COMMUNITIES AS DYNAMIC COLLABORATIVE FILTERS: MODELING TECHNOLOGY, KNOWLEDGE, AND STRUCTURAL CHANGE IN E-MAIL BASED ONLINE COMMUNITIES

Brian S. Butler[1]

Abstract

Online communities are an increasingly important part of efforts to promote knowledge sharing among individuals within and between organizations. Towards this end, significant resources are being invested in technology infrastructures designed to increase the visibility of and access to knowledge within distributed populations. However, while there is a growing body of practice-based literature about online communities, there is little explicit theory regarding the interaction of information technology and population characteristics in the development of these communities. Based on theories of natural groups and voluntary associations, a model of online community dynamics is presented. Using computational modeling techniques this work examines how developers' efforts to alter the costs of participation and contribution might interact with the distribution of knowledge needs in the target population to determine the viability of an online community as a tool for providing access to and visibility for the knowledge resources that are potentially available.

[1] Please direct all questions or correspondence regarding this paper to Brian Butler via e-mail at bbutler@katz.pitt.edu, phone at 1-412-648-1614, or mail at 226 Mervis Hall, University of Pittsburgh, Pittsburgh, PA, 15260.

In their 1998 paper examining the role of computer aided knowledge sharing systems in the enhancement of organizational learning, Goodman and Darr (1998) report that while examining the implementation of a centrally managed system designed to facilitate the sharing of best practices, they identified several geographically distributed communities which served a similar purpose. In this work they compared and contrasted the use of 'distributed communities' with the use and adoption of a formally managed best-practices library. They noted that the informal communities served the same purposes and had in place mechanisms that were conceptually comparable to the design features of the formal computer-aided knowledge sharing system. They concluded that these self-designed communities served an important role in the transfer of best practices in the organization and that researchers and practitioners interested in knowledge management systems should consider developing computer-aided systems for focused communities rather than for organizations as a whole. This study is one of many in which research on learning and innovation has highlighted the role that communities play in the transfer of knowledge both within and between organizations. Brown and Duguid (1991) highlight the role that communities of practice can play in individual and organizational learning. In a case study of Xerox, Storck and Hill (2000) considered the strategic implications of community structure within a large corporate environment. Similarly, von Hippel (1988) noted the role of professional communities in facilitating the transfer of knowledge between organizations.

The growing awareness, among both practitioners and researchers, of the role of such communities, coupled with the emergence of communication technologies such as e-mail and the Web, has lead many organizations to make significant investments in infrastructures for supporting computer-based, or online, communities. As the general public became aware of the Internet, a great deal of attention was devoted to the potential of newsgroups and other "spontaneous" online communities as structures that supported knowledge sharing among diverse, widely distributed populations (Rheingold 1993; Baym 2000). As e-mail and the Web became more widely available, businesses began to consider how these technologies might be used to support knowledge sharing among suppliers or customers (Hagel and Armstrong 1997) as well as within their organizations. Yet as organizations have begun to work with online communities, they have discovered that achieving the success seen in the early "spontaneous" Internet communities is more difficult it had seemed. From this realization has developed substantial practice-based literature (and consulting) focused on the challenges of developing and supporting online communities.

However, while there is a growing body of research literature focused on the importance of communities in knowledge sharing and learning in organizations, and a developing body of practice-based knowledge about the mechanics of online communities, there is little explicit theory regarding the interaction of information technology infrastructures and social processes in the development of interest or knowledge-based communities. It is recognized by both practitioners and researchers that communities are emergent social structures, which developers may support and influence, but ultimately cannot deterministically direct. Yet discussion of online community infrastructures often focuses on the immediate impact of individuals' behaviors without considering the second-order effects that may occur as changes in individuals' behaviors feedback into the processes of community development. Within models that can account for both individual and community level effects it is difficult, if not impossible, to determine how infrastructure design choices will ultimately affect the usefulness of an online community as a vehicle for increasing the visibility of and access to knowledge within a distributed population.

In this paper we develop of a model of online communities that allows us to explicitly consider the consequences of infrastructure design choices at a community level. Drawing from studies of natural groups and voluntary associations we focus on individual interests, the costs of participation and contribution, and membership choice to model online communities as structures which emerge from the interaction of infrastructure characteristics, individual behavior, and features of the target population. Applying computational modeling techniques, we then examine variants of the model in order to develop propositions regarding the implications of alternative infrastructures for the emergence of online communities that provide visibility for and access to knowledge within a distributed community. The model analysis is followed by a discussion of limitations of the model and the approach taken here, directions for future research, and implications for practice.

1. A STRUCTURATION THEORY OF ONLINE COMMUNITIES

Within the proposed model an interest community consists of two components, an infrastructure and a target population. The infrastructure supports communication among the individuals who choose to participate in the online community. To model the emergence of online interest communities, we apply structuration theory (Giddens 1984) to characterize the interplay of individuals' perceptions and behavior and the structural characteristics of the online community (Figure 1). To this foundation is added the expectation that developers' infrastructure design decisions will affect the process of community emergence by altering factors that individuals take into account when making decisions regarding contribution to or continued participation in the online interest community[2].

---

[2] In this model structuration is used to characterize the relationship between individual action/perception and the social structure of an online community. It is not used as a basis for conceptualizing the link between individual behaviors, social structure, and the use of information technology, as has been done in the recent applications of structuration in the information systems research literature (Orlikowski and Robey 1991; DeSanctis and Poole 1994). Rather it is assumed that a online community developer alone has the ability to make changes to the infrastructure and do so in a way that individuals in the target population are presented only a binary participation decision (i.e. whether or not to participate). More complex models of the technology/behavior/structure relationship were omitted not because they are irrelevant, but rather because they are secondary to the overall research question of understanding the consequences of infrastructure features for online community characteristics. While it is beyond the scope of this paper to consider all aspects of infrastructure, individual behavior, and community development, the framework and computational model presented here provides a basic theoretical foundation for examining the role of infrastructure features in the development of interest communities – whether that infrastructure is a result of developer fiat or an emergent result of a communal process.

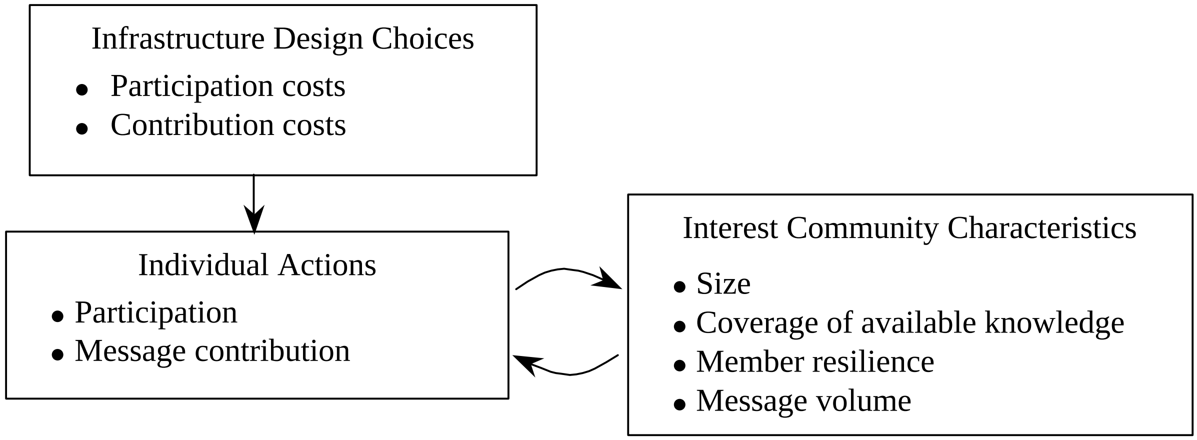

Figure 1: Structuration Model of Online Interest Communities

Individuals in the target population make two choices regarding the online community: participation and contribution. First, individuals must decide whether they are going to participate in the online community. Participation is defined as engaging with the online community infrastructure in such a way that the individual is exposed to the communication activity generated by other members of the community. This definition of participation (similar to the idea of adoption discussed by Goodman and Darr (1998)) is based on the principle that knowledge cannot be transferred unless an individual is at least a receiver of communication activity. This behavioral definition of membership was selected as being more conservative than the psychological definitions often considered in discussions of online community because it is unclear, at least in the context of knowledge management, why it would beneficial for individuals to consider themselves members of an online community and yet not be in a position to receive the messages generated by it. Thus, while perceptions of membership may be important for individual learning, individual exposure to the communication activity can be seen as a minimum threshold for achieving the types of goals that developers often have for online community development efforts. Second, individuals are in a position to choose whether they are willing to share their knowledge with other members of the community, through the contribution of messages. Again, while there are many ways that the act of message contribution can be conceptualized, ranging from single messages or complex models of online dialogue, we focus here on the basic decision by an individual to create and distribute a message to the other members of the online community. These two actions, participation and contribution, characterize how an individual in the target population can act with respect to an online community.

Because participating in an online community requires time, attention, and resources, it is a costly activity. As a result, individuals can be seen as making decisions regarding their participation based on expectations about the costs and benefits of continued involvement (Moreland and Levine 1982). The proposed model of interest community development posits that individuals develop beliefs regarding the expected benefits and costs based on their initial beliefs, the communication activity they observe, their interests, and the costs that they incur processing community message. Initial beliefs are assumed to be characteristic of the target population, determined by factors that exist outside the online community. The communication activity that an individual has seen as a result of earlier participation provides information about the knowledge available through the online community. Individuals' interests determine whether or not messages are useful to them and the costs of attending to and processing messages are the primary cost of participation. Together these factors drive each individual's assessment of the expected costs and benefits of future participation.

It is from the aggregation of individuals' contribution and participation choices that an online community's characteristics emerge. The number of participating members, coverage of the available knowledge, the structure's ability to withstand a stream of costly uninteresting activity (i.e. member resilience), and the volume of community communication are all features of an online community that emerge from individuals' perceptions and actions. At the same time each aspect of an online community is likely to be of interest to a developer interested in fostering knowledge sharing within a distributed population. The number of participating members is a rough indication of the number of people who are in a position to benefit directly from the online community – and hence benefit from investments made in the community infrastructure. If an online community is intended to serve as an access point to the knowledge of the target population it is necessary for the knowledge of the community members to reflect the knowledge available in the population as a whole. Unless strict, and costly, control of message content is exercised, an online community is subject to the threat of messages that are not of interest the participants. Whether this 'noise' takes the form of advertising, public grievances, conflict between members, or persistent low-quality contributions, most community builders hope to develop online communities in which the participants have a reasonably high tolerance of sporadic noise. Finally, developers interested in fostering knowledge sharing are interested in online communities that result in communication activity. Simply put, if individuals participate in an online community, but little or nothing is said, it is likely that the impact of that investment on individual and organization learning is minimal.

Individuals contribute messages, read messages sent by others, develop beliefs about the costs and benefits, and regularly re-evaluate their participation in the online community. Individuals affect an online community by choosing to enter it, and hence receive the communication distributed within the infrastructure; by deciding to contribute messages; or by choosing to leave, foregoing exposure to the community communication activity. From the aggregation of these actions, community characteristics, such as size, coverage of the available knowledge, member resilience, and message volume emerge – characteristics that affect individuals' future actions. It is this development process that an online community builder affects when he or she makes community infrastructure investments. By altering the costs of processing messages, the developer may affect individuals' participation decisions. Similarly, by changing the costs of creating and sending messages the community builder may affect individuals' willingness and propensity to contribute messages.

In the remainder of this paper we will describe a multi-agent computational model of online community dynamics based on the structuration framework outlined above. Analysis of this model focuses on how online community builders' choices about the cost of participation and the cost of contribution interact with the knowledge structure of the target population in the emergence of an online community. Specifically, we will consider how these factors impact the resultant community's size, coverage of the available knowledge, member resilience, and message volume.

2. A COMPUTATIONAL MODEL OF E-MAIL BASED ONLINE COMMUNITIES

The structuration framework provides insight into the link between a developer's infrastructure choices and the characteristics of the resulting online community. However, as noted above, the emergent nature of those characteristics, arising from the feedback loop involving individual actions and community characteristics, makes it difficult to specify a priori how infrastructure decisions are likely to affect community-level outcomes. To develop specific, testable propositions based on the conceptual framework presented above a computational model was created. Computational models are valuable theory development tools when the focus is on aspects of emergent phenomena and when there are feedback loops or other potentially non-linear aspects in the processes of interest (Burton and Obel 1995). The analysis of a computational model allows for the systematic investigation of the proposed relationships, a step that can be used to guide further empirical and theoretical studies. In this work, computational modeling provides a bridge between an emergent process model of community development and a variance model in which infrastructure choices are related to online community characteristics. This section describes a computational model that instantiates the proposed theory of online interest community development.

The online community model has two components: an infrastructure and a collection of individual agents who comprise the target population. The infrastructure determines how individuals can communicate with each other, the per message cost incurred by an individual when processing a message, and, implicitly, the cost of contributing a message. The individuals are each characterized in terms of their initial perceptions of the online community, their knowledge/interests, and the processes by which they choose to participate in and contribute to the online community. An infrastructure and a population of individuals are modeled as a system from which the emergent characteristics of the community (size, knowledge coverage, member resilience, and message volume) emerge.

The theoretical framework outlined above does not specify a particular infrastructure; rather, it posits that the infrastructure influences the costs incurred by individuals participating or contributing to the online community. However, within the computational model it is necessary to specify not just the costs, but also how the particular infrastructure handles the distribution of messages to individuals who have elected to participate in the community. In the baseline computational model the infrastructure organizes community activity in episodes consisting of three stages: participant identification, message collection, and message distribution (Figure 2).

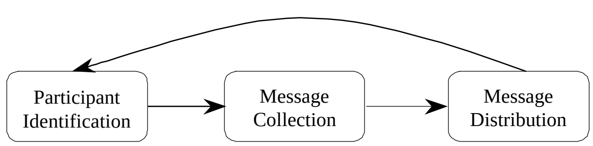

Figure 2: Infrastructure Processes

In the first stage, individuals decide whether they are going to participate in the online community. In the message collection stage individuals can choose to contribute messages to the community. Within the computational model, each message is limited to a single topic and is only distinguishable from others by that topic. Message topics are designated by points on a circle with a circumference of 1 that represents the target population's knowledge space. After all individuals have decided whether to contribute a message, the infrastructure shifts to message distribution in which all new messages are sent to all participants in the online community. The three-stage infrastructure process is performed as a series of online community episodes.

Individuals in the target population are characterized in terms of their knowledge and interests, their beliefs about the online community, and the process by which they choose to participate in and contribute to the online community. An individual's knowledge and interests determine the message topics that he or she is capable of contributing to the online community. Knowledge and interests also describe the message topics that provide some benefit for the individual -- as opposed to messages that provide no benefit for the individual (i.e. noise). In the computational model, an individual's knowledge and interests are described by a continuous interval ($[I_L, I_H]_i$) within the target population's knowledge space. Interesting messages each provide an average benefit (b) that in the model is normalized to 1 for all individuals. Knowledge and interests specified in this way are personal, describing each individual's, and not a community's, ability to contribute and derive benefit from messages on a variety of topics.

Individuals' assessment of the expected net benefit of future membership underlies their willingness to participate in the online community (Moreland and Levine 1982). Potential participants, who have not yet been exposed to community messages, make their decision to enter based on expectations of the costs and benefits, as has been found in traditional groups (Brinthaupt, Moreland et al. 1991). Likewise, individuals who previously participated also choose to participate based on the expected net benefit, an evaluation that combines their initial expectation with the information they have derived from their exposure to community messages. The computational model characterizes an individual's beliefs about the future of the online community in terms of their expectations about the content (C) and volume (V) of future message activity. Future message content beliefs are described in terms of the expected overlap between the set of topics discussed and the individual's interests. These content expectations (C_{it}) are represented by values between 0 and 1 that indicate individuals *i*'s beliefs during episode *t* regarding the probability that a message will be beneficial for them. Modeling content expectations as abstracted from information about particular topics avoids restrictive assumptions about message topic and individual interest distributions. It is also consistent with prior studies of social interaction that suggest individuals maintain generalized beliefs about interactions, not detailed histories of past social activity (Wasserman & Faust 1994, p. 57). Volume expectations (V_{it}) are represented by positive values indicating individual *i*'s beliefs about the likely message volume in episode *t*. These beliefs, the perceived likelihood of a message having interesting content (C_{it}) and the expected volume of communication in each episode (V_{it}), are combined with information about the average processing costs (c) in each individual's assessment of the expected net benefits (E_{it}) of future membership.

$$E_{it} = (1-c)C_{it}V_{it} - c(1-C_{it})V_{it} \quad (1a)$$

$$= (C_{it} - c)V_{it} \quad (1b)$$

The first term of equation 1a determines the expected net benefit of interesting messages; the second term is the expected costs due to uninteresting messages. In both terms, the relative message processing cost (c) is a factor that describes the average cost of attending to and processing a community message relative to the average benefit received when the message is beneficial.

In the computational model individuals evaluate the opportunity to participate in the online community by comparing the benefit expected from the community messages and the costs of processing those messages. If the expected net benefit is positive, the individual chooses to participate; otherwise he or she chooses to forego participation. This is the first step in the process by which individuals engage the infrastructure upon choosing to participate in the online community (Figure 3).

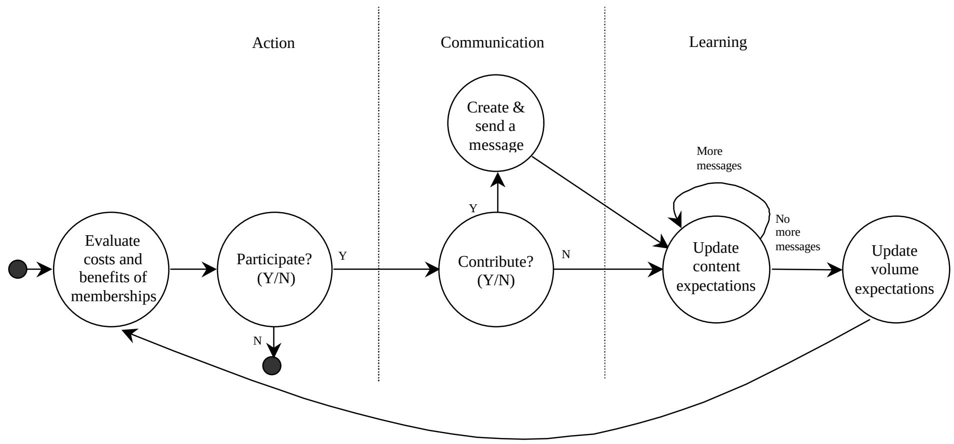

Figure 3: Individual Processes

This three stage cyclic model of communication, learning, and action is based on those proposed by Turner (1988) and Carley (1991) in their studies of social structure development, and is similar to the model of group membership and socialization proposed by Moreland and Levine (1982). Individuals, based on their expectations of costs and benefits, decide whether or not to participate in the online community (action). While participating they engage in two activities, contribution (communication) and processing of messages (learning).

Individuals contribute to the online community by constructing messages that are sent to other participants via the infrastructure. Following prior work on discussion groups (Skvoretz 1988) and information sharing in decision-making teams (Wittenbaum and Stasser 1996), each individual's message contribution behavior is modeled as the result of an independent stochastic process. An individual contributes, at most, one message per episode with a given probability (p_i), that varies among individuals but does not change over time. Upon choosing to contribute, an individual creates a message by randomly selecting a topic from his knowledge and interest range. While it does not represent all aspects of group communication, this basic model captures a fundamental link between individuals' knowledge and interests, their decisions, and the content of community messages.

As participants are exposed to messages, their beliefs about the community change. When individuals process messages they receive direct benefits, in the form of topical information or social interaction. At the same time, messages also provide information about what the volume and content of future messages might be. Based on this information, individuals can adjust their beliefs about the costs and benefits of participating in the online community. This learning behavior involves updating the individual's message content and volume expectations. Content expectation updating is characterized as a reinforcement process (Hunter, Danes et al. 1984). Changes in an individual's content expectations (ΔC_{it}) are the result of combining prior beliefs about the probability of a message being interesting (C_{it-1}) and whether the new message's topic is of interest to the individual[3]. This change is calculated as follows:

$$\Delta C_{it} = \begin{cases} wC_{it-1}(1-C_{it-1}) & \text{if the message topic is within the individual's interest range} \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

When multiple messages are received in a single episode, the change in an individual's content expectations is determined separately for each message. After all messages for an episode have been processed, volume expectations (V_{it}) are updated based on the mean message volume for all episodes the individual has observed.

Using this modeling framework, an online community is represented as a social system consisting of an infrastructure and a target population of individuals, components that are modeled as distinct, but inter-linked, entities (Figure 4).

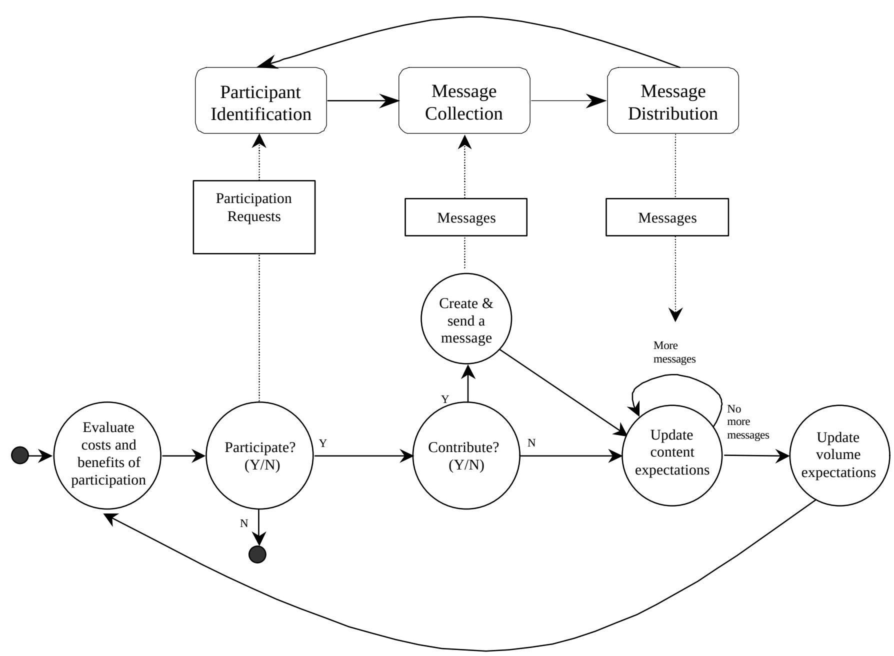

Figure 4: Model Online Community Processes

Individuals engage in evaluation, communication, and learning (Figure 3). The infrastructure engages in the activities of participant identification, message collection, and

---

[3] The process also involves assigning a weight to the new information about community message content (w). In the current model this value is set at 0.02 for all individuals. This value was identified as part of the model calibration process as being appropriate for modeling the processes of e-mail based Internet communities.

message distribution (Figure 2). Within the computational model, the infrastructure organizes activity in staged episodes. All individuals in the target population are given the opportunity to perform an action before any one moves to the next stage. The individuals take action in a fixed order and infrastructure actions are not intermingled with individuals' actions within a given stage. All individuals decide about participation before the infrastructure begins accepting new messages. New messages are distributed to all participants before any single individual begins updating his or her expectations.

3. ANALYSIS

The analysis presented here focuses on how infrastructure design decisions interact with characteristics of the target population to affect emergent characteristics of an online community. A sample of 96 e-mail based Internet communities, also known as listservs, was used to provide ranges and anchor points for parameter selection (For more information about the calibration sample and process see Butler 1999).

In the computational model several individual parameters can be manipulated to reflect the impact of infrastructure features on contribution and participation costs: probability of contribution (p_i) and relative message processing cost (c). As message contribution costs go up (down) the likelihood of any particular individual contributing a message (p_i) decreases (increases). Infrastructure features that lower average message processing costs or raise the average message benefit from an interesting message would both have the effect of reducing the relative message processing cost (c) faced by individuals. In the analysis presented here contribution probability is manipulated by independently varying the contributor/non-contributor ratio within the target population (C/NC Ratio) and the probability that a participating contributor will be willing to contribute a message in a given episode (CProb). The structure was chosen to reflect the contribution structure often seen in online communities, and to support analysis of the impact of developer efforts to increase the number of contributors separate from those aimed at increasing the volume or frequency of messages generated by individuals who are already willing to contribute to the online community. C/NC ratio and CProb for a modeled online community were determined by randomly selecting values between 0 - 0.75 and 0 – 0.15 the ranges of values observed in the calibration sample of e-mail based Internet communities. Relative message processing cost was systematically varied to represent low (0.33), medium (0.66), and high (0.99) cost infrastructures where the low cost value (0.33) was based on the estimate derived from the calibration sample.

In a modeled online community individuals in the target population can differ with respect to their initial content and volume expectations (C_{i0} and V_{i0}) and their knowledge and interests ($[I_L,I_H]$). The overall magnitude of the initial content and volume expectations reflects the degree to which individuals in the target population are favorable to the idea of an online community. A lower (higher) mean level of initial content expectation in the population as a whole indicates that individuals are, in general, less (more) convinced that the online community will ultimately give them access or exposure to knowledge that they value. Whether they came from interaction with the developer, word-of-mouth, or prior experience with other online communities, these initial expectations are formed prior to any direct exposure to the online community. In the analysis presented below, target populations are constructed by randomly selecting individuals' initial content expectations from a uniform distribution between 0.75 and 1, suggesting populations of individuals who initially believe that at least ¾ of all messages in the online community will be of interest to them. This reflects prior studies of individuals' expectations regarding traditional groups and clubs that found that individuals tend to be optimistic about the expected benefits of participation (Brinthaupt, Moreland et al. 1991). To reflect prior experience, volume expectations in the modeled target populations were set based on the mean message rate for the calibration sample, a set of e-mail based, Internet communities.

Target populations are also characterized in terms of the overall distribution and structure of an individual's knowledge and interests within the population's knowledge space. Individuals differ in terms of the type of knowledge and interests they have (placement) and the proportion of the populations' knowledge that they are aware of and interested in (range). In the analysis that follows, target populations are modeled as collections of individuals who have systematically varied ranges of knowledge and interests that are uniformly distributed in the population's knowledge space. Populations in which individuals have focused, mid-range, or broad knowledge and interest ranges are modeled by randomly selecting the length of individuals' knowledge and interest interval from between 0 – 0.33, 0.33 – 0.66, and 0.66 – 1.00 respectively. After an individual's knowledge and interest range is set, its placement was determined by randomly choosing an anchor point in the population's knowledge space.

The current work considers the following emergent community outcomes: size, knowledge coverage, member resilience, and message volume. Size is assessed by counting the number of individuals who would still choose to participate in the online community after the final cycle of the simulation. Knowledge coverage, or the degree to which the knowledge and interests of the population as a whole are reflected in the knowledge and interest of the members of the online community, is assessed by sampling the knowledge space at 20 regular intervals and determining if there is at least one member of the online community whose knowledge/interest range contains that point (i.e. they know about and are interested in that topic). If there is at least one community participant who is interested in and knowledgeable about the topic, then the raw knowledge coverage score is increased by 1. A higher knowledge coverage score (reported as a percentage) indicates that a greater portion of the knowledge/interests present in the population is represented in the online community. Member resilience, or members' willingness to continue to participate in the online community even in the presence of unequivocally useless message activity, is assessed by calculating the minimum and mean values for individuals' content expectations. The minimum indicates the lowest expectations that are present in the online community, and provides an indicator of the degree to which the least satisfied participant's beliefs would need to change before he or she would stop participating. The mean indicates the central tendency of all participants and hence provides a measure of the magnitude of belief change that would needed to result a significant proportion of the online community's participants ending their involvement. The message volume, or overall level of communication activity in the online community, is assessed in terms of the total number of messages during the modeled period.

To characterize the predictions of the model with respect to the relationship between the infrastructure choices, target population characteristics, and the community outcomes the computational model was used to simulate a sample of online communities. The infrastructure parameters and population characteristics were systematically varied (Table 1).

Construct	Model Parameter	Values
Infrastructure Parameters		
Relative Message Processing Cost	c	Low (0.33 – Estimated from calibration dataset) Medium (0.66) High (0.99)
Contributor/Non Contributor Ratio	C/NCRatio	0 – 75%
Contribution Probability	CProb	0-0.15 / per episode
Target Population Characteristics		
Mean Interest/Knowledge Range	KIRange	Low (0.15) Medium (0.50) High (0.85)

Table 1: Manipulated Infrastructure Parameters and Population Characteristics

For each combination of relative message processing cost and population interest/knowledge range, 500 online communities were modeled. Each community was modeled in a target population of 100 individuals with the assumption that all members of the target population were aware of the online community from the start. The community outcomes were assessed after the modeled community was cycled through 1000 episodes. The resulting data set from 4500 modeled online communities was then analyzed with MANOVA analyses in order to characterize the relationship between the infrastructure parameters, target population characteristics, and the emergent community characteristics.

4. RESULTS & DISCUSSION

Because the data is from an empirically grounded computational model and not an empirical data collection effort, the results presented here should be seen not as conclusions but as conjectures. The MANOVA analysis relating the infrastructure parameters (relative message processing cost, contributor/non-contributor ratio, and contribution probability) and the target population characteristics (knowledge/interest range) to the emergent features of the community (size, knowledge coverage, member resilience, and message volume) indicates that, except for a few of the higher order (3-way) interactions each of the factors is significantly related to the outcomes. It is the relationships, their direction, and relative magnitude that provide the basis for deriving conjectures from the computational model.

4.1 Infrastructure Parameters

The main effects of relative processing costs (c) suggest that online communities based on infrastructures that enable lower per message participation costs will have more members, who are (at the minimum and on average) less positive about the community (i.e. less resilient). In addition these communities will see a greater volume of activity and have members who provide more complete coverage of the knowledge and interests present in the target population.

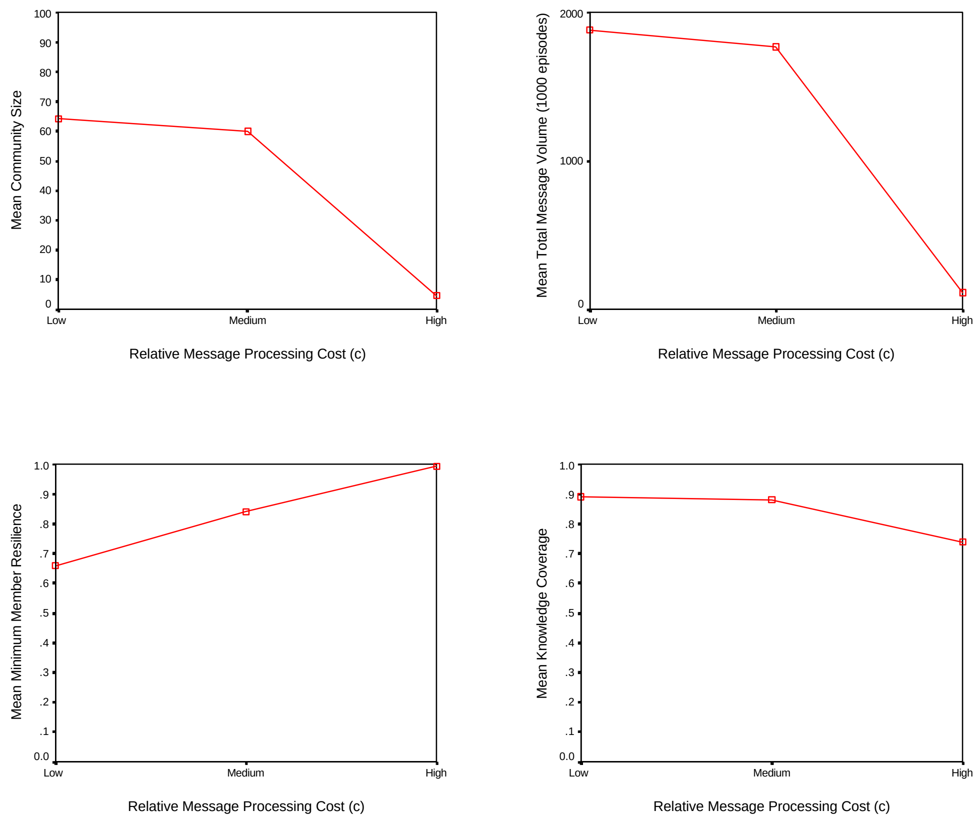

Figure 5: Effects of Relative Message Processing Cost (c)

While there are statistically significant changes associated with each level of message processing cost, there is a much greater change in community size (and activity) when moving from the high cost infrastructure to the medium range infrastructure.

There are also changes in emergent community features associated with different levels of contribution. These analysis results suggest that, as would be expected, infrastructures that increase the proportions of the population that is willing to contribute messages and/or the frequency with which contributors provide messages will see a greater message volume and have more resilient members. However, these communities with their higher levels of contribution will also tend to be smaller and have less complete coverage of the population's knowledge space. Thus, while reducing costs of participation results in larger, more active, less resilient communities, lowering the costs of contribution is predicted to have a different effect, increasing the level of activity while decreasing the size and increasing member resilience.

In addition to the main effect associated with the infrastructure parameters, there are also significant interactions between them. For example, reduction in community size, associated with increased contribution levels, was significantly lower in cases with high participation costs – a consequence of participation costs' overwhelming effects (Figure 6).

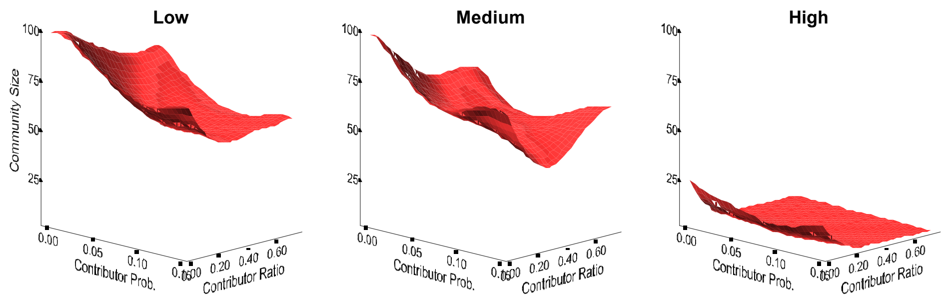

Figure 6: Interaction of Participation and Contribution Effects on Community Size

Similar effects are seen with member resilience, message volume, and knowledge coverage. When participation cost is high relative to the benefit received, the model suggests that an online community developer has one primary focus – to reduce the costs of participation (or increase the benefits). In this situation, adapting the infrastructure to encourage wider or greater levels of contributions is predicted to have little or no effect. However, as the relative cost of message processing decreases, the model suggests that an online community developer must shift from the singular goal of participation cost reduction to consider the more complex challenge of balancing contribution and participation behavior to achieve the desired levels of message activity, size, member resilience, and knowledge coverage.

4.2 Target Population Characteristics

The problem faced by online community developers is further complicated when characteristics of the target populations' interests and knowledge are considered. In the analysis considered here, communities developing in target populations comprised of individuals with relatively narrow knowledge and interest ranges were compared with those operating in populations of individuals with mid and wide knowledge and interest ranges. The computational model predicts that online communities operating in the narrow populations will tend to be smaller, have less message activity, have less resilient members, and provide less coverage of the overall knowledge space.

Combining the model results regarding the infrastructure parameters and target population characteristics begins to suggest the nature of the complex problem space faced by online community developers. Considering community size (Figure 7) we see that, as noted above, when participation costs are high they tend to overwhelm all other influences resulting in communities that are small, including the nature of interests and knowledge in the population. As the costs of participation drop, the other factors (contribution effects and population knowledge and interest structures) emerge as significant influences on community size (as well as the other community features). Among target populations comprised of individuals interested in and knowledgeable about a relatively high proportion of the overall knowledge space, the effects of contribution factors are not significant. It is only in cases of low/moderate participation costs and low/moderate interest and knowledge ranges that the proportion of contributors in the population and their frequency of contribution begin to play a role. Within those contexts, the analysis suggests that efforts to reduce contribution costs will result in smaller online communities. Thus the theoretical model, and its computational instantiation, supports the recommendation made by Goodman and Darr (1998), while adding an important caveat. That is, developers would do well to develop technology and techniques for supporting the development of communities in focus populations (i.e. those in which individuals are knowledgeable about and interested in a higher proportion of the relevant knowledge space). In doing so, there is the potential to increase the information flow (message volume) among a wider range of individuals (larger size communities) through the reduction of participation and contribution costs. However, the present model also reminds us that in populations with narrow knowledge the recommendation becomes a self-fulfilling prophecy. If a developer develops an infrastructure which reduces participation and contribution costs, and yet fails to select a target population in which individuals are relatively knowledgeable and interested, then the communities that result will tend to be significantly smaller (if more active), have less resilient members, and provide far less coverage of the available knowledge.

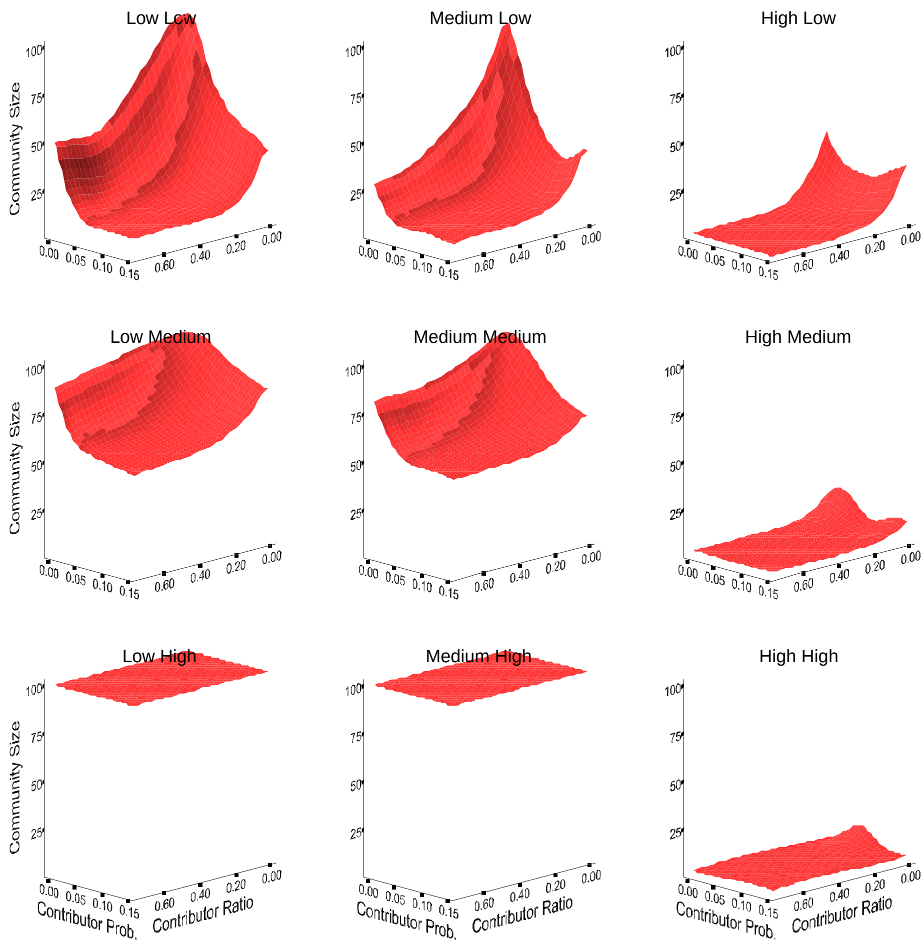

Figure 7: Interacting Effects of Infrastructure and Target Population Characteristics on Community Size

Overall, the analysis of this computational model suggests that online community developers are faced with a complex problem, not only because they are working with an emergent phenomena, but also because as they are attempting to navigate a problem space in which changes in the solution (i.e. the community infrastructure and/or the target population) may alter the problem that they are faced with. Even in this conceptually simplified model, the results suggest that developers must contend with a hierarchy of effects -- a hierarchy in which changes in the infrastructure that appear to be improvements (i.e. reducing the costs of participation) have the potential to create new challenges to the development and maintenance of online communities that provide visibility and access to knowledge in distributed populations.

5. Directions for Future Research

The computational model considered in this paper is presented for purposes of theory development. As with all simulation studies, this work should be judged in terms of the balance between the purpose (theory generation), the relationship of the model features to that purpose, and the use of empirical data (for calibration and validation) relative to that purpose (Burton and Obel 1995). Although inclusion of the additional community or infrastructure features within the computational model might increase its "realism", the features included in the current instantiation of the theory support the primary purpose of integrating member development, structural change, and communication costs into a model of online community development. Further development of the contribution and target population models, explicit consideration of the interpersonal networks that exist around online communities, and examination of the operation of larger systems of multiple communities would build on the foundation presented here to further the study of the role of community infrastructures in the development of effective, beneficial online communities.

The structuration framework presented here suggests that integration of individual and structural change processes provide a theoretical foundation for examining the consequences of community infrastructure design decisions, as well as presenting a foundational process model of online community development. This theory focuses on the dynamic aspects of social structures that play an important role in the flow of information within and between organizations (von Hippel 1988; Goodman and Darr 1998). It also considers the impact of alternative technologies on these processes. As organizations, both public and private, invest in information technology with the goal of increasing access and visibility of knowledge in distributed populations, it becomes increasingly important to understand how an online community's infrastructure and social processes interact.

As network infrastructures such as the Internet become widely available, there is a tendency to assume that providing the ability to efficiently communicate necessarily encourages social interaction and an associated flow of information. Prior research on computer-mediated communication has focused primarily upon demonstrating ways that individuals *can* interact effectively through networked infrastructures, and thus has begun to explore the possible community infrastructures. However, this foundational work has failed to highlight a fundamental pitfall of online community development - while individuals *can* behave in many ways, how they *will* behave is the result of the complex interaction of technology, individual behaviors, and the emergent community characteristics. While the immediate effects of new technologies on mechanistic efficiency may be most visible, it is incorrect to assume that they are necessarily the most important consequence of introducing new communication infrastructures (Sproull and Kiesler 1991). However, without dynamic models that take into account individual, structural, and technological features of online communities, our ability to understand, and ultimately predict the impact of new communication technologies is limited. By combining empirical data and computational modeling, this paper develops a theoretical framework for examining online communities and provides a foundation for future research about how the availability technology has, and will continue, to affect the development of communities both within and outside organizations.

6. REFERENCES

Baym, N. K. (2000). Tune in, log on : soaps, fandom, and online community. Thousand Oaks, Calif., Sage Publications.

Brinthaupt, T. M., R. L. Moreland, et al. (1991). "Sources of Optimism Among Prospective Group Members." Personality and Social Psychology Bulletin **17**(1): 36-43.

Brown, J. S. and P. Duguid (1991). “Organizational Learning and Communities of Practice:Toward a Unified View of Working, Learning, and Innovation.” Organization Science **2**(1): 40-57.
Burton, R. M. and B. Obel (1995). “The Validity of Computational Models in Organizational Science: From Model Realism to Purpose of the Model.” Computational and Mathematical Organizational Theory **1**(1): 57-71.
Butler, B. S. (1999). The Dynamics of Cyberspace: Examining and Modelling Online Social Structure. Graduate School of Industrial Administration. Pittsburgh, Carnegie Mellon University**:** 142.
Carley, K. (1991). “A Theory of Social Stability.” American Sociological Review **5-6**: 331-354.
DeSanctis, G. and M. S. Poole (1994). “Capturing the Complexity in Advanced Technology Use:Adaptive Structuration Theory.” Organization Science **5**(2): 121-148.
Giddens, A. (1984). The constitution of society : outline of the theory of structuration. Berkeley, University of California Press.
Goodman, P. S. and E. D. Darr (1998). “Computer-Aided Systems and Communities:Mechanisms for Organizational Learning in Distributed Environments.” MIS Quarterly **22**(4): 417-440.
Hagel, J. and A. Armstrong (1997). Net gain: expanding markets through virtual communities. Boston, Harvard Business School Press.
Hunter, J. E., J. E. Danes, et al. (1984). Mathematical Models of Attitude Change. New York, Academic Press.
Moreland, R. L. and J. M. Levine (1982). “Socialization in Small Groups: Temporal Changes in Individual-Group Interactions.” Social Psychology **15**.
Orlikowski, W. J. and D. Robey (1991). “Information Techology and the Structuring of Organizations.” Information Systems Research **2**(2): 143-169.
Rheingold, H. (1993). The virtual community : homesteading on the electronic frontier. Reading, Mass., Addison-Wesley Pub. Co.
Skvoretz, J. (1988). “Models of Participation in Status-Differentiated Groups.” Social Psychology Quarterly **51**(1): 43-57.
Sproull, L. and S. Kiesler (1991). Connections: New ways of working in the networked organization. Cambridge, MA, MIT Press.
Storck, J. and P. A. Hill (2000). “Knowledge Diffusion through "Strategic Communities".” Sloan Management Review: 1-12.
Turner, J. H. (1988). A Theory of Social Interaction. Standford, CA, Stanford University Press.
von Hippel, E. (1988). The Sources of Innovation. New York, Oxford University Press.
Wittenbaum, G. M. and G. Stasser (1996). Management of Information in Small Groups. What's Social About Social Cognition?: Research in Social Shared Cognition in Small Groups. J. L. Nye and A. M. Browser. Thousand Oaks, CA, Sage Publishing.