Evolution of community structure in knowledge work: A case study in open source

Mehmet Gençer, Beyza Oba

Abstract

Open source software development communities work over the Internet and provide a rich record of community interaction available for research. Structure of these communities evolve as they grow, forming specialized sub-communities to sustain knowledge intensive work and innovation. In this paper we report research results on the structural evolution of a large scale OSS community, using records of e-mail groups for a seven year time period. Our findings indicate that birth rate of sub-communities is proportional to community size, regardless of its stability. We have further found that evolution of and behavior in sub-communities which sustain knowledge discovery and knowledge creation processes are significantly different. While knowledge discovery favors large and stable groups even if they lack focus, knowledge creation is sustained by formation of small and focused expert groups as needed and exhibits more intense communication among participants.

1 Introduction: Organising for Innovation

The ongoing quest for efficiency in organizing work paves the way for emergence of specialized sub-units in organisations. In the famous example of Smith's needle factory (2006), for instance, the efficiency of work rests on the specialization that each worker or team of workers performs a specific operation repetitively and efficiently, rather than each learning a variety of operations for finishing a needle from raw materials and switching tools and materials to perform those operations.

In most situations, managerial intervention in organisational re-structuring is problematic due to the disturbance it creates on the emergent organisation. The state of organisation research is considered at its infancy in terms of providing an organisation design perspective (Jelinek et al., 2008). The problem is highly visible in knowledge intensive work organisations and when innovation is the core matter. In such situations managerial intervention in structural change is required to let innovation happen, rather than plan or control it. However, organisation studies provide limited insight and managerial advice for the matter.

Software engineering is a relatively new field of practice and industry in which the bulk of activity is innovation, not manufacturing. Scaling of such knowledge work in developing software has always been problematic (Brooks, 1995). This situation has created a lot of interest around the Open source software (OSS) approach to developing software which emphasizes evolution and bottom-up processes instead of top-down planning (von Hippel and von Krogh, 2003; Pisano, 2006). As a consequence body of research on OSS has been growing quickly in recent years. However, interest in organizational knowledge and learning is rather recent in management studies (Spender and Grant, 1996) and the rich and 'open' corpus of data related to OSS communities rarely received due attention in this respect.

In this paper we report results of a study on the structural evolution of a large scale OSS community, called Eclipse. Eclipse is an inter-organizational OSS project and community initiated by IBM in 2001 to produce a software technology collaboratively. IBM has contributed an existing software base they owned to the project, under an open source license, and seek participation of other firms in the computing industry to collaboratively develop the technology. Over time the project community has enjoyed large scale participation from several firms in the computing and related industries who embrace OSS as a business strategy, as well as participation of many non-profit organizations and individuals. The aim of the project was to develop a versatile software technology for use in enterprise software development. The corresponding industry segment had an oligopolistic market structure at the time and IBM, with its recent interest in open source strategy, believed that they can establish an inter-organizational ecosystem to offer a more competitive and innovative alternative (Capek et al., 2005). The initiative was a significant success, and the Eclipse Foundation was formed in 2004 to support the growing Eclipse community and oversee the technology's progress. In time, hundreds of firms and non-profit organizations joined the foundation, supporting development of the Eclipse software both financially and by providing dedicated software developers. The community's growth is not limited to formal members of the Eclipse Foundation, however, and thousands of individuals from numerous business or non-profit organizations, as well as independent software professionals and university students are now part of the community.

The members of the globally dispersed Eclipse community mainly collaborate by means of e-mail groups; although there are several conferences, printed or on-line journals about the Eclipse technology which provide additional channels of communication. Daily work in the community takes place in the e-mail groups where a variety of users from all over the world exchange messages to seek help, to report software errors they have discovered, or to solve problems in coordination. The participants of the community range from naive users who seek help for using software, to dedicated programmers who need to work together to develop new features or solve problems. Community e-mail traffic is divided into mail-groups to facilitate more efficient communication. Furthermore, the mail-groups for 'users' who seek help from one another for deploying and using software are separate from those mail-groups reserved for 'experts' who are involved in discussions for coordinated modification of software code to fix errors, to implement new features, or to handle issues related to software packaging and version management. It is not uncommon that users, after developing experience, join the ranks of experts. As common in OSS communities management of such shifts are meritocratic, and fitting-into expert groups is related to an individual's development of the expertise on the software and community norms (von Krogh et al., 2003). Thus participation to the community is dynamic where individuals may participate in multiple mail groups, leave and join groups as their interests and expertise changes. As the community has grown and the software it produces became large and complex, the organisation of work became divided into many sub-communities and corresponding e-mail groups, each of which focuses on different aspects or parts of work on development of Eclipse software.

Our aim in this study is to understand evolution of the community structure in relation to knowledge processes. For this purpose we have formed several hypotheses regarding the community structure and its evolution, which are explained in section 2 in relation to theoretical framework we adopted. We have collected a data set using records of e-mail groups for a seven year time period starting with community's initiation, which is described in section 3. Our analysis methods are largely based on testing of our hypothesis, where some parts remain exploratory. The analysis and findings are summarized in section 4. We discuss implications of these findings in section 5, followed by conclusions.

2 Theoretical Background and Hypothesis

The increasing importance of innovation for business success in recent decades paralleled the volume of academic research on inter and intra-organisational knowledge and communication. As Nonaka et.al. (2006) note, by interacting and sharing knowledge with others, the individual "*enhances the capacity* to define a situation or problem"(Nonaka et al., 2006, 1182). Where innovation, rather than routine work, is the business focus, this acknowledgement makes principles of scientific management and hierarchical division of business organisation (and effectively its knowledge) questionable. On the other hand, reflection of concepts regarding individual knowledge processes on organisational knowledge often result in "anthropomorphizing the legal entity called the corporation"(Schein, 1996, p230). Unlike the coherence we attribute to individual knowledge, there are multiple work groups with their corresponding occupational cultures in an organisation. Each group have its own norms, language, etc. (Brown and Duguid, 2001; Orr, 1986), often conflicting or incompatible with one another, thus requiring a communication sphere of its own to work efficiently. As a consequence, there is the dual requirements of information hiding and sharing when organising for knowledge work. Until we develop a firm understanding of communication in relation to organisational knowledge and problem solving, however, organisation design remains a managerial craft rather than a science.

A related research thread of inter-organisational knowledge and innovation appears to be largely focused on extending the competitive business logic with collaborative strategies. This research thread, which is recently converging under the label of 'open innovation', emphasises the potential of inter-organisational collaborative innovation in the service of business strategy (Jorde and Teece, 1989; Chesbrough and Appleyard, 2007). Various research can be related to this emerging thread which focus on inter-organisational networks and specialisation (Langlois and Robertson, 1992; Kogut and Metiu, 2001). As Langlois and Robertson assert "a modular system ... enlists the division of labor in the service of innovation"(1992, p302). In this sense, modularity and dynamism of collaborative networks is seen as an important aspect of inter-organisational innovation. On the other hand empirical research in this thread is almost entirely focuses on organisational level, and void of studies regarding micro and meso levels of inter-organisational collaboration.

OSS phenomenon is frequently cited within open innovation research. However, business adoption of OSS model is rather recent (Fitzgerald, 2006). As a consequence the business organisation is not a common unit of analysis in existing OSS research. There are few studies concerning specialization in OSS communities (e.g. von Krogh et al. 2003), however they tend to focus on the individuals level rather than communities or organisations as their unit of analysis.

A different thread of research on the communities of practice (CoP) can be seen as complimentary to intra-organisational research (Lave and Wenger, 1991; Brown and Duguid, 2001; Orr, 1986). CoP framework provides a grounded perspective on individual and group level processes in organisational knowledge, with empirical studies mostly on intra-organisation level. Nevertheless, the CoP framework makes visible the contextual differences between (sub)communities within an organisation in relation to knowledge processes. For example, Orr (1986) underlines the gap between the two groups of managers and field workers regarding the nature of work. Handling of daily tasks is not only different from the managerial perspective and prescriptions, but also relies heavily on communication and information exchange between sub-community members, which in turn requires shared norms, understanding, and code of conduct. Cop research usually emphasises the invisible boundaries of sub-communities within an organisation, and underlines the importance of socialisation as the basis of new member's integration into these communities (Lave and Wenger, 1991).

Articulating this line of thinking in the CoP framework, we expect certain differences between the two broad groups of 'user' and 'expert' mail-groups within an OSS community like Eclipse. Reading of sampled e-mails from both types of mail-groups hinted that expert communities are more tightly knit groups where people are familiar with others. In regard to their orientations, expert communities are expected to have relatively smaller populations to facilitate coordination of collective action, and user communities to have relatively higher populations to facilitate exploitation of a larger knowledge source. In addition expert communities are those which share a deeper practice, compared to user communities. As a consequence, we expect expert communities to be smaller, and exhibit tighter boundaries. Therefore it is expected that population entries and exits are relatively low in those compared to user communities. These expectations can be summarised in the following hypotheses, which extend the CoP perspective in relation to contextual differences of communities:

Hypothesis 1 *Community population is larger in user communities than it is in expert communities.*

Hypothesis 2 *:*

a) Ratio of population entries to population is lower in expert communities compared to user communities.

b) Ratio of population exits to population is lower in expert communities compared to user communities.

The same line of thinking can be followed if one wishes to extend the analysis of contextual differences into communication characteristics in two types of communities: How responsive are community members to each other? How easily they communicate? How well they participate in collective action? A rather trivial extension of the CoP emphasis on socialisation is that in smaller, tightly knit groups, individuals are expected to demonstrate higher participation in the collective action, and be more responsive to each other's communication:

Hypothesis 3 *Participation degree is higher in expert communities compared to user communities.*

Hypothesis 4 *Responsiveness is higher in expert communities compared to user communities.*

The expectation articulated in the above section that populations of expert communities are more stable (i.e. low entries and exits) has important consequences for the communication that takes place in them. Since new participants create disturbance to the community, which is relieved only after a rather lengthy process of socialization, user communities with their relatively more unstable populations are expected to demonstrate consequences of this disturbance in the communication. Following this argument, in user communities we expect longer conversations and lower responsiveness compared to expert communities, because of the need to accommodate relatively lower communication capability in user communities due to disturbance created by high number of population entries. The second finding is that expert groups in Eclipse have a focus on coordinating activity among several experts who are related to software issues under discussion. For this reason we expect expert communities to have relatively higher conversations breadth (number of participants to a conversation) and participation degree, compared to user communities. These expectations can be formulated as the following hypotheses:

Hypothesis 5 *Conversation depth is lower in expert communities compared to user communities.*

Hypothesis 6 *Conversation breadth is higher in expert communities compared to user communities.*

We now turn from differences of communities in terms of static qualities, to their evolution in time. Aldrich's (1999) organisational evolution framework provides a basis for approaching developmental dynamics. However, most organisation research on evolution considers the populations of business organisations, in relation to institutionalisation and emergence of organisational population (1999), or inter-organizational innovation and specialization (Jorde and Teece, 1989; Langlois and Robertson, 1992; Kogut and Metiu, 2001). Studies such as Langlois (1990), Langlois and Robertson (1992), Carlsson and Stankiewicz (1991), and Kogut (2000), for example, focus on historical co-evolution of networks and systems of specialized firms or groups of firms.

Although the findings of evolutionary research regarding the business systems level are not directly relevant for the emergence of specialized communities like those in Eclipse, they point to the evolutionary perspective as a promising approach for use in a case like Eclipse where hierarchical command is replaced by network coordination. Within the hierarchy of business organizations, formation of sub-units appear to be an instrument to accommodate (by means of compartmentalising) the expanding labor force or increasing variety of activities. On the other hand in evolutionary thinking variation happens regardless of such changes. The amount of this variation is proportional to the population, since it is the locus of mutations. If we follow this line of thinking, we expect births of specialized sub-communities in an open source community like Eclipse to be effected by the population of community, and not effected by its changes (i.e. entries to or exits from the population):

Hypothesis 7 *Concerning the births of specialized sub-communities:*

a) The higher the population of an open source community, the higher the births of specialized sub-communities.

b) Births of specialized sub-communities in an open source community is independent of entries to the population.

c) Births of specialized sub-communities in an open source community is independent of exits from the population.

It is worth noting that the above hypothesis captures only an endogenous (albeit major) factor affecting community births. There are possibly other endogenous and exogenous factors at work, such as leadership discretion to channel community work into certain strategic directions, environmental changes which spontaneously increase or decrease participation to Eclipse project, or changes in the design of software product itself which create new areas of work. For practical reasons, we limit our empirical research to the population factor within the context of this study.

A natural extension of our interest on evolution of community structure is how two types of communities may possibly differ in terms of growth, following their birth. However, we cannot derive any hypothesis concerning this aspect, within the available theoretical framework. Therefore, we adopt an exploratory approach in evaluating the empirical evidence in relation to community growth, whose results are presented and discussed in the relevant sections below.

3 Research data and analysis

Archives of the Eclipse community mail-groups are publicly available and we have used custom computer programs to retrieve all available archives for the time period from April 2001 (initiation of the project) to October 2008. Among the retrieved mail-groups 137 were groups which were used by users, and 94 were groups which were used by experts/developers. A total of 602.202 e-mail records (486.507 from user groups and 115.695 from developer groups) were retrieved. Due to various problems with their content some of messages (less than %1) were not usable. Remaining 595.882 records for a total of 230 mail groups were in a usable state. We have identified 55.340 different personal e-mail addresses as sources of these e-mails.

Several pieces of information are available from an e-mail record. The most basic ones are the mail group that the e-mail belongs, date and time of sending, and sender's name and e-mail (although some are obfuscated). In addition there is referencing information in each e-mail record which makes possible to identify an e-mail conversation (or mail 'thread' as they are commonly referred to) that spans several e-mails. Each e-mail message is given a unique identification number by e-mail programs, and when an e-mail is sent as a response to a previous one (in contrast to starting a fresh conversation), identification number of the message replied to is included. This reference information in e-mail records make it possible to identify e-mail conversations.

By analyzing the e-mail records using a custom computer program we have obtained the following information:

- The number of people who have participated to each mail-group (sub-community size).
- The time each mail-group has started (birth date of sub-community).
- The first and last activity of each individual, where such activity possibly spans multiple mail-groups. In this way the total population of the Eclipse community, and number of newcomers and departers are found for each monthly period (although the method fails towards the end of time frame as it labels all members as leaving the community).
- For each mail-group, number of e-mails sent by each individual in the mail-group (i.e. participation degree data for each sub-community). This data was generated as a panel data per time period.
- For each mail-group, the ratio of e-mails that received a response (i.e. responsiveness level in the sub-community). This data was generated as a panel data per time period.

- For each mail-group, the average depth and breadth of e-mail conversations (i.e. the number of messages and number of people involved in conversations). This data was generated as a panel data per time period.

In addition to the above, we have evaluated the contents of randomly sampled e-mails from various groups in order to develop an understanding of the differences of work context in sub-communities within Eclipse. Although our research design does not aim a content analysis of the whole data corpus, this evaluation has been instrumental in developing a research focus.

4 Summary of findings

Analysis findings of communities in this section is presented around two axes of investigation: (1) differences in population and communication characteristics of 'user' and 'expert' type communities in Eclipse, in relation to their contextual differences, and (2) birth and development of specialised communities in relation to life-cycle of the greater Eclipse community, and community types. The findings are summarised in the two separate sections below.

4.1 Structure and communication differences between community types

Reading of sampled e-mails from a variety of mail-groups at the beginning of this research revealed certain differences in the contexts of two types of sub-communities: (1) the *user* communities, majority of whose participants are end-users of the Eclipse software seeking help from one another regarding problems they face in employing the software for their needs, and (2) the *expert* communities whose participants are software developers who try to coordinate their efforts in improving the Eclipse software. Expert communities are innovation oriented and a major contextual difference between the two types of communities is that between knowledge processes. In user communities the common motivation for participation is exploiting the existing experience in the community to find answers relevant to one's problems encountered while using the software (knowledge discovery). On the other hand participants of expert groups are interested in exploring possible actions and their consequences in the pursuit of improving software and eventually concerting collective action for improvement of software (knowledge creation).

Based on these observations, the two types of sub-communities in the data set are labeled accordingly to facilitate analysis. There are 136 user communities, and 94 expert communities in the data sample. This section explores the differences between user and expert communities in terms of their population and characteristics of communication activity within them.

4.1.1 Population of communities

Analysis of variance is used to test hypotheses 1 and 2 regarding population of two types of communities in Eclipse, whose results are shown in Table 1. The table also shows mean values of measures. The results support all hypotheses above.

Insert table 1 here

To further explore population differences of user and expert communities, Figures 1 and 2 show overall distribution of community populations and separate quantile plots of user and expert communities. The group of crowded communities at the far end of the histogram in Figure 1 was further investigated. All of these are found to be user groups of most popular Eclipse software components, in line with the above results. On the other hand the difference of size distributions demonstrated in Figure 2 shows that the first quantile for user communities is lower than that of expert communities, which is rather contrary to findings above and calls for further attention. In fact the median values for community sizes are 36 and 47, for user and expert mail groups, respectively, which has the reverse magnitude relationship compared to mean values. This indicates that while in average user communities have higher populations compared to expert communities, most user communities with very low population and which can be considered inactive or underdeveloped. Most user communities simply fail to grow after their birth.

Insert figure 1 here

Insert figure 2 here

4.1.2 Differences of communication characteristics

In testing hypothesis 3 through 6 regarding communication characteristics of the two types of communities, we use the panel data for participation degree, responsiveness and conversations for the time frame April 2001 to April 2008, thus leaving out some of the data set towards the end which is likely to give false results about ongoing conversations. To reduce bias as a result of outliers, periods of very low activity in panel data were filtered out prior to analysis.

The results of analysis of variance tests are shown in Table 2 along with mean values of measures for user and expert groups.

Insert table 2 here

The results support hypotheses 3 and 5, but not the other two. Regarding the conversation breadth (hypothesis 5), we have checked for possible correlation between conversation depth and breadth to investigate the results contrary to our hypothesis. When the ratio of the two is computed we see that breadth/depth ratio is higher for expert communities (0.824) compared to user communities (0.632). In other words long conversations in expert communities involve more people than a similar length conversation in user communities. This resonates well with the difficulty of problems tackled in expert communities hence requiring involvement of more people, but not with our hypothesis that expert groups communicate more efficiently hence conversations should be shorter.

On the other hand regarding the case of responsiveness (hypothesis 4) we cannot offer a full explanation for the empirical finding except our observation that there is frequent use of announcement type messages in expert groups which require no response from other community members.

4.2 Development of specialisation

Sub-communities in Eclipse came into being at different points in time within the frame of greater Eclipse community's life-cycle and they are of varying sizes (i.e. number of people in the sub-community). The change of number of communities born through Eclipse life-cycle is shown in Figure 3 which displays the number of births in five month time slices from April 2001 to April 2008. Number of people joining he Eclipse community in the same time span is given in Figure 4.

Insert figure 3 here

Insert figure 4 here

During the periods at the beginning and after period 40 there is an increase in births of specialized communities, which seem to coincide with community growth following establishment of the Eclipse Foundation which stimulated this growth. What is rather interesting about the community births is that they continue to stay relatively high towards the later stages of community at the right hand side of the plot, roughly after period 60, despite the fact that the community growth is lower during that time.

For an analysis of concomitant features of population growth and community births, it is necessary to combine the data for the greater Eclipse community with the data for the communities level. Since community births is a higher level and slow process compared to communication of individuals a redundant dataset is generated which contains population measures for five month periods from April 2001 to April 2008, instead of monthly periods. In this new dataset population is taken as the average of population over five corresponding monthly periods from the dataset for the greater Eclipse community, and entries (newcomers) and exits (departers) from the population are found by summing the same monthly periods. Number of community births in each of the five month periods are found by selecting those mail group communities from the communities data set whose period of birth falls in the period of measurement. Interaction of the measures in this new dataset is shown in Figure 5. The plots seem to support the hypothesized relation between population and community births (hypothesis 7a-c). However, since entries and exits are also high when the population is high, statistical inference is necessary to establish impact of growth measures.

Insert figure 5 here

Since the dependent variable, community births, is bounded (i.e. non-negative) generalized linear models is used to find impact of population growth variables. The results of modeling is shown in Table 3[1]. All factors are scaled to have unit mean to compare their impact. Results of statistical inference support hypotheses 7a through 7c. Among population growth variables, only

[1]Model residuals are checked against non-stationarity using Kwiatkowski-Phillips-Schmidt-Shin test.

population itself is significant, at $p < .05$ level. Furthermore the intercept of the model is insignificant, indicating that community births is directly proportional to population. The model has a considerably high explanatory power as can be seen from comparison of the residual deviance of the model and that of the null hypothesis.

Insert table 3 here

These findings confirm that at the communities level, speed of specialization process, in the form of emergence of new sub-communities, is proportional to the size of population as suggested by our evolutionary argument. Yet, one significant deviance in the observations is the low number of births prior to week 40 (approximately the establishment of Foundation in 2004). This time interval is one which community leadership was in a strategic turning point and concentrated on restructuring the community and finding ways to provide industry support to the project. Lack of stimulation from management level must have had an impact on communities level. Such impact, however, cannot be investigated with the data used in this research.

4.2.1 Developmental differences

Our analysis regarding developmental differences of user and expert communities is purely exploratory and not driven by any hypothesis. This investigation is necessary to explore community development, but our theoretical framework does not provide any basis to develop any hypothesis.

We have seen in the above section that increasingly more communities were born as the greater community has grown. The findings in the previous section about population distributions shows that many user communities fail to grow and live, while it is not the case for expert communities. To explore growth of different types of communities, populations of communities versus their age is shown separately for user and expert communities in Figures 6 and 7.

Insert figure 6 here

Insert figure 7 here

In the case of user communities younger communities appear to have failed to attract many members. However, in the case of expert groups one sees a pattern of gradual expansion of communities, setting aside the natural size differences due to their subject nature. To better see the phenomenon in relation to Eclipse life-cycle, Figure 8 shows population of user communities versus their time of birth. It is more clear in this plot that only a few 'user' type communities which were born after a certain time have managed to grow, and only to a limited size. The rest have failed to develop.

Insert figure 8 here

5 Discussion of findings, and conclusions

Empirical findings from a large scale inter-organisational community reported in this study highlight certain features of communication and community development, and how these features differ across knowledge discovery and knowledge creation oriented sub-communities. Empirical data is rarely available to approach some of these features, as well as their variance with respect to type of collective knowledge processes. In this section we summarise the results and discuss their implications for theory and practice.

First of the two group of our findings reported here concerns the differences of population and differences in communication behaviour between two types of specialised sub-communities in the Eclipse case. ‘User’ type sub-communities mostly consist of naive users whose aim in joining the community is to take advantage of its software technology, and their interaction with other community members is mostly related to seek existing experience about problems they encounter in deploying the software. As a result the focus of activity in user communities is accessing a great variety of knowledge (i.e. collective knowledge discovery). ‘Expert’ type sub-communities, on the other hand, consist of software developers who are involved in fixing software errors or implementing new features, thus in pursuit of intense collaboration with relevant peers in the community (i.e. collective knowledge creation). By borrowing concepts from CoP theory, we have hypothesised that smaller community size is an advantage for expert communities in contrast to user communities, as it will enable stronger socialisation which provides a more extensive basis of norms and shared language necessary for expert communities to work. Analysis confirmed that ‘user’ communities indeed have larger populations compared to ‘expert’ communities. User communities also have a high ratio of population entries and exits. Our results indicate that knowledge discovery processes (‘user’ communities) favour larger community size to specialisation. Knowledge creation (‘expert’) communities are home to a deeper practice, which favours smaller and more stable populations.

The differences in communication behaviour were explored in terms of several measures such as responsiveness of community, participation degree of members to communication, and depth and breadth of conversations. The results confirmed some of our hypothesised expectations that expert communities communicate more efficiently and have more active members.

This first group of findings show that our extension of CoP framework to comparative treatment of different communities with respect to two stereotypes of knowledge processes generally worked well. On the other hand some contradictory results such as lower responsiveness observed in expert communities indicate that there is much room for improving these rather new measures and their interpretation.

The second group of findings concern development of specialisation, reflected in birth and development of sub-communities. By reflecting on the organisational evolution theory we have hypothesised that speed of structural evolution (i.e. birth of new sub-communities) is proportional to size of population. Analysis confirmed this group of hypotheses. Taken in more general terms, these findings parallel what Kauffman (1990) terms as living at the “edge of chaos”. In using the term, Kaufman suggests that complex systems which maintain a level of complexity higher than what is necessary for survival in the current conditions (but without dissolving into chaos) are the norm in nature. Because such ‘hot’ systems have a better chance of discovering incremental changes that make it possible for them to adapt to and survive in changing environmental conditions. In the Eclipse community, the tendency for structural change observed as the births of specialized communities should be considered as an important organizational element that nurtures adaptability.

This part of our study also addressed how specialised communities of different types differ in terms of how they develop, using exploratory methods. Findings show that in the case of user communities, once the user population was dispersed into a number of communities, newly formed communities cannot develop. This is not the case for expert communities, however, where new communities can develop well. In interpreting these findings, we have extended the evolutionary logic by borrowing from CoP theory to address different behavioural orientations of communities. In the ‘user’ communities the concern of participants is finding help from the community about the problems encountered when using software. As a result the focus of activity in user communities is accessing a great variety of knowledge. In the second type, *expert* communities, participants are mostly developers and their concern is coordinating their efforts to improve software. Although numerous user communities exist in Eclipse, very few of those which are formed after a certain point of Eclipse life-cycle have managed to develop to a sustainable population. A likely reason for this is that established user communities offer their new participants access to a large knowledge base of its existing large population, hence rendering new under-populated communities relatively unattractive. On the other hand expert communities are much more amenable to division since communication oriented towards coordination and generation of new knowledge favors more intense and focused interaction among a smaller number of peers, rather than large crowds. New expert communities do not face the developmental shortcoming of new user communities. Although not addressed in this study, it is worth noting that communication behavioural patterns such as information brokerage possibly come into play to address possible negative effects of such high compartmentalisation in expert communities of Eclipse.

There is little research on the factors which stimulate or suppress formation of sub-units in such inter-organisational contexts. Within business organizations sub-unit formation is regarded as dictated by the management through the top-down hierarchy of the business enterprise. Sub-units in business organizations are at best subjected to performance comparison in research (e.g. McGrath 2001 on top-management groups’ performance differences). On the other hand in an open source inter-organisational community like Eclipse, emergence of specialized sub-units is not controlled in the same way as in a hierarchical business organization. Even though the Eclipse community has a formal leadership represented by the Eclips Foundation, the leadership concedes, as much as or perhaps more than it leads, the formation of specialized communities. Or better, following Jelinek’s phrasing, the leadership at best “invites it to emerge”(Jelinek, 2004, p115), but division of organization remains emergent rather than planned. The Eclipse case suggests that managerial discretion to adjust community structure (e.g. create new user sub-communities) is largely disconnected from community level experience; an issue often voiced in CoP research. In this sense, appropriate knowledge management systems which exploit electronic repositories present opportunities to improve managerial practice. While there is an increasing amount of research on such methods in other disciplines (e.g. engineering research on social networks trend analysis), there is little penetration of such methods into management research.

6 Conclusion

In this paper we have presented the results of an empirical study on the structure of a large scale inter-organisational software development community, changes in this structure through time, and differences of population and communication behaviour in different types of sub-communities within the larger community. We have developed several hypotheses by combining and extending organisational evolution theory with communities of practice research framework. Two types of sub-communities, one oriented towards knowledge discovery, and other towards knowledge creation were addressed separately. Our findings generally confirm that such differentiation proves useful in comparative understanding of structure, structural evolution, and communication characteristics in different types of sub-communities.

References

Aldrich, H. (1999), *Organizations Evolving*, Sage.

Brooks, F. P. (1995), *The mythical man-month*, Addison Wesley Longman.

Brown, J. S. and Duguid, P. (2001), 'Knowledge and organization: A social-practice perspective', *Organiation Science* **12**(2), 198–213.

Capek, P. G., Frank, S. P., Gerdt, S. and Shields, D. (2005), 'A history of ibm's open-source involvement and strategy', *IBM Systems Journal* **44**(2), 249–257.

Carlsson, B. and Stankiewicz, R. (1991), 'On the function, nature and compositon of technological systems', *Journal of Evolutionary Economics* **1**, 93–118.

Chesbrough, H. W. and Appleyard, M. M. (2007), 'Open innovation and strategy', *California Management Review* **50**(1), 57–76.

Fitzgerald, B. (2006), 'The transformation of open source software.', *MIS Quarterly* **30**(3), p587 – 598.

Jelinek, M. (2004), *Managing as Designing*, Stanford, chapter Managing Design, Designing Management.

Jelinek, M., Romme, A. G. L. and Boland, R. J. (2008), 'Organization studies as a science for design: creating collaborative artifacts and research', *Organization Studies* **29**(03), 317–329.

Jorde, T. M. and Teece, D. J. (1989), 'Competition and cooperation: Striking the right balance', *California Management Review* pp. 25–37.

Kauffman, S. A. (1990), 'The sciences of complexity and "origins of order"', *Proceedings of the Biennial Meeting of the Philosophy of Science Association* **2**, 299–322.

Kogut, B. (2000), 'The network as knowledge: Generative rules and the emergence of structure', *Strategic Management Journal* **21**, 405–425.

Kogut, B. and Metiu, A. (2001), 'Open-source software development and distributed innovation', *Oxford Review of Economic Policy* **17**(2), 248–264.

Langlois, R. N. (1990), 'Creating external capabilities: Innovation and vertical disintegration in the microcomputer industry', *Business and Economic History* **19**, 93–102.

Langlois, R. N. and Robertson, P. L. (1992), 'Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries', *Research Policy* **21**, 297–313.

Lave, J. and Wenger, E. (1991), *Situated learning: legitimate peripheral participation*, Cambridge University Press.

McGrath, R. G. (2001), 'Exploratory learning, innovative capacity, and managerial oversight', *Academy of Management Journal* **44**(1), 118–131.

Nonaka, I., von Krogh, G. and Voelpel, S. (2006), 'Organizational knowledge creation theory: Evolutionary paths and future advances', *Organization Studies* **27**(8), 1179–1208.

Orr, J. E. (1986), Narratives at work: story telling as cooperative diagnostic activity, *in* 'CSCW '86: Proceedings of the 1986 ACM conference on Computer-supported cooperative work', ACM, New York, NY, USA, pp. 62–72.

Pisano, G. (2006), 'Profiting from innovation and the intellectual property revolution', *Research Policy* **35**(8), 1122–1130.
URL: *http://dx.doi.org/10.1016/j.respol.2006.09.008*

Schein, E. H. (1996), 'Culture: The missing concept in organization studies', *Administrative Science Quarterly* **41**, 229–240.

Smith, A. (2006), *Ulusların zenginliği (An Inquiry into the nature and causes of the wealth of nations)*, Palme.

Spender, J. C. and Grant, R. M. (1996), 'Knowledge and the firms: Overview', *Strategic Management Journal* **17**, 5–9.

von Hippel, E. and von Krogh, G. (2003), 'Open source software and the "private-collective" innovation model: Issues for organization science', *Organization Science* **14**(2), 209–223.

von Krogh, G., Spaeth, S. and Lakhani, K. (2003), 'Community, joining, and specialization in open source software innovation: a case study', *Research Policy* **32**, 1217–1241.

Table 1: Differences of sub-community demographics with respect to community type

	AOV	User comm.	Expert comm.
Population	$p < 0.01$	561.70	93.00
Newcomers/Population	$p < 0.001$	0.62	0.25
Departers/Population	$p < 0.001$	0.65	0.25

Table 2: Relation of sub-community type and communication characteristics.

Variables	AOV	Mean values	
		User groups	Expert groups
Partic.Degree	$p < 0.001$	4.112	13.030
Responsiveness	$p < 0.001$	0.578	0.325
Conv. Depth	$p < 0.001$	3.831	2.821
Conv. Breadth	$p < 0.001$	2.421	2.324

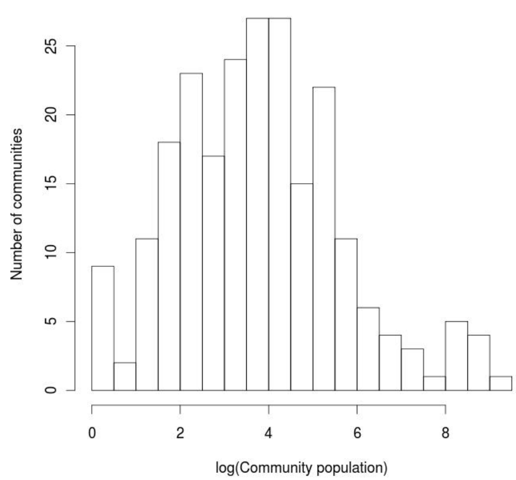

Figure 1: Distribution of community sizes

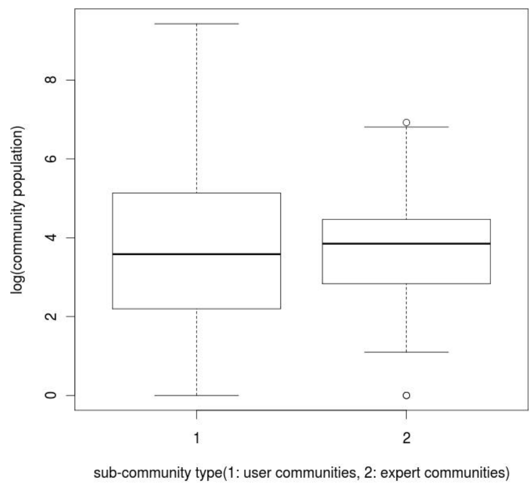

Figure 2: Size differences between user and expert groups

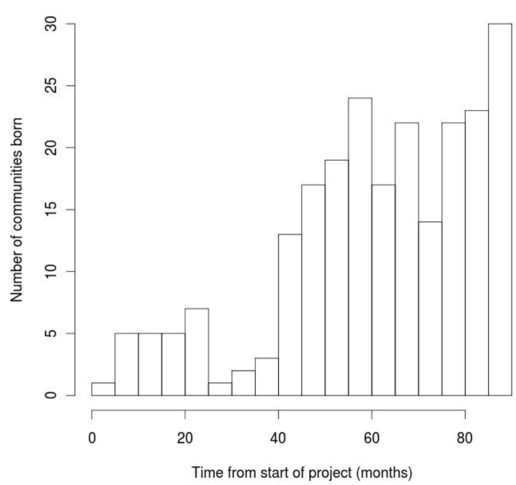

Figure 3: Number of sub-communities born through time in Eclipse community

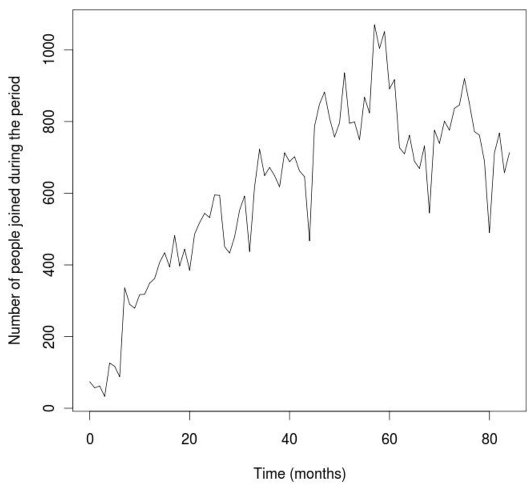

Figure 4: Number of people joining the Eclipse community through time

Figure 5: Interaction of sub-community births with population growth

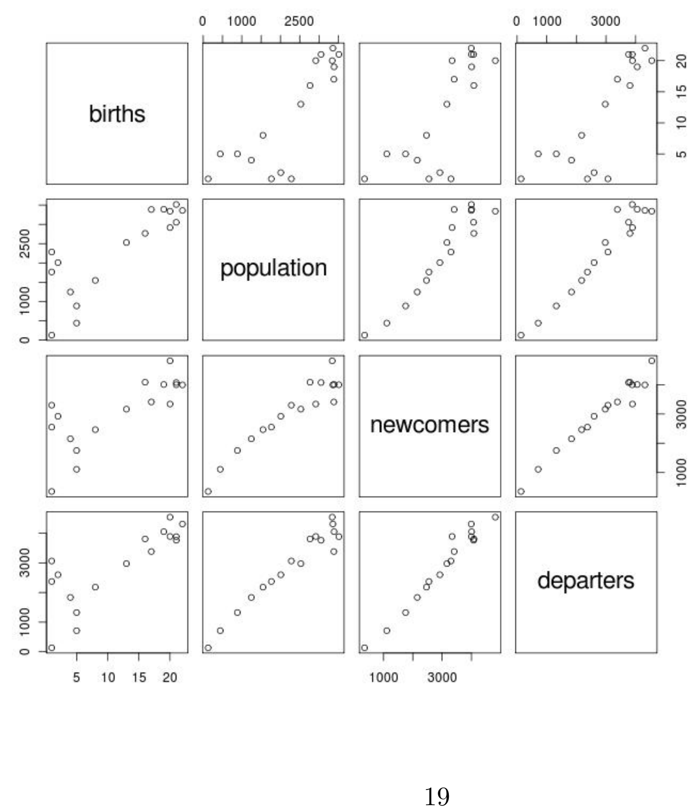

Table 3: Models for sub-community birth rate in relation to population growth.

Covariates	Model Coefficients Community births
Population	1.3683*
Newcomers	-0.6956
Departers	1.2782
Intercept	
	0.4626
Null deviance	
	115.955
Residual deviance	
	32.685

* $p < .05$

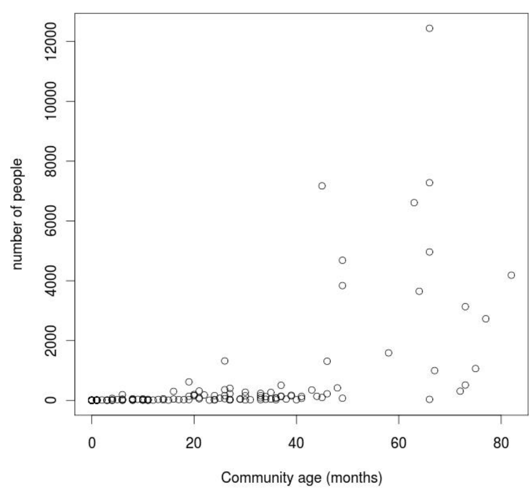

Figure 6: Sub-community population versus age for user communities.

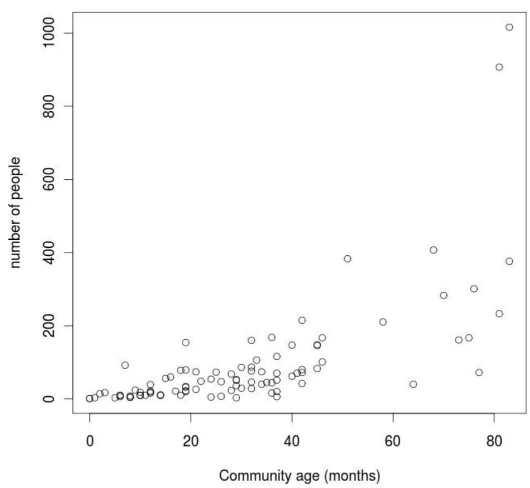

Figure 7: Sub-community population versus age for expert communities.

Figure 8: Population versus time of birth for user communities