A Bayesian Computational Model of Social Capital in Virtual Communities

Ben Kei Daniel, Juan-Diego Zapata-Rivera, Gordon McCalla
University of Saskatchewan, Department of Computer Science, ARIES Laboratory
{Ben.Daniel, Diego.Zapata, Gordon.McCalla}@usask.ca

Abstract. The theory of social capital (SC) is frequently discussed in the social sciences and the humanities. There is a plethora of research studies, which seek to define and empirically test the idea of SC in a number of ways. This growing body of research has only supported the significance of (SC) in physical communities. While many attempts have been made to examine different forms of social capital in physical communities, its application to other types of communities remains open to research. Recent interest in computer science and information systems in studying virtual communities (VCs) and the value these communities provide to information exchange and knowledge construction makes examination of SC in these communities relevant. We begin our understanding of SC in VCs by mapping out different variables that constitute SC based on qualitative experts' knowledge of SC. We then develop an initial computational model of SC, and generate conditional probability tables (CPTs) that can be refined using real world case scenarios developed by experts in virtual communities. The Bayesian model seems to represent the situations mentioned in the paper adequately. This model provides a useful tool for understanding of SC in VCs.

Introduction

Recently a substantial body of research has sought to define, test and apply the theory of social capital (SC) in physical communities. Studies show that building social capital requires continuous positive interaction that enables people to identify common goals, achieve shared understanding and norms, build trust, and commit themselves to each other (Prusak & Cohen 2001; Putnam, 2000; World

Bank, 1999). Lesser (2002) has pointed out that positive interactions, which occur between individuals who form networks, can lead into the formation of SC.

Further, SC can also be developed and fostered when individuals believe that their actions will be appropriately reciprocated, and that each member of a community will meet expected obligations and abide by available social norms. Thus, issues around trust, shared norms and values, obligations and expectations are critical in developing social capital among members of a group (Nahapiet & Ghoshal, 1998; Putnam, 2000; Prusak & Cohen 2001).

Research mainly in social science and the humanities has repeatedly examined the notion of SC in physical communities. For example, SC is used to address the problems of lack of civic engagement (Putnam, 1993; 2000), the role of SC and civic virtue (Sirianni & Friedland, 1995); SC in community and school achievement (World Bank, 1999); SC in development (Gittell & Vidal, 1998); SC and organizational development (Prusak & Cohen, 2001) and SC as a means of addressing the issue of the "digital-divide and digital-dividend" (i.e. the social disparity gap created by lack of technological skills in society and the benefits to those who possess such skills) (Resnick, 2002).

Despite this plethora of research seeking to examine and understand the nature and value of SC in physical communities, little has been done to extend this understanding to other forms of communities. This paper contributes by extending the understanding of SC to virtual communities (VCs). This understanding also serves as basis for constructing and testing a computational model of SC. We argue that modelling SC enables systems' designers to build tools and systems that support the analysis, formation, and sustainability of SC in VCs. We use Bayesian Belief Networks (BBNs) (Pearl 1988) to create a model of SC. This model is constructed out of qualitative descriptions of the degree of influence between any two related variables. Moreover, the variables that constitute to SC are driven from experts' knowledge of SC and VCs. We have run a series of experiments to validate the initial model. The results of the experiments also help us to determine the most essential variables within SC.

The rest of the paper is organized as follows. We first examine the meaning and nature of SC. Second, we review related work on SC in physical communities. This leads us to a discussion on the relevance of the concept in VCs. Third, we describe the approach and methods we used in constructing the Bayesian model and its conditional probability tables (CPTs). Fourth, we present our experiments and the results. Finally, conclusions are drawn and future work is presented.

The Nature of Social Capital

The notion of SC is different from other forms of capital such as human capital, financial capital and physical capital. Unlike these other forms of capital, SC is a stock of active connections among people, which covers the trust, mutual understanding, and shared values and behaviours that bind people as members of human networks and communities (Cohen & Prusak, 2001). This stock of capital like human capital cannot be easily translated into dollar value. This difficulty is one of the reasons why SC is not commonly embraced and valued especially in business. A fundamental difference between SC and the other forms of capital is quality rather than quantity. Simply put, SC reveals the presence or absence of a set of relationships among people. These relationships can be productive when they are based on a common set of expectations, a set of shared values, and a sense of trust among people. The quality of this set of relationships enables people to work together and accomplish common goals.

It is also possible to notice the lack of SC in a community, through the determination of the absence of productive relationships, a lack of shared understanding, presence of distrust, which, normally results in conflicting values and goals. For example a community that has difficulties collaborating or working together exhibits a lack of SC.

Related Work

SC resides in individuals and within a community. It enables individuals to collaborate, work, and learn as a community. This resource is usually available to any member in a particular community; it can also be an individual's private good. For instance a community with individuals that are well connected to each other manifest a high public or common SC, while a single individual who is well connected to other individuals in a community might own a high private SC compared to others who are loosely connected.

The value of SC in physical communities is well founded. Previous research shows that SC allows people to resolve collective problems more easily (World Bank, 1999). That is people are normally well off if they cooperate with each other. But in a case where individuals benefit more by shirking their responsibility, hoping others will do the work for them, normally there are mechanisms—social sanctions for coping with such a breach in social norms.

Putnam (2000) has observed that SC greases the wheel that allows communities to advance smoothly. Prusak & Cohen (2001) have also suggested that when people trust each other and preserve continuous interaction, they sustain social capital. In this case, presence of trust makes every day business easy and fun. Without trust however, conducting business will become complex and difficult.

A network that constitute SC also serve as a conduit for helpful information dissipation that will contribute to achievement of personal as well as community goals. For instance, people who are well connected usually get good news first. In addition, SC can also help to preserve social norms in a community and reduce delinquent or egoistic behaviour. It is common to notice that people who are well connected in a community and have active trusting connections with others are likely to behave within the prescribed social order of their community. To illustrate this point, take for example any religious leader who will normally be constrained to behave in specific ways because of the expectation the community puts on them. This can be the case with other respected professions in communities, including teachers, professors, doctors, police, lawyers etc.

The benefits of SC have been extended to educational realms. Researchers at the World Bank (1999) use a number of statistics to support the case for SC in education. They argue that schools are more effective when parents and local communities are actively involved. Teachers are more committed and students have high-test scores. Also parents better utilize school facilities in communities with a high stock of SC and citizens take active interests in children's educational achievement (Coleman and Hoffer 1987; Braatz and Putnam, 1996; Francis et al 1998).

Further, the World Bank researchers argue that poverty in some countries can be caused by lack of connections to the formal economy including material and informational resources, the poor people have limited SC, and the little they have is primarily derived from family and neighbours. Therefore, increases in SC of the poor help them to transcend their closed networks in order to access additional resources.

We further argue that in any multicultural settings with little shared understanding, SC can help break down traditionally deeply held belief systems and hence can increase shared understanding within a group. For instance, imagine an *agent x* who stereotypically believes that *agent q* is lazy because *agent q* belongs to a family of *agents z,* and all members of the *agent z* family have a history of being lazy. Suppose *agent x* has a trustworthy friend, *agent p*, who has interacted with agent q from the *agent z* family, and *agent p* suggests to *agent x* that some members of the *agent z* family are not necessarily lazy. *Agent x* is likely to change his/her degree of belief in the laziness of all *agents* in the *agent z* family thus increasing shared understanding between *agent x*, and the *agent z* family. Though another possibility might be *agent x* starts to distrust agent p, depending on how seriously he/she believes in *agent z*'s degree of laziness.

In collaborative learning environments, SC can act as a pipeline for exchange and sharing of "tacit knowledge" (non-formal knowledge that is accumulated through personal experiences). As Prusak & Cohen (2001) have pointed out, SC can promote better knowledge sharing, when trusting relationships are established. For instance, individuals can easily share experiences and knowledge when they connect with other individuals in a community. In addition, individuals who are well connected to other individuals in a community are the ones that are likely to obtain and benefit from peer support. By the same token those who are well connected in a community are the ones who are likely to offer peer support when necessary because they might carry sense of obligations to their community.

Despite a wide range of research evidence in support of the value of SC in physical communities, there is a lack of research that examines the notion of SC in VCs. Our goal is to extend this understanding of SC to VCs. Most proponents of SC argue that SC in physical communities produces trust (Putnam, 2000; World Bank, 1999). However, in virtual communities, where most individuals barely know each other, it is difficult to establish and develop trusting relationship. This implies that the development of SC, in most cases which is dependent on trust in virtual communities is influenced by a number of variables such as individuals being aware of each other backgrounds and the goals of the community, the nature of interaction, attitudes of individuals interacting etc.

Creating computational models out of such variables has been a daunting and imprecise activity. Bayesian Belief Networks (BBNs) (Pearl 1988) have become accepted and used widely to model uncertain reasoning situations and cause–effect and other probabilistic relationships. BBNs have been used in such areas as: diagnosis of medical problems, diagnosis of malfunctioning systems, planning in uncertain domains, speech recognition, user modelling and story understanding. The causal information encoded in BBNs facilitates the analysis of action sequences, observations, consequences, and expected utility. Related variables are connected in a DAG (Directed Acyclic Graph). Conditional probabilities are attached to each variable (node) based on its direct dependencies. Using BBNs it is possible to integrate new evidence and propagate it through the model.

Although BBNs originated in the AI/CS community as an effective computational tool for manipulating joint distributions of many variables, BBNs are beginning to be seen by some philosophers and social scientists (e.g. Cartwright) as providing a complete treatment of path analysis (Sewell Wright), which has significant application in the social sciences.

Social Capital in Virtual Communities

Our discussion of SC in VCs begins with basic understanding of the concept of community and carries on to distinguish major forms of VCs. A community is a group of people who socially relate to each other to achieve some common interests or goals. These relationships often reinforce one another (rather than being one-on-one). Fundamental to this notion of community is a measure of commitment to a set of shared values, norms, and meanings, and a shared history, and identification within a particular culture.

A physical community resides in one fixed locale and its members usually often meet, talk and know each other pretty well. While a virtual community is a composite of people, the space where they interact, their goals, and the technologies that they use to communicate, collaborate, and work together to achieve their goals as a community.

There are varied forms of virtual communities, most of which are organized around temporal community models and may share certain elements in common. In this paper we distinguish two major forms of virtual communities a virtual learning community (VLC) (McCalla, 2000; Schwier, 2001), and a virtual community of practice (VCoP) (Wenger & Lave, 1991). Though similar to the notion of CoPs, our notion of VCoPs conducts most of its activities in cyberspace.

A virtual learning community is a group of people who gather in cyberspace with key intention of pursuing learning goals (Daniel, McCalla & Schwier, 2002). While all virtual communities have an element of learning in them not every community can be called a learning community. A learning community is so called when all its members have explicit goals involving learning. A virtual learning community taken as whole is greater than the sum of its parts, that is, highly skilled or knowledgeable individuals in a community is a necessary, but often not a sufficient condition for a community to be termed as “learning community.” Such individuals should join those who are less knowledgeable so that they all learn together continuously as a community.

In the corporate sector, communities are mainly organized around specific work related activities normally referred to as communities of practice (CoP). Wenger & Snyder (2000) define CoP as a group of people who are informally bound together by shared expertise and passion for a joint enterprise. Such a group is also characterized by tight-knit individuals working towards common goals and who are willing to collaborate to solve common problems, share best practices, support each other, and have a common identity.

In an ideal world, virtual communities hold enormous potential for development of SC. In fact, the spectacular rise of electronic mail, Internet services, and telecommunications offers unprecedented opportunities to access instant information and connect with individuals either within the same community or in other communities.

Paradoxically, instant access to information also implies information overload and hence a growing need to deal with the overload. There are a number of strategies individuals can deal with information overload. McCalla (2000) for instance has noted that when individuals are faced with information overload in cyberspace, they drastically constraint their interaction, forming “electronic villages”. This suggests that individuals make connections with other individuals in other “electronic villages”, forming a bonded type of SC in the terms of Putnam (2000). Bonding SC can help individuals manage and determine what information is relevant to communicate and how to present this information in useful ways to other individuals in their own “electronic villages”.

In the corporate sector, using information and communication technologies that support the infrastructure of virtual communities, various companies can now make connections and establish relationships with customers, suppliers, and other contractors fairly quickly. In fact, companies that can easily develop SC with their consumers are likely to succeed in turbulent competitive markets. For instance, they can acquire instant feedback to improve their products from current consumers and while new consumers can quickly review other consumers' feedback to make decision. Today, companies like Amazon and eBay apply similar strategies to develop and maintain relationships with their customers.

SC can also foster competition among companies, and hence, companies can constantly strive to produce better products and services in order for them to stay on top of their competitors. In fact, companies that have high SC, with their competitors (bridging--social capital) have access to relevant information about their competitors and are likely to understand the quality of products and services produced and provided by their competitors. Such an understanding enables them to produce high quality products and services.

Though there are numerous possible benefits of SC in VCs, the concept is still ill defined. We based our definition of SC the definition provided by Prusak & Cohen (2001). We assume that SC in VCs is the outcome of trusting relationships, and that a number of variables affect SC through trust. This definition of SC is the basis of our model of SC.

Social Capital and Bayesian Beliefs Networks

BBNs are useful for representing imprecise, incomplete and uncertain knowledge. However, knowledge engineering effort required to create conditional probability tables per each of the variables given its parents in the model has prevented many researchers to use them in different areas. Although, there are algorithms to learn prior and conditional probabilities from data, this is not always possible which makes it necessary to elicit these probabilities from experts. In order to overcome this issue qualitative description of conditional probabilities has been proposed. Related research in this area includes qualitative probability networks proposed by Wellman (1990) and extended by other authors such as Druzdzel and Henrion (1993).

Although in qualitative probability networks degrees of influence between each pair of variables in the network are assigned, they do not map these degrees of influence to actual CPTs. Instead, qualitative probability networks implement their own algorithms to propagate evidence using just the degrees of influence (i.e. positive or negative, strong, medium and weak). The approach presented later in the next section maps qualitative degrees of influence to CPTs so that any of all the existing algorithms for BBNs can be used. Lacave & Diez (2002) used a similar approach. This approach generates an initial network that could be further refined by using systems such as cbCPT (Zapata-Rivera 2002). In fact, the model of SC presented in this work has been refined in different ways.

SC is an imprecise concept, and it is also a multidimensional concept, incorporating different levels and units of analysis. Trust, civic engagement, and community involvement are generally seen as ways to measure social capital (Putnam, 1993; Putnam, 2000; Prusak & Cohen, 2001). Since SC is multivariate concept, some variables may be more significant than others. We use BBNs to model the interactions among the variables constituting SC. In the next section we present procedures and methods for the construction of an initial computational model of SC.

Methods and Procedures

In this study we have undertaken a naturalistic qualitative research method. In particular, we have used participant observation, which involve immersion and observation of various activities in the communities studied. We have also used unstructured interview technique to gather further data on the experiences of other participants of different virtual communities. As Hammersley & Atkinson (1983) noted that a participant observation approach requires an observer to become some kind of a member of the observed group, and establish a role within the group. More over an observer needs to maintain a critical role in the group being observed. And hence, should establish physical presence and familiarity with the social protocols of the group or community (habits, use of language, non-verbal communication). The interpretation of the data driven through this approach was transcribed by explaining the meaning of experiences of particular individuals observed through experiences of the observer.

More specifically, in order to gain deep understanding of the nature of interaction and social relationships that constitute SC in VCs, we first interviewed two experts about the nature of SC, and what are the critical variables, which constitute SC. The experts interviewed included one researcher in the area of SC in temporal communities and one other expert in both SC and VCs. We asked experts to define main variables constituting SC and to describe critical relationships among the variables. Second, one of the researchers participated and was actively involved in various learning activities of more than three virtual communities for a period of at least three years. His involvement further enables him to observe and acquire various experiences in virtual communities. An unstructured interview was also administered on fifteen participants of three different VCs. Participants were asked to describe their experiences on interaction and social relationships in VCs. We also draw secondary data based on the results of a case study of eight participants in a virtual learning community (Dykes, 2003). The main goal of the study was to explore students learning experiences in virtual learning communities.

In collaboration with the experts, an initial Bayesian structure for SC showing relationships among variables using different degrees of influence was created (see Figure 1.). Further from the results of the participants' interviews, several case scenarios were extracted and used to test the behaviour of the model created by the experts and the researchers.

Variables in the model include type of community, attitudes, interactions, shared understanding, demographic cultural awareness, professional cultural awareness, task knowledge awareness, individual capability awareness, knowledge about norms, trust, and social capital. We represent various degrees of influence by the letters S (strong), M (medium), and W (weak). And the signs + and - represent positive and negative relationships. The experts and the researchers then collaboratively work together to test the Bayesian model using different pieces of evidence from the case scenarios.

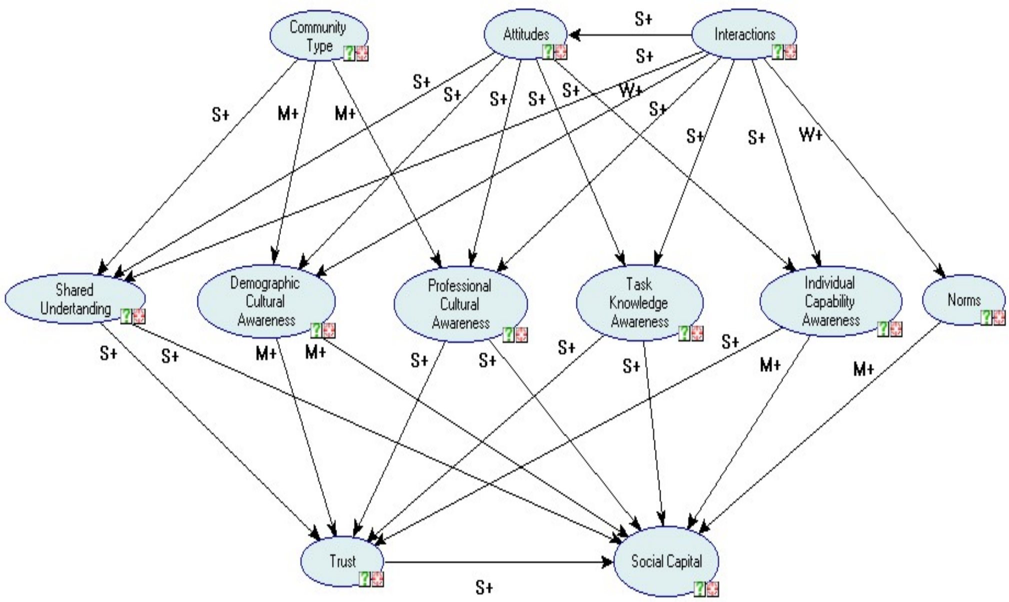

Figure 1: A Bayesian Model of Social Capital in Virtual Communities

We explain how conditional probabilities are obtained from qualitative descriptions of the influence between variables using a simple example. The Bayesian model shows that *Attitudes* (a binary variable, states: positive and negative) and *Interactions* (a binary variable, states: positive and negative) influence in a positive strong manner *TaskKnowledgeAwareness* (a binary variable, states: high and low). Depending on the kind of evidence coming from a parent (i.e. the observed parent's state) and how it affects the child, positive and negative relationships are defined. In our case, for example, a positive relationship between *Attitudes* and *TaskKnowledgeAwareness* means that in the presence of evidence of positive *Attitudes* in the community, the probability of high levels of *TaskKnowledgeAwareness* will increase. Similarly, evidence of negative *Attitudes* will increase the probability of low levels of *TaskKnowledgeAwareness.* On the other hand, a negative relationship between *Attitudes* and *TaskKnowledgeAwareness* means that there is an inverse relationship between the variables.

In order to create conditional probabilities out of qualitative descriptions of the strength of the relationship, our approach adds weights based on the number of parents and the kind and the strength of the relationship. Following with our example and assuming positive relationships, we determine the weights associated to each degree of influence. First, a threshold value is associated with each degree of influence. These threshold values correspond to the highest probability value that a child could reach under certain degree of influence of its parents (i.e. assuming that *Attitudes* and *Interactions* have positive strong relationships with *TaskKnowledgeAwareness*, evidence of positive interactions and positive attitudes will produce a conditional probability value of 0.98 of *TaskKnowledgeAwareness* = high). The corresponding weights are obtained by subtracting a base value (1 / number of parents, 0.5 in our case with two parents) from the threshold value associated to the degree of influence (i.e. threshold value for strong = 0.98) and dividing the result by the number of parents (i.e. (0.98 - 0.5) / 2 = 0.48 / 2 = 0.24). Table 1 presents threshold values and weights used in our example. The value α = 0.02 leaves some room for uncertainty when considering evidence coming from positive strong relationships.

Degree of influence	**Thresholds**	**Weights**
Strong	1-α = 1 - 0.02 = 0.98	(0.98-0.5) / 2 = 0.48 / 2 = 0.24
Medium	0.8	(0.8-0.5) / 2 =0.3 / 2 = 0.15
Weak	0.6	(0.6-0.5) / 2 =0.1 / 2 = 0.05

Table 1. Threshold values and weights with two parents

In our case, let's assume that Attitudes and Interactions have positive strong relationships with *TaskKnowledgeAwareness* and there is evidence of positive *Attitudes* and *Interactions* in a particular community. That is, weights will be added to the conditional probability table of *TaskKnowledgeAwareness* every time *Attitudes* = positive or *Interactions* = positive. For example, the conditional probability value associated with *TaskKnowledgeAwareness* given that there is evidence of *Attitudes* = positive and *Interactions* = positive is 0.98. This value is obtained by adding to the base value the weights associated to *Attitudes* and *Interactions* (0.24 each). Table 2 shows a complete conditional probability table for this example.

	Attitudes	***Positive***		***Negative***	
	Interactions	***Positive***	***Negative***	***Positive***	***Negative***
TaskKnowledge Awareness	***High***	0.98	0.74	0.74	0.5
	Low	0.02	0.26	0.26	0.5

Table 2. CPT for two parents with positive strong relationships

P(*TaskKnowledgeAwareness*= high | *Attitudes*= positive & *Interactions*= positive) = 0.5 + 0.24 + 0.24 = 0.98

P(*TaskKnowledgeAwareness*= low| *Attitudes*= positive & *Interactions*= positive) = 1 - 0.98 = 0.02

P(*TaskKnowledgeAwareness*= high| *Attitudes*= positive & *Interactions*= negative) = 0.5 + 0.24 = 0.74

P(*TaskKnowledgeAwareness*= low | *Attitudes*=positive & *Interactions*= negative) = 1 - 0.74 = 0.26

P(*TaskKnowledgeAwareness*= high | *Attitudes*= negative & *Interactions*= positive) = 0.5 + 0.24 = 0.74

P(*TaskKnowledgeAwareness*= low | *Attitudes*= negative & *Interactions*= positive) = 1 - 0.74 =0.26

P(*TaskKnowledgeAwareness*= high | *Attitudes*= negative & *Interactions*= negative) = 0.5 **

P(*TaskKnowledgeAwareness*= low |*Attitudes*= negative & *Interactions*= negative)=1-0.5=0.5

Since the expert has not provided any information about what to do when there is evidence of *Attitudes* = negative and *Interactions* = negative, no value has been added to the base value (0.5 **). However, one expects to get a high conditional probability value of *TaskKnowledgeAwareness* = negative when *Attitudes* = negative and *Interactions* = negative, a possible alternative would be to use P(*TaskKnowledgeAwareness* = positive | *Attitudes* = negative & *Interactions* = negative) = 0.02 and P(*TaskKnowledgeAwareness* = negative | *Attitudes* = negative & *Interactions* = negative) = 0.98 assuming that a positive strong relationship also occurs when *Attitudes* = negative and *Interactions* = negative. Table 3 shows this conditional probability table.

	Attitudes	Positive		Negative	
	Interactions	***Positive***	***Negative***	***Positive***	***Negative***
TaskKnowledge Awareness	***High***	0.98	0.74	0.74	0.02
	Low	0.02	0.26	0.26	0.98

Table 3. Alternative CPT for two parents with positive strong relationships

Using this approach it is possible to generate conditional probability tables (CPTs) for nodes with any number of parents and different degrees of influence assuming that the expert defines degrees of influence using one kind of evidence (i.e. evidence coming from one of the parent's states). However, when the expert defines degrees of influence for more than one of the parents' states, adding weights could result in ties, which could generate inconsistent CPTs. In such cases, one could ask the expert which parent should be used in case of ties (i.e. which variable prevails). This extra parameter can be used to solve conflicting situations.

Using this approach we have generated CPTs for all the variables in the Bayesian model of SC. It is important to mention that without this approach eliciting conditional probabilities for most of the variables especially for *Trust* (5 parents) and *Social Capital* (7 parents) would have been a difficult task. Next sections present some of the scenarios used to refine and test this initial Bayesian model of SC.

Descriptions of Case Scenarios

In this section we describe three different communities that will be used to test and to refine the initial Bayesian model of SC. Cases were randomly selected from two main types of virtual communities: virtual learning communities (VLCs) and communities of practice (CoPs).

Community A

Community A is a virtual learning community of graduate students learning fundamental concepts and philosophies of educational technology for a period of six months, from September 2001 through April 2002 at the University of Saskatchewan, Canada. They come from diverse professional training and cultural backgrounds. Participants come from Africa, Asia and North America most of them are practising teachers teaching in different domains at secondary and primary schools levels.

The researchers observed that though some individuals in the community have extensive experiences with educational technologies, they are not exposed to each other and thus are not aware of each other's talents and experiences.

This community has a strict formalised structure and all the individuals know explicit goals. There is a common goal to all individuals, that is successfully completing the class but also individuals might have different individual goals when it comes to how much they are willing to learn about the domain. We observed positive interaction in this community, but as the interactions progressed, flaming attitudes were observed. Individuals began to disagree more on the issues under discussion either because they disagreed with their colleagues' opinions, or they did not understand the issues.

Community B

Community B is what can be referred to as "a community of practice" for software engineers who gather in cyberspace to discuss issues around software development. Their general goals were sharing information, and providing knowledge and peer-support. Members in this community share common concerns and are drawn from all over the world. They can be categorized into two groups; highly experienced software developers and novices.

Despite diversity in demographic culture, community members are aware that all members come from the same professional practice (culture) and share the same goals. This was evident because of the nature of interaction, i.e. use of the same frame of reference, questions asked etc. At first, the community does not know individual's diversity in skills and demographic culture. But after a considerable period of interaction, individuals were exposed to each other and began connecting and exchanging personal information. It was also observed that some individuals who seemed to be more vocal offered a lot of help and became well known in the community. There was reasonable level of shared understanding but formal social norms for interaction were not stated. Activities in this community were informally organized.

Community C

This community consists of a group of individuals learning fundamentals of programming in Java. It is an open community whose members are geographical distributed and they have diverse demographic and professional culture but are not aware of each other's backgrounds. The researchers were able to infer as the individuals interacted, that individuals had diverse programming experiences, skills and knowledge. Participants range from novice to experienced students and from corporate experts to amateurs. It was interesting to observe that though these individuals at first did not know each other, they were willing to offer help and to support each other in learning Java.

Though there were no formal norms of interaction, individuals interacted as if they were some kinds of set down rules. This community had no defined goals, but rather individuals are mainly concerned about asking or answering questions.

Experimental Design

In order to test our initial Bayesian model of SC, each case scenario is analysed looking for evidence regarding the variables in the model. Once evidence is added to the model (i.e. observing a particular state of a variable), it is propagated to the rest of variables following the structure of the Bayesian model. This process generates a set of new marginal probabilities for the variables in the model. We are especially interested in analysing the levels of trust and SC on each of the communities. In addition, it is possible to make inferences based on the probabilities generated on any of the variables (nodes) of the model.

Results

These results are based on the nature of the cases described. It should be noted the cases themselves represent general characteristics of a community given specific types of observed interactions. It might be possible to come up with more cases and variables to replicate the experiment. In the next section, we describe the results from each of the case scenarios.

Community A

Community A is a virtual learning community (Community Type = *VLC*.) Based on the case description shared understanding is set to *low* and professional knowledge awareness is set to *doesnotexist*. Individuals in this community are familiar with their geographical diversity and so demographic cultural awareness is set to *exists*. There is well established formal set of norms set affront by the teacher (Norms = *known*.) Figure 2 shows the Bayesian model after the evidence from community A has been added (shaded nodes) and propagated through the model.

Propagating the evidence resulted into a low level of trust (P (Trust=*low*)=0.633) and a corresponding *low* probability level of SC (P (SC=*low*)=0.593). Several explanations can be provided for the drop in the levels of SC and trust. Based on the model the probabilities of interactions and attitudes being negative are *high*. This can be explained by the lack of shared understanding in the community and the lack of professional cultural awareness.

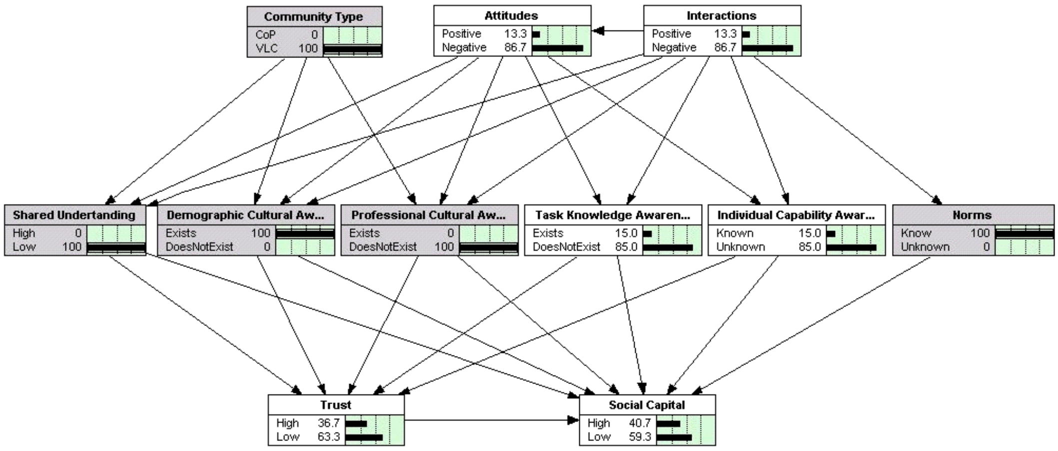

Figure 2. A Bayesian model of SC when evidence from community A has been added and propagated through the model

Drawing from the model it can be observed that the probability of task knowledge awareness being *doesnotexist* is high (0.85) and individual capability awareness is *unknown* (0.85). It be can inferred that lack of shared understanding and professional cultural awareness have further affected the levels of task knowledge awareness and individual capability awareness through negative interactions and attitudes. For instance, opinions of those individuals in the community that are more knowledgeable can be ignored, which makes them to pursue more of their personal goals other than cooperating towards the achievement of community goals. Finally, we can conclude that the drop in the levels of trust and SC are attributed to the lack of awareness and the presence of negative interaction and attitudes in this community.

Community B

Variables observed in this case include community type which has been set to community of practice (*CoP*), professional awareness culture was set to *exists*, since after interaction, we were able to determine that individuals in that community became aware of their individuals talents and skills, individual's capability awareness and task awareness were set to *exists* as well. Individuals in this community share common concerns and frame of reference, and so shared understanding has been set to *high*. Figure 3 shows the Bayesian model after the evidence from community B has been added (shaded nodes) and propagated through the model.

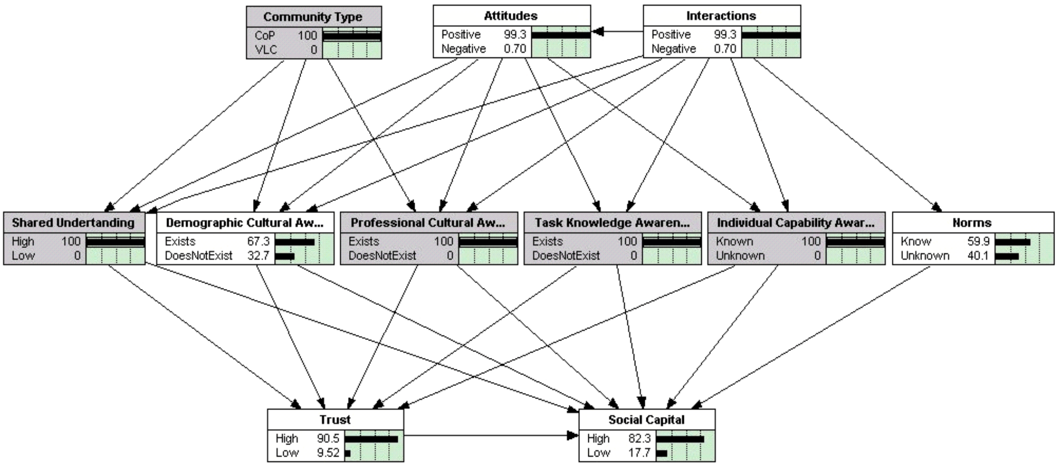

Figure 3. A Bayesian model of SC when evidence from community B has been added and propagated through the model

Propagating this set of evidence, we observe high levels of trust and SC (P(Trust=*high*)=0.905 and P(SC=*high*)=0.823) . Given the evidence, we observed that the interactions and attitudes in the model are positive which, influences in a positive way demographic cultural awareness and norms.

Further, the presence of shared understanding and the high degrees of different kinds of awareness and knowledge of norms in this community have resulted into significantly high levels of trust and SC. Based on this scenario we can see that demographic cultural awareness has little influence on the level of trust in communities of practice. Subsequently, it does not affect SC much. This simply relates to the fact that professional culture is more valued than demographic culture in CoPs. For instance, people in CoPs mainly build and maintain social relations based on common concerns other than geographical distribution. This indicates that geographical distribution matters less in this case.

Community C

Variables extracted from this case scenario include, community type (*VLC*), shared understanding *unknown,* and professional cultural awareness, demographic cultural awareness, individual's capability awareness and task awareness were set to *exists*. Figure 4 shows the Bayesian model after the evidence from community C has been added (shaded nodes) and propagated through the model.

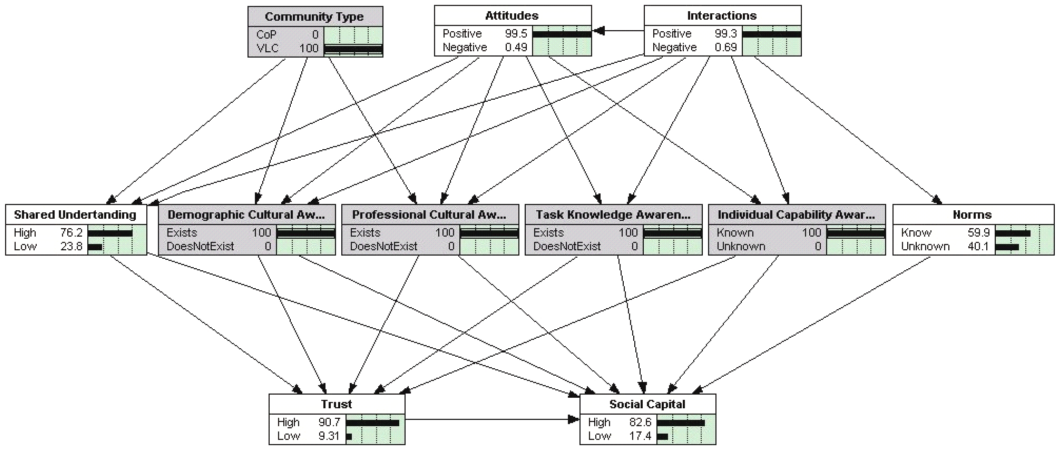

Figure 4. A Bayesian model of SC when evidence from community C has been added and propagated through the model

In community c, we observe high levels of trust and SC (P (Trust=*high*)=0.907 and P (SC=*high*)=0.826). These high levels of trust and SC can be explained by the fact that the community is focused on a particular domain. This issue can be observed in the model by looking at the high levels of shared understanding and different forms of awareness. Finally, the model indicates it is more likely that interactions and attitudes are positive in this community given the observed evidence.

Conclusion and Future Work

The theory of SC is well researched in physical communities. Though there are many benefits associated with SC in physical communities, no attempt has been made to examine this stock of social capital in VCs. The results of our experiments show that SC is a multifaceted concept, formed out of different variables. We have also shown that trust, shared understanding and different forms of awareness have significant influence on the levels of SC in VCs. Furthermore, different variables that are positively related to trust affect SC directly. These variables in turn behave differently at different levels, depending on the type of community. In addition, we have inferred that VLCs that are organized within a specific domain manifest high levels of trust and SC given the presence of shared understanding and different forms of awareness.

The Bayesian model described in this paper seems to represent the situations mentioned in the case scenarios adequately. However, this model or variations of it could be used to represent similar cases. It is further possible that this particular Bayesian model can be updated to adapt to different situations. It could be the case that several Bayesian models involving different variables and/or relationships among the nodes are needed for different kinds or groups of scenarios.

This paper offers three major contributions. First we extend the notion of SC to VCs. Second, we have developed an initial computational model of SC that can be tested and refined using real world case scenarios of VCs developed by experts. Third, we have developed and shown that it is possible to use a computational approach and to create CPTs out of a qualitative description of the strengths between variables. This can then be used to model imprecise and incomplete knowledge in a variety of domains in the social sciences and the humanities.

Though the results of this study are limited to the scenarios described, different scenarios can be developed and tested against the model, it is possible to replicate the experiments and refine the model. Our future work, however, will involve development of more formal measures of each of the variables in the model, and investigation of how to improve the levels of SC and trust using different strategies. We will also continue working closely with experts in VCs to develop more case scenarios and collect more data on various VCs and use this data to validate the model. Once a Bayesian model of SC, which can satisfactorily represents a group of virtual communities has been achieved, it could be updated dynamically as time and interaction in a community progresses. We will also explore more communities with more participants and identify and build tools, which support the creation and sustainability of SC in VCs.

References

Cohen, D. & Prusak, L. (2001). In Good Company. How Social Capital Makes Organizations Work. Massachusetts: Harvard Business School Press.

Daniel, B., McCalla, G. Schwier, R. (2002). A Process Model for Building Social Capital in Virtual Learning Communities. Proceedings of the International Conference on Computers in Education (ICCE) 2002: p. 574-p.577.

Dykes, M.E. (2003). A Case Study of Student Led Online Discussion in a Graduate Level Course. Master’s Thesis. Department of Curriculum Studies/Educational and Communications Technologies. University of Saskatchewan.

Druzdzel, M. J. & Henrion, M.: Efficient Reasoning in Qualitative Probabilistic Networks. AAAI 1993: 548-553.

Hammersley, M & Atkinson, P. (1983). Ethnography. Principles in Practice. London: Tavistock

McCalla, G. (2000). The Fragmentation of Culture, Learning, Teaching and Technology: Implications for Artificial Intelligence in Education Research. International Journal of Artificial Intelligence (11)(2), pages 177-196.

Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, San Mateo, CA.

Putnam R (2000). Bowling Alone: The Collapse and Revival of American Community, New York: Simon Schuster.

Putnam R (1995) ‘Tuning in and tuning out: the strange disappearance of social Capital in America’, political science & Politics, December, pages 664-683

Putnam R D (1993) Making Democracy Work: civic traditions in modern Italy, Princeton NJ: Princeton University Press.

Lacave, C. and Diez, F. J.. (2002). Explanation for causal Bayesian networks in Elvira. In Proceedings of the Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2002), Lyon, France.

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

Resnick, P. Beyond Bowling Together: SocioTechnical Capital. Chapter 29 in "HCI in the New Millenium", Carroll, J. M. (ed.). Addison-Wesley. 2002, pages 247 272.

Rheingold, H. (1993). The Virtual Community: Homesteading on the Virtual Frontier. New York: Addison-Wesley.

Sirianni, C. and Friedland, L. (1995) 'Social capital', Civic Practices Network, [http://www.cpn.org/sections/tools/models/social_capital.html]

Schwier, R. A. (2001). Catalysts, Emphases, and Elements of Virtual Learning Communities. Implication for Research. The Quarterly Review of Distance Education, 2(1), pages 5-18.

Wiseman, R. L, Hammer, M. R., & Nishida, H. (1989). Predictors of intercultural communication competence. International Journal of Intercultural Relations, pages 13, 349-370.

The World Bank (1999). Social capital Research Group. [http://www.worldbank.org/poverty/scapital/].

Wellman, M. P. (1990). Fundamental concepts of qualitative probabilistic networks. Artificial Intelligence, 44:257-303.

Wenger, E., & Snyder, W. (2000). Communities of Practice: the Organizational Frontier. Harvard Business Review. (January-February 2000), 139-145.

Wenger, E. (1998). Communities of Practice. Learning Meaning and Identity. Cambridge: Cambridge University Press.

Zapata-Rivera. J.D. (2002) cbCPT: Knowledge Engineering Support for CPTs in Bayesian Networks. Canadian Conference on AI 2002: 368-370