# Strategic Interaction and Knowledge Sharing in Open Source Software Communities

George Kuk

Nottingham University Business School
University of Nottingham, United Kingdom
g.kuk@nottingham.ac.uk

## Abstract

This paper argues that open source software (OSS) developers often strategically select to work with other highly resourceful individuals to form a small but better organized network to manage a large but potentially chaotic OSS community. The small network arrangement affords individual developers to engage in a reciprocal interdependent relationship, which not only brings them closer to the centre of the OSS development but also expedites the releases and exchanges of individual knowledge resources. Based on separate analyses of two OSS development mailing lists of over 1500 messages, the present study indicates that the OSS developers that belonged to the network not only regularly contribute content and code but also intensively engage in knowledge reuse and combination. Implications of the present findings are discussed.

**Keywords:** strategic interaction, reciprocal interdependence, knowledge resourcefulness

**Suggested track**: Communities of practice, knowledge networks and networking

## Introduction

Research in the dynamics of Usenet postings suggests that online discussion is not for mass interactions where large numbers interact, but instead dominated by a few highly verbose individuals (Smith, 1999; Whittaker, Terveen, Hill & Cherny, 1998). This so-called participation inequality has led to the view that online communities will never become a true community for the masses and will offer limited utility values (e.g. Nielsen, 1997). The recent success of Open Source Software (OSS) specifically the process of OSS development has shifted this view even though most of the creation of content and codes is concentrated on a small network of developers. The fact that participation inequality does not seem to exert the same effect on OSS communities as it does on Usenet presents a gap in our understanding of the logic of being small, specifically how this rises to the challenges of the knowledge intensive programming work and the requirements of OSS development. This gap is particularly pertinent to the fact that most recent research in OSS has mainly focused on the private incentives underlying the individual collective actions, which are actions carried out individually by a large number of developers, because they share similar motivations, resources and face similar software needs (von Hippel & von Krogh, 2003).

Yet there is a paucity of research on the interplay between the structural characteristics of OSS communities (in addition to the virtual settings) and individual factors (in addition to shared motives and goals). The aim of this paper is to examine how individual developers behave and interact in a knowledge-sharing context within a social structure offered by the OSS communities. The specific questions include: How do the OSS communities differ from the Usenets in terms of posting behaviors, contents and structures? Specifically, what are the key structural characteristics and their relations to the interaction dynamics underpinning the OSS development? What criteria other than shared motives and goals are there to drive developers to join a specific developer's mailing list? How do developers enter and select other developers in forming a collaborative relationship? And finally, what are the implications of the interplay between structural and individual factors on the mobilization and releases of individual knowledge resources? The first two questions seek to identify the structural factors that characterize the OSS development. The next two questions examine the impact of individual factors on the developers’ joining and collaborative behaviors. The last question addresses how a small network of developers can sustain the OSS development.

The rest of the paper is organized as follows. I first review the nature of participation inequality and its wider impact on the utilization and consumption of personal and network resources; and then introduce a research model to examine the interplay between individual and structural factors and their relations to the formation of a small but better organized network that effectively mobilizes and releases the needed knowledge resources to support the OSS development.

## Participation Inequality

Recent research has suggested that regardless of whether interaction happens in cyberspace or physical collocation, large or small groups, collective interests seldom produce collective actions. In analyzing 500 newsgroups of over 2 million Usenet messages posted from just under 660,000 posters, Whittaker et al. (1998) found a very small number of frequent posters contributed to the majority of the messages, with the top 2.9% contributed an average 25% of the total posts in each newsgroup, and 27% of the postings were from singleton posters who only participated once. Similar results were also found in another large-scale study of Usenet postings (Smith, 1999). The findings reiterate the view that online discussion is not for mass interactions where large number of individuals interact, but instead dominated by a few highly verbose individuals.

The distribution of postings follows an exponential decay curve, which is a typical observation in small group interactions. Empirical evidence largely based on examination of speaking turns suggests that the distribution of participation is far from even, and participation inequality is the expected norm. In earlier work, Bales and his associates (Bales, 1950, 1953; Bales & Slater, 1955) observed that, even in equal-status and same gender groups, differentiation of participation quickly developed, with few members dominating the discussion. Stephan and Mishler (1952) observed a similar distribution in their groups. They found that the distribution of participation was not random but hierarchical, and the difference in participation rate between the most and the least talkative members increased as group size increased. This contradicts the view that participation inequality is a problem in all but the smallest groups (Olson, 1965; Oliver & Marwell, 1988).

The uneven distribution statistics underlie a key concern that most applied researchers have, that is participation inequality may dampen the intellectual rigor and the depth of discussions, and the incentives for learning and knowledge sharing (Ho, 2002). The recommended remedial strategy is by assigning someone to moderate interaction (Kuk, 2000), and to reinforce the Netiquette by weeding out any noncompliance behaviors and uninteresting materials (Nielsen, 1997). So far the remediation has mixed success, specifically the almost zero effect that online moderation has on the depth of threaded discussion (Smith, 1999; Whittaker et al., 1998).

Seemingly, participation inequality is symptomatic of a social dilemma. Olson (1965) attributed the underlying prime cause to the effect of free riding suggesting that everyone seeks to evade the cost of participation while benefiting from others' participation. He further claimed that to instigate collective action, selective incentives have to be provided to individuals as inducement for their contribution to the production of public goods. In case of OSS postings, because much of the underlying incentives for joining the OSS discussions can be attributed to private and collective incentives (e.g. elevating individual and group reputation), we could expect this to be sufficient in providing the right set of conditions that would favor more equalitarian participation amidst less of an occurrence of flame wars and/or postings of uninterested materials.

And yet participation inequality is still very common in OSS development. Table 1 summarizes the findings from some of the reported research. So in terms of the theorized impact of private incentives on participation, Olson's theory seems to be at odd with the observed phenomenon.

**Table 1**. Top contributors to OSS projects

| Study | OSS project | Contributions |
|---|---|---|
| Franke & von Hippel (2003) | Apache Usenet Forum | A total of 1371 postings by 563 different participants; the most prolific 1% contributed 20% of the postings, and the top 20% contributed 61% of the messages. |
| Ghosh & Prakash (2000) | FLOSS survey | Based on 1271 contributors, 10% of the total accounted for 72.3% of the total code base; the top 0.08% alone contributed 19.8% of the code base. |
| Koch & Schneider (2002) | GNOME | 52 programmers out of 301 contributed 80% of the code. |
| von Krogh et al. (2003) | Freenet | Participation in the development list concentrated on 4 individuals, 1.1% of the population accounted for 50% of the email list traffic. |

Most researchers go further to invalidate Olson's claim by suggesting that some of the underlying assumptions of Olson are unrealistic, specifically decisions by interest group members are independent (Marwell & Oliver, 1993). The next section presents a synthesis of the relevant and resource mobilization literature from multiple disciplines to explain the logic of participation inequality in OSS communities. The specific point to advance is that participation inequality is a structural characteristic of collective action afforded by strategic interaction, which relies upon reciprocal interdependence as a mechanism to enhance the mobilization and the releases of personal and network resources.

## Strategic Interaction and Reciprocal Interdependence

The concentration on a small subset of the population is an emergent property of collective action in the production of public goods. It addresses the start-up dilemma commonly encountered in most collective actions. That is, at the outset, a small number of highly motivated and resourceful individuals have to volitionally bear the start-up cost of collective action. As there are no immediate returns, these forerunners have to be cautious and strategic in the way they select projects and partners. Oliver and Marwell (2001, p.296) suggest that "in cases of high jointness of supply and heterogeneous groups, a collective good could actually be provided by a fewer contributors in a larger interest groups than in a smaller one, assuming that the two groups had the same distributional properties." They further advance the idea that individuals are highly strategic when it comes to the recruitment and selection of collaborators (Prahl et al., 1991). And far from being random, organizers (core developers and the elders in case of OSS communities) recruit and select only the most highly resourceful and motivated individuals in order to warrant future success. This claim is backed by research in social movements indicating that Olson's assertion of selective incentives as a prerequisite of collection action is only applicable to special situations, and non-participation can be instead attributed to the anticipatory fear of failure in collective action (Klandermans,1988; Oegema & Klandermans, 1994).

With the potentially unlimited supplies of talents, individual developers can afford to be strategic when it comes to the selection decision of which mailing lists to join and which collaborators to work with. To further enhance their chance of success, the interactions among developers have to be tightly coupled to an extent that there is reciprocal interdependence with the outputs of each developer become inputs for the others and vice versa (Markus, 1987).

Thompson (1967) suggests that reciprocal interdependence is the most complex among all other types of interdependence. The reciprocities, inter-linkages and exchanges among individual developers constitute the conversational interactivity embedded within threaded messages. Jones and his associates (2004) find that one of the dearest costs is in maintaining interaction within threaded discussions. It has been suggested that in contrast to a reply to a single message, individuals find it more cognitively taxing to reply to a threaded discussion, as their responses have to be coherently developed to take account of the ebb and flow of earlier postings, and more pedantic to compensate for the deficiencies in doing knowledge intensive programming in a lean rather than a rich communication medium (Knock, 2001). And also any disrupted turn adjacency, which is most likely in larger groups, may incur additional costs and extra effort for repairing the turn-taking organization. According to Brookes' Law, the cost of complexity and communication is proportional to the square of the number of developers while the collective output only increases linearly. A small group can therefore be viewed as a way of managing the complexity of software development, and importantly mitigating any potential loses in communication and coordination.

In relation to the OSS communities, strategic interaction has two inadvertent outcomes on the embedded structures therein. Firstly, it provides a selection mechanism in determining the formation and composition of a strategic network of interaction, and has inadvertently resulted in the concentration of code development on a small network of highly resourceful and highly motivated developers. Secondly, for such a small network to be viable and to mitigate any communication and coordination loses, it provides an interaction mechanism to bind members in a reciprocal interdependent relationship.

Figure 1 schematically summarizes the interplay between individual factors (knowledge resources and unique personal experience) and the network structural characteristic as a result of strategic interaction, and their relations on the network composition. Knowledge resources include the ability for knowledge reuse and combination, and network composition includes individual developers that drive and sustain the OSS development.

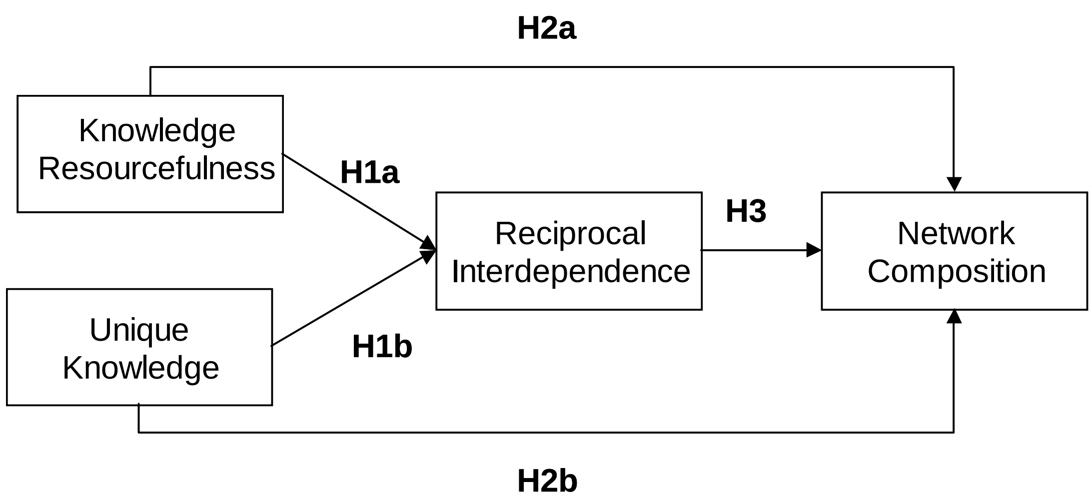


**Fig. 1**. Research model

First, individuals who not only possess the subject knowledge but also have the ability of integrating knowledge will most likely to benefit others in a reciprocal interdependent relation (e.g. Majchrzak, Cooper & Neece, 2004; Nissen, 2002). When it comes to the selection decision, resourceful individuals will enter into a collaborative relationship with similar others; and this will further enhance and develop the reciprocal interdependent relationship. In effect, reciprocal interdependence serves to strategically lock in developers to increase the mobilization and releases of individual knowledge resources. This has led to the following hypothesis:

*H1a. High knowledge resourcefulness will increase the level of reciprocal interdependence as a result of strategic interaction.*

Second, considering that most of the OS initiatives start with private incentives (von Hippel & von Krogh, 2003), unique personal acquired knowledge will be highly sorted specifically when it addresses other developers' software needs (Sacchi, 2004). One of the commonly reported benefits from the developers is the learning opportunity that participation in code development provides through peer review. Hence, individuals who have a lot of unique personal experience and/or knowledge to offer are most likely to appeal to similar others and to engage in a reciprocal interdependent relationship. This has led to the following hypothesis:

*H1b. High unique experience and knowledge will positively increase the level of reciprocal interdependence*

Third, recent research has suggested that the OSS communities are a status-conscious hierarchy whereby individuals have to earn their status and to follow special projects' joining scripts, which require participants to appease the existing community of developers by modestly contributing code and/or by offering feature gifts (von Krogh et al., 2003). Hence, regardless of whether individuals are highly resourceful or have a lot of unique experience and knowledge to contribute, as part of the joining process, individual developers will have to follow the joining script by first engaging in a reciprocal interdependent relationship with others before they earn a place in the final network. The process of joining underlies the mediator role of reciprocal interdependence between individual factors and network composition. The following two hypotheses are formulated.

*H2a. The effect of knowledge resourcefulness on network composition will be mediated by reciprocal interdependence*

*H2b. The effect of unique personal experience and knowledge on network composition will be mediated by reciprocal interdependence*

Prior research in reciprocal interdependence has shown its critical role in accumulating critical mass and in increasing network externalities for technology adoption (Markus, 1987). Here, in view of strategic interaction, reciprocal interdependence provides the needed mechanism for the formation of the network that in turn sustains the OSS development. The following hypothesis is formulated.

*H3. Reciprocal interdependence will exert a direct positive impact on the formation and composition of the network*

Prior to the hypothesis testing, three more empirical tests are conducted: to ascertain whether OSS communities exhibit participation inequality similar to Usenet; to examine strategic interaction by mapping out the reciprocal interdependence; and to evaluate the effectiveness of strategic interaction in mobilizing knowledge resources by examining the knowledge production within a network arrangement.

## Methodology

*Overview.* The present study was based on the separate analyses of two datasets, which were derived from two mailing lists used by the developers' community of the KDE desktop environment for Linux and other UNIX based architectures. KDE as the default desktop environment for most of the Linux distributions had received numerous awards and was one of the most successful OSS projects. According to www.kde.org, there were over 750 core developers who held concurrent versioning system (CVS) accounts for adding and/or changing code, artwork and translations. The first mailing list was the kde-devel available at: http://lists.kde.org/?l=kde-devel&r=1&w=2; and the second mailing list was the kde-core-devel available at: http://lists.kde.org/?l=kde-core-devel&r=1&w=2.

The first mailing list supported for discussions concerning the development of KDE. It had free read and write access, and was used by both developers and power users for testing unstable development versions. Whereas the second mailing list was restricted to the core developers only. Data were collected prior to the "freeze" period of the release schedule. The freeze period was when all the parts of the code for the new release were submitted and the only changes allowed were bug fixes. Through the freeze period the developers ensured that publicized new release had as few known bugs as possible.

In the process of data collection, I only selected discussion threads that were concerned with code development. For the first mailing list, out of 240 threads reviewed, 128 threads amounted to 867 messages were identified to provide the first data corpus. For the second mailing list, out of 164 discussion threads, 109 threads amounted to 628 messages were identified for the second data corpus. The average number of messages per thread was 6.8 and 5.8 respectively for dataset 1 and 2. All these were significantly higher than previously reported findings (see Smith, 1999; Whittaker et al., 1998).

## Measures

**Development of a coding classification**. Open coding methodology (Strauss & Corbin, 1990) was used to develop categories for classifying the message contents discussed in each thread. Saturation of concepts (i.e. no more concepts were identifiable) occurred approximately after the first 60 threads were analysed. Three specific types of knowledge sharing categories emerged. They were: reusing the existing public domain knowledge; revealing personal uniquely acquired experience and knowledge; and recombining new knowledge by integrating with the existing knowledge. In relation to the knowledge-centred literature, the three Rs (i.e. reuse, reveal and recombine) contributed to knowledge creation in the following ways: that the first R was paramount to avoid any wastage in duplicating previous effort and in reconstructing existing information and knowledge (Feldman, 2004); that the second R prevented only sharing and reinforcing what was already known to the community members but aimed to increase knowledge flow by continuously contributing fresh ideas and new information (Nonaka, 1994; Robertson et al., 1996; Szulanski, 2000); and that the third R allowed knowledge integration for propelling further knowledge reuse and learning (Newell et al., 2000; Markus, 2001).

To establish the reliability of the coding scheme, three coders were deployed, and the following procedure was adopted. A total of 50 messages were selected for training and another set of 50 messages for testing. The messages were chosen from an entirely different source and were not used in the present study. During training, coders were first provided with the instructions of the meaning and examples of the three knowledge categories and were then given unlimited time to do the coding. The results from the training phase were then discussed to enable the participants to better understand the coding system and to identify the underlying causes for disagreement. The discussion identified the main reason for disagreement was due to a misunderstanding of the highly technical context of the messages including terminology, jargon, and abbreviations. Thus for the testing phase a small list of abbreviations and terms was compiled and provided to the coders for reference. This resulted in a considerable improvement with an average agreement of 74%. The final coding of the two datasets was carried out by one of the coders.

*Knowledge resourcefulness*. I used the measures of knowledge reuse and combination to define knowledge resourcefulness as the capability of individual developers in terms of their subject knowledge and their ability to integrate existing knowledge to create new knowledge. As the measure depicted the frequency in which individual developers engaged in knowledge reuse and combination, the measure was logarithmically transformed.

*Unique experience and knowledge*. I used the knowledge category of revealing personal acquired experience and knowledge, which was also subject to logarithmic transformation.

*Reciprocal interdependence*. In the literature, there was no single established way in measuring reciprocal interdependence. Here I deployed the in-degree and out-degree measures (in terms of the number of in- and out-links incident to a specific developer) as a proxy of the reciprocal interdependence. Cyram Netminer (version 1.1.5) was used to generate the network measures based on a first order transition matrix, which essentially measured the sequential relationships embedded within threaded messages. For example, with a linear threaded discussion with messages in the order of “ABCAC”, the first order transition was as follows: first, A was followed by B, then B was followed by C, and C was followed by A and lastly A was followed by C. This information was then tallied in a matrix format to provide the first order transition matrix. The final reciprocal interdependent measure was a mean composite of in- and out-degree for each developer.

*Network composition*. Each developer was classified according to whether he or she belonged to a clique. To determine the size of the clique, the Lambda set approach was first carried out. The approach effectively ranked each network member according to the flow in terms of the number of links that pass through. The flow information allowed a further classification of developers into subsets to signify the level of disruption that it would cost to the flow among all other developers if disconnected (Wasserman & Faust, 1994). A plot of the flow against the number of members of each subset offered a form of “elbow test” similar to the scree test in factor analysis. Both plots for dataset 1 and 2 indicated 5 members as an optimal solution. I then used a clique of a grouping of 5 members to identify the final network composition. For dataset 1, eight cliques were identified and for dataset 2, 12 cliques were found; and in terms of the number of members, there were 14 and 16 respectively for the final networks 1 and 2. The final network composition could be viewed as overlapping networks of cliques.

*Status*. Developers who had CVS accounts were dummy coded to represent as their status as core developers. For dataset 1, 24.5% of the participants were classified as core developers whereas for dataset 2 as the mailing list was for core developers, all the participants were core developers. The status served here as a control variable in the testing of the hypotheses for dataset 1.

**Analysis**

The individual level was used as the unit of analysis. In total, there were 151 developers in the first data corpus, and 50 core developers in the second. I tested the hypotheses based on a series of regression analysis using SPSS release 12.0.1. For the testing of H2a and H2b, three regression equations were used to test the mediation effect. The first regressed the reciprocal interdependence (the proposed mediator) on knowledge resourcefulness and unique experience; the second regressed network composition on knowledge resourcefulness and unique experience; and the third regressed network composition capability on knowledge resourcefulness, unique experience and reciprocal interdependence. The three regressions in the form of y = a + b(x) were represented as follows:

*1$^{st}$ regression*: Reciprocal Interdependence = $a_1$ + $b_1$(knowledge resourcefulness) + $c_1$(unique experience)

*2$^{nd}$ regression*: Network = $a_2$ + $b_2$(knowledge resourcefulness) + $c_2$(unique experience)

*3$^{rd}$ regression*: Network = $a_3$+ $b_3$(knowledge resourcefulness) + $c_3$(unique experience) + $d_3$(reciprocal interdependence)

According to Barron and Kenny (1986), full mediation has to meet all the following criteria: that $b_1$ and/or $c_1$ is statistically significant in the 1$^{st}$ regression; that $b_2$ and/or $c_2$ has to be statistically significant in the 2$^{nd}$ regression; that $d_3$ in the 3$^{rd}$ regression has to be also statistically significant; and finally the level of significance of $b_3$ and/or $c_3$ has to be non-significant. For dataset 1, the variable status was also entered in the 2$^{nd}$ and 3$^{rd}$ regression equations (not included in the above equations) to control a possible confounding effect of status on the network composition, but not for dataset 2, as all the developers had the same “core developer” status.

## Results

To ascertain the presence of participation inequality, the production function of the present study (only dataset 1 was presented here) was first compared to that of Usenet postings (Whittaker et al., 1998) and estimates based on two other published OSS studies. Figure 2 displays the production function curves. The production curve of present study (the second from the left) was closely matched with Usenet (the third from the left). This indicates that the OSS postings were concentrated a small number of developers. The distribution of both postings followed a logarithmic function, with the OSS postings, $y = 20.74 \mathit{ln}(x) + 7.37$, $R^2 = 0.99$; and with the Usenet postings (Whittaker et al., 1998), $y = 20.67 \mathit{ln}(x) + 3.09$, $R^2 = 1$.

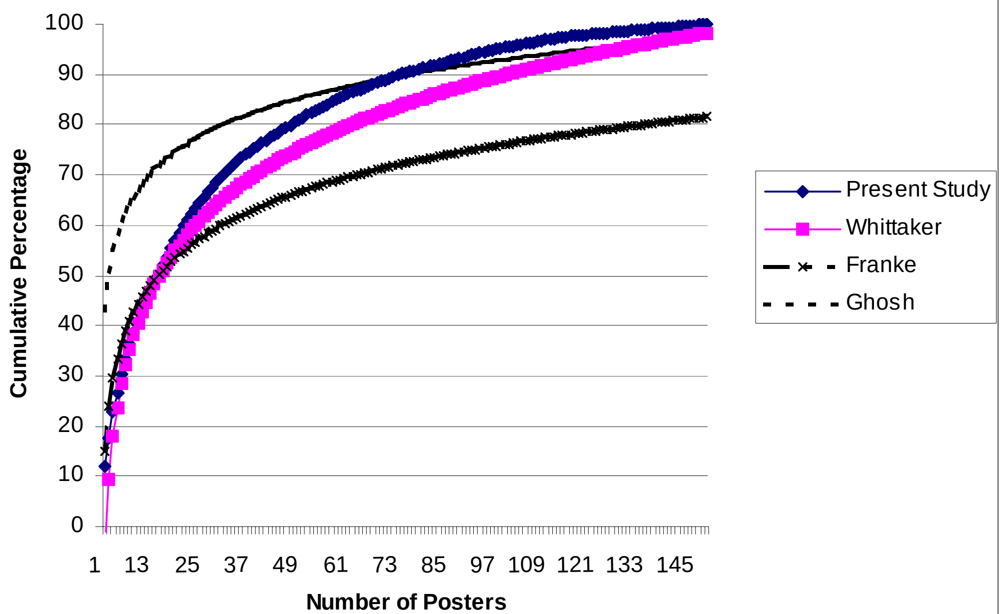


**Fig. 2.** Production Function Curves

For the first dataset, 60% of the postings were contributed by 16% of the developers (26 out of 151 developers). Among the top 16% contributors, 11 belonged to the network and the other 14 were outside the network. In comparing the production function curves between network and non-network members of the top contributors of the first dataset, Figure 3a clearly shows that top 11 contributors within the network obtained a higher level of resourcefulness than others outside the network. But in terms of unique experience (as shown in Figure 3b), the two production curves between network and non-network were similar. The trends (not shown here) of the second dataset were also similar to the first dataset. In sum, although not all network members were the top contributors, the developers within the network exhibited a higher level of resourcefulness than their counterparts outside the network. This suggests that simply contributing a lot did not guarantee a place in the final network, but developers who entered into the network have to not only contribute regularly but also invest their resources heavily in the network.

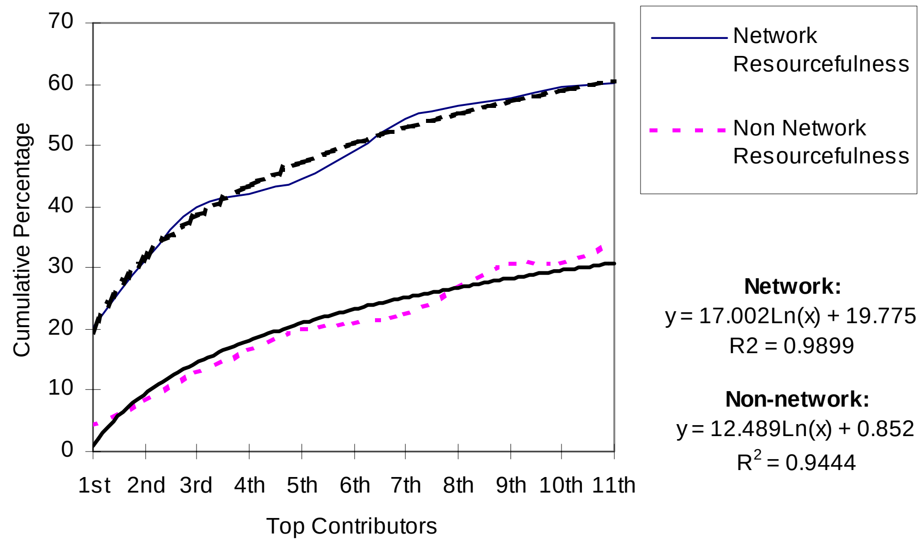


**Fig. 3a**. Resourcefulness (network vs non-network)

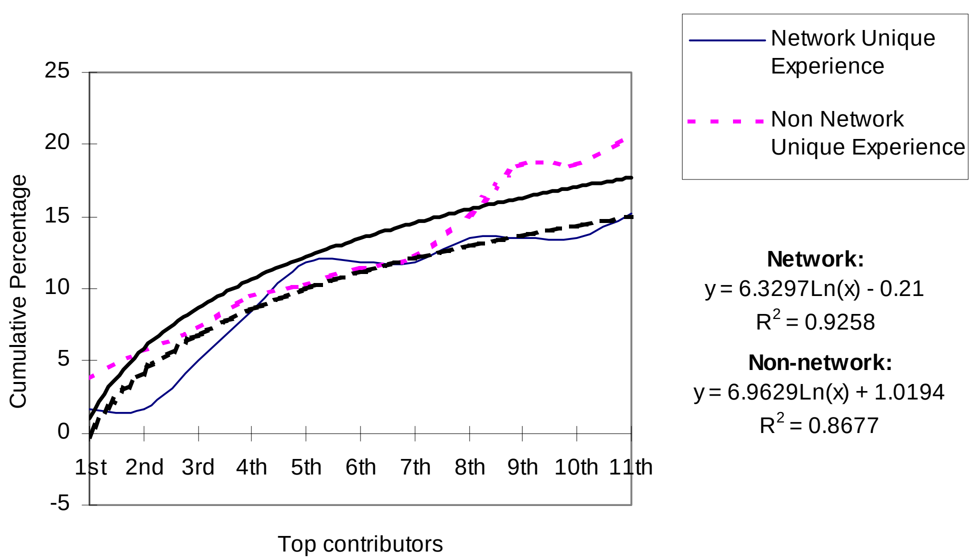


**Fig.3b**. Unique Experience (network vs non-network)

Figure 4 further displays the linkages among the 14 developers within the small network therein the first dataset. The display illustrates how reciprocal interdependence manifested itself within the network. Three specific key developers occupied the centre of reciprocal interdependence (indicated by the thicker links). They belong to the set of developers with the highest Lambda score. That is, most flow of interactions pass through the three individuals; their removal would cause the most disruption to the overall flow among all other developers.

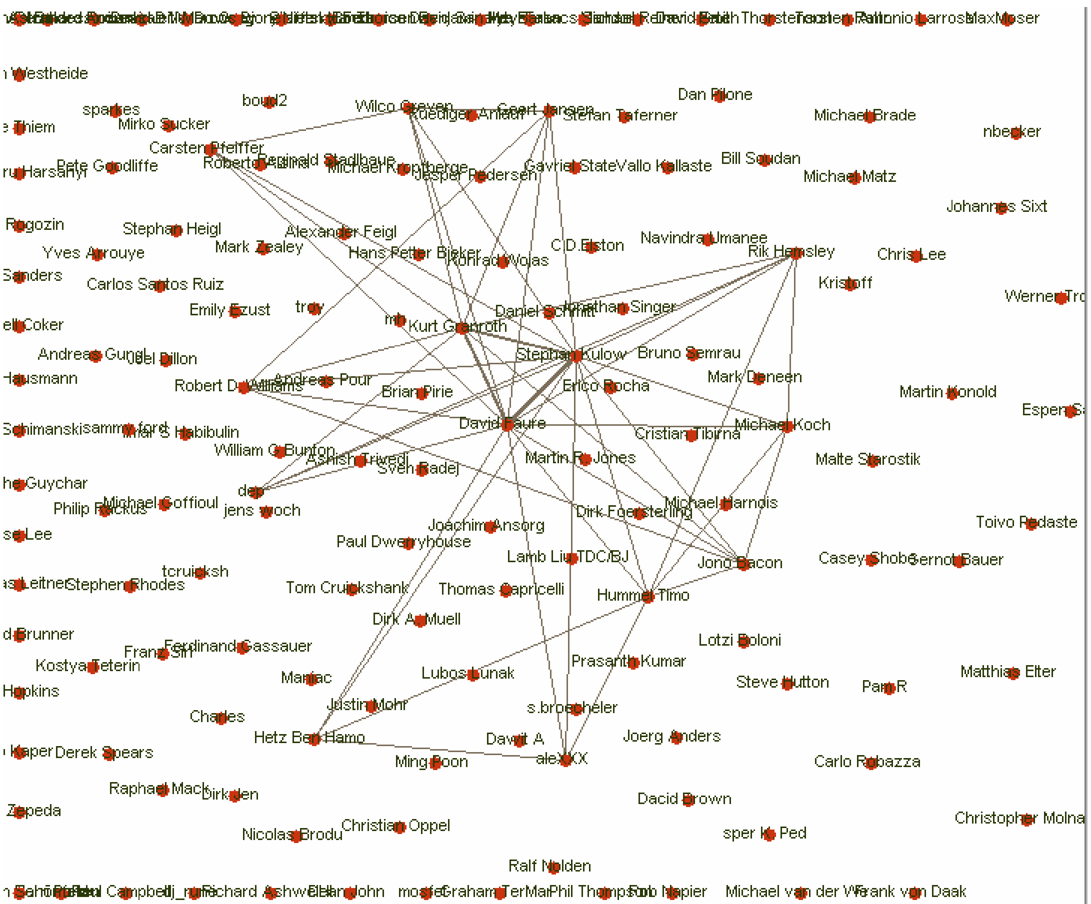


**Fig. 4.** A network illustration of reciprocal interdependence

## Hypothesis Testing

Table 2 shows the intercorrelations among all the testing variables. Almost all them were significant apart from the intercorrections of unique experience with network and reciprocal interdependence in the second dataset. Significant correlations ranged from 0.17 to 0.80 for the first dataset, and from 0.33 to 0.63 for the second dataset.

**Table 2** . Means, standard deviations and zero-order intercorrelations among variables

| | Mean | SD | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| *Dataset 1* | | | | | | | |
| *1. Network* | 0.09 | 0.29 | 1.00 | | | | |
| 2. Reciprocal Interdependence | 0.02 | 0.04 | 0.63 *** | 1.00 | | | |
| 3. Knowledge Resourcefulness | 0.53 | 0.86 | 0.58 *** | 0.80 *** | 1.00 | | |
| 4. Unique Experience | 0.07 | 0.26 | 0.39 *** | 0.64 *** | 0.40 *** | 1.00 | |
| 5. Status (Control Variable) | 0.25 | 0.43 | 0.35 *** | 0.41 *** | 0.53 *** | 0.17 * | 1.00 |
| *Dataset2* | | | | | | | |
| 1. Network | 0.36 | 0.48 | 1.00 | | | | |
| 2. Reciprocal Interdependence | 0.10 | 0.09 | 0.63 *** | 1.00 | | | |
| 3. Knowledge Resourcefulness | 1.01 | 1.00 | 0.52 *** | 0.48 *** | 1.00 | | |
| 4. Unique Experience | 0.25 | 0.49 | 0.18 ns | 0.25 ns | 0.33 * | 1.00 | |

*Note.* N1 = 151; N2 = 50; *** p < .001.

Table 3 displays the results of the first and second series of regression analyses. With the first dataset, all the criteria for full mediation were met, the results support all the hypotheses indicating that knowledge resourcefulness (H1a), unique experience (H1b) and reciprocal interdependence (H3) all positively influenced the network composition, and the effect of knowledge resourcefulness and unique experience were mediated by reciprocal interdependence (H2a and H2b). With the second dataset, the results support H1a, and H3, but only partially support H2a. The partial mediation suggests that resourceful individuals exerted a positive impact on the network composition, and there existed two routes for entering the network.

**Table 3.** Regression equations for mediation

| | **Regression Equations** | **b** | **c** | **d** | $R^2$ | ***F*** |
|---|---|---|---|---|---|---|
| *Dataset 1* | | | | | | |
| 1 | Reciprocal Interdependence = $a_1$ + $b_1$(Resourcefulness) + $c_1$(Unique Experience) | 0.64*** | 0.39*** | ——— | 0.76 | 235.722 *** |
| 2 | Network = $a_2$ + $b_2$(Resourcefulness) + $c_2$(Unique Experience) | 0.52*** | 0.19** | ——— | 0.37 | 28.44 *** |
| 3 | Network = $a_3$+ $b_3$(Resourcefulness) + $c_3$(Unique Experience) + $d_3$(Reciprocal Interdependence) | 0.18ns | 0.01ns | 0.45*** | 0.42 | 25.91 *** |
| *Dataset 2* | | | | | | |
| 1 | Reciprocal Interdependence = $a_1$ + $b_1$(Resourcefulness) + c1(Unique Experience) | 0.44*** | 0.11ns | ——— | 0.24 | 7.23 ** |
| 2 | Network = $a_2$ + $b_2$(Resourcefulness) + $c_2$(Unique Experience) | 0.52*** | 0.01ns | ——— | 0.27 | 8.67 *** |
| 3 | Network = $a_3$+ $b_3$(Resourcefulness) + $c_3$(Unique Experience) + $d_3$(Reciprocal Interdependence) | 0.30* | 0.05ns | 0.50*** | 0.46 | 13.06 *** |

*Note* . ns = non-significant; *** p < .001, ** p < .01, * p < .05

Finally, the overall model (as in Figure 1) in terms of the amount variance accounted by the testing variables is summarized in Table 4. For dataset 1, a total of 41% of the variance observed in the network composition were accounted for, and 46% for the second dataset. The overall fit of the model was considered to be extremely good according to Cohen (1977); a small effect is represented by $R^2$ value of 1%, a medium by value of 6%, and a large effect by a value of 15%.

**Table 4.** Results of regressing network composition on knowledge resourcefulness, unique experience and reciprocal interdependence

| | Dependent variable: Network | | |
|---|---|---|---|
| Predictors | *Beta* | *R2* | *F* |
| *Dataset 1* | | | |
| Status (control variable) | 0.07 | 0.42 | 27.91 *** |
| Knowledge Resourcefulness | 0.18 | | |
| Unique Experience | 0.10 | | |
| Reciprocal Interdependence | 0.45 *** | | |
| *Dataset 2* | | | |
| Status (control variable) | 0.30 * | 0.46 | 13.06 *** |
| Knowledge Resourcefulness | 0.05 | | |
| Unique Experience | 0.50 *** | | |
| Reciprocal Interdependence | | | |

*Note* . N1 = 151; N2 = 50; * p < .05, ** p < .01, *** p < .001.

## Discussion

This study was motivated by the two specific needs: to better understand why and how the OSS development can flourish based on a small network of a selected few despite its potentially unlimited access to the global R&D resources; and to better understand the dynamics by examining the interplay between the structural characteristics of such a small network and the underlying individual factors (including ability, knowledge and resources). The premise is that selective incentives focused on motives are incomplete to explain the observed dynamics in the OSS development, and that strategic intention provides an important mechanism through which the smaller but better organized groups manage the larger but potentially chaotic communities in the OSS development, and importantly serve to attract highly resourceful developers. The results of this study provide support for our premise specifically resourceful developers enter into a reciprocal interdependent relationship as a result of strategic interaction, and developers who establish a high level of reciprocity with others are mostly likely to be at the center of the OSS development.

Before examining the results of the study and their implications, some of its key limitations must be evaluated. The first limitation is only the KDE desktop environment was examined though a quantitative approach based on a single OSS project was not uncommon in this field of study. To compensate for the generalizability of the present findings, a second dataset was used to serve as a confirmatory test The second limitation concerns the measures, especially measuring the constructs such as knowledge resourcefulness and reciprocal interdependence, is challenging and further refining the measures would be a worthwhile endeavour. Nevertheless, considerable effort was expended to rely on reliable and objective rather than subjective measures, and the supportive results have provided the measures with an extremely high degree of predictive validity with an average of 44% of the amount of variance accounted for in the overall model.

The present study applies the notion of strategic interaction to elucidate the underlying rationale of why people choose to enter into some form of strategic partnership as a means to mitigate any wasted effort and to further enhance the chance of future success; and uses reciprocal interdependence to provide the underlying mechanism of not only binding individual developers but also mobilizing and releasing individual resources. In line with the above logics, the present findings reveal that resourceful individuals choose to interact with similar others through a structural relationship characterized by reciprocal interdependence. The concentration on a small network of developers is essential as a means to maximize the releases of individual resources; and the reciprocal interdependence effectively locks highly resourceful developers into knowledge sharing. This selectivity generally follows the findings that people often self-select into work situations consistent with their personality and goals structure (Schneider, 1987; Schneider, Goldstein & Smith, 1995). And in the context of a social structure offered by OSS communities, the selectively coupled with reciprocal interdependence requires additional input of individual knowledge resources and ability in order to make the whole OSS development viable.

Individual developers who successfully engaged with other developers in a reciprocal interdependent relation also earned their place in the final network. This is particularly true in situations when the status of being a core developer has to be earned. As indicated in the first dataset, individual developers have to regularly impart their unique experience and knowledge, and most notably to engage in knowledge reuse and combination before they become part of the final network. However, in situations where all the developers have proven their abilities by their acquired status as core developers, it is possible for direct entry to the final network without engaging and expending extraneous resources in a reciprocal interdependent relationship.

**Implications for Research**

In contrast to the question of why so many thousands of OSS developers contribute freely to OSS development, which has been the focus of much prior research, the present study examines the logic of why the main bulk of actual OSS development has been concentrated on a few OSS developers.

The phenomenon of participation inequality has been widely observed in the literature. Yet most research only acknowledges its existence without offering any theoretical explanation despite the fact that participation inequality seems to be at odds with theories that premise on selective incentives (Olson 1965). Klandermans (1988) is one of the main opponents of the selective incentives argument. He regards selective incentives as incomplete in accounting for social movement. Marwell and Oliver (1993) further suggest that selective incentives are applicable to special circumstances characterized by a decelerating production function, where free riding has its stronghold, but “free riding is not the problem in an accelerative case, unless all public goods dilemma are said to be free riding by definitional fiat” (p. 182). Because initially no one will perceive any gain from contribution, this presents a start-up dilemma to most collective action.

In relation to OSS, most scholars contend that because it has the backing of private investment and other collective incentives (e.g. von Hippel & von Krogh, 2003), the initial set of conditions encountered in OSS is fairly different. Yet selective incentives cannot explain away the phenomenon of participation inequality especially how OSS development can flourish with a small number of individuals. The main contention here is that selective incentives concern mainly with the underlying motives but miss out the “ability” factor, which is critically important when it comes to mobilizing human and social capital for effective knowledge sharing and problem solving (Alder & Kwon, 2002; Cross et al., 2002). Simply put, an OSS developer who has the right motives but lacks the ability does not add significant value to the creation of knowledge. The challenge therefore is in the recruitment and selection of not only the highly motivated but also the highly competent individuals amidst tens and thousands of OSS developers.

The motivation-ability argument works in tandem with the underlying intent of strategic interaction. Future research should develop the motivation-ability argument further specifically examining the combined effect of motivation and ability across a range of OSS projects. A few variants of our question include: Are there any differences in terms of knowledge resourcefulness, unique experience and reciprocal interdependence between successful and not so successful OSS projects? Are the characteristics of structural interaction more prominent in successful than unsuccessful OSS projects. What criteria do individual developers use to determine others' motives and abilities? Would these criteria change over time and /or differ across different stages of OSS development?

**Implications for Practice**

The findings reported here are by no means implicating that the use of smaller groups in place of larger groups, and that smaller groups should be detached from the wider community. The logic is not just being small but "being small, global and connected". This can be only affordable when the smaller groups can tap on the larger groups for high jointness of supply of talents and resources (Marwell & Oliver, 1993, 2001). Whether being small, global and connected presents a faster, better and cheaper approach to software development will undoubtedly require further research.

Whilst no physical mortal firms are able to match the scale of OSS development and to exploit the global R&D resources as well as the OSS communities, some successful Internet portals have proven the importance of having a dedicated group from the public regardless of how small it is to sustain their private initiatives (Damsgaard, 2002; Damsgaard et al., 2004). Also in spite of the fact that knowledge is one kind of public goods, the flow and growth of knowledge in OSS development are for the utilization and consumption of the power users, the core developers and so forth. Hence, whether other types of organization can effectively adopt the model of OSS at least has to consider the role and significance of whether the collective action carries any R&D values, and needless to say the challenge in the recruitment and selection of the best people. Firms therefore have to be cautious when they come to apply the present findings in view of the fact that strategic interaction similar to selective incentives might only work in specific circumstances. Nevertheless, the preliminary results of the present study are encouraging and should provide firms with some good guidelines regarding what and how firms can do to facilitate the releases of individual knowledge resources.

### Conclusion

This study makes significant contributions to the literature on the interplay between individual factors and the structural characteristics of OSS development as a result of strategic interaction, particularly given that many existing insights have been derived from the selective incentives argument. The underlying dynamics of strategic interaction through reciprocal interdependence highlights the greatest strategic asset accrued to OSS communities from the small groups of highly motivated and resourceful individuals, and hopefully provides stimulation for future research in this area.

## References


Adler, P. S., & Kwon, S. W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, *27*, 17-40.

Bales, R. F. (1950). A set of categories for the analysis of small group interaction. American Sociological Review, *15*, 257-263.

Bales, R. F. (1953). The equilibrium problem in small groups. In T. Parsons, R. F. Bales, E. A. Shils (Eds.), *Working Papers in the Theory of Action* (pp. 111-161). Glencoe, IL: Free Press.

Bales, R. F., & Slater, S. P. (1955). Role differentiation in small decision-making groups. In T. Parsons, P., & P. E. Slater (Eds.), *The Family, Socialization and Interaction Processes* (pp. 259-306). Glencoe, IL: Free Press.

Baron, R., & Kenney, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, *51*, 1173-1182.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. NY: Academic Press.

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*, 25-46.

Damsgaard, J. (2002). Managing an Internet portal. *Communications of the Association for Information Systems*, *9*, 408-420.

Damsgaard, J., Horsti, A., & Nilsson, O. (2004). Sustainable evolution of business models: Cases from Scandinavian Internet portal market, *Proceedings of ECIS Profession in the Global Networking Environment*. Turku, Finland.

Franke, N., & von Hippel, E. (2003). Satisfying heterogeneous user needs via innovation toolkits: The case of Apache security software. *Research Policy*, *32*, 1199-1215.

Ghosh, R., & Prakash, V. V. (2000). The Orbiten free software survey. First Monday. Available: http://www.firstmonday.dk/issues/issue5_7/ghosh/ [7 Sept 2003].

Ho, S. (2002). Encouraging online participation. Proceedings of the 11th Annual Teaching Learning Forum, 5-6 February 2002. Perth, Edith Cowan University; available at http://www.ecu.edu.au/conferences/tlf/2002/pub/docs/Ho.pdf

Jones, G., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, *15*, 194-210.

Klandermans, B. (1988). Union action and the free-rider dilemma. In L. Kriesberg, B. Misztal (Eds.), *Research in Social Movements, Conflict and Change*, vol 10, *Social Movements as a Factor of Change in the Contemporary World*. Greenwich: JAI Press.

Knock, N. (2002). Compensatory adaptation to a lean medium: An action research investigation of electronic communication in process involvement groups. *IEEE Transactions on Professional Communication*, *44*, 267-285.

Koch, S., & Schneider, G. (2002). Effort, cooperation and coordination in an open source software project: GNOME. *Information Systems Journal*, *12*, 27-42.

Kuk, G. (2000). When to speak again: self-regulation under facilitation. *Group Dynamics: Theory, Research and Practice*, *4*, 291-306.

Leifer, R., & Delbecq, A. (1978). Organizational/environmental interchange: A model of boundary spanning activity. *Academy of Management Review*, *3*, 40-50.

Majchrzak, A., Cooper, L. P., & Neece, O. E. (2004). Knowledge reuse for innovation. *Management Science*, *50*, 174-188.

Markus, M. L. (1987) Toward a "critical mass" theory of interactive media: Universal access, interdependence, and diffusion. *Communication Research*, *14*, 491-511.

Markus, M. L. (2001). Toward a theory of knowledge reuse: Situations and factors in reuse success. *Journal of Management Information Systems*, *18*, 57-93.

Marwell, G., & Oliver, P. (1993). *The Critical Mass in Collective Action*. Cambridge: University Press.

Newell, S., Swan, J. &Galliers, R. (2000). A knowledge-focused perspective on the diffusion and adoption of complex information technologies: The BPR example. *Information Systems Journal*, *10*, 239-259.

Nielsen, J. (1997). Community is dead; long live mega-collaboration. Available: http://www.useit.com/alertbox/9708b.html

Nissen, M. E. (2002). An extended model of knowledge flow dynamics. *Communications of the Association of Information Systems*, *8*, 251-266.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5, 14-37.

Oegema, D., & Klandermans, B. (1994). Why social movement sympathizers don’t participate: Erosion and nonconversion of support. *American Sociological Review*, *59*, 703-722.

Oliver, P. E., & Marwell, G. (1988). The paradox of group size in collective action: A theory of the critical mass, II. *American Sociological Review*, *53*, 1-8.

Oliver, P. E., & Marwell, G. (2001). Whatever happened to critical mass theory? A retrospective and assessment. *Sociological Theory*, *19*, 292-311.

Olson, M., Jr. (1965). *The Logic of Collective Action: Public Goods and the Theory of Goods*. Cambridge, MA: Harvard University Press,.

Prahl, R., Marwell, G., & Oliver, P. E. (1991). Reach and selectivity as strategies of recruitment for collective action: A theory of the critical mass, V. *Journal of Mathematical Sociology*, *16*, 137-164.

Robertson, M., Swan, J., & Newell, S. (1996). The Role of networks in the diffusion of technological innovation. *Journal of Management Studies*, 33, 335-361.

Scacchi, W. (2004). Free and open source development practices in the game community. *IEEE Software*, 59-66.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40,* 437-454.

Schneider, B., Goldstein, H. W., & Smith, D. B.. (1995). The ASA framework: an update. *Personnel Psychology*, *48*, 747-773.

Smith, M. (1999). Invisible crowds in cyberspace: Measuring and mapping the social structure of USENET. In M. Smith, P. Kollock, (Eds.), *Communities in Cyberspace*. London: Routledge Press.

SPSS for Windows. Release 11.0.1. SPSS Inc.

Stephan, F. F., & Mishler, E. G. (1952). The distribution of participation in small groups: An exponential approximation. *American Sociological Review*, *17*, 598-608.

Strauss, A. K., & Corbin, J. (1990). *Basics of Qualitative 683 Research: Grounded Theory Procedures and Techniques*. CA, Newbury Park: Sage Publications.

Szulanski, G. (2000). The Process of knowledge transfer: A diachronic analysis of stickiness. *Organizational Behavior and Human Decision Processe*s, 82, 9-27.

Thompson, J.D. (1967). *Organizations in Action: Social Science Bases of Administrative Theory*. NY: McGraw Hill.

von Hippel, E., & von Krogh, G. (2003). Open source software and the private-collective innovative model. *Organizational Science*, *12*, 209-223.

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

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Whittaker, S., Terveen, L., Hill, W., & Cherny L. (1998). The dynamics of mass interaction. *Proceedings of Conference on Computer Supported Cooperative Work*. NY: ACM Press.