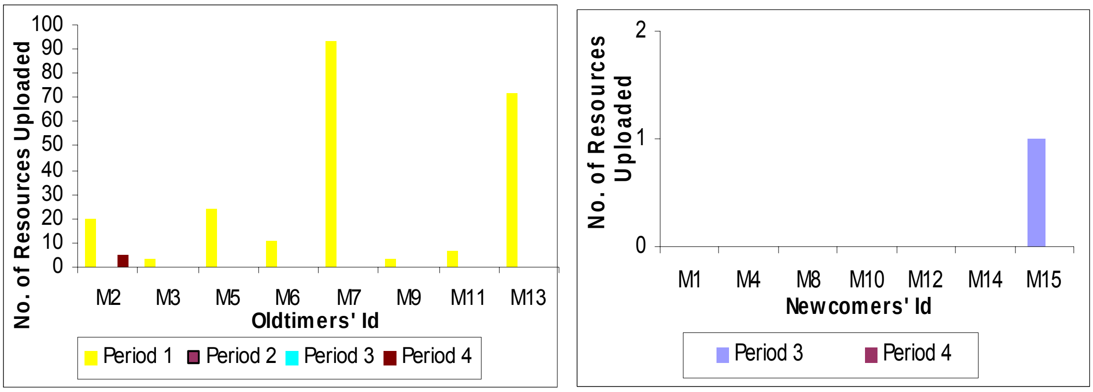


Figure 8.4 Uploading activity in the VC from the creation of the VC until the end of the experimental study. The left figure shows the oldtimers activity and the right the newcomers’ activity.

Downloading (Figure 8.5) took place during all periods except the second period. During the third period, when 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 (M9, M2, M13), as well as newcomers (M14, M15) had downloaded resources from the VC.

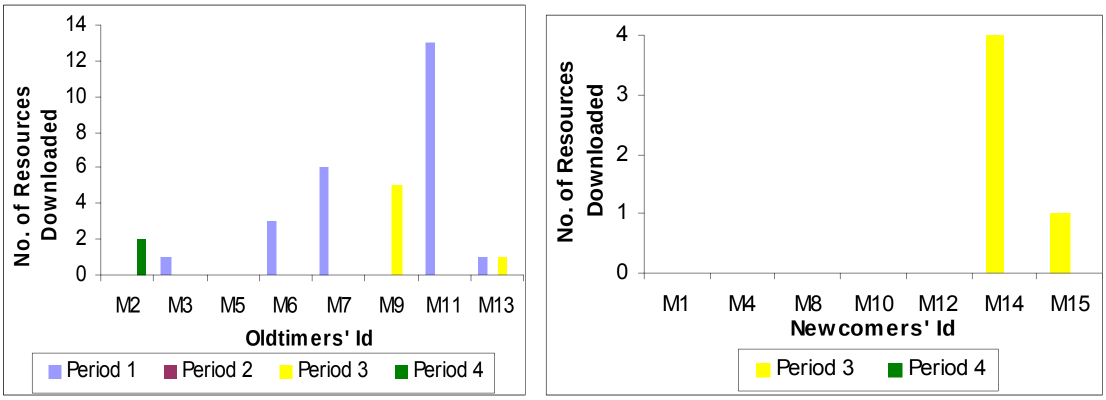


Figure 8.5 Downloading activity of VC members from the creation of the VC until the end of the experimental study. The left figure shows the downloading of oldtimers, and the right figure shows the downloading activity of newcomers.

CCen was shifting between members during the experimental study. Figure 8.6 shows the variations of CCen and how it shifted between VC members. Since the activity of all oldtimers dropped (see activity tables above), and resumed during the experimental study, CCen also dropped. The important fact is that after notifications were generated members began to gain CCen (CCen at the second period was zero for all members). For example M2 has 0 CCen at the second period since M2 started uploading and downloading again in the fourth period he had CCen = 1.016. M13 used to have CCen = 6.1, which dropped to 0 in the second period, while in the fourth period the CCen of M13 was 1.27. M11 had CCen = 5.3 during the pre-study period, then CCen of M11 was 0 during the second and third periods, and increased to 0.7 in the fourth period. 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 indicate related members or cognitively central members (i.e. M9 did not have TM about this community). In the second questionnaire, M9 had an opinion and identified that M7 was one of the CCenM (and indeed M7 was an influential member during the pre-study period). This demonstrates that after the notifications were generated, M9 became more active and aware of what was happening in the VC.

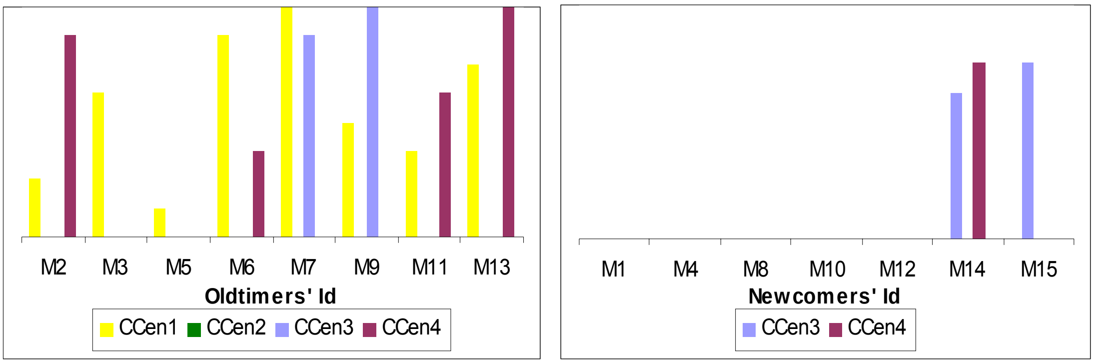


Figure 8.6 CCen variations during the experimental study. The bars show the CCen rank of the centrality of each member during each period. The bigger the bar the higher the CCen. The left figure shows the CCen of oldtimers and the right figure shows the CCen of newcomers. In both groups, members' centrality starts to improve after the first set of notifications was sent during the third and fourth periods (CCen3 and CCen4, respectively).

What change patterns can be recognised after notifications were sent? After the notifications were sent, the CM algorithms were applied to the tracking data and a new CM extracted for the VC activities. The pattern algorithms presented in Chapter 6 were applied to extract changes during the corresponding study periods. Due to the short time of the experimental study and the specific VC, some of the change patterns (see Chapter 6) could not be identified.

Change Pattern 1(a) (*A member is moving to the periphery due to his cognitive centrality as time passes*) was identified during the third period for members M2, M3, M5, M6, M7, M9, M11, M13. These are all the oldtimers who used to be active before the study and became inactive (less active) during the first two periods of the study. After the second set of notifications, this pattern applied only to M15 who was moving to the periphery. M15 is a newcomer who appears to download and upload after he received the notifications (third period) but became inactive in the fourth period.

Observations relevant to CCen: Cognitive centrality is a crucial part of the functioning of a closely-knit knowledge sharing VC. During the study, CCen shifted from member to member three times. 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 and M2 (see Figure 8.6). The shifting of CCen to different members in the VC shows a fairly dynamic VC were different members engage at different times. The fact that the CCen is changing from one member to another after each set of notifications, and taking into account that prior notifications the CCen was 0, demonstrates that the notification had some positive effect on the VC functioning as a whole. Member M2 was peripheral during the first three periods but after the second set of notifications was sent, M2 became cognitively central during the fourth period. M14 who was a newcomer was peripheral at the beginning of the third period but gained centrality at the end of this period. The CCen of M2 increased due to his uploading to the VC, as well as downloading. 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 CCenM. M2 and M13 read resources uploaded by M11. The above detections show that notifications acted as a trigger to promote the resources uploaded by central members, and hence helped CPerM to identify relevant resources they were unaware of.

8.4.2 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 TM, SMM, CCen.

One member uploaded but not downloaded from the VC and according to his reply, this member a) has difficulties to identify in which folder resources relevant to his interests are stored, and b) he is only interested in resources uploaded by specific members. For this member notifications can be useful to help him identify where important resources for him are stored and also what specific members (relevant to that member) are uploading in the VC.

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 as one of the CCenM, although that member has not made valuable contribution to the VC. This assumption was made since this member was one of the established researchers participating in the community. Furthermore, this shows that members make the assumption that CCenM in the VC are the senior/established researchers in the overall community which in fact was not the case. It is interesting to see that during the fourth period, after M2 was notified that people read papers uploaded by this member, M2 indeed contributed to the VC and became central in the VC:

"I realised that I was cognitively central and tried to upload papers of general interest. […]" (M2, Questionnaire 3)

From the second questionnaire (after the first set of notifications was sent) we can see that the opinion of members' changed 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 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 in the first notification message sent to M11). 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. For example, M3 selected the two main supervisors (M2, M6), M6 selected M2 and M13 (the creator of the VC), and M7 selected M2 and M8 (a very active member in the physical community). An important observation was made that in the third questionnaire the replies were closer to what was extracted in the CM, since one of the central members in the physical community (M2) became central also in the VC uploading resources following the notifications.

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 8.7. 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*" answer, and this shows that all members have had an opinion on the purpose of this VC.

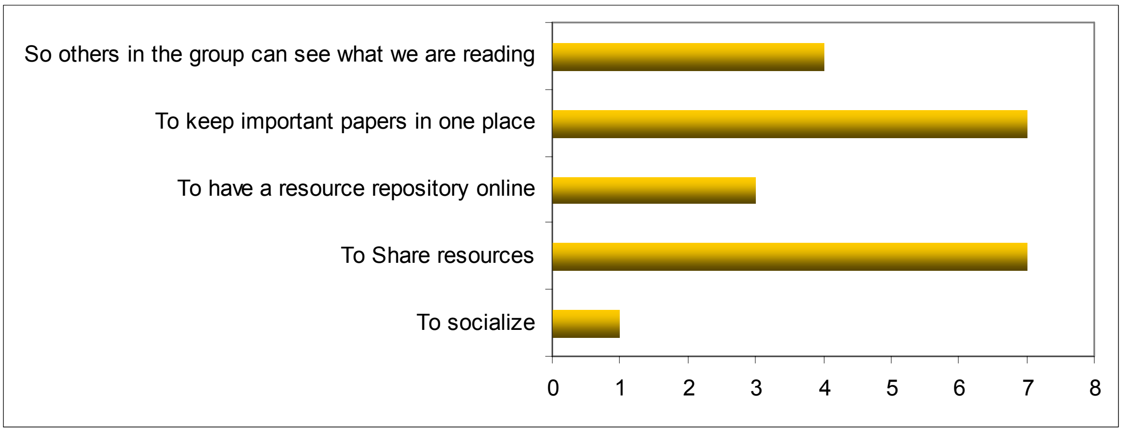


Figure 8.7 Answers to the question "*Can you please state in your own understanding, why the Personalisation & Intelligent Knowledge Management VC has been created?*"

8.4.3 Effect of Notifications on Oldtimers

The effect of the notifications on oldtimers was evaluated through the questions defined in Section 8.2.3. The second and third questionnaires for oldtimers was combined with data stored in the CM, and was analysed to answer the questions in this section. Two sets of notifications have been generated to members on a different format according to the description in section 7.4. The first format of notifications which included the general links to the VC has been evaluated using the second questionnaire (given at the end of period 2), and the second format of notifications has been evaluated using the third and last questionnaire (given at the end of period 4). The messages generated can be found in Appendix D.

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. Two members due to lack of time, one member stated that he had not noticed the links in the notification and one member stated that the information was not relevant to him. One member provided further comments on why he did not follow the links provided: *"I got distracted while I looked at the BSCW space, once interrupted, I did not come back to the virtual community; I have been extremely busy the last 2 weeks."(M2, Questionnaire 3)*

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 relevant only to the member receiving the notification. 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. Since, the information included in the notification messages was extracted from the CM, the response to this question can be regarded as a way of validating the CM.

The information received through the first set of notifications (the beginning of the second period) helped VC members in different ways (Figure 8.8). All members responded that the messages helped them identify people with similar interests, 7 members agree that the messages helped identify people they may contact for information, as well as to identify who was uploading similar resources. In addition, 6 out of 8 members replied that the messages helped them identify where resources that could be useful for them were located:

"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 8.8). 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, Questionnaire 3)

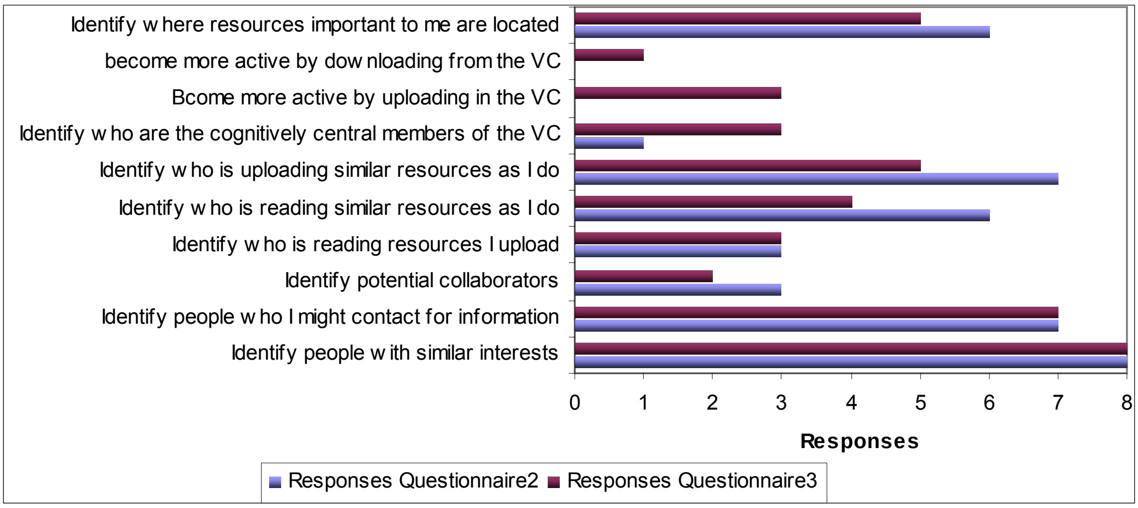


Figure 8.8 How information received with the notifications helped members

From people replies in the questionnaires after the notifcications were sent, we can see that oldtimers at first tend to believe that the notifications provided a way of creating awareness in the VC, while after the second round of notifications members believed that the notifications were not only providing awareness but encouraged members to remain/become active.

Have the notifications motivated oldtimers?: Although, messages do not seem to motivate members in becoming more active in the community by either uploading or downloading (Figure 8.8), 6 members agree that the first set of notifications helped them remain in general active by visiting the VC space. 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, 1 member believed the notifications motivated him to download and 3 members believed that the messages would motivate members to upload resources. 6 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:

"The message was helpful as it showed me resources some of which were interesting. There were some resources, which I found interesting (and was unaware of)."(M2, Questionnaire 3)

Have oldtimers become more confident to contribute? 4 oldtimers agree that receiving notifications in both formats would boost their confidence in contributing to the VC.

"I now know that there are people who have read what I uploaded previously" (M13, Questionnaire 2)

4 members remained neutral 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)

“Receiving the notification messages didn't really impact my confidence to contribute to the virtual community.” (M5, 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 4 members stated that they felt more self-assured in uploading and be 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**:** *Q8) may have similar research interests, Q9) may read similar resources, Q10) may upload similar resources.* Following the method described in Section 8.3.3, the data was statistically analysed and compared among the three questionnaires. With regard to Q8 and Q10 statistical 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 notification sets. On the other hand, the results for Q9 show a marginal statistical difference between the questionnaire answers (Table 8.2), i.e. before notifications and after the two rounds of notifications were generated.

Table 8.2 Wilcoxon signed non-parametric test results for Q9. Results extracted from the questionnaires compared to the data in the CM for all members. The results show marginal statistical significant difference between the replies of the three questionnaires.

Wilcoxon Signed Non-Parametric Test for Q9		
F1 metric for Q9 Replies	**Z**	**Sig. (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 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 chose 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), 6 out of 8 members 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 M2, M1 and M8 were 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 the CM. Furthermore, we have M14 appearing in the CM in the third and fourth periods but no one selected that member in their answers. In general the overlap between members' selections and what was extracted from the CM increased between the first and third questionnaires ($p= 0.025$, Table 8.2).

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 select M8 and M1 to upload similar resources to him. 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 neither 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 as 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 CM (e.g. M11 did not change his opinion throughout the study and was influenced entirely by what was happening in the physical community). 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)*

How does the activity of oldtimers changed after the notifications? Based on the data extracted from the questionnaires, two oldtimers have downloaded resources from the VC after the first set of notifications generated. One oldtimer downloaded and one uploaded after the second set of notifications. Five members have remained inactive even after they have received both notifications. The main reason for their inactivity is *lack of time due to other work commitments*. It is important to mention that oldtimers had not uploaded any resources to the VC before period 3 and before they received the notification messages. Although the experimental study ran for very limited time, the data show that the notifications had slight impact on the activity of oldtimers motivating them to be active in the VC.

8.4.4 Findings from the Newcomers' First Questionnaire

The first questionnaire was sent to the 7 newcomers 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 extract their individual user models. In addition, this questionnaire helped in assessing issues related to newcomers' awareness.

Out of the 7 newcomers, 5 did not participate at all in the VC prior to the generation of notifications. One member uploaded and downloaded and one member only uploaded from the VC (See Figure 8.4). 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 CM indicated similarities between M12 and members M15, M6 and M9 (these members are in M12's research group). This example illustrates other factors which influence the participation in a VC – personal style or cultural dimensions (e.g. some people would prefer to work alone rather than to share with others). 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 personal style but also to a lack of awareness how a newcomer could contribute to the community. 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 (5 out of 7) as the central members. 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 (M14 might have realised that although his supervisor was the influential member in the physical community, this was not the case in the VC). Similarly, M15 initially considered M2 and M6 as the two CCenM in the VC (these are the two lead supervisors) but after the notifications were sent M15 thought that M6 and M9 were central (note that the notification sent to M15 pointed at similarity with M15 and directed to resources by that member). 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.

With respect to SMM, newcomers were asked to state why the specific VC had been created (see Figure 8.9). Several options were given from which the members could select several. 6 (out of 7) newcomers chose the option *"To Share resources"* and 3 (out of 7) 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.

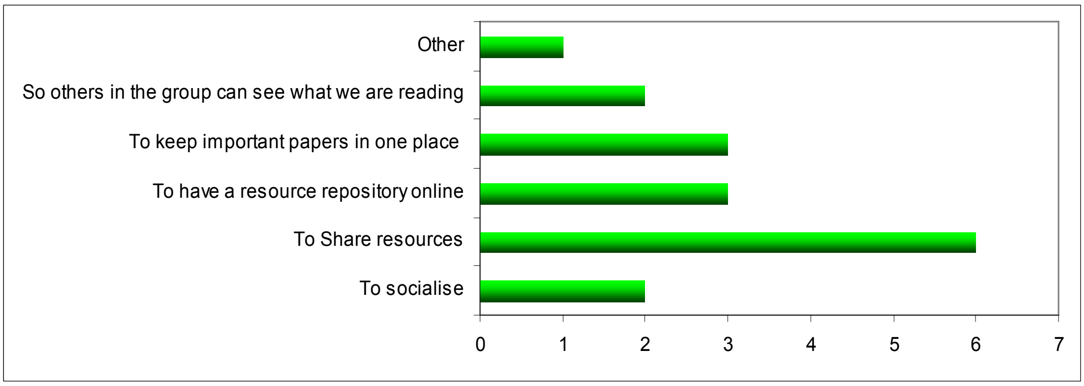


Figure 8.9 Answers to the question "Can you please state in your own understanding why the Personalisation & Intelligent Knowledge Managament VC has been created?" One member replied "Other " and he specified: *"To improve my awareness of the field & get up to date information about the field"*.

8.4.5 Effect of Notifications on Newcomers

This section presents and discusses the findings with respect to the effects of notifications on newcomers. Only one set of notifications was sent to newcomers, at the end of the third period, following the second format of notifications (see Appendix D). The notifications to newcomers was delivered as a welcome message providing relevant information (members and resources) according to individual interests of each member extracted from the newcomers' questionnaire.

Have newcomers followed the notifications sent? 3 (out of 7) newcomers followed the notifications and 2 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 but was unaware of this. Although the notifications did bring this similarity with others to M12's attention, this member did not consider such information valuable since he was not part of the physical community. This is an example that people's connections in real life influence their participation in the VC (which is manifested strongly in closely-knit communities within organisational settings). 4 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, 3 newcomers (out of 7) 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 (Figure 8.10). One member suggested that the information he received was not directly relevant to his PhD research.

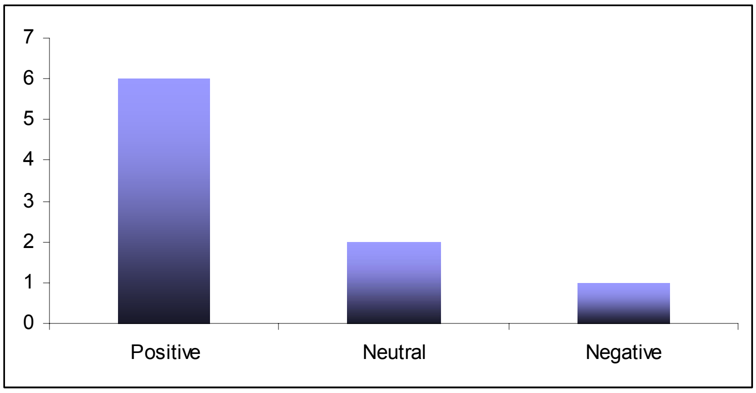


Figure 8.10 Responses to the following questions: In your opinion the information you received through the notification messages was relevant to you? (Six members replied positively and one negative) , Have the notifications motivated newcomers? (Five members agree that the messages motivated them to remain active while two members have selected neutral), Have newcomers become more confident to contribute after they received the notifications? (Five members agree that the information helped their confidence and two members have selected neutral option).

According to the information collected through questionnaire 3 (Figure 8.11) notifications helped newcomers identify where resources important to them are located, identify people with similar interests, become more active by downloading and identify people they may contact for information. 2 members agree that the information they received allowed them to become more active by uploading, identify who is uploading similar resources as they do, 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 how the notifications could help them integrate in 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)

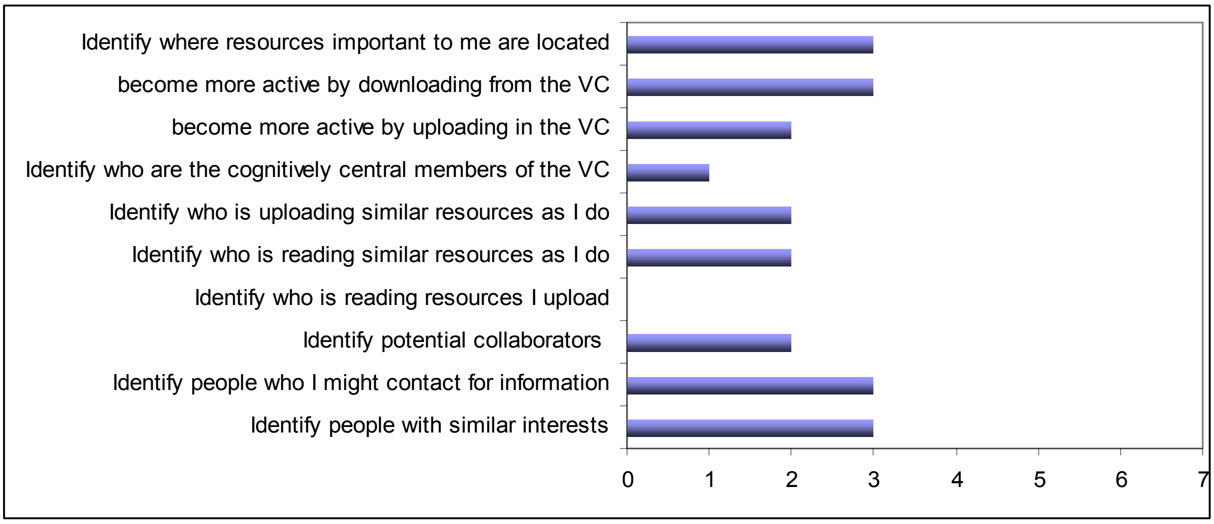


Figure 8.11 How the information received with the notifications helped members?

Although only 3 members have followed the notifications and only 2 of them had an activity in the VC (Figure 8.10), 5 members agreed that the information they received would motivated them to remain in general active.

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

"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. […]." (M8, Questionnaire 3).

Two of seven members selected the neutral option and one of them explains:

"I am not a central member of this community and have different research interests." (M12, Questionnaire 3)

Newcomers see the notifications as an awareness feature that helps them identify their similarities with others in the VC, and who are the CCenM. 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 contribute to a VC (Brazelton and Gorry, 2003), and it is even harder if their research interests do not fit with the VC (as M12 in our case). The fact that 5 out of 7 newcomers in this VC find the notification messages motivational, gives encouraging support that notification messages 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 (see Figure 8.10).

"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)

Two members have selected the neutral option (Figure 8.10) and commented that:

"As mentioned earlier I have different research interests from people in the VC" (M12, Questionnaire 3)

"I don't generally contribute to the community because I don't feel I have anything that can be useful for the other members... which hasn't got anything to do with the notifications really!" (M4, Questionnaire 3)

According to 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 possible connection between notifications and newcomers' confidence.

Have the notifications had an effect on TM or SMM of newcomers? 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: **Q8) may have similar research interests to them**, **Q9) may read similar resources to them, Q10) may upload similar resources to them.** Following the method described in Section 8.3.3, 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 8.3) for both questionnaires (see Section 8.3.3). This can be due to that 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 in the CM 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 M2, M1 and M13 at the first questionnaire. After he received the notifications M14 selects M1, M13 but now he chooses M7. Similarly for Q10 the CM 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 (who selected by M15 as similar in uploading) were members from his research group supervised by the same supervisor.

There are interesting observations with respect to Q8. 5 out of 7 members had changes in their opinion who had similar interests to them. In the first CM 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 CM extracted, there was a greater overlap between the members selected by newcomers in the second questionnaire and what has been extracted in the CM (Table 8.3). 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 CM. M4 selected only M2 in the first questionnaire, but in the second questionnaire M4 also selects M3 (and similarities were also detected in the CM). M8 had selected M2, M7 and M3 at the beginning but after the notifications M8 added M1, which is a link present in the CM 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 8.3 Wilcoxon non-parametric test results for Q8. Results extracted from the questionnaires compared to the data in the CM 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**	**Sig. (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 (Chapter 4). Figure 8.6 summarises the CCen variations, 3 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 notifications. The CCen of M14 dropped but the $CCen \neq 0$ shows that M14 remained active.

8.5 Discussion

8.5.1 Effect of notifications in general

Community awareness: This research started with the foundations looking at TM, SMM and CCen as important for the VC to grow and sustain. By providing notification messages we aimed at creating awareness among members with respect to who the CCenM are and how VC members relate to each other. Based on the results extracted from the questionnaires and the CM we can conclude that notification messages might not have a significant impact on the behaviour of community members but had some 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. 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. Although there were members who benefited from the notifications, there were members who were not influenced by the notifications.

Newcomers' integration: We expected the newcomers' participation in the VC to be more influenced by notifications but only 2 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 time that the study took place. One of the newcomers was on time off at some point during the study and also the Christmas period overlapped with the experimental study. On the other hand, one of the inactive newcomers became active after the end of the study. The inactivity of newcomers can be due to only provided one round of notifications to newcomers and this has not provided enough time for them to integrate properly within the VC. Nevertheless, some of them integrate and became aware of the similarities they had with others.

8.5.2 Feasibility of the Study

Use of Tracking Data: The approach followed in this PhD is based on analysis of tracking data from the VC. Obviously, there are elements that can be captured from this kind of data but others cannot. The advantage of extracting a CM based on tracking data is that it represents the actual interaction of members with the resources available in the VC and data cannot be manipulated in any way. Furthermore, if members are working explicitly online, then a CM 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, the current approach cannot capture the personal interactions through tracking data. 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 of before. Although the notifications did not influenced radically the behaviour of members, there was indication of positive effect 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 CM will not represent the current interests of the members. Opening the CM 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 CM 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 members to gather information about their interest. However, when members did not download/upload papers, the extract CM included only interest similarity graphs. This is especially an issue when the VC is voluntary like in this PhD and there is no explicit incentive for people to participate (e.g. relevant studies have been conducted in course settings where users were given credits to participate in the VC). In our case however, people would have to see personal benefits for them in order the approach to succeed. This makes the notification text crucial, as also confirmed by the study. We only considered simple types of notifications; future research would be needed to examine a range of notification forms and their impact on the VC.

Community stages: Considering the main stages of a VC (Figure 2.1), we can conclude that the adaptive notifications approach is more suitable for the Grow and Sustain stages of a community where members are trying to make connections and keep the community active. For example with the VC used in this study we could not identify most of the change patterns identified in the VC used in the study in Chapter 6 (the VC under study had minimal activity before the study begun, thus changes could not be captured). On the other hand, there was some evidence from this study that the notification messages were beneficial for newcomers and helped some of them to integrate.

Cultural differences: Cultural differences have also had an impact on the results obtained in this study. People from different cultures have different styles and prefer to work closest to what they are used to. This study though is not big enough either in terms of numbers or variety of people or in terms of duration to draw any conclusions related to cultural differences. Nevertheless, individual members' cultural background is affecting their sharing and integration in the VC. Furthermore the organisational culture is also important. For example, people have a personal way of organising their resources that might not be reflected in the way the VC is organising resources. We did not aim to change how people worked or to force people to share. However, it was noted that notifications could be adapted to members' cultural behaviour in order to be more effective.

Use of notifications: Some participants in our study saw the approach of receiving notification messages as interruption of their practice. On the other hand, some members found the targeted notification messages a useful reminder. 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 uploading in the VC. We can conclude that the study found out that notifications would be a useful approach to influence the community as a whole (though there is a caution that the approach would not be uniformly accepted). It was also clear that the physical and virtual community could differ which could impact the effectiveness of the notifications. A further improvement could be instead of relying only on tracking data for modelling a community and deriving notification messages, it could also be good to have a systematic approach of collecting members' opinions about connections they may have with other members in both the physical and virtual community. Looking at the discrepancies between the physical and virtual community, we can build more targeted and effective notifications.

8.5.3 Applicability of the Results

Type of community: This study used a specific community but at the same time it represents a typical community of researchers (some people working on projects together, many of the projects not overlapping, members remotely located and people based in the same institution). This is a completely voluntary community and no incentives were used for participation. Since the community had a physical presence and a virtual presence discussions have not been examined (in fact, most of the discussions in this community happened in the physical context). Discussion forums can be helpful when people are remotely located and virtual presence is their only way of communicating. However, using discussion data gives a biased view of the community since many people exhibit lurking behaviour.

Limited study period: The study was conducted within three months which was a relatively short period. Nevertheless, we observed some positive influence on the community related to TM, SMM and CCen. If the duration of the study was longer, we would expect stronger results, e.g. a better integration of newcomers (some newcomers remained active in the community even after the study ends). Longer period would provide larger corpus of tracking data, and thus the change patterns algorithms would have been employed and extract certain patterns that could not be detected in the limited period of this study. Consequently, notifications linked to these patterns would have been generated to make notification messages more appealing to individual members. This would allow us to evaluate what this study suggested that the more targeted the notifications the better for VC members. As a result, members would develop better awareness of what was happening in the community. There is a tendency based on the results obtained that people become more motivated after they have received the notifications. However, we cannot conclude with confidence that this would be manifested in a longer study (e.g. people may ignore notifications or find them distracting).

Use of ontology: Although the community under study was created under a common title, the expertise of people was heterogeneous. Different results might have been obtained if the topic of the VC was more closed. The effect of awareness in such VCs would have been different and the connections among people would have been stronger (every member would have been connected with every other member). The heterogeneous style of this VC allowed us to exploit a semantic structure (ontology) that showed the benefits of this approach (all members agree the notifications are relevant). It is important to note that without the ontology it would have been difficult to extract accurate connections between people.

BSCW system: This VC was using the BSCW system with all its advantages and disadvantages. BSCW is a robust system and the functionality is stable and well-tested. BSCW allowed us to keep the tracking data used as generic as possible. However, there were some negative aspects associated with BSCW. The system has its own style of interaction, many of the members were not used to this style. Most members had not used BSCW before and also, during the study, were not using BSCW for any other activities in their practice. For users this was yet another system to learn:

"I already use other bookmarking/reference tracking system and I didn't see the advantage of adding another place to keep track of papers"

A number of members commented that they had not used the system because they do not like BSCW style: *"[…] also the community folder structure does not appeal to me"*

We can argue here that some of the negative results obtained can be attributed to the BSCW platform following studies that people tend to perform best when the tools are similar to what they have used to and also what appeals to their working style (Uruchrutu et al., 2005).

Experimenter as a member of the community: Finally the experimenter was a member of this community, thus some remarks must be noted. The questionnaires were given before the CM was extracted in each period in order to mitigate the influence on the experimenter's behaviour. When participants know the experimenter, they might reply in the questionnaire in a biased way (for example one member noted that the purpose of this community is: *"for Stella's PhD"*). Having this issue in mind the researcher 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 fixed in order to please the experimenter since they did not know what could be a correct answer. The experimenter's replies to the questionnaires differ from what has been extracted in the CM. On the other hand being a member of the VC allowed the experimenter to be part of what was actually happening and being able to interpret the results in a more meaningful way.

8.6 Summary

This chapter presented and discussed the summative evaluation of the notification messages. Questionnaires and statistical approaches have been used - qualitative and quantitative data were collected and analysed. The results presented in Section 8.4 support the hypothesis that notification messages can have a positive effect on members (both newcomers and oldtimers).

The second message format (personalised information for each member with links to resources in the VC), was preferred by members. All oldtimers reported they had followed the links included in the notification messages. In all cases members rated the notification messages as relevant to them. This verifies that the information kept in the CM with respect to VC members was realistic and relevant. Furthermore, there was some evidences that notifications motivated members to visit the VC space and in the case of newcomers to upload and download resources. In general, notification messages can be used for motivating members to keep active in the VC.

The notifications had some positive effect on the confidence of VC members. Both oldtimers and newcomers felt more confident to contribute after they received the notification messages. Although statistical analysis did not show sufficient evidence for the development of TM or SMM for either oldtimers or newcomers, the comments of VC members show a slow development of TM and SMM since members are reporting they are becoming aware of the resources and people available in the VC. The pre study questionnaire with oldtimers and the first questionnaire with newcomers show evidences of SMM among members. Some newcomers and oldtimers had increase in their activity after they received notification messages. Members have either uploaded or downloaded due to the information they received with notifications. Finally, there is evidence that monitoring the CCen of members can be used to support the knowledge sharing in a closely knit VC.

The questions addressed in the experimental study presented in this chapter have not been directly derived from any previous studies since there are no evaluation studies with VCs focusing on TM, SMM and CCen. Although this evaluation is driven by clear aims and questions, the potential impact of the notifications on members and the VC in general requires a long term study which is not feasible in the time limit of a PhD (considering that a community has to be established, start functioning, a series of interventions can be done, and the long term effect observed).

Chapter 9

Conclusion

9.1 Introduction

This thesis presents research in the broad area of employing User Modelling and Adaptation techniques to provide support to a VC as whole and improve knowledge sharing among VC members. Specifically, this research is relevant to designing and building algorithms capable of extracting and maintaining a community model that can assist with identifying problematic knowledge sharing patterns in a VC. We have focused on supporting knowledge sharing in closely-knit VCs following organisational psychology processes - TM, SMM and CCen - and generating notifications containing personalised information about VC members. The research developed a computational framework employing a community model containing information about individual community members and relationships between members, based on which notification messages are generated targeting individual members, but at the same time, aiming to support the VC as a whole. Knowledge sharing patterns have been defined to inform a more targeted and personalised support for knowledge sharing. The main contribution of this work lies primarily in the definition and development of a community model based on TM, SMM and CCen and the exploitation of that model to provide intelligent adaptive support for knowledge sharing in a VC as a whole.

This chapter provides a synopsis of the outcomes of this research. Firstly, we will discuss the achievements and generality of our work, and will sketch out the contributions to the relevant research fields. Secondly, we will provide a reflection on the key decisions taken during the research journey. Finally, we will discuss immediate and long-term future research directions.

9.2 Synopsis

9.2.1 Summary of the Work Conducted

This research work has proposed and implemented a computational framework for extracting a community model based on VC tracking data and following TM, SMM and CCen. The exploitation of the CM enabled us to provide intelligent support for knowledge sharing to VCs.

In Chapter 1 we proposed three research questions. This research addressed these questions in the following way:

(i) *How to extract a computational model to represent the functioning and evolution of the community as a whole by using semantically enhanced tracking data?*

We have (a) formalised the input data to capture essential information about members, activities and resources, which will be represented in the CM; (b) developed algorithms to extract a CM based on tracking data and semantically enriched this data by using an ontology;

(ii) *By using that model how can intelligent functionality be provided to support the development of TM, building of SMM and monitoring of CCen?*

We have (c) developed graph-based algorithms to analyse the extracted CM and identify knowledge sharing patterns, both static and time-dependent; (d) employed the CM and the algorithms for detecting knowledge sharing patterns in a mechanism for generating notification messages aimed in supporting knowledge sharing in a VC.

(iii) *How can intelligent support of the above processes affect the functioning of the community?*

We ran an evaluation (e) in a real, active knowledge sharing VC. Results evidence that by supporting the development of TM, building of SMM and monitoring CCen in a VC it can be beneficial for knowledge sharing in a VC.

The rest of this section will discuss in more detail how the above questions have been addressed in this thesis.

(a) Input Formalisation (Chapter 4). Formalisation of the input took into account the simplicity and generality of the approach so it can be used in other knowledge sharing applications. Input data include, 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.

(b) Community model extraction mechanism (Chapter 4). 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 manually detected, and provided as the basis for automatic detection of community patterns and dynamic community-tailored support.

(c) Definition of knowledge sharing patterns (Chapter 5 and Chapter 6). Static knowledge sharing behaviour patterns in a VC have been defined, following selected processes (TM, SMM, CCen) important for the effective functioning of closely-knit communities. We have demonstrated with a study how these patterns can be detected and used to provide community-tailored support. Static knowledge sharing patterns can be useful in identifying problematic cases, especially during the start-up phase of a VC. Furthermore, the CM have been employed for detecting community change patterns to identify when intelligent support is needed to support a community to sustain. The results from a study conducted based on archival data show how pattern detection can be used to generate notifications that may help a VC to sustain.

(d) Generation of notification messages (Chapter 7). Adaptive notification generation mechanism has been defined aiming at supporting CCenM, CPerM, the development of TM and the establishment of SMM. The formalisation of adaptive notification mechanism defines why and how a notification is generated, according to detected knowledge sharing patterns.

The formalisation of the above aspects inform the generation of notifications that can be adapted in different closely-knit VCs. In this thesis, we have demonstrated how tracking data extracted from a widely used knowledge sharing system - BSCW - can be used in designing and extracting a CM and providing community-tailored support. Archival data from an existing VC have been used to validate the algorithms for extracting a CM and detecting knowledge sharing patterns.

(e) Experimental Study (Chapter 8). Following the framework defined in this thesis we have used tracking data of a real and active VC to validate the notification generation mechanism. Hence, we have extracted a CM, employed the algorithms to extract knowledge sharing patterns, and used the detections as input for generation of personalised notification messages to VC members. An experimental study has been performed to identify the effect of the notification mechanism, so it can be employed in supporting knowledge sharing in VCs. The results of the evaluation (Chapter 8) 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 generated 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 closely knit VC. The evaluation also pointed out improvements and possible applications of the framework, which are discussed in Section 9.5.

9.2.2 Generality of the Proposed Approach

The generality of the approach presented in this thesis can be discussed following the main components of the proposed framework:

Input formalisation: The input data considered has been kept in a generic format to be in line with any conventional knowledge sharing system. The proposed data descriptions are easily applied to knowledge sharing systems that have members sharing and rating resources and providing keywords (tags) and/or some metadata associated with each resource. The resource metadata followed the Dublin Core[13] metadata schema, which is a conventional standard for metadata description. The implementation of the algorithms included input data stored in a MySQL database. The tables can be directly used to store the same data format in any domain. The ontology developed reflects the topics of interest in the community. The ontology using for the algorithms in this thesis was built using Protégé[14] and encoded in OWL. The ontology can be reused in any other system or exploited by algorithms that can reason through an OWL ontology. A WordNet similarity measure was employed as an input for the CM to be extracted. The algorithm, taken from a third party, has not been purpose built for this work. It has been extended to detect similarity between resource keywords. The algorithm and the proposed extension are written in Java, and can be used straightaway in any other Java based application.

Community model extraction mechanism: The community modelling extraction mechanism contains four parts, namely: individual user models, relationships model, peripheral/central topics, list of CCenM. Although the algorithms for extracting the CM depend on people sharing resources in the VC and the provided keywords of shared resource, they are generic and can be used by a different system that allows resource sharing. The general structure of the relationship graphs and the use of a relational database modelling approach for their representation, facilitate their implementation and easy integration in different systems. The algorithms used to extract the relationship graphs are depended on the tracking data extracted from BSCW that provided information about who has downloaded a resource and who has

[13] http://dublincore.org/
[14] http://protege.stanford.edu/

uploaded the resource. This does not affect the generality of the approach since the framework describes how similar relationship types can be defined and implemented when having different format of tracking data, using different programming languages and technologies. In the IUM, the interests of members have been extracted here according to keywords of the resources members have read and/or uploaded. Even if keywords are not available other representations (e.g. tags) can be used to represent the interests of a member. To extract the CCen of a member, the relationship types have been used. Having tracking data from a different platform, suitable relationship types can be defined and used for the extraction of the CCen.

Definition of knowledge sharing patterns: Graph-based patterns were employed in defining community knowledge sharing patterns. The definitions were kept generic; however they are strongly depended on the relationship types defined. The types considered in this thesis are applicable broadly to any closely-knit community for knowledge sharing. The graph-based pattern detection approach is applicable to other relationships detected in VCs. An exhaustive list of knowledge sharing patterns has not been provided, and is beyond the scope of this work. Knowledge sharing patterns can vary from one community to another. Using the graph-based approach presented in this research, further patterns can be defined.

Generation of notification messages: Similarly to the detection of knowledge sharing patterns, notifications have been formalised in a generic way. Although the generation of notifications is depended on the detection of specific patterns, the approach is general, given that suitable patterns have been defined relevant to the VC. The definitions of the notifications provide the foundations upon relevant notifications to be defined according to the specific community.

9.3 Reflection on Decisions Made and the Methodology Used

Different factors have influenced the decision making during this research work. General hypothesis, assumptions about the framework, time restrictions and the use of human sample in evaluating part of the components have all had a role on how decisions throughout the study were taken. This section will reflect on limitations, constraints and lessons learned, and how these relate to the research objectives of this work.

9.3.1 Hypotheses and Assumptions that Informed the Framework

In this section the main hypotheses and assumptions stated at Chapter 1 will be revisited.

Hypotheses

Providing intelligent support tailored to the community as a whole and supporting TM, SMM, and monitoring CCen can be beneficial for knowledge sharing and community functioning. The application of intelligent support that is tailored to the community rather than the individual member has enabled us to design a more comprehensive awareness for VC members. On the other hand, the summative evaluation study showed that notifications could have been formulated in such a way that would be more personal by allowing members to see “what’s in it for me”(even if it aimed to support the whole community).

We have taken the decision to design support that promotes the development of SMM and TM and the monitoring of CCen. This decision was followed throughout the thesis and underlined the algorithms developed. The results of both formative and summative evaluations support the selection of these three processes as the basis for community support, these being problematic patterns that can be supported and which are relevant to the above processes have been successfully identified. Nevertheless, long term studies are necessitated for any affirmative claims about the positive effect of using these processes for intelligent community support. These three processes are all cognitive processes. Accordingly, members needed a long time to conceptualise what was happening in the VC and permit this to be reflected in their behaviour in the VC.

Monitoring static and time depended patterns of knowledge sharing behaviour of members can enable a more targeted support to be generated and help the community to share knowledge more efficiently, in a more sustainable manner and for a longer period of time. Monitoring patterns of knowledge sharing behaviour of members that enable the generation of a more targeted support proved to be a useful approach. They provided a personalised support to community members despite being heavily depended on having sufficient activity in the community. If members were not participating by uploading and downloading many patterns would not have be discovered.

Assumptions

TM, SMM and monitoring CCen within a community are important for the functioning of a VC (Ilgen et al., 2005). Development of TM and SMM and monitoring of CCen in the community have been identified in the literature as being significant for the functioning of a VC. Our decision to adopt these three processes as a basis for supporting knowledge sharing in the VC proved to be the right one (based on both formative and summative evaluations), even though this study demonstrated that a long time is needed for TM and SMM to be conceptualised and developed inside the community.

Resources shared by community members represent the topics of interest of the VC members and correspond to the knowledge a given member holds (Song et al., 2005; Cheng and Vassileva, 2006). This assumption played a crucial role in the development of the framework. Since we are considering only the tracking data of a community this represented the only means by which to extract the members’ interests. A different approach would be to require members to provide keywords and/or phrases that will represent their interests upon joining the community. Interests based on resources read and uploaded seem feasible and indeed the only possible ones when tracking data is used. Nonetheless, our evaluation study showed that the virtual space does not always correspond to the actual interests in the real community and consequently some extension of the input data would have to be considered (e.g. monitoring discussions members participate in, open user models where members are asked to amend their profiles).

9.3.2 Use of Existing Technologies

During this research several algorithms have been developed. These algorithms rely on existing technologies and were used in such a way so as to semantically enhance the proposed framework. In this section the discussion focuses on the performance and quality of these technologies. Furthermore, it includes any implications that might arise from the use of such technologies on the overall approach.

All the algorithms defined in Chapter 4 for extracting the CM require the use of the *WordNet Similarity Measure* in order to define the semantic similarity between two sets of words. As already mentioned in Section 4.2.3, this comparison is undertaken by adapting the original algorithm, developed by Seco, Veale and Hayes (2004) in order to accept compounds (phrases). It is important to acknowledge that the accuracy by which the semantic relationships between members are extracted in the VC model is very much depended on the performance of the original WordNet similarity measure (Seco et al., 2004). This algorithm has been validated by Seco, Veale and Hayes and the results were presented at the European Conference of Artificial Intelligence in 2004. Based on their reported results, the WordNet similarity measure that they have developed outperformed other information theoretic approaches. Consequently, it can be regarded as an accurate approach for measuring semantic similarity.

Jena Owl Reasoner is a semantic web framework for Java that offers reasoners for OWL, RDF and RDFS ontologies. In this research, Jena API has been employed in order to extract relationships between concepts. These appear in the OWL ontology developed and applied in the algorithms presented in Section 4.2.3. The reason the decision was taken to use Jena in the present research was that we were utilizing Java both as a programming environment and an OWL ontology. Jena provides an OWL API for reasoning using Java programming and an

OWL ontology. The use of Jena does not restrict the approach since it offers reasoner types that work with different types of ontologies. Thus, if someone wants to extent/modify the present approach and use an RDF ontology, Jena can still be used if the RDF reasoner is integrated accordingly. Where a different OWL ontology is used there is no need for any further modification of the Java program code.

9.3.3 Human Factors

Technological outcomes can be predicted most of the time. What cannot be predicted are human behaviours and the external factors influencing humans to function in a certain way. In this thesis, we focused on supporting the sharing of knowledge among members who are using the web as a medium for communication. However, in most of the cases what is happening online in a VC is part of a larger environment, usually within a physical community. 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. This can be considered as a limitation of the approach when applied to a community of people who are using the virtual space as an extension of their physical interaction. Nevertheless, when the community has only virtual interaction and members have not a physical contact, the approach proposed in this theis would be a feasible way for modelling a VC using tracking data capturing the interaction of community members. An alternative way that considers also the physical dimension of the community can be to open the derived CM to the community members and allow people to interact, modify and add elements that reflect their physical conceptualisations and interactions. For example, members might consider an influential person from their physical community to be a CCenM in the VC. However, this might not represent the virtual activity of that member. In the way we are modelling the community, this will not be captured in the tracking data, and hence will be missed in the CM. The open community model though will allow members to make this addition and create a more comprehensive model of their virtual and physical community.

Furthermore, people categorise and organise their resources differently according to specific characteristics, diverse conceptualisations, search habits, and technologies used (Berlin et al., 1993; Indratmo and Vassileva, 2005). Although our approach partly supported the development of awareness regarding duplication of resources, we cannot claim we are providing support for resource organisation. However, this is an issue that affects the participation of VC members in the community. As it has also been noted in the responses of members in the evaluation study (Chapter 8), the BSCW style of organising resources was not preferred by some members and thus their participation in the VC was minimal. In addition, there are people who just do not want to share with others what they are reading. In this case whatever support is provided members will hardly change their beliefs and behaviour. Furthermore, culture plays a vital role on how people interact, their social behaviour and their willingness to share. This work does not consider cultural preferences, future work is required to take into account for culture could affect the adaptation mechanism. Along this line, initial work on culture and adaptation, e.g. (Reinecke and Bernstein, 2009), can provide helpful insights.

9.3.4 Constraints of the Methodology

Chapter 8 stressed the fact that evaluation of adaptive systems has been challenging especially when evaluating the benefits of the system as a whole. Different circumstances affected the way the formative and summative evaluation studies were conducted, and consequently the results obtained. This is summarised below.

Physical Vs Virtual Community

Most of the members participated in the summative evaluation study (Chapter 8) were not using BSCW as part of their every day practice. Instead, the community was functioning in the physical space and members were attending weekly seminars together. At the planning stage of the VC (when the VC was created for the sake of this evaluation study), all members were invited to join along with members outside the physical community (e.g. USA). After the VC was created, the process of sharing resources was transferred online and the VC started to function as such. Some of the members were collaborating on research projects, two other members were organising an international workshop together, and through these activities they were sharing resources using the VC. The fact that the community was functioning as a physical community prior to the creation of the VC had an impact on the evaluation of the framework.

For example, members had already a conceptualisation on who the influential members in the physical community were. Although these were not the same as the influential members in the VC, members' answers in the questionnaires reflected their view of the physical community. Consequently, we believe that the existence of the physical community prior the creation of the VC had an impact on members' replies and affected the results of the evaluation.

Duration of the Final Evaluation Study

One of the main constraints of the summative evaluation (Chapter 8) is the duration of the study. During the three months of the evaluation study, we identified some problematic cases with the knowledge sharing among members. However, because the duration of the study was limited we could not capture long term patterns, thus not all knowledge sharing change patterns could be discovered. Furthermore, due to the limited length of the study, issues relevant to TM, SMM, and CCen had not enough time to develop in the community.

Use of Open Questions

A limitation of the evaluation study can be the formulation of the open questions in all four questionnaires (Appendix B). 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 as explained above), and we had also three members (closest to a member), extracted in the CM the value for Precision, Recall and F1 was the same (Appendix E). In a more general approach where the selection number is not fixed, the precision, recall, and F1 metrics would be more informative.

Use of Tracking Data

The CM extracted was based on community tracking data. This approach allowed us to extract a model only based on the interactions of members. Consequently, we managed to extract similarities and relations that members were not aware of. Although these can be considered as important advantages of using tracking data, there also are limitations. Firstly, all the information extracted in the community model reflected only the virtual interaction of members. Any information relevant to the physical part of the community could not be captured through the tracking data. Secondly, if a member has changed his interests and has not shared any resource to the VC (thus, the change of interest could not be captured through the tracking data), then we have incomplete data and thus, support will not be sufficiently personalised. Thirdly, this is also linked to the cold-start problem. For example, just after the community has been created, or when a new member joins the community, there is no information regarding this member in the tracking data. Consequently, interests of that member cannot be captured. This is the main reason for using a questionnaire at the beginning of the study and when newcomers joined. This was needed so we can extract initial interests and develop the IUM of members.

Use of Archived Data

A component based evaluation of the framework took place in chapters 4, 5 and 6. For this purpose, we have used a large corpus of authentic data over a long time span. Since our work had a time limit (usual for a PhD project), we took advantage of the availability of authentic archival data. Using archival data is a common way of validating personalisation algorithms (e.g. often used for validating recommender algorithms). In this way, we could identify and validate all the graph based algorithms developed for extracting static and dynamic knowledge sharing patterns (chapters 5 and 6). The downside of this approach is that we could not validate further any results from the studies since members of the VC were not available (the community stopped functioning). Furthermore, the validation of the notifications could not be done using the same VC.

Use of Parameters and Thresholds

Three parameters, θ, σ_{Pop} and σ_{Per}, have been used during the experimentation for extracting the CM. In addition a parameter has also been used when extracting the list of members to whom a notification should be sent. This section will discuss the subjectivity of the choice of values for the thresholds used within these parameters.

In extracting the interests of members a parameter used in order to determine when a keyword/tag can be considered as an interest for a given member. Within this parameter, threshold θ has been defined in Chapter 4, to be the frequency in which a keyword appeared in the keyword/tag list extracted for a given member. The value of θ was decided on the basis of empirical experimentation where the goal was to be able to extract interest similarity (InterestSim) relationship between active VC members, based on their personal list of interests. In the formative evaluation (Chapter 4, Chapter 5 and Chapter 6) where the VC was larger than the VC used in the summative evaluation (both in terms of people and resources), the threshold θ set to be greater than five ($\theta > 5$). This means that a keyword/tag had to appear with a frequency of more than five times in the list of keywords extracted for a given member in order to be added in that member's list of interests. The value of five considered as appropriate since it allowed only keywords that were representing at least five resources uploaded/downloaded by a given member to be considered, and it also allowed interests to be extracted for all active users in the VC. In the summative evaluation (Chapter 8) the value of θ was set to be greater than two ($\theta > 2$). Thus, only keywords that appear with a frequency of more than two were added in the list of interests of a given member. The value of two selected in this case since in this VC the activity of uploading and downloading resources, hence the number of keywords extracted for each member, were less than the activity in the first VC. Not having a large number of resources, thus keywords shared within the VC, means that a value of θ greater than two was resulting in interests not to be extracted for most members, and a value of θ less than two to allow all keywords extracted for a given member to be added to his list of interests. Before we experimented with different values we were not in a position to comprehend that if the threshold was set to be above five (in the formative evaluation) or above two (in summative evaluation) the algorithms would not have picked any interests for most members nor it would have been possible to capture any InterestSim relationships.

Similarly to above, a parameter used in order to determine the popular and peripheral topics of interest of a VC. Two thresholds have been defined σ_{Pop} and σ_{Per} (Chapter 4), that determined the frequency by which a topic of interest appears in the VC. A topic of interest can be classified as a σ_{Pop} if it is a popular topic, or a σ_{Per} if it is a peripheral topic in the VC and it would be added in the corresponding list. In addition, the topics added in each list (L_{Pop} and L_{Per}) should reflect the topics of interest of the cognitively central and peripheral members. An analogous approach to above was followed in deciding the appropriate value for σ_{Pop} and σ_{Per}. After experimentation the values for the SW VC used in the formative evaluation were set to be $\sigma_{Pop} \geq 5$ and $\sigma_{Per} \leq 3$. These numbers selected as allowed topics that represented the interests of the cognitively central and peripheral members accordingly to appear in the corresponding lists. Following the same procedure, and having in mind that the VC used in the summative evaluation was smaller than the SW VC, 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.

A parameter has also been used in the algorithms when constructing the list of target members $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$, to whom a notification should be sent (Chapter 7). Due to the closeness (in terms of topic similarity), of a closely-knit VC, a relationship graph shows most active members to be connected to each other. Thus a strategy was needed to identify what message will be sent to a given member in such way that not all members will receive all messages. In formative evaluation, this parameter has not been considered given the fact that we have not generated any notifications. In summative evaluation, the threshold value was set to be three. Hence, the three most similar members were added to the target members list $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$ that used in sending the notification messages. The selected value was identified as the most appropriate after experimentation with different values. 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. A different approach to the one we have employed in this research, would have been to consider the actual values of the relationships between members and use a threshold that will allow only "strong" relations to be extracted. This was not possible to be used in our case since the values of the relationships between members appear to be very close. We experimented with different relationship values but these were allowing in some cases too many members to be on the list $\{M_{i_1}, M_{i_2}, \ldots, M_{i_n}\}$, and in other cases members were appearing with no connections since they only had "weak" relationships. The approach followed here overcomes both of these problems.

Although the use of parameters might be considered as a limitation of the approach, it can be asserted that in many cases it constitutes an advantage since it allows the adaptation of the overall approach to different settings. For example, there is no need to have restrictions on the size of the closely-knit VC, the density of the relationships which appear in a graph, the interests extracted for a given member and the lists of popular and peripheral topics as long as a threshold value is defined where is needed.

9.4 Contributions

The work presented in this thesis resulted in a number of original contributions. In this section we will outline the significance of the achievements with respect to the related research areas.

9.4.1 Contribution to User Modelling and User Adaptive Systems

There is a growing interest in providing adaptive support for teams, groups and communities. Along this line, personalisation and adaptation can play a crucial role, as illustrated by recent user-modelling approaches (Cheng and Vassileva, 2005; Song et al., 2005). A number of approaches, such as visualizations, notifications, and community ratings, have been exploited to facilitate community/group awareness, motivate participation, and improve community knowledge sharing. However, existing adaptation techniques focus mainly on supporting individual members, rather than supporting the community to function as a whole (Bretzke and Vassileva, 2003; Farzan et al., 2009; Sankaranarayanan and Vassileva, 2009). We proposed a novel approach for community-tailored support which aimed at facilitating processes related to the effectiveness and sustainability of VCs and is based on a community model derived from analysis of log data.

To the best of our knowledge, there are no such holistic community modelling approaches. More specifically we have proposed (a) a novel framework for holistic adaptive support in virtual communities, (b) a mechanism for extracting and maintaining a semantic community model based on the processes identified, and (c) deployment of the community model to identify static patterns and provide holistic support to a virtual community.

The community and relationship model in (Bretzke and Vassileva, 2003) is the closest to ours but there is a crucial difference. Users' interests are modelled in (Bretzke and Vassileva, 2003) based on how frequently and how recently users have searched for a specific area from the ACM taxonomy, and user relationships are derived based on any successful download or service that took place between two users. In contrast, our approach employs the metadata of the resources shared in the community along with the ontology and derives a semantically relevant list of interests for every user. Furthermore, the CM extracted in our case is semantically richer and theoretically underpinned. Recently research on modelling communities employed graph theory to model relationships between members (Kay et al., 2006) or members' interactions in general (Falkowski and Spiliopoulou, 2007). The key contribution of our approach to community modelling is the considering of semantic relationships, i.e. an edge connecting two members represents their semantic similarity to each other, and the relevance of this link to the community's domain.

9.4.2 Contribution to Computer Supported Cooperative Work

CSCW research has exploited different approaches to facilitate group work and knowledge sharing, such as visualisations, notifications and awareness techniques (Ackerman and McDonald, 1996; Zacklad, 2003; Gouvea et al., 2006; Wang et al., 2007). Most of these approaches have been applied to particular settings where positive aspects have been observed. In this line, we are contributing to the CSCW community with a novel approach for providing semantically enriched community awareness. Semantically enriched algorithms have been developed that inform the generation of personalised notifications to VC members.

Visualisation techniques are another approach for providing awareness of what is happening in a community, and thus, supporting participation and collaboration in a VC. For example, graphical representations are used to make people aware of the relevance to the activity or to the position of a particular member in the group (Kay et al., 2006) or to show the status (or popularity) of a resource (Wang et al., 2007). The key limitation of visualisation techniques is their passive influence on the functioning of the community, e.g. while examining graphical representations members may not be able to see how their contribution could be beneficial for the community. In contrast, our approach provides notification messages that explicitly making aware people of how they relate to others in the community.

9.4.3 Contribution to Research in Social Networks

Analysis of community evolution refers to different approaches for detecting changes over time in large or small people networks represented as graphs. Existing approaches are examining mainly structural changes of social networks (e.g. density, degree distribution, average distance, clustering coefficient) by comparing the characteristics of graph instances at given time points (Leskovec et al., 2005; Falkowski et al., 2006; Asur et al., 2007; Falkowski and Spiliopoulou, 2007; Lin et al., 2007; Palla et al., 2007; Lin et al., 2008). In the context of identifying patterns of changes (evolution) in the community we are contributing to the social networks area with a semantically enriched approach for modeling change patterns in a closely-knit VC.

To the best of our knowledge, there is no other approach which examines community evolution with regard to TM, SMM, CCen. In contrast with existing work that monitors how the network/graph under investigation is evolving over time in order to get an insight of the community (Song et al., 2005; Kumar et al., 2006; Falkowski and Spiliopoulou, 2007), in this work the purpose of detecting changes is to exploit the extracted information in order to provide intelligent support to the community as a whole. Along the same line, a principle difference from the existing work is that we aim to detect change patterns connected to specific processes related to effective functioning and sustainability of a VC and not just to model a VC. In contrast with the existing methods, which consider simple indicators for a relationship (e.g. direct connection), (Leskovec et al., 2005; Falkowski et al., 2006; Falkowski and Spiliopoulou, 2007), we exploit semantic techniques (such as resource meta-data and ontological reasoning) to derive possible relationships between members.

9.5 Improvements and Future Work

In the previous section we have outlined the main achievements and contributions of this work. This section will note possible applications of the current framework and draw immediate improvements, followed by a discussion of further long term research directions.

9.5.1 Further Application of the Current Framework

Long term evaluation study

The framework evaluation has been discussed in Chapter 8 and pointed at the effects of the notifications on VC members and on the VC as a whole. The results pointed at the limited duration of the experimental study and the need for a longer evaluation that would validate the effects of the notification on the development of TM, SMM and of the monitoring of CCen.

As immediate application of the approach, we are planning to continue monitoring the VC by extracting static and change patterns that will allow the identification of possible knowledge sharing problems with respect to TM, SMM, and CCen within the VC. This will inform the generation of appropriate notifications. We have noticed that a newcomer of the VC became active shortly after the end of the evaluation study. This shows that even though the community was brought together for the purpose of the specific study, the VC became part of some people's practice. It will be interesting to investigate the phases the community will go through and for how long will the community sustain given that we are providing the appropriate defined support. Furthermore, the experimental design can be improved further. For example a better feedback mechanism could be employed that will allow members to immediate provide feedback on the suitability and usefulness of the notifications they received. Furthermore, interviews can be conducted with selected members that will allow a deeper analysis of the effects of the notifications on the VC as a whole.

Using different forms of notifications

During the experimental study (Chapter 8) 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. Other approaches on designing notification messages can be considered, implemented and validated using the current framework. Social theories can inform designing 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 easy to facilitate participation in a community of students when they receive a reward towards their final mark (Cheng and Vassileva, 2005). When communities are voluntary, further motivation, incentives and persuasion 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 closely-knit VC.

Application in different community types

A community of researchers is one of the possible applications of the current framework. A further application in different types of closely-knit communities is an appealing continuation of this work. CoPs are an attractive possible application of our framework. It will be interesting to examine whether our approach will have a benefit to CoPs members and extend their 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 CoPs are usually having 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 CoPs. 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?

9.5.2 Improvements and Immediate Research Directions

Evolving expertise of VC members

Although the current research interest of members can be extracted through the keywords of the resources uploaded and downloaded by VC members, the current framework does not consider evolution of the expertise of VC members. There are studies following similar approaches as ours (graph based approaches) that can observe expertise evolution (Song et al., 2005). In our case, graph based approaches can be combined with the ontology to examine expertise evolution based on the interests of members and the timestamps associated with these interests.

Improve the scalability of RM algorithms

The algorithms developed for extracting relationships among people exploit all possible options, and thus scalability would be an issue. For example the algorithms will be very slow if employed to extract relationships in a very large corpus of resources (e.g. more than 1000). Thus, a possible solution to this problem is to consider only the *n* latest resources uploaded or downloaded by members when extracting similarities and or relations among them. In this way, the current interests of members will be considered, as well as semantic relationships will be extracted much quicker. A different approach would be to consider only the direct sub-classes and super-classes (instead of all the sub-classes and super-classes) of a given class in the ontology when semantically enhancing the keywords extracted from resources.

Overcome the “cold-start” problem

One of the main downsides of using tracking data for modelling a community is that no relations can be modelled unless interaction in the VC happens. This problem appeared also in our case, e.g. we could not extract any similarity unless members actually downloaded/ uploaded a resource in the VC. This is also a problem with newcomers. A newcomer has no log of interactions in the community, thus no relation can be extracted for that member. Other community systems prompt users to provide initial information about their interests, for example select favourite movies (Harper et al., 2007). This can be a direction to overcome this problem in our approach. For example we can ask members to provide a number of keywords/phrases upon registration to the community that will represent their interests and can be used as initial input to the CM algorithms. A similar approach will be to provide a list of resources or all the resources available in the VC and ask members to select resources that best represent their interests and use those as initial input to the community modelling algorithms.

Combine different ways of community support and awareness

Notification messages are only one way of providing support to VC members. Using the current framework, and through the information extracted in the CM, one can generate visualisations that will complement the notifications sent to members. There are currently approaches that are using visualisations as their main way of providing support and awareness (Bretzke and Vassileva, 2003; Kay et al., 2006; Upton and Kay, 2009). Visualisations can also be used as a medium to open the CM to VC members and use it as a medium of creating awareness among VC members (Upton and Kay, 2009).

9.5.3 Longer Term Future Research Directions

Applicability of the approach in different stages of the community lifecycle

Section 2.2.2 presented a lifecycle that communities usually follow during their existence. We have argued that with our framework we support three stages of this lifecycle. Further research can look for a systematic way to examine how the community evolves as a whole through each stage. For example, what are the factors and processes that need to be supported at each stage and how can these processes be promoted in the community; what methods can be used in examining the influence of these processes to the VC as a whole? Furthermore, different methods exist in supporting VCs. It will be helpful to examine what methods can be employed in generating support for VC members at each stage of the lifecycle. The above questions can be answered through a long term research study that will systematically analyse a VC throughout its life from planning all the way to close.

Improved motivational approach

Although in this research we argued that by supporting the sharing of knowledge we can facilitate collective knowledge sharing, personal and cultural resistance to sharing has a negative impact in collective knowledge sharing. Section 9.3.3, discussed how different personal organisation habits can influence the sharing of knowledge in a community (Indratmo and Vassileva, 2005). Resistance to sharing with others constitute a problem for knowledge sharing in virtual communities. Different methods employed by researchers to motivate and facilitate participation (Agostini et al., 2003; Preece et al., 2004; Cheng and Vassileva, 2005; Harper et al., 2007; Ardissono et al., 2009). Persuasion theories can be encompass to inform the modelling and in extend suitable support generation (Cialdini, 1993). The following aspect can be considered in designing appropriate mechanisms for persuading members to participate:

- Reciprocity – people returning a favour;
- Commitment – people honour the commitment they perceive as the right one;
- Social Proof – people will do things that they see other people are doing;
- Authority – people tend to obey authority figures;
- Liking – people are easily persuaded by other people whom they like.

Similarly long time studies can examine how technology build based on motivational and incentive theories can influence the knowledge sharing in a VC. A possible research direction along this line is the development of models based on cultural aspects (Reinecke and Bernstein, 2009) of people in a community, and how can these inform technologies for supporting knowledge sharing.

And finally, we have started this research with the view that by providing support for knowledge sharing in a VC as a whole, based on TM, SMM and CCen, we will be able to help a virtual community to sustain. We have argued that by modelling the community and providing support to the community as a whole rather than the individual member, it will be possible to facilitate knowledge sharing and to improve the experience of VC members. The diversity of aspects we examined and the knowledge we gained through the several theoretical areas made our work an exciting research journey. We are confident that future research on supporting VCs will benefit from the framework presented in this thesis. Potential extensions include the integration of the framework as a community modelling service, in an existing tool, that will inform support to users. Furthermore, an immediate application of the framework has been considered in the BRAIN project funded by JISC and involves the University of Leeds and Coventry University in the UK. The peer reviewed publications presenting this research demonstrate that the framework can be considered as a valuable contribution in the research community.

References

Ackerman, M. (1998): Augmenting Organizational Memory: A Field Study of Answer Garden, *ACM Transactions on Information Systems*, vol. 16, no. 3, pp.224.

Ackerman, M. and McDonald, D. (1996): *Answer Garden 2: merging organizational memory with collaborative help*, CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work, ACM: Boston, Massachusetts, United States

Agostini, A., Albolino, S., De Michelis, G. D., De Paoli, F. D. and Dondi, R. (2003): *Stimulating knowledge discovery and sharing*, GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work, ACM: Sanibel Island, Florida, USA

Alicia, I. and Gondy, L. (2009): A life-cycle perspective on online community success, *ACM Computing Surveys (CSUR)*, vol. 41, no. 2, pp.1-29.

Applet, W. (1999): *WWW Based Collaboration with the BSCW System*, SOFSEM'99, Springer-Verlag: Milovy, Czech Republic

Ardichvili, A., Page, V. and Wentling, T. (2003): Motivation and barriers to participation in virtual knowledge-sharing communities of practice, *Journal of Knowledge Management*, vol. 7, no. 1, pp.64-77.

Ardissono, L., Bosio, G. G. A. and Petrone, G. (2009): *Context-aware notification management in an integrated collaborative environment*, In Proceedings of International Workshop on Adaptation and Personalization for Web 2.0 (AP-Web 2.0 2009) at UMAP’09: Trento, Italy

Asur, S., Parthasarathy, S. and Ucar, D. (2007): *An event-based framework for characterizing the evolutionary behavior of interaction graphs*, KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM: San Jose, California, USA

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

Bajanki, S., Kaufhold, K., Bek, A. L., Dimitrova, V., Lau, L., O'Rourke, R. and Walker, A. (2009): *Use of Semantics to Build an Academic Writing Community Environment*, Proceeding of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, IOS Press: Brighton, UK

Berlin, L., Jeffries, R., O'Day, V., Paepcke, A. and Wharton, C. (1993): *Where did you put it? Issues in the design and use of a group memory*, CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM: Amsterdam, The Netherlands

Bieber, k., Engelbart, D., Furuta, R., Hiltz, S. R., Noll, J., Preece, J., Stohr, E. A., Turoff, M. and Walle, B. V. D. (2002): Towards Virtual Community Knowledge Evolution, *Journal of Management Information Systems*, vol. 18, no. 4, pp.11 - 35.

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

Brusilovsky, P., Karagiannidis, C. and Sampson, D. (2004): Layered evaluation of adaptive learning systems, *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 14, no. 4, pp.402-421.

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

Chanal, V. and Kimble, C. (2010): *Born to be Wild: Using Communities of Practice as a Tool for Knowledge Management*, Ethicomp 2010: The'Backwards, Forwards and Sideways' changes of ICT Tarragona, Spain

Cheng, R. and Vassileva, J. (2005): *Adaptive Reward Mechanism for Sustainable Online Learning Community*, International Conference on Artificial Intelligence in Education 2005: Amsterdam, NL

Cheng, R. and Vassileva, J. (2005): *User and Community Adaptive Rewards Mechanism for Sustainable Online Community*, 10th International Conference on User Modelling 2005: Edinbourgh, Scotland

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.

Constant, D., Kiesler, S. and Sproull, L. (1994): What's Mine Is Ours, or Is It? A Study of Attitudes about Information Sharing, *INFORMATION SYSTEMS RESEARCH*, vol. 5, no. 4, pp.400-421.

Davies, J. and Duke, A. (2004): OntoShare - An Ontology -Based Knowledge sharing System for Virtual Communities of Practice, *The Journal of Universal Computer Science*, vol. 10, no. 3, pp.262 - 283.

Davies, J., Duke, A. and Sure, Y. (2003): *OntoShare: a knowledge management environment for virtual communities of practice*, K-CAP '03: Proceedings of the international conference on Knowledge capture, ACM: Sanibel Island, FL, USA

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.

Dennis, A. R., Pootheri, S. K. and Natarajan, V. L. (1998): Lessons from the early adopters of Web groupware., *Journal of Management Information Systems*, vol. 14, no. 4, pp.65-86.

Dillenbourg, P. (1999). What do you mean by 'Collaborative Learning' ? *Collaborative-learning: Cognitive and Computational Approaches*. Dillenbourg, P., Elsevier, Oxford**:** pp. 1-19.

Edmondson, A. (1999): Psychological safety and learning behavior in work teams, *Administrative Science Quarterly*, vol. 44, no. 2, pp.350 - 383.

Engestrom, Y. (2001): Expansive Learning at Work: toward an activity theoretical reconceptualization, *Journal of Education and Work*, vol. 14, no. 1, pp.133 - 156.

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

Eunyoung, C. and JoongHo, A. (2009): *Virtual community 101: know your virtual community and members*, Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, ACM: Suwon, Korea

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., Jorg, B. and Myra, S. (2006): *Mining and Visualizing the Evolution of Subgroups in Social Networks*, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006), IEEE Computer Society: Hong Kong, China

Falkowski, T. and Spiliopoulou, M. (2007): *Users in Volatile Communities: Studying Active Participation and Community Evolution*, Proc. of the International Conference on User Modeling 2007, 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. (2001): *Communities of Interest: Learning through the Interaction of Multiple Knowledge Systems*, 24th Annual Information Systems Research Seminar in Scandinavia (IRIS'24): Ulvik, Norway

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

Goecks, J. and Cosley, D. (2002): *NuggetMine: intelligent groupware for opportunistically sharing information nuggets*, IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces, ACM: San Francisco, California, USA

Golder, S. and Huberman, B. (2005). "The Structure of Collaborative Tagging Systems." from http://arxiv.org/abs/cs.DL/0508082

Gouvea, M. T. A., Motta and Santoro (2006): *Recommendation as a Mechanism to Induce Participation in Communities of Practice*, Computer Supported Cooperative Work in Design II: 9th International Conference, CSCWD 2005, Springer Berlin / Heidelberg: Coventry, UK

Gouvea, M. T. A., Motta, C. and Fm, S. (2005): *Designing Support to Participation in Communities of Practice through Recommender Systems*, Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies, IEEE Computer Society: Kaohsiung, Taiwan

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

Gruber, T. (1993): A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, vol. 5, no. 2, pp.199-220.

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

Hartley, J. R. (1999): Effective Pedagogies for Managing Collaborative Learning in On-line Learning Environments, *Journal of Educational Technology & Society*, vol. 2, no. 2

He, J. (2004): *Knowledge Impacts of User Participation: A Cognitive Perspective*, Proceedings of the 2004 SIGMIS conference on Computer personnel research: Careers, culture, and ethics in a networked environment (SIGMIS '04), ACM: Tucson, Arizona, 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.

Hoadley, C. M. and Kilner, P. G. (2005): Using Technology to Transform Communities of Practice into Knowledge Building Communities, *SIGGROUP Bulletin*, vol. 25, no. 1, pp.31 - 40.

Hollingshead, A. B. (1998): Communication, Learning and Retrieval in Transactive Memory Systems, *Journal of Experimental Social Psychology*, vol. 34, no. 5, pp.423 - 442.

Hollingshead, A. B. (2000): Perceptions of Expertise and Transactive Menory in Work Relationships, *Group Processes and Intergroup Relations*, vol. 3, no. 3, pp.257 - 267.

Hollingshead, A. B. and Brandon, D. P. (2003): Potential Benefits of Communication in Transactive Memory systems, *Human Communication Research*, vol. 29, no. 4, pp.607 - 615.

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

Hung, D. W. L. and Der-Thanq, C. (2001): Situated Cognition, Vygotskian Thought and Learning from the Communities of Practice Perspective: Implications for the Design of Web-Based E-Learning, *Education Media International*, vol. 38, no. 1, pp.3 - 12.

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.

Indratmo and Vassileva, J. (2005): *Human and Social Aspects of Decentralized Knowledge Communities*, Proc. Workshop "The Semantic Desktop", at the International Semantic Web Conference: Galway, Ireland

Jameson, A. (2003). Adaptive Interfaces and Agents. *Human-Computer Interaction Handbook*. Jacko, J. A. and Sears, A., Erlbaum, Mahwah, NJ**:** pp. 305 - 330.

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.

Karagiannidis, C. and Sampson, D. (2000): *Layered Evaluation of Adaptive Applications and Services,* AH '00: Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer-Verlag: Trento, Italy

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

Kerr, N. L. and Tindale, R. S. (2004): Group Performance and Decision Making, *Annual Review of Psychology,* vol. 55, no., pp.623 - 655.

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

Kim, A. J. (2000): *Community Building on the Web: Secret Strategies for Successful Online Communities*. London: Addison Wesley.

Klein, J. H., Connell, N. A. D. and Meyer, E. (2005): Knowledge characteristics of communities of practice, *Knowledge Management Research & Practice,* vol. 3, no. 2, pp.106-114.

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.

Koper, R., Pannekeet, K., Hendriks, M. and Hummel, H. (2004): Building Communities for the Exchange of Learning Objects: Theoretical Foundations and Requirements, *Alt- J Research in Learning Technology,* vol. 12, no. 1, pp.21 - 35.

Kosba, E., Dimitrova, V. and Boyle, R. (2007): Adaptive feedback generation to support teachers in web-based distance education, *User Modeling and User-Adapted Interaction,* vol. 17, no. 4, pp.379-413.

Kreijns, K., Kirschner, P. A. and Jochems, W. (2002): The Sociability of Computer-Supported Collaborative Learning Environments, *Educational Technology & Society,* vol. 5, no. 1, pp.8 - 22.

Kumar, R., Novak, J. and Tomkins, A. (2006): *Structure and evolution of online social networks*, KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM Press: Philadelphia, PA, USA

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.

Lau, L., Dimitrova, V., Bajanki, S., Kaufhold, K., Le Bek, A., O'Rourke, R. and Walker, A. (2009): *Semantic social scaffolding for communities of practice in dissertation writing*, TEL-CoPs'09: 3rd International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice to be held in conjunction with the 4th European Conference on Technology Enhanced Learning: Nice, France

Laurillard, D. (2002): *Rethinking University Teaching*. London: Routledge Falmer.

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

Lazar, J. and Preece, J. (2002). Social Considerations in Online Communities: Usability, Sociability, and Success Factors. *Cognition in the Digital World*. Oostendorp, H. v., Lawrence Erlbaum Associates Inc. Publishers, Mahwah: NJ

Leskovec, J., Kleinberg, J. and Faloutsos, C. (2005): *Graphs over time: densification laws, shrinking diameters and possible explanations*, KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, ACM Press: Chicago, Illinois, USA

Lewis, D. and Allan, B. (2005): *Virtual Learning Communities: a Guide to Practitioners*. Berkshire: Open University, McGraw Hill.

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

Lindstaedt, S. (1996): *Towards Organizational Learning: Growing Group Memories in the Workplace*, Conference on Human Factors in Computing Systems: Vancouver, British Columbia, Canada

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

Mark, M. and Greer, J. (1993): Evaluation methodologies for intelligent tutoring systems, *Journal of Artificial Intelligence in Education*, vol. 4, no. 2-3, pp.129-53.

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).

McMillan, D. W. and Chavis, M. D. (1986): Sense of Community: a Definition and Theory, *Journal of Community Psychology*, vol. 14, no., pp.6 - 23.

Menges, F., Mishra, B. and Narzisi, G. (2008): Modeling and simulation of e-mail social networks: a new stochastic agent-based approach, *WSC '08: Proceedings of the 40th Conference on Winter Simulation*, Winter Simulation Conference Miami, Florida.

Merali, Y. and Davies, J. (2001): Knowledge Capture and Utilization in Virtual Communities, *In Proceedings of the First International Conference on Knowledge Capture* Victoria, British Columbia, Canada.

Millan, E. and Perez-De-La-Cruz, J. (2002): A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation, *User Modeling and User-Adapted Interaction*, vol. 12, no. 2, pp.281-330.

Miller, G. (1995): WordNet: a lexical database for English, *Commun. ACM*, vol. 38, no. 11, pp.39-41.

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.

Mulder, I., Swaak, J. and Kessels, J. (2002): Assessing Group Learning and Shared Understanding in Technology-Mediated Interaction, *Educational Technology & Society*, vol. 5, no. 1, pp.35 - 47.

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.
Noy, N. and McGuinness, L. D. (2001). "Ontology Development 101: A Guide to Creating Your First Ontology." Retrieved 23 July, 2010, from http://protege.stanford.edu/publications/ontology_development/ontology101.html.
Olson, D. L. and Delen, D. (2008): *Advanced Data Mining Techniques*. Berlin: Springer.
Palla, G., Barabasi, A.-L. and Vicsek, T. (2007): Quantifying social group evolution, *Nature*, vol. 446, no. 7136, pp.664-667.
Paramythis, A. and Weibelzahl, S. (2005): *A Decomposition Model for the Layered Evaluation of Interactive Adaptive Systems*, International Conference on User Modeling 2005, Springer: Edinburgh, Scotland, UK
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.
Preece, J. (2001). Online Communities: Usability, Sociability, Theory and Methods. *Frontiers of Human-Centred Computing, Online Communities and Virtual Environments*. Earnshaw, R., Guedj, R., Dam, A. v. and Vince, T., Springer Verlag, Amsterdam**:** pp. 263 - 277.
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
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.
Rashid, A., Ling, K., Tassone, R., Resnick, P., Kraut, R. and Riedl, J. (2006): *Motivating participation by displaying the value of contribution*, CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM: Montréal, Québec, Canada
Reinecke, K. and Bernstein, A. (2009): *Tell Me Where You've Lived, And I'll Tell You What You Like: Adapting Interfaces to Cultural Preferences*, Proceedings of User Modeling, Adaptation, and Personalization (UMAP), 2009, Springer, Verlag: Trento, Italy
Resnik, P. (1995): *Using Information Content to Evaluate Semantic Similarity in a Taxonomy*, In Proceedings of the 14th International Joint Conference on Artificial Intelligence: Montreal, Quebec, Canada
Rheingold, H. (1993): *The Virtual Community: Homesteading on the Electronic Frontier*. Reading, MA: ADDISON-WESLEY PUBLISHING COMPANY.
Romm, C., Pliskin, N. and Clarke, R. (1997): Virtual communities and society: Toward an integrative three phase model, *International Journal of Information Management*, vol. 17, no. 4, pp.261-270.
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.

Shen, K., Yu, A., Khalifa, M., Shen, K., Khalifa, M. and Yu, A. (2006): *Supporting Social Interaction in Virtual Communities: Role of Social Presence*, Proceedings of the Twelfth Americas Conference on Information Systems: Acapulco, Mexico

Shlomo, B., Ariel, G., Tsvi, K. and Larry, M. (2007): *Evaluating User Model Effectiveness by Simulation*, Workshop on Personalized Access on Cultural Heritage at 11th international conference on user modeling: Corfu, Greece

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

Stocker, R. and Larkin, H. (2008): Association in multi-agent simulations of dynamic random social networks, *BIONETICS '08: Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) Hyogo, Japan.

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.

Tian, Y., Huang, T. and Gao, W. (2001): *Algorithms of integrated student modeling in online virtual educational community*, Proceedings of the Info-tech and Info-net, 2001 conference, IEEE: Beijing

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

Vanlehn, K., Ohlsson, S. and Nason, R. (1996): Applications of simulated students: An exploration, *Journal of Artificial Intelligence in Education*, vol. 5, no. 2, pp.135 - 175.

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

Vygotsky, L. (1978). Mind in Society:The development of higher psychological processes. Cole, M., John-Steiner, V., Scribner, S. and Souberman, E., Harvard University Press, Cambridge**:** pp. 86.

Wang, Y., Graether, W. and Prinz, W. (2007): *Suitable notification intensity: the dynamic awareness system*, Proc. of the int. ACM conf. on Supporting Group Work: GROUP 2007, ACM: Sanibel Island, Florida, 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. (1998): Communities of Practice: Learning as a Social System, *Systems Thinker,* vol. 9, no. 5, pp.2 - 3.

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

Wenger, E. (2001): *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Wenger, E., McDermott, R. and Snyder, M. W. (2002): *Cultivating Communities of Practice: A Guide to Managing Knowledge* Harvard Business School Press.

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

Wu, H. and Gordon, M. D. (2004): Collaborative Filing in a Document Repository, *SIGIR '04* Sheffield, South Yorkshire, UK.

Zacklad, M. (2003): *Communities of action: a cognitive and social approach to the design of CSCW systems*, GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work, ACM: Sanibel Island, Florida, USA

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

Zhdanova, A. (2005): *Towards a community-driven ontology matching*, K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture, ACM: Banff, Alberta, Canada

Appendix A

Ontology Used as Community Context

This appendix will give the ontology developed and used in the studies on Chapters 4, 5, 6 and 8. The ontology is presented here in XML format. The ontology was build based on the titles of the hierarchy of folders created in the VC. This was done in order to represent the context of the community.

A.1 Ontology


```
  <?xml version="1.0" ?>
  <!DOCTYPE rdf:RDF (View Source for full doctype...)>
- <rdf:RDF xmlns="http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#"
    xml:base="http://www.comp.leeds.ac.uk/stellak/semwebonto.owl"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl11="http://www.w3.org/2006/12/owl11#"
    xmlns:owl11xml="http://www.w3.org/2006/12/owl11-xml#"
    xmlns:semwebonto="http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <owl:Ontology rdf:about="" />
          - <!--


          ///////////////////////////////////////////////////////////////////////////////
          //////
              //
              // Classes
              //

          ///////////////////////////////////////////////////////////////////////////////
          //////

    -->
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation
    -->
- <owl:Class rdf:about="#Adaptation">
  <rdfs:subClassOf rdf:resource="#Application" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Application_Service
    -->
- <owl:Class rdf:abou="#Adaptation_Application_Service">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Education_Hypermedia
    -->
- <owl:Class rdf:about="#Adaptation_Education_Hypermedia">
  <rdfs:subClassOf rdf:resource="#Adaptation_Hypermedia" />
  <rdfs:subClassOf rdf:resource="#Adaptation" />
  <rdfs:subClassOf rdf:resource="#Education" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Hypermedia
    -->
- <owl:Class rdf:about="#Adaptation_Hypermedia">
```

```
  <rdfs:subClassOf rdf:resource="#Adaptation" />
  <rdfs:subClassOf rdf:resource="#Hypermedia" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Hypermedia_Service
    -->
- <owl:Class rdf:about="#Adaptation_Hypermedia_Service">
  <rdfs:subClassOf rdf:resource="#Adaptation_Hypermedia" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Hypermedia_System
    -->
- <owl:Class rdf:about="#Adaptation_Hypermedia_System">
  <rdfs:subClassOf rdf:resource="#Adaptation_Hypermedia" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Intelligence_Technology
    -->
- <owl:Class rdf:about="#Adaptation_Intelligence_Technology">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Internet_Course
    -->
- <owl:Class rdf:about="#Adaptation_Internet_Course">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
  <rdfs:subClassOf rdf:resource="#Education" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Learning_System
    -->
- <owl:Class rdf:about="#Adaptation_Learning_System">
  <rdfs:subClassOf rdf:resource="#Learning" />
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_System
    -->
- <owl:Class rdf:about="#Adaptation_System">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Technique
    -->
- <owl:Class rdf:about="#Adaptation_Technique">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Web
    -->
- <owl:Class rdf:about="#Adaptation_Web">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Adaptation_Web_Site
    -->
- <owl:Class rdf:about="#Adaptation_Web_Site">
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Agent
    -->
- <owl:Class rdf:about="#Agent">
  <rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Application
    -->
- <owl:Class rdf:about="#Application">
  <rdfs:subClassOf rdf:resource="#Semantics_Web" />
    </owl:Class>
  - <!--
```

```
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Author_Adaptation_Hypermedia
-->
- <owl:Class rdf:about="#Author_Adaptation_Hypermedia">
<rdfs:subClassOf rdf:resource="#Adaptation_Hypermedia" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Cognition_Centrality
-->
- <owl:Class rdf:about="#Cognition_Centrality">
<rdfs:subClassOf rdf:resource="#Organisation_Psychology" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Cognition_Consensus
-->
- <owl:Class rdf:about="#Cognition_Consensus">
<rdfs:subClassOf rdf:resource="#Organisation_Psychology" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Cognition_Engineering
-->
- <owl:Class rdf:about="#Cognition_Engineering">
<rdfs:subClassOf rdf:resource="#Engineering" />
<rdfs:subClassOf rdf:resource="#Social_Cognition_Theory" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Collaboration
-->
- <owl:Class rdf:about="#Collaboration">
<rdfs:subClassOf rdf:resource="#Knowledge_Management" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Community
-->
- <owl:Class rdf:about="#Community">
<rdfs:subClassOf rdf:resource="#Application" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Community_Practice
-->
- <owl:Class rdf:about="#Community_Practice">
<rdfs:subClassOf rdf:resource="#Community" />
<rdfs:subClassOf rdf:resource="#Knowledge_Management" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Concept
-->
- <owl:Class rdf:about="#Concept">
<rdfs:subClassOf rdf:resource="#Ontology" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Concept_Modelling
-->
- <owl:Class rdf:about="#Concept_Modelling">
<rdfs:subClassOf rdf:resource="#Concept" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Content_Adaptation
-->
- <owl:Class rdf:about="#Content_Adaptation">
<rdfs:subClassOf rdf:resource="#Adaptation_Technique" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Content_Management
-->
- <owl:Class rdf:about="#Content_Management">
<rdfs:subClassOf rdf:resource="#Information_Management" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Context
-->
- <owl:Class rdf:about="#Context">
<rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
```

```
  <rdfs:subClassOf rdf:resource="#Ontology" />
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Description_Logic
  -->
- <owl:Class rdf:about="#Description_Logic">
  <rdfs:subClassOf rdf:resource="#Logic" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Description_Subsumption
  -->
- <owl:Class rdf:about="#Description_Subsumption">
  <rdfs:subClassOf rdf:resource="#Description_Logic" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Digitalisation_Gazetteer
  -->
- <owl:Class rdf:about="#Digitalisation_Gazetteer">
  <rdfs:subClassOf rdf:resource="#Library_Information_Science" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Digitalisation_Library
  -->
- <owl:Class rdf:about="#Digitalisation_Library">
  <rdfs:subClassOf rdf:resource="#Library_Information_Science" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Document_Management
  -->
- <owl:Class rdf:about="#Document_Management">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Domain_Ontology
  -->
- <owl:Class rdf:about="#Domain_Ontology">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education
  -->
- <owl:Class rdf:about="#Education">
  <rdfs:subClassOf rdf:resource="#Application" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Hypermedia
  -->
- <owl:Class rdf:about="#Education_Hypermedia">
  <rdfs:subClassOf rdf:resource="#Hypermedia" />
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Materialism
  -->
- <owl:Class rdf:about="#Education_Materialism">
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Metadata
  -->
- <owl:Class rdf:about="#Education_Metadata">
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Semantics_Web
  -->
- <owl:Class rdf:about="#Education_Semantics_Web">
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Web_Service
```

```
    -->
- <owl:Class rdf:about="#Education_Web_Service">
  <rdfs:subClassOf rdf:resource="#Education" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Education_Working_Group
    -->
- <owl:Class rdf:about="#Education_Working_Group">
  <rdfs:subClassOf rdf:resource="#Education" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Electronics_Community
    -->
- <owl:Class rdf:about="#Electronics_Community">
  <rdfs:subClassOf rdf:resource="#Community" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Electronics_Learning
    -->
- <owl:Class rdf:about="#Electronics_Learning">
  <rdfs:subClassOf rdf:resource="#Learning" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Electronics_Learning_Course
    -->
- <owl:Class rdf:about="#Electronics_Learning_Course">
  <rdfs:subClassOf rdf:resource="#Learning_Management_System" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Electronics_Learning_Environment
    -->
- <owl:Class rdf:about="#Electronics_Learning_Environment">
  <rdfs:subClassOf rdf:resource="#Electronics_Learning" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Element_Set
    -->
- <owl:Class rdf:about="#Element_Set">
  <rdfs:subClassOf rdf:resource="#Metadata" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Emergence_Semantics
    -->
- <owl:Class rdf:about="#Emergence_Semantics">
  <rdfs:subClassOf rdf:resource="#Semantics" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Engineering
    -->
- <owl:Class rdf:about="#Engineering">
  <rdfs:subClassOf rdf:resource="#Application" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Evaluation
    -->
- <owl:Class rdf:about="#Evaluation">
  <rdfs:subClassOf rdf:resource="#Adaptation_System" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Evaluation_Adaptation_System
    -->
- <owl:Class rdf:about="#Evaluation_Adaptation_System">
  <rdfs:subClassOf rdf:resource="#Evaluation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#First_Order_Logic
    -->
- <owl:Class rdf:about="#First_Order_Logic">
  <rdfs:subClassOf rdf:resource="#Logic" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Geography_Ontology
```

```
  -->
- <owl:Class rdf:about="#Geography_Ontology">
  <rdfs:subClassOf rdf:resource="#Domain_Ontology" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Group_Modelling
  -->
- <owl:Class rdf:about="#Group_Modelling">
  <rdfs:subClassOf rdf:resource="#User_Modelling" />
  <rdfs:subClassOf rdf:resource="#Community" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Heuristic_Evaluation
  -->
- <owl:Class rdf:about="#Heuristic_Evaluation">
  <rdfs:subClassOf rdf:resource="#Evaluation" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Horn_Logic
  -->
- <owl:Class rdf:about="#Horn_Logic">
  <rdfs:subClassOf rdf:resource="#Logic" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Human_Computer_Interaction
  -->
- <owl:Class rdf:about="#Human_Computer_Interaction">
  <rdfs:subClassOf rdf:resource="#Application" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Hypermedia
  -->
- <owl:Class rdf:about="#Hypermedia">
  <rdfs:subClassOf rdf:resource="#Application" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Hypermedia_Design
  -->
- <owl:Class rdf:about="#Hypermedia_Design">
  <rdfs:subClassOf rdf:resource="#Hypermedia" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Hypermedia_Learning
  -->
- <owl:Class rdf:about="#Hypermedia_Learning">
  <rdfs:subClassOf rdf:resource="#Education_Hypermedia" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Inference_Web
  -->
- <owl:Class rdf:about="#Inference_Web">
  <rdfs:subClassOf rdf:resource="#Reasoning_Semantics_Web" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Information_Management
  -->
- <owl:Class rdf:about="#Information_Management">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Information_Retrieval
  -->
- <owl:Class rdf:about="#Information_Retrieval">
  <rdfs:subClassOf rdf:resource="#Library_Information_Science" />
  <rdfs:subClassOf rdf:resource="#Knowledge_Engineering" />
  </owl:Class>
- <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Information_Retrieval_Process
  -->
- <owl:Class rdf:about="#Information_Retrieval_Process">
  <rdfs:subClassOf rdf:resource="#Information_Retrieval" />
```

```
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Information_Share
   -->
- <owl:Class rdf:about="#Information_Share">
  <rdfs:subClassOf rdf:resource="#Collaboration" />
  <rdfs:subClassOf rdf:resource="#Community" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Integration_Rule_Ontology
   -->
- <owl:Class rdf:about="#Integration_Rule_Ontology">
  <rdfs:subClassOf rdf:resource="#Ontology_Representation" />
  <rdfs:subClassOf rdf:resource="#Rule_Language" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Intelligence_User_Interface
   -->
- <owl:Class rdf:about="#Intelligence_User_Interface">
  <rdfs:subClassOf rdf:resource="#User_Modelling" />
  <rdfs:subClassOf rdf:resource="#Human_Computer_Interaction" />
  <rdfs:subClassOf rdf:resource="#Adaptation" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Interaction_System
   -->
- <owl:Class rdf:about="#Interaction_System">
  <rdfs:subClassOf rdf:resource="#Human_Computer_Interaction" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Interaction_Television
   -->
- <owl:Class rdf:about="#Interaction_Television">
  <rdfs:subClassOf rdf:resource="#Multimedia" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Interface_Design
   -->
- <owl:Class rdf:about="#Interface_Design">
  <rdfs:subClassOf rdf:resource="#Human_Computer_Interaction" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Acquisition
   -->
- <owl:Class rdf:about="#Knowledge_Acquisition">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  <rdfs:subClassOf rdf:resource="#Knowledge_Engineering" />
  <rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Base
   -->
- <owl:Class rdf:about="#Knowledge_Base">
  <rdfs:subClassOf rdf:resource="#Ontology" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Building
   -->
- <owl:Class rdf:about="#Knowledge_Building">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Community
   -->
- <owl:Class rdf:about="#Knowledge_Community">
  <rdfs:subClassOf rdf:resource="#Community" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Engineering
   -->
- <owl:Class rdf:about="#Knowledge_Engineering">
```

```
<rdfs:subClassOf rdf:resource="#Engineering" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Flow
  -->
- <owl:Class rdf:about="#Knowledge_Flow">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Gap
  -->
- <owl:Class rdf:about="#Knowledge_Gap">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Management
  -->
- <owl:Class rdf:about="#Knowledge_Management">
  <rdfs:subClassOf rdf:resource="#Application" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Representation_Reasoning
  -->
- <owl:Class rdf:about="#Knowledge_Representation_Reasoning">
  <rdfs:subClassOf rdf:resource="#Application" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Knowledge_Sharing
  -->
- <owl:Class rdf:about="#Knowledge_Sharing">
  <rdfs:subClassOf rdf:resource="#Knowledge_Management" />
  <rdfs:subClassOf rdf:resource="#Learning" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Layer_Evaluation
  -->
- <owl:Class rdf:about="#Layer_Evaluation">
  <rdfs:subClassOf rdf:resource="#Evaluation_Adaptation_System" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning
  -->
- <owl:Class rdf:about="#Learning">
  <rdfs:subClassOf rdf:resource="#Education" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Community
  -->
- <owl:Class rdf:about="#Learning_Community">
  <rdfs:subClassOf rdf:resource="#Community" />
  <rdfs:subClassOf rdf:resource="#Learning" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Environment
  -->
- <owl:Class rdf:about="#Learning_Environment">
  <rdfs:subClassOf rdf:resource="#Learning" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Management_System
  -->
- <owl:Class rdf:about="#Learning_Management_System">
  <rdfs:subClassOf rdf:resource="#Electronics_Learning" />
  <rdfs:subClassOf rdf:resource="#Learning" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Object
  -->
- <owl:Class rdf:about="#Learning_Object">
  <rdfs:subClassOf rdf:resource="#Learning" />
  <rdfs:subClassOf rdf:resource="#Library_Information_Science" />
```

```
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Object_Metadata
  -->
- <owl:Class rdf:about="#Learning_Object_Metadata">
  <rdfs:subClassOf rdf:resource="#Learning_Object" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Learning_Technology
  -->
- <owl:Class rdf:about="#Learning_Technology">
  <rdfs:subClassOf rdf:resource="#Learning" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Library_Information_Science
  -->
- <owl:Class rdf:about="#Library_Information_Science">
  <rdfs:subClassOf rdf:resource="#Application" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Logic
  -->
- <owl:Class rdf:about="#Logic">
  <rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Logic_Programming
  -->
- <owl:Class rdf:about="#Logic_Programming">
  <rdfs:subClassOf rdf:resource="#Logic" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Metadata
  -->
- <owl:Class rdf:about="#Metadata">
  <rdfs:subClassOf rdf:resource="#Learning_Object" />
  <rdfs:subClassOf rdf:resource="#Web_Data_Integration" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Metadata_Schema
  -->
- <owl:Class rdf:about="#Metadata_Schema">
  <rdfs:subClassOf rdf:resource="#Metadata" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Mobile_Learning
  -->
- <owl:Class rdf:about="#Mobile_Learning">
  <rdfs:subClassOf rdf:resource="#Learning" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Multimedia
  -->
- <owl:Class rdf:about="#Multimedia">
  <rdfs:subClassOf rdf:resource="#Hypermedia" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Multimedia_Information_System
  -->
- <owl:Class rdf:about="#Multimedia_Information_System">
  <rdfs:subClassOf rdf:resource="#Multimedia" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Multimedia_Learning_Object
  -->
- <owl:Class rdf:about="#Multimedia_Learning_Object">
  <rdfs:subClassOf rdf:resource="#Multimedia" />
  <rdfs:subClassOf rdf:resource="#Learning_Object" />
    </owl:Class>
  - <!--
  http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Multimedia_Metadata
  -->
```

```
- <owl:Class rdf:about="#Multimedia_Metadata">
  <rdfs:subClassOf rdf:resource="#Multimedia" />
  <rdfs:subClassOf rdf:resource="#Learning_Object" />
  <rdfs:subClassOf rdf:resource="#Web_Data_Integration" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology
  -->
- <owl:Class rdf:about="#Ontology">
  <rdfs:subClassOf rdf:resource="#Semantics_Web_Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Engineering
  -->
- <owl:Class rdf:about="#Ontology_Engineering">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Evolution
  -->
- <owl:Class rdf:about="#Ontology_Evolution">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Generation
  -->
- <owl:Class rdf:about="#Ontology_Generation">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Language
  -->
- <owl:Class rdf:about="#Ontology_Language">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Learning
  -->
- <owl:Class rdf:about="#Ontology_Learning">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Map
  -->
- <owl:Class rdf:about="#Ontology_Map">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Model
  -->
- <owl:Class rdf:about="#Ontology_Model">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Population
  -->
- <owl:Class rdf:about="#Ontology_Population">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Reason
  -->
- <owl:Class rdf:about="#Ontology_Reason">
  <rdfs:subClassOf rdf:resource="#Reasoner" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Ontology_Representation
  -->
- <owl:Class rdf:about="#Ontology_Representation">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  <rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
```

```
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Organisation_Memory
-->
- <owl:Class rdf:about="#Organisation_Memory">
<rdfs:subClassOf rdf:resource="#Knowledge_Management" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Organisation_Psychology
-->
- <owl:Class rdf:about="#Organisation_Psychology">
<rdfs:subClassOf rdf:resource="#Semantics_Web_Social_Impact" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Pedagogy_Agent
-->
- <owl:Class rdf:about="#Pedagogy_Agent">
<rdfs:subClassOf rdf:resource="#Education" />
<rdfs:subClassOf rdf:resource="#Agent" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Personal_Technique
-->
- <owl:Class rdf:about="#Personal_Technique">
<rdfs:subClassOf rdf:resource="#User_Modelling" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Predicate_Calculus
-->
- <owl:Class rdf:about="#Predicate_Calculus">
<rdfs:subClassOf rdf:resource="#Logic" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Predicate_Logic
-->
- <owl:Class rdf:about="#Predicate_Logic">
<rdfs:subClassOf rdf:resource="#Logic" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Prior_Knowledge
-->
- <owl:Class rdf:about="#Prior_Knowledge">
<rdfs:subClassOf rdf:resource="#Knowledge_Management" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Query_Language
-->
- <owl:Class rdf:about="#Query_Language">
<rdfs:subClassOf rdf:resource="#Ontology_Language" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Reasoner
-->
- <owl:Class rdf:about="#Reasoner">
<rdfs:subClassOf rdf:resource="#Ontology" />
<rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Reasoning
-->
- <owl:Class rdf:about="#Reasoning">
<rdfs:subClassOf rdf:resource="#Reasoner" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Reasoning_Language
-->
- <owl:Class rdf:about="#Reasoning_Language">
<rdfs:subClassOf rdf:resource="#Reasoner" />
</owl:Class>
- <!--
http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Reasoning_Semantics_Web
-->
```

```
- <owl:Class rdf:about="#Reasoning_Semantics_Web">
  <rdfs:subClassOf rdf:resource="#Reasoning" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Recommendation_System
    -->
- <owl:Class rdf:about="#Recommendation_System">
  <rdfs:subClassOf rdf:resource="#Adaptation_System" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Recommendation_System_Adaptation
    -->
- <owl:Class rdf:about="#Recommendation_System_Adaptation">
  <rdfs:subClassOf rdf:resource="#Recommendation_System" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Rule_Language
    -->
- <owl:Class rdf:about="#Rule_Language">
  <rdfs:subClassOf rdf:resource="#Rule_Logic" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Rule_Logic
    -->
- <owl:Class rdf:about="#Rule_Logic">
  <rdfs:subClassOf rdf:resource="#Logic" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Rule_Markup
    -->
- <owl:Class rdf:about="#Rule_Markup">
  <rdfs:subClassOf rdf:resource="#Rule_Logic" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics
    -->
- <owl:Class rdf:about="#Semantics">
  <rdfs:subClassOf rdf:resource="#Semantics_Web" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Grid
    -->
- <owl:Class rdf:about="#Semantics_Grid">
  <rdfs:subClassOf rdf:resource="#Semantics" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Network
    -->
- <owl:Class rdf:about="#Semantics_Network">
  <rdfs:subClassOf rdf:resource="#Semantics" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Search
    -->
- <owl:Class rdf:about="#Semantics_Search">
  <rdfs:subClassOf rdf:resource="#Semantics" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Web
    -->
- <owl:Class rdf:about="#Semantics_Web">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Web_Agent
    -->
- <owl:Class rdf:about="#Semantics_Web_Agent">
  <rdfs:subClassOf rdf:resource="#Agent" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Web_Ontology
    -->
```

```
- <owl:Class rdf:about="#Semantics_Web_Ontology">
  <rdfs:subClassOf rdf:resource="#Semantics_Web" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Web_Service
  -->
- <owl:Class rdf:about="#Semantics_Web_Service">
  <rdfs:subClassOf rdf:resource="#Semantics_Web" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Semantics_Web_Social_Impact
  -->
- <owl:Class rdf:about="#Semantics_Web_Social_Impact">
  <rdfs:subClassOf rdf:resource="#Semantics_Web" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Share_Mentality_Model
  -->
- <owl:Class rdf:about="#Share_Mentality_Model">
  <rdfs:subClassOf rdf:resource="#Organisation_Psychology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Social_Cognition_Theory
  -->
- <owl:Class rdf:about="#Social_Cognition_Theory">
  <rdfs:subClassOf rdf:resource="#Semantics_Web_Social_Impact" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Social_Network
  -->
- <owl:Class rdf:about="#Social_Network">
  <rdfs:subClassOf rdf:resource="#Community" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Student_Modelling
  -->
- <owl:Class rdf:about="#Student_Modelling">
  <rdfs:subClassOf rdf:resource="#User_Modelling" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Taxonomy
  -->
- <owl:Class rdf:about="#Taxonomy">
  <rdfs:subClassOf rdf:resource="#Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Temporal_Logic
  -->
- <owl:Class rdf:about="#Temporal_Logic">
  <rdfs:subClassOf rdf:resource="#Logic" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Time_Ontology
  -->
- <owl:Class rdf:about="#Time_Ontology">
  <rdfs:subClassOf rdf:resource="#Domain_Ontology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Topic_Map
  -->
- <owl:Class rdf:about="#Topic_Map">
  <rdfs:subClassOf rdf:resource="#Knowledge_Representation_Reasoning" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Transaction_Memory
  -->
- <owl:Class rdf:about="#Transaction_Memory">
  <rdfs:subClassOf rdf:resource="#Organisation_Psychology" />
  </owl:Class>
 - <!--
 http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Upper_Ontology
  -->
```

```
- <owl:Class rdf:about="#Upper_Ontology">
  <rdfs:subClassOf rdf:resource="#Ontology" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Use_Evaluation
    -->
- <owl:Class rdf:about="#Use_Evaluation">
  <rdfs:subClassOf rdf:resource="#Evaluation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#User_Interface
    -->
- <owl:Class rdf:about="#User_Interface">
  <rdfs:subClassOf rdf:resource="#Human_Computer_Interaction" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#User_Model
    -->
- <owl:Class rdf:about="#User_Model">
  <rdfs:subClassOf rdf:resource="#User_Modelling" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#User_Model_Ontology
    -->
- <owl:Class rdf:about="#User_Model_Ontology">
  <rdfs:subClassOf rdf:resource="#User_Model" />
  <rdfs:subClassOf rdf:resource="#Ontology" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#User_Modelling
    -->
- <owl:Class rdf:about="#User_Modelling">
  <rdfs:subClassOf rdf:resource="#Application" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#User_Modelling_System
    -->
- <owl:Class rdf:about="#User_Modelling_System">
  <rdfs:subClassOf rdf:resource="#User_Modelling" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Validation
    -->
- <owl:Class rdf:about="#Validation">
  <rdfs:subClassOf rdf:resource="#Evaluation" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Video_Retrieval
    -->
- <owl:Class rdf:about="#Video_Retrieval">
  <rdfs:subClassOf rdf:resource="#Information_Retrieval" />
  <rdfs:subClassOf rdf:resource="#Multimedia" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Data_Integration
    -->
- <owl:Class rdf:about="#Web_Data_Integration">
  <rdfs:subClassOf rdf:resource="#Library_Information_Science" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Distance_Education
    -->
- <owl:Class rdf:about="#Web_Distance_Education">
  <rdfs:subClassOf rdf:resource="#Web_Education" />
    </owl:Class>
  - <!--
   http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Education
    -->
- <owl:Class rdf:about="#Web_Education">
  <rdfs:subClassOf rdf:resource="#Education" />
    </owl:Class>
  - <!--
```

```
    http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Education_Tool
    -->
- <owl:Class rdf:about="#Web_Education_Tool">
  <rdfs:subClassOf rdf:resource="#Web_Education" />
  </owl:Class>
 - <!--
    http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Electronics_Learning
    -->
- <owl:Class rdf:about="#Web_Electronics_Learning">
  <rdfs:subClassOf rdf:resource="#Learning" />
  <rdfs:subClassOf rdf:resource="#Web_Distance_Education" />
  </owl:Class>
 - <!--
    http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#Web_Intelligence
    -->
- <owl:Class rdf:about="#Web_Intelligence">
  <rdfs:subClassOf rdf:resource="#Semantics" />
  </owl:Class>
 - <!--
    http://www.comp.leeds.ac.uk/stellak/semwebonto.owl#WordNet
    -->
- <owl:Class rdf:about="#WordNet">
  <rdfs:subClassOf rdf:resource="#Taxonomy" />
  </owl:Class>
 - <!--
    http://www.w3.org/2002/07/owl#Thing
    -->
  <owl:Class rdf:about="http://www.w3.org/2002/07/owl#Thing" />
  </rdf:RDF>
```

Appendix B

Questionnaires Given to Virtual Community Members

This appendix contains the questionnaires given to VC members during the final experimental study presented in Chapter 8. B.1 is the first questionnaire sent to oldtimers. The second questionnaire answered by oldtimers used to assess the effect of the first set of notifications to oldtimers, and is presented in B.2. Newcomers' initial questionnaire can be found in B.3. Finally, B.4 shows the final questionnaire answered by all members.

It is important to mention that in questions where the names of people in the community were given, the names have been replaced with codes to comply with data protection regulations

B.1 Questionnaire 1 (Answered Only by Oldtimers)

This questionnaire is the first in a series that will allow me to proceed with the evaluation of my PhD. I would appreciate your help with this by answering the following questions related to the BSCW Virtual Community (VC) that was created for our group. Please note that providing your name is essential otherwise the results will not be possible to be analysed. For those of you who do not have the link to the VC space please see below: http://public.bscw.de/bscw/bscw.cgi

Thank you in advance for your help!

Enter your information:

Name, Surname:

Email address:

1. Please provide 5 – 10 keywords/phrases that best describe your research interests (e.g. Knowledge management, adaptation).

2. Cognitively central members are those who share the most important resources for the whole community and have the strongest influence on the knowledge sharing process in the community. Select two members from the list below who you believe are the most cognitively central for this community. Justify why. M2, M3, M5, M6, M7, M9, M11, M13

3. Select three members from the list below who you may contact for advice. M2, M3, M5, M6, M7, M9, M11, M13

4. Select three members from the list below who you may contact for information. M2, M3, M5, M6, M7, M9, M11, M13

5. Select three members from the list below who may read resources you upload. M2, M3, M5, M6, M7, M9, M11, M13

6. Select three members from the list below who may benefit from what you know. M2, M3, M5, M6, M7, M9, M11, M13

7. Select three members from the list below who may upload resources you would read. M2, M3, M5, M6, M7, M9, M11, M13

8. Select three members from the list below who may have similar research interests to you. M2, M3, M5, M6, M7, M9, M11, M13

9. Select three members from the list below who may read similar resources to you. M2, M3, M5, M6, M7, M9, M11, M13

10. Select three members from the list below who may upload similar resources to you. M2, M3, M5, M6, M7, M9, M11, M13

11. Can you please state in your own understanding, why the Personalisation & Intelligent Knowledge Management VC has been created? (many answers possible)

- ☐ To socialise
- ☐ To Share resources
- ☐ To have a resource repository online
- ☐ To keep important papers in one place
- ☐ So others in the group can see what we are reading
- ☐ I don't know
- ☐ Other, please specify

12. Have you ever downloaded/uploaded resources in the VC?

- ☐ Only downloaded
- ☐ Only uploaded
- ☐ Uploaded and Downloaded

- [ ] Not participated at all

13. Please select more than one if needed. The reason you participated to the VC by only uploading resources is: (many answers possible)

- [ ] another member asked me to upload a resource
- [ ] people from the group can benefit from what I upload
- [ ] I like sharing resources I find interesting
- [ ] others can see what I am working on
- [ ] it is a good place for storage resources I will revisit
- [ ] I want to become popular in the VC by uploading
- [ ] I have checked and not found any interesting resources to download
- [ ] I don't have time to ckeck for resources in the VC
- [ ] I didn't know there are resources available for download in the VC
- [ ] I don't know in which folder resources relevant to my interests are stored
- [ ] I am only interested in resources uploaded by specific members
- [ ] I am a new member and I don't know where I can start downloading
- [ ] Other, please specify in the next page

14. Please select more than one if needed. The reason you participated in the VC by downloading resources only is (many answers possible)

- [ ] I have found interesting resources in the VC to download
- [ ] another member asked me to download some resources
- [ ] I downloaded resources out of curiosity
- [ ] I was just browsing and downloaded resources
- [ ] accidentally I clicked at a resource's title and I downloaded it
- [ ] I usually check at the VC when I need a resource
- [ ] I was new at this VC and I was trying to see what resources were available to download
- [ ] I downloaded a resource it was uploded by a member I value his/her opinion
- [ ] I don't think others will be interested in what I am reading
- [ ] I don't really work in similar areas with any of the other members
- [ ] I am a new member and don't know what others are interested in
- [ ] I don't like sharing resources
- [ ] I don't have time to upload any resources
- [ ] Other, please specify in the next page

15. Please select all that apply. I participated by uploading and downloading resources beacuse: (many answers possible)

- ☐ another member asked me to upload a resource
- ☐ people from the group can benefit from what I upload
- ☐ I like sharing resources I find interesting
- ☐ others can see what I am working on
- ☐ it is a good place for storage resources I will revisit
- ☐ I want to become popular in the VC
- ☐ I downloaded resources I found interesting
- ☐ I downloaded resources out of curiosity
- ☐ I was just browsing and downloaded resources
- ☐ accidentally I clicked at a resource's title and downloaded it
- ☐ I usually check at the VC when I need a resource to download
- ☐ I was new at this VC and I was trying to see what resources were available to download
- ☐ I downloaded resources that was uploded by a member I value his/her opinion
- ☐ Other, please specify in the next page

16. Please select all that apply. I have not participated yet in the VC because: (many answers possible)

- ☐ I have checked and not found any interesting resources
- ☐ I don't have time to ckeck for resources in the VC
- ☐ I didn't know there are resources available for download in the VC
- ☐ I don't know in which folder resources relevant to my interests are stored
- ☐ I am only interested in resources uploaded by specific members
- ☐ I am a new member and I don't know where I can start
- ☐ I don't have time to upload any resources
- ☐ I don't think others will be interested in what I am reading
- ☐ I don't really work in similar areas with any of the other members
- ☐ I am a new member and don't know what others are interested in
- ☐ I don't like sharing resources
- ☐ Other, please specify in the next page

17. Please specify

18. Thank you for taking the time to answer this questionnaire. Please provide any further comments below or just type "No comments"

B.2 Questionnaire 2

This questionnaire is the second in a series that will allow me to proceed with the evaluation of my PhD. I would appreciate your help with this by answering the following questions related to the BSCW Virtual Community (VC) that was created for our group and also related to the emails you received earlier this month about your participation in the BSCW VC. Please note that providing your name is essential otherwise the results will not be possible to be analysed. For those of you who do not have the link to the VC space please see below: http://public.bscw.de/bscw/bscw.cgi

Thank you in advance for your help!

Enter your information:

Name, Surname:

Email address:

At the begining of the month you have received some notification messages indicating relationships you might have developed or information with regard to your membership in BSCW VC. The following questions will concern those notification messages.

1. Cognitively central members are those who share the most important resources for the whole community and have the strongest influence on the knowledge sharing process in the community. Select two members from the list below who you believe are the most cognitively central for this community. Justify why. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

2. Select three members from the list below who you may contact for advice. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

3. Select three members from the list below who you may contact for information. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

4. Select three members from the list below who may read resources you upload. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

5. Select three members from the list below who may benefit from what you know. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

6. Select three members from the list below who may upload resources you would read. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

7. Select three members from the list below who may have similar research interests to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

8. Select three members from the list below who may read similar resources to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

9. Select three members from the list below who may upload similar resources to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

10. In your oppinion the information you received through the notification messages was relevant to you?

☐ Yes

☐ No, please specify why

11. The infromation I received help me to: (many answers possible)

☐ Identify people with similar interests

☐ Identify people who I might contact for information

☐ Identify potential collaborators

☐ Identify who is reading resources I upload

☐ Identify who is reading similar resources as I do

- ☐ Identify who is uploading similar resources as I do
- ☐ Identify who are the cognitively central members of the VC
- ☐ become more active by uploading in the VC
- ☐ become more active by downloading from the VC
- ☐ Identify where resources important to me are located

12. Have you followed the links provided in the notifications?

- ☐ Yes
- ☐ No

13. Have you upload or download from the VC because of the notifications you received?

- ☐ Upload
- ☐ Download
- ☐ I have not upload nor download

14. I have not followed the links in the notifications because: (many answers possible)

- ☐ I have no time
- ☐ I am not interested
- ☐ The information provided were not relevant to me
- ☐ I haven't noticed the links provided in the message
- ☐ Other please specify

15. Can you please state why you have not upload or download after you followed the links provided in the notification message?

16. The messages I received motivate me to remain active in the virtual community:

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Neither agree nor disagree
- ☐ Agree
- ☐ Strongly Agree

17. Please provide any comments relevant to your previous answer:

18. Since I received the notification messages I feel more confident to contribute to the virtual community:

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Neither agree nor disagree
- ☐ Agree
- ☐ Strongly Agree

19. Please provide any comments relevant to your previous answer:

20. Thank you for taking the time to answer this questionnaire. Please provide any further comments below or just type "No comments"

B.3 Newcomers' Questionnaire

This questionnaire is the first in a series that will allow me to proceed with the evaluation of my PhD. I would appreciate your help with this by answering the following questions related to the BSCW Virtual Community (VC) that was created for our group. Please note that providing your name is essential otherwise the results will not be possible to be analysed. For those of you who do not have the link to the VC space please see below: http://public.bscw.de/bscw/bscw.cgi

Thank you in advance for your help!

Enter your information:

Name, Surname:

Email address:

1. Please provide 5 – 10 keywords/phrases that best describe your research interests (e.g. Knowledge management, adaptation).

2. Cognitively central members are those who share the most important resources for the whole community and have the strongest influence on the knowledge sharing process in the community. Select two members from the list below who you believe are the most cognitively central for this community. Justify why. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

3. Select three members from the list below who you may contact for advice. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

4. Select three members from the list below who you may contact for information. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

5. Select three members from the list below who may read resources you upload. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

6. Select three members from the list below who may benefit from what you know. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

7. Select three members from the list below who may upload resources you would read. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

8. Select three members from the list below who may have similar research interests to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

9. Select three members from the list below who may read similar resources to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

10. Select three members from the list below who may upload similar resources to you. M1, M2, M3, M4, M5,M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

11. Can you please state in your own understanding, why the Personalisation & Intelligent Knowledge Management VC has been created? (many answers possible)

- ☐ To socialise
- ☐ To Share resources
- ☐ To have a resource repository online
- ☐ To keep important papers in one place
- ☐ So others in the group can see what we are reading
- ☐ I don't know
- ☐ Other, please specify ______

12. Have you ever downloaded/uploaded resources in the VC?

- ☐ Only downloaded
- ☐ Only uploaded
- ☐ Uploaded and Downloaded

- [ ] Not participated at all

13. Please select more than one if needed. The reason you participated to the VC by only uploading resources is: (many answers possible)

- [ ] another member asked me to upload a resource
- [ ] people from the group can benefit from what I upload
- [ ] I like sharing resources I find interesting
- [ ] others can see what I am working on
- [ ] it is a good place for storage resources I will revisit
- [ ] I want to become popular in the VC by uploading
- [ ] I have checked and not found any interesting resources to download
- [ ] I don't have time to ckeck for resources in the VC
- [ ] I didn't know there are resources available for download in the VC
- [ ] I don't know in which folder resources relevant to my interests are stored
- [ ] I am only interested in resources uploaded by specific members
- [ ] I am a new member and I don't know where I can start downloading
- [ ] Other, please specify in the next page

14. Please select more than one if needed. The reason you participated in the VC by downloading resources only is (many answers possible)

- [ ] I have found interesting resources in the VC to download
- [ ] another member asked me to download some resources
- [ ] I downloaded resources out of curiosity
- [ ] I was just browsing and downloaded resources
- [ ] accidentally I clicked at a resource's title and I downloaded it
- [ ] I usually check at the VC when I need a resource
- [ ] I was new at this VC and I was trying to see what resources were available to download
- [ ] I downloaded a resource it was uploded by a member I value his/her opinion
- [ ] I don't think others will be interested in what I am reading
- [ ] I don't really work in similar areas with any of the other members
- [ ] I am a new member and don't know what others are interested in
- [ ] I don't like sharing resources
- [ ] I don't have time to upload any resources
- [ ] Other, please specify in the next page

15. Please select all that apply. I participated by uploading and downloading resources beacuse: (many answers possible)

- [ ] another member asked me to upload a resource
- [ ] people from the group can benefit from what I upload
- [ ] I like sharing resources I find interesting
- [ ] others can see what I am working on
- [ ] it is a good place for storage resources I will revisit
- [ ] I want to become popular in the VC
- [ ] I downloaded resources I found interesting
- [ ] I downloaded resources out of curiosity
- [ ] I was just browsing and downloaded resources
- [ ] accidentally I clicked at a resource's title and downloaded it
- [ ] I usually check at the VC when I need a resource to download
- [ ] I was new at this VC and I was trying to see what resources were available to download
- [ ] I downloaded resources that was uploded by a member I value his/her opinion
- [ ] Other, please specify in the next page

16. Please select all that apply. I have not participated yet in the VC because: (many answers possible)

- [ ] I have checked and not found any interesting resources
- [ ] I don't have time to ckeck for resources in the VC
- [ ] I didn't know there are resources available for download in the VC
- [ ] I don't know in which folder resources relevant to my interests are stored
- [ ] I am only interested in resources uploaded by specific members
- [ ] I am a new member and I don't know where I can start
- [ ] I don't have time to upload any resources
- [ ] I don't think others will be interested in what I am reading
- [ ] I don't really work in similar areas with any of the other members
- [ ] I am a new member and don't know what others are interested in
- [ ] I don't like sharing resources
- [ ] Other, please specify in the next page

17. Please specify

18. Thank you for taking the time to answer this questionnaire. Please provide any further comments below or just type "No comments"

B.4 Questionnaire

This questionnaire will allow me to proceed with the evaluation of my PhD. I would appreciate your help with this by answering the following questions related to the BSCW Virtual Community (VC) that was created for our group and also related to the emails you received earlier this month about your participation in the BSCW VC. Please note that providing your name is essential otherwise the results will not be possible to be analysed. For those of you who do not have the link to the VC space please see below: http://public.bscw.de/bscw/bscw.cgi

Thank you in advance for your help!

Enter your information:

Name, Surname:

Email address:

Earlier this month you have received some notification messages indicating relationships you might have developed or information with regard to your membership in BSCW VC. The following questions will concern those notification messages.

1. Cognitively central members are those who share the most important resources for the whole community and have the strongest influence on the knowledge sharing process in the community. Select two members from the list below who you believe are the most cognitively central for this community. Justify why. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

2. Select three members from the list below who you may contact for advice. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

3. Select three members from the list below who you may contact for information. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

4. Select three members from the list below who may read resources you upload. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

5. Select three members from the list below who may benefit from what you know. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

6. Select three members from the list below who may upload resources you would read. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

7. Select three members from the list below who may have similar research interests to you. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

8. Select three members from the list below who may read similar resources to you. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

9. Select three members from the list below who may upload similar resources to you. M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, M13, M14, M15

10. In your oppinion the information you received through the notification messages was relevant to you?

- ☐ Yes
- ☐ No, please specify why

11. The infromation I received help me to: (many answers possible)

- ☐ Identify people with similar interests
- ☐ Identify people who I might contact for information
- ☐ Identify potential collaborators
- ☐ Identify who is reading resources I upload
- ☐ Identify who is reading similar resources as I do
- ☐ Identify who is uploading similar resources as I do
- ☐ Identify who are the cognitively central members of the VC
- ☐ become more active by uploading in the VC
- ☐ become more active by downloading from the VC

☐ Identify where resources important to me are located

12. Have you followed the links provided in the notifications?

☐ Yes
☐ No

13. Have you upload or download from the VC because of the notifications you received?

☐ Upload
☐ Download
☐ I have not upload nor download

14. I have not followed the links in the notifications because: (many answers possible)

☐ I have no time
☐ I am not interested
☐ The information provided were not relevant to me
☐ I haven't noticed the links provided in the message
☐ Other please specify

15. Can you please state why you have not upload or download after you followed the links provided in the notification message?

16. The messages I received motivate me to remain active in the virtual community:

☐ Strongly Disagree
☐ Disagree
☐ Neither agree nor disagree
☐ Agree
☐ Strongly Agree

17. Please provide any comments relevant to your previous answer:

18. Since I received the notification messages I feel more confident to contribute to the virtual community:

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Neither agree nor disagree
- ☐ Agree
- ☐ Strongly Agree

19. Please provide any comments relevant to your previous answer:

20. Thank you for taking the time to answer this questionnaire. Please provide any further comments below or just type "No comments"

Appendix C

The Virtual Community Environment

Screenshots of the BSCW VC environment used in the final evaluation, (Chapter 8), are given in this section. Figure C.1 shows the environment that a member can see when he logs into the VC. The folders shown have been created by VC members, who named and added descriptions for each folder. A member can change the name and description of a folder and he can also delete a folder. The number appearing on the right of the folder name shows how many folders or resources that folder contains. The name of the person created the folder can also be seen. The names have been replaced with codes to comply with data protection regulations. For awareness purposes the date a folder last modified is also provided.

In the last column there are two awareness icons. The footprint icon (Figure C.3) shows modifications that have been done in the VC environment (e.g. members deleted or joined, folders created/deleted, description of folders changed). The glasses icon (Figure C.4) shows a reading history of VC members, who has read what resource and when. Detailed timestamp is also provided. The data in this area is what we have extracted and used to represent the relationships between members.

A part of the folder hierarchy can be found in Figure C.2. A member when uploads a resource has to provide a name of the resource, which does not have to be the title (as appears in the metadata), a description and can also add tags that describe the resource’s content. Awareness icons are available for all folders individually -contain awareness information for a specific folder- and also for the community in general.

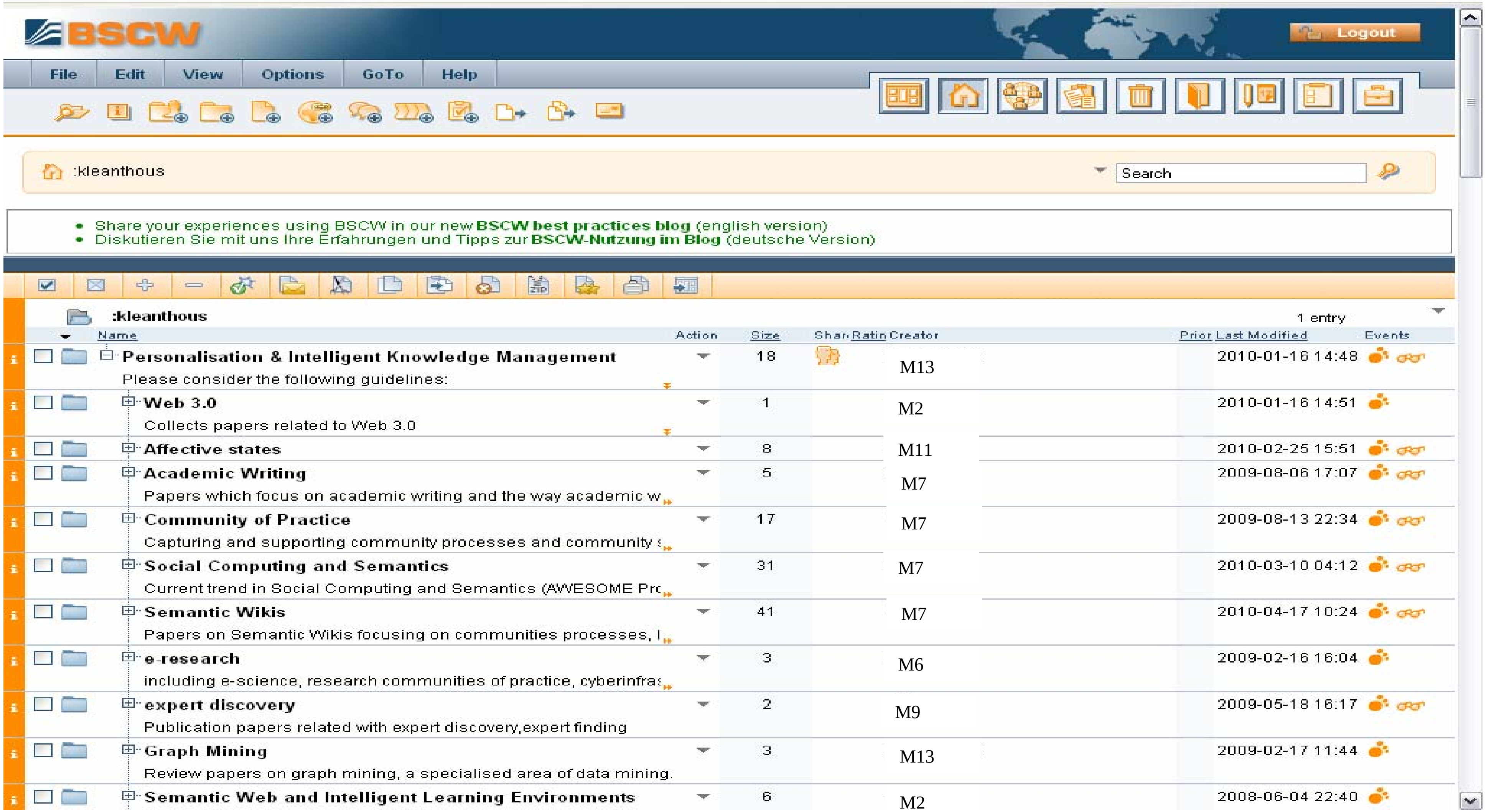


Figure C.1 What a member sees when he enters the VC

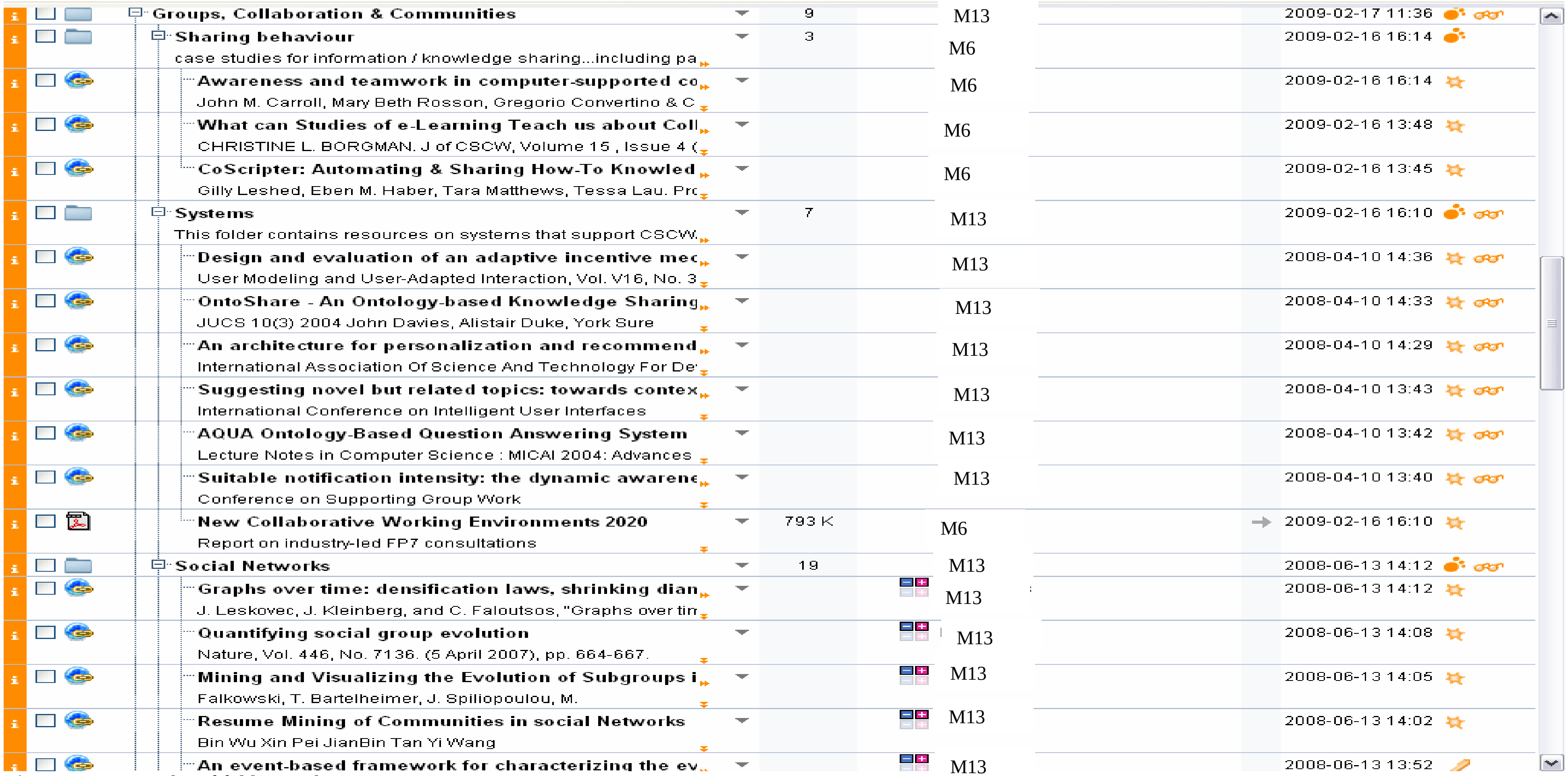


Figure C.2 Hierarchy of folders in the BSCW VC

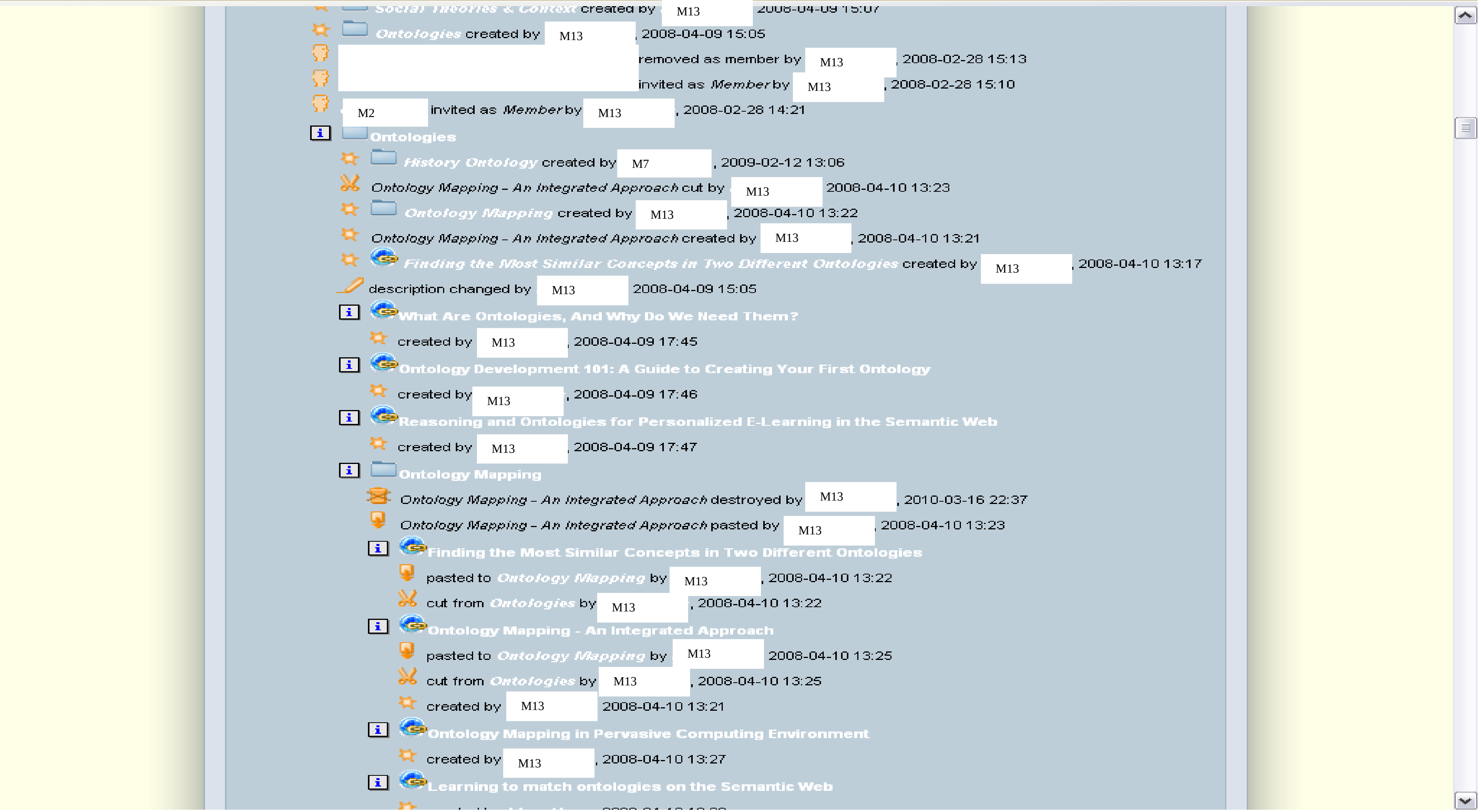


Figure C.3 Modifications history of (and inside) the VC

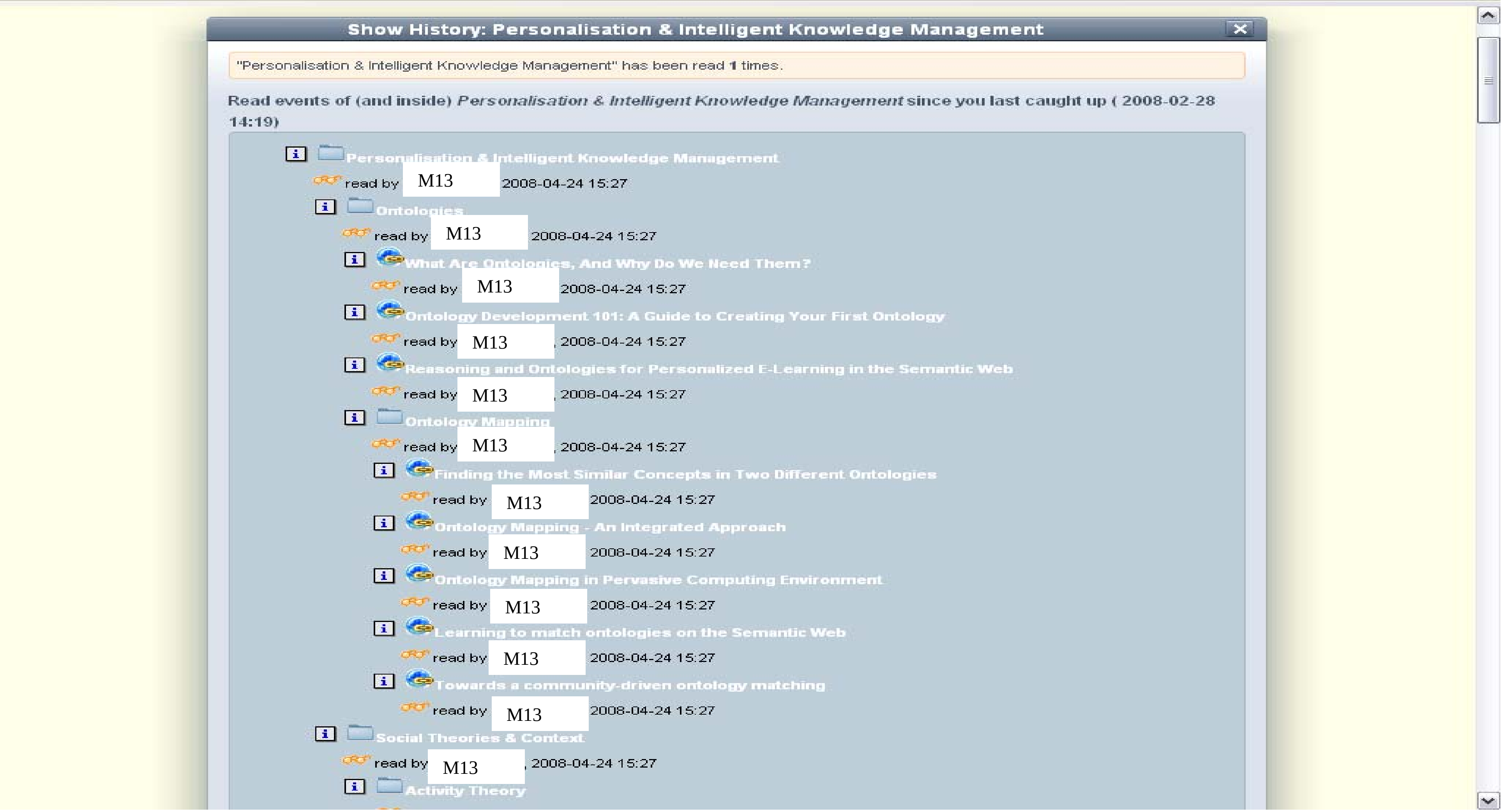


Figure C.4 Read events history of (and inside) the VC

Appendix D

Notifications Sent to Members

This section contains all the notification messages sent to VC members during the experimental study presented in Chapter 8. Newcomers received only one set of notifications, which was based on the second format of notification messages that included more personalised and targeted information (Table D.1).

Oldtimers had two rounds of notifications. The first one included general links to the VC space that directed members to explore the VC in order to find resources and people relevant to them. The second format included the more personalised information and links to specific resources, thus were more targeted than the first.

The names of members have been replaced with codes to comply with data protection regulations

Table D.1 Notifications generated during the experimental study (Chapter 8) to newcomers

Mid	Notification Type	Notification Messages to Newcomers
M1	**N3-3**	*Welcome to the VC! Based on the information you have provided, the following members M3, and M11, might have uploaded resources that can be of your interest. Use the links below to navigate through their resources.* *Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states* *http://public.bscw.de/bscw/bscw.cgi/d100633675/Conveying%20mood%20and%20emotion%20in%20instant%20messaging%20by%20using%20a%20two-dimensional%20model%20for%20affective%20states* *Influences of mood on information seeking behavior* *http://public.bscw.de/bscw/bscw.cgi/d99089575/Influences%20of%20mood%20on%20information%20seeking%20behavior* *Resources by M3* *Activity-based Adaptive Mobile Learning in Fire and Rescue Services* *http://public.bscw.de/bscw/bscw.cgi/d93638731/Activity-based%20Adaptive%20Mobile%20Learning%20in%20Fire%20and%20Rescue%20Services.pdf* *Resources by M11: http://public.bscw.de/bscw/bscw.cgi/99076334*

Mid	Notification Type	Notification Messages to Newcomers
M4	**N3-3**	*Welcome to the VC! Based on the information you have provided, the following members M5, M11 and M3 might have uploaded resources that can be of your interest. Use the links below to navigate through their resources."* *Activity-based Adaptive Mobile Learning in Fire and Rescue Services* *http://public.bscw.de/bscw/bscw.cgi/d93638731/Activity-based%20Adaptive%20Mobile%20Learning%20in%20Fire%20and%20Rescue%20Services.pdf* *Resources by M11* *http://public.bscw.de/bscw/bscw.cgi/99076334* *Resources by M5* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576*
M8	**N3-3**	*Welcome to the VC! Based on the information you have provided, the following members M2, M15 and M11 might have uploaded resources that can be of your interest. Use the links below to navigate through their resources."* *Web 3.0: Merging Semantic Web and Social Web (Workshop at Hypertext 2009)* *http://public.bscw.de/bscw/bscw.cgi/d100645544/Web%203.0%3a%20Merging%20Semantic%20Web%20and%20Social%20Web%20(Workshop%20at%20Hypertext%202009)* *Ontology Construction from Online Ontologies* *http://public.bscw.de/bscw/bscw.cgi/d94141430/Ontology%20Construction%20from%20Online%20Ontologies* *Formal Approach to Reconciliation of Individual Ontologies* *http://public.bscw.de/bscw/bscw.cgi/d94141348/Formal%20Approach%20to%20Reconciliation%20of%20Individual%20Ontologies* *Ontology Evolution: Not the Same as Schema Evolution* *https://commerce.metapress.com/content/yjnvvn0vx5bcqnfj/resource-secured/?target=fulltext.pdf&sid=0kx1jv551mkuft55cq2jtb45&sh=www.springerlink.com* *Open Provenance Model* *http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf* *Resources by M11: http://public.bscw.de/bscw/bscw.cgi/99076334*

Mid	Notification Type	Notification Messages to Newcomers
M10	**N3-3**	*Welcome to the VC! Based on the information you have provided, the following members M15 and M11 might have uploaded resources that can be of your interest. Use the links below to navigate through their resources.* *Open Provenance Model* *http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf* *Antecedent-Consequent Relationships and Cyclical Patterns between Affective States and Problem Solving Outcomes* *http://public.bscw.de/bscw/bscw.cgi/d99089160/Antecedent-Consequent%20Relationships%20and%20Cyclical%20Patterns%20between%20Affective%20States%20and%20Problem%20Solving%20Outcomes* *Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states* *http://public.bscw.de/bscw/bscw.cgi/d100633675/Conveying%20mood%20and%20emotion%20in%20instant%20messaging%20by%20using%20a%20two-dimensional%20model%20for%20affective%20states* *Resources by M11* *http://public.bscw.de/bscw/bscw.cgi/99076334*
M12	**N3-3**	*Welcome to the VC! Based on the information you have provided, the following members M5 and M15 might have uploaded resources that can be of your interest. Use the links below to navigate through their resources.* *Open Provenance Model* *http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf* *Ontologies, Applications Integration and Support to Users in Learning Objects Repositories* *http://compsci.wssu.edu/iis/swel/SWEL07/swel07-aied07-program.html* *Learning Object Context on the Semantic Web* *http://ieeexplore.ieee.org/iel5/10997/34637/01652531.pdf?isnumber=34637∏=CNF&arnumber=1652531&arSt= 669&ared= 673&arAuthor= Jovanovic, J.; Knight, C.; Gasevic, D.;Richards,G.* *Resources byM5* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576*

Mid	Notification Type	Notification Messages to Newcomers
M14	**N3-3, N2-5**	*Welcome to the VC! Based on the information you have provided, the following members M3, M13, and M5 might have uploaded resources that can be of your interest. Use the links below to navigate through their resources."* *Share your knowledge with these members by start uploading resources. They will benefit from what you share with them as you are benefiting from what they share with you* *Resources by M5* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576* *Activity-based Adaptive Mobile Learning in Fire and Rescue Services* *http://public.bscw.de/bscw/bscw.cgi/d93638731/Activity-based%20Adaptive%20Mobile%20Learning%20in%20Fire%20and%20Rescue%20Services.pdf* *Resources by M13* *http://public.bscw.de/bscw/bscw.cgi/93431763* *http://public.bscw.de/bscw/bscw.cgi/93431848* *http://public.bscw.de/bscw/bscw.cgi/93429625*
M15	**N3-3, N2-5**	*Welcome to the VC! Based on the information you have provided, the following members Vania Dimitrova, Lydia Lau, and Siraya Sitthisarn might have uploaded resources that can be of your interest. Use the links below to navigate through their resources.* *Share your knowledge with these members by start uploading resources.* *They will benefit from what you share with them as you are benefiting from what they share with you* *Resources by M2* *Knowledge Provenance: An Approach to Modeling and Maintaining The Evolution and Validity of Knowledge* *http://public.bscw.de/bscw/bscw.cgi/d94141003/Knowledge%20Provenance%3a%20An%20Approach%20to%20Modeling%20and%20Maintaining%20The%20Evolution%20and%20Validity%20of%20Knowledge* *An Architecture for Provenance Systems* *http://public.bscw.de/bscw/bscw.cgi/d94141260/An%20Architecture%20for%20Provenance%20Systems* *Issues in Building Practical Provenance Systems*

		http://public.bscw.de/bscw/bscw.cgi/d94141227/Issues%20in%20Building%20Practical%20Provenance%20Systems *Resources by M6* *Scaling System-Level Science: Scientific Exploration and IT Implications* *http://public.bscw.de/bscw/bscw.cgi/d97119876/Scaling%20System-Level%20Science%3a %20Scientific%20Exploration%20and%20IT%20Implications* *The Human Infrastructure of Cyberinfrastructure* *http://public.bscw.de/bscw/bscw.cgi/d97117571/The%20Human%20Infrastructure%20of%20Cyberinfrastructure* *Special issue on collaboration in e-Research* *http://public.bscw.de/bscw/bscw.cgi/d97117087/Special%20issue%20on%20collaboration%20in%20e-Research* *Resources by M9* *Expert Finding by Capturing Organisational Knowledge from Legacy Documents* *http://public.bscw.de/bscw/bscw.cgi/d97013937/Expert%20Finding%20by%20Capturing %20Organisational%20Knowledge%20from%20Legacy%20Documents.pdf*

Table D.2 Notifications generated to oldtimers during the experimental study (Chapter 8). Two sets of notifications have generated in two different formats

Mid	Notification Type	Notifications to Oldtimers
First Format of Notifications		
M2	**N1-1, N1-2**	*You appear to have similar interests with M6, M7, M13, M11 and M9. You may find it helpful to see the resources these members are uploading and downloading.* *You appear to upload similar resources to M5, M7, and M9. You may find it helpful to see the resources these members are uploading and downloading.* *Use the link provided below to navigate through the resources these members have uploaded and/or downloaded:* *http://public.bscw.de/bscw/bscw.cgi/92756529?op=showevents&type=ReadEvent&id=92756529_92756851* *http://public.bscw.de/bscw/bscw.cgi/92756529?client_size=1024x598*
M6	**N1-1, N1-2, N1-4**	*You appear to have similar interests with M2, M7, M13 and M9. 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:* *You appear to read similar resources with M7 and M13. 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:* *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:* *Did you know you have an undiscovered upload similarity with M5? Check out the resources these members are reading and uploading.* *Use the link provided below to navigate through the resources these members have uploded and/or downloaded:* *http://public.bscw.de/bscw/bscw.cgi/92756529?op=showevents&type=ReadEvent&id=92756529_92756851* *http://public.bscw.de/bscw/bscw.cgi/92756529?client_size=1024x598*

Mid	Notification Type	Notifications to Oldtimers
M7	**N1-2**	*You appear to have similar interests with M6, M2, M13,M9 and M11. 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:* *You appear to read similar resources with M6 and M13. 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:* *You appear to upload similar resources with M2. 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:* *Use the link provided below to navigate through the resources these members have uploded and/or downloaded:* *http://public.bscw.de/bscw/bscw.cgi/92756529?op=showevents&type=ReadEvent&id=92756529_92756851* *http://public.bscw.de/bscw/bscw.cgi/92756529?client_size=1024x598*
M9	**N1-2, N1-6**	*You appear to have similar interests with M6, M7, M13 and M2. 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:* *You appear to upload similar resources with M6, M7, M11 and M2. 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:* *M6 finds what you are uploading very interesting. You may find what this member is uploading interesting and useful. Follow the links to navigate through resources this member is uploading* *Use the link provided below to navigate through the resources these members have uploded and/or downloaded:* *http://public.bscw.de/bscw/bscw.cgi/92756529?op=showevents&type=ReadEvent&id=92756529_92756851* *http://public.bscw.de/bscw/bscw.cgi/92756529?client_size=1024x598*
M11	**N2-1**	*You appear to have similar interests with M6, M13, and M7. 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:* *Use the link provided below to navigate through the resources these members have uploded and/or downloaded:* *http://public.bscw.de/bscw/bscw.cgi/92756529?op=showevents&type=ReadEvent&id=92756529_92756851* *http://public.bscw.de/bscw/bscw.cgi/92756529?client_size=1024x598*

Mid	Notification Type	Notifications to Oldtimers
Second Format of Notifications		
N2	**N2-1, N2-2**	*Resources you have previously uploaded have been very useful to other members they have read your resources. Continue sharing with others and keep your centrality up.* *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* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576* *Resources by M11* *http://public.bscw.de/bscw/bscw.cgi/99076334*
M3	**N2-1, N2-2**	*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* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576* *Resources you have previously uploaded have been very useful to other members. They have read your resources. Continue sharing with others and keep your centrality up.*
M5	**N2-3**	*Resources you have previously uploaded have been very useful to other members. They have read your resources. Continue sharing with others and keep your centrality up.*

Mid	Notification Type	Notifications to Oldtimers
M6	**N2-1, N2-2**	*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 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* *Expert Finding by Capturing Organisational Knowledge from Legacy Documents* *http://public.bscw.de/bscw/bscw.cgi/d97013937/Expert%20Finding%20by%20Capturing%20Organisational%20Knowledge%20from%20Legacy%20Documents.pdf* *Resources by M5* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576* *Resources by M2* *http://public.bscw.de/bscw/bscw.cgi/94051630*
M7	**N2-1**	*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.*
M9	**N2-1**	*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.*
M11	**N2-1, N2-2**	*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 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* *http://public.bscw.de/bscw/bscw.cgi/94055242* *http://public.bscw.de/bscw/bscw.cgi/94046302* *http://public.bscw.de/bscw/bscw.cgi/94046576* *Resources by M2: http://public.bscw.de/bscw/bscw.cgi/94051630* *Folders that might be of your interest: http://public.bscw.de/bscw/bscw.cgi/93429584*

Appendix E

Sample Results from Questionnaires

E.1 Oldtimers' Responses

In the following tables we can find sample responses of oldtimers from both questionnaires. The Id of the users is provided here for anonymity purposes. The Member Id (MId) is given along with his replies. What has been extracted in the CM with respect to the particular question is also provided in the tables. Sets A, B, C are extracted and metrics of Precision. Recall and F1 are also provided.

Table E.1 Select two members from the list below who you believe are the most cognitively central for this community

MId	Mname	Replies Questionnaire1		Ccen as extracted in the CM		B	A	C	Precision	Recall	F1
2	M2	M7	M13	M7	M6	1	2	2	0.5	0.5	0.5
3	M3	M2	M6	M7	M6	1	2	2	0.5	0.5	0.5
5	M5	M2	M7	M7	M6	1	2	2	0.5	0.5	0.5
6	M6	M2	M13	M7	M6	0	2	2	0	0	0
7	M7	M2	M8	M7	M6	0	2	2	0	0	0
9	M9	-	-	M7	M6	-	2	2	-	-	-
11	M11	M2	M6	M7	M6	1	2	2	0.5	0.5	0.5
13	M13	M6	M5	M7	M6	1	2	2	0.5	0.5	0.5
MId	**Mname**	**Replies Questionnaire2**							**Precision**	**Recall**	**F1**
2	M2	M5	M7	M7	M9	1	2	2	0.5	0.5	0.5
3	M3	M2	M6	M7	M9	0	2	2	0	0	0
5	M5	M2	M3	M7	M9	0	2	2	0	0	0
6	M6	M2	M13	M7	M9	0	2	2	0	0	0
7	M7	M2	M8	M7	M9	0	2	2	0	0	0
9	M9	M13	M7	M7	M9	1	2	2	0.5	0.5	0.5
11	M11	M7	M13	M7	M9	1	2	2	0.5	0.5	0.5
13	M13	M6	M7	M7	M9	1	2	2	0.5	0.5	0.5
MId	**Mname**	**Replies Questionnaire3**							**Precision**	**Recall**	**F1**
2	M2	M2	M5	M2	M13	1	2	2	0.5	0.5	0.5
3	M3	M2	M6	M2	M13	1	2	2	0.5	0.5	0.5
5	M5	M2	M7	M2	M13	1	2	2	0.5	0.5	0.5
6	M6	M2	M13	M2	M13	2	2	2	1	1	1
7	M7	M2	M8	M2	M13	1	2	2	0.5	0.5	0.5
9	M9	M13	M7	M2	M13	1	2	2	0.5	0.5	0.5
11	M11	M13	M7	M2	M13	0	2	2	0	0	0
13	M13	M5	M6	M2	M13	0	2	2	0	0	0

Table E.2 Select three members from the list below who may have similar interests to you

MId	Mname	Replies Questionnaire1			InterestSim extracted in the CM			B	A	C	Precision	Recall	F1
11	M11	M2	M6	M13	M7	M6	M13	2	3	3	0.66	0.66	0.66
7	M7	M2	M8	M3	M13	M3	M6	1	3	3	0.33	0.33	0.33
6	M6	M13	M9	M2	M9	M2	M13	3	3	3	1	1	1
9	M9	-	-	-	M6	M7	M13		3	3	-	-	-
5	M5	M2	M13	M7	M13	M7	M9	2	3	3	0.66	0.66	0.66
2	M2	M5	M13	M3	M6	M13	M7	1	3	3	0.33	0.33	0.33
3	M3	M2	M9	M7	M11	M13	M7	1	3	3	0.33	0.33	0.33
13	M13	M2	M11	M6	M6	M7	M9	1	3	3	0.33	0.33	0.33
MId	**Mname**	**Replies Questionnaire2**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M6	M13	M1	M3	M4	0	3	3	0	0	0
7	M7	M2	M9	M8	M15	M9	M4	1	3	3	0.33	0.33	0.33
6	M6	M13	M9	M15	M15	M9	M4	2	3	3	0.66	0.66	0.66
9	M9	M2	M5	M13	M15	M2	M13	2	3	3	0.66	0.66	0.66
5	M5	M2	M7	M6	M15	M12	M2	1	3	3	0.33	0.33	0.33
2	M2	M6	M13	M5	M15	M5	M12	1	3	3	0.33	0.33	0.33
3	M3	M2	M1	M6	M4	M1	M11	1	3	3	0.33	0.33	0.33
13	M13	M6	M4	M2	M15	M11	M4	1	3	3	0.33	0.33	0.33
MId	**Mname**	**Replies Questionnaire3**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M6	M13	M15	M6	M9	1	3	3	0.33	0.33	0.33
7	M7	M2	M8	M9	M15	M6	M11	0	3	3	0	0	0
6	M6	M2	M5	M13	M15	M11	M7	0	3	3	0	0	0
9	M9	M13	M7	M6	M15	M6	M11	1	3	3	0.33	0.33	0.33
5	M5	M2	M7	M6	M15	M12	M6	1	3	3	0.33	0.33	0.33
2	M2	M7	M5	M13	M6	M15	M11	1	3	3	0.33	0.33	0.33
3	M3	M2	M1	M8	M13	M11	M9	0	3	3	0	0	0
13	M13	M2	M3	M4	M15	M6	M11	0	3	3	0	0	0

Table E.3 Select three members from the list below who may read similar resources to you

MId	Mname	Replies Questionnaire1			ReadSim extracted in the CM			B	A	C	Precision	Recall	F1
11	M11	M2	M6	M7	M3	M6	M13	1	3	3	0.33	0.33	0.33
7	M7	M2	M6	M3	M6	M3	M13	2	3	3	0.66	0.66	0.66
6	M6	-	-	-	M3	M7	M13	-	3	3	-	-	-
9	M9	-	-	-	M6	M2	M7	-	3	3	-	-	-
5	M5	M2	M13	M7	M11	M2	M3	1	3	3	0.33	0.33	0.33
2	M2	M6	M11	M13	M9	M5	M7	0	3	3	0.33	0.33	0.33
3	M3	M2	M6	M8	M13	M7	M6	1	3	3	0.33	0.33	0.33
13	M13	M2	M6	M11	M3	M7	M6	1	3	3	0.33	0.33	0.33
MId	**Mname**	**Replies Questionnaire2**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M6	M7	M15	M6	M7	2	3	3	0.66	0.66	0.66
7	M7	M2	M9	M8	M15	M6	M14	0	3	3	0	0	0
6	M6	M5	M2	M13	M15	M11	M7	0	3	3	0	0	0
9	M9	M6	M7	M13	M15	M6	M7	2	3	3	0.66	0.66	0.66
5	M5	M2	M13	M6	M11	M2	M3	1	3	3	0.33	0.33	0.33
2	M2	M5	M6	M13	M14	M13	M6	2	3	3	0.66	0.66	0.66
3	M3	M2	M1	M6	M15	M13	M6	1	3	3	0.33	0.33	0.33
13	M13	M2	M6	M11	M15	M3	M6	1	3	3	0.33	0.33	0.33
MId	**Mname**	**Replies Questionnaire3**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M6	M7	M15	M6	M7	2	3	3	0.66	0.66	0.66
7	M7	M2	M6	M11	M15	M11	M6	2	3	3	0.66	0.66	0.66
6	M6	M5	M13	M2	M15	M11	M7	0	3	3	0	0	0
9	M9	M13	M7	M6	M15	M11	M6	1	3	3	0.33	0.33	0.33
5	M5	M2	M7	M6	M2	M9	M11	1	3	3	0.33	0.33	0.33
2	M2	M6	M13	M7	M14	M13	M6	2	3	3	0.66	0.66	0.66
3	M3	M2	M1	M8	M2	M6	M1	2	3	3	0.66	0.66	0.66
13	M13	M2	M11	M6	M14	M2	M11	2	3	3	0.66	0.66	0.66

Table E.4 Select three members from the list below who may upload similar resources to you

MId	Mname	Replies Questionnaire1			UploadSim extracted in the CM			B	A	C	Precision	Recall	F1
11	M11	M2	M7	M13	M9	M6	M7	1	3	3	0.33	0.33	0.33
7	M7	M2	M3	M8	M11	M9	M6	0	3	3	0	0	0
6	M6	M13	M2	M9	M11	M9	M7	1	3	3	0.33	0.33	0.33
9	M9	-	-	-	M11	M6	M7	-	3	3	-	-	-
5	M5	M13	M7	M2	M11	M9	M2	1	3	3	0.33	0.33	0.33
2	M2	M7	M5	M13	M11	M9	M5	1	3	3	0.33	0.33	0.33
3	M3	M2	M7	M8	M11	M9	M5	0	3	3	0	0	0
13	M13	M2	M6	M11	M11	M9	M6	2	3	3	0.66	0.66	0.66
MId	**Mname**	**RepliesQuestionnaire2**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M7	M13	M9	M6	M7	1	3	3	0.33	0.33	0.33
7	M7	M2	M8	M9	M11	M9	M6	1	3	3	0.33	0.33	0.33
6	M6	M13	-	-	M11	M9	M7	0	3	3	-	-	-
9	M9	M6	M13	M7	M11	M6	M7	2	3	3	0.66	0.66	0.66
5	M5	M2	M6	M8	M2	M7	M9	1	3	3	0.33	0.33	0.33
2	M2	M5	M7	M13	M5	M7	M11	2	3	3	0.66	0.66	0.66
3	M3	M2	M1	M6	M11	M9	M2	1	3	3	0	0	0
13	M13	M2	M6	M11	M11	M9	M6	2	3	3	0.66	0.66	0.66
MId	**Mname**	**Replies Questionnaire3**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
11	M11	M2	M7	M13	M9	M6	M7	1	3	3	0.33	0.33	0.33
7	M7	M8	M9	M1	M9	M11	M6	1	3	3	0.33	0.33	0.33
6	M6	M2	M5	M13	M11	M9	M7	0	3	3	0	0	0
9	M9	M13	M7	M6	M11	M6	M7	2	3	3	0.66	0.66	0.66
5	M5	M2	M7	M13	M2	M11	M9	1	3	3	0.33	0.33	0.33
2	M2	M5	M13	M7	M5	M9	M13	2	3	3	0.66	0.66	0.66
3	M3	M2	M1	M8	M2	M5	M11	1	3	3	0.33	0.33	0.33
13	M13	M2	M6	M7	M9	M11	M6	1	3	3	0.33	0.33	0.33

E.2 Newcomers' Responses

In the following tables we can find sample responses of newcomers from both questionnaires. The Id of the users is provided here for anonymity purposes. The Member Id (Mid) is given along with his replies. What has been extracted in the CM with respect to the particular question is also provided in the tables. Sets A, B, C are extracted and metrics of Precision. Recall and F1 are also provided.

Table E.5 Select two members from the list below who you believe are the most cognitively central for this community

					Sets					
MId	Replies Questionnaire1		CCenM as extracted in CM		B	A	C	Precision	Recall	F1
M1	M2	M13	M7	M9	0	2	2	0	0	0
M4	M2	M13	M7	M9	0	2	2	0	0	0
M8	M2		M7	M9	-	2	2	-	-	-
M10	M6	M9	M7	M9	1	2	2	0.5	0.5	0.5
M12	M7	M2	M7	M9	1	2	2	0.5	0.5	0.5
M14	M2	M13	M7	M9	0	2	2	0	0	0
M15	M6	M2	M7	M9	0	2	2	0	0	0
	Replies Questionnaire2				B	A	C	Precision	Recall	F1
M1	M2	M6	M2	M13	1	2	2	0.5	0.5	0.5
M4	M2	M5	M2	M13	1	2	2	0.5	0.5	0.5
M8	M2	M6	M2	M13	1	2	2	0.5	0.5	0.5
M10	M2	M7	M2	M13	1	2	2	0.5	0.5	0.5
M12	M2	M8	M2	M13	1	2	2	0.5	0.5	0.5
M14	M13	M7	M2	M13	1	2	2	0.5	0.5	0.5
M15	M6	M9	M2	M13	0	2	2	0	0	0

Table E.6 Select three members from the list below who may read similar resources to you.

MId	Replies Questionnaire1			ReadSim extracted in the CM			B	A	C	Precision	Recall	F1
M1	M2	M3	M13	-	-	-	-	-	3	-	-	-
M4	M2			-	-	-	-	-	3	-	-	-
M8	M2	M7	M9	-	-	-	-	-	3	-	-	-
M10	M7	M9	M12	-	-	-	-	-	3	-	-	-
M12	M15	M3	M9	-	-	-	-	-	3	-	-	-
M14	M13	M2	M1	M15	M7	M6	0	3	3	0	0	0
M15	M9	M12	M10	M7	M6	M11	0	3	3	0	0	0
MId	**Replies Questionnaire2**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
M1	M3	M13	M9	-	-	-	-	-	3	-	-	-
M4	M3	M2	M5	-	-	-	-	-	3	-	-	-
M8	M2	M7	M3	-	-	-	-	-	3	-	-	-
M10	M6	M9	M12	-	-	-	-	-	3	-	-	-
M12	M15	M9	M10	-	-	-	-	-	3	-	-	-
M14	M1	M13	M7	M15	M13	M7	2	3	3	0.66	0.66	0.66
M15	M9	M10	M12	M9	M7	M6	1	3	3	0.33	0.33	0.33

Table E.7 Select three members from the list below who may have similar interests to you

MId	Replies Questionnaire1			InterestSim			B	A	C	Precision	Recall	F1
M1	M2	M13	M8	M3	M4	M11	0	3	3	0	0	0
M4	M2			M3	M1	M11	-	3	3	0	0	0
M8	M2	M7	M3	M15	M12	M2	1	3	3	0.33	0.33	0.33
M10	M6	M9	M12	M15	M11	M13	0	3	3	0	0	0
M12	M6	M9	M10	M5	M15	M2	0	3	3	0	0	0
M14	M13	M2	M1	M15	M3	M5	0	3	3	0	0	0
M15	M9	M12	M6	M9	M2	M5	1	3	3	0.33	0.33	0.33
MId	**Replies Questionnaire2**						**B**	**A**	**C**	**Precision**	**Recall**	**F1**
M1	M2	M3	M13	M3	M2	M13	3	3	3	1	1	1
M4	M3	M2		M5	M2	M3	2	3	3	0.66	0.66	0.66
M8	M2	M7	M1	M5	M2	M1	2	3	3	0.66	0.66	0.66
M10	M6	M9	M12	M6	M7	M14	1	3	3	0.33	0.33	0.33
M12	M15	M9	M10	M5	M10	M2	1	3	3	0.33	0.33	0.33
M14	M13	M1		M15	M11	M9	0	3	2	0	0	0
M15	M9	M12	M6	M6	M11	M9	2	3	3	0.66	0.66	0.66

Table E.8 Select three members from the list below who may upload similar resources to you

MId	Mname	Replies Questionnaire1			UploadSim			B	A	C	Precision	Recall	F1
1	M1	M2	M3	M13									
4	M4	M2											
8	M8	M2	M7	M9									
10	M10	M6	M9	M12									
12	M12	zul	M3	M9									
14	M14	M13	M2	M1									
15	M15	M9	M12	M10	M7	M6	M11	0	3	3	0	0	0
MId	**Mname**	**Replies Questionnaire2**											
1	M1	M3	M13	M9									
4	M4	M3	M2	M5									
8	M8	M2	M7	M3									
10	M10	M6	M9	M12									
12	M12	M15	M9	M10									
14	M14	M1	M13	M7									
15	M15	M9	M10	M12	M9	M11	M6	1	3	3	0.33	0.33	0.33