“Network tie structure causing OSS group innovation and growth”

AUTHORS	Stefan Kambiz Behfar Ekaterina Turkina Thierry Burger-Helmchen

ARTICLE INFO	Stefan Kambiz Behfar, Ekaterina Turkina and Thierry Burger-Helmchen (2017). Network tie structure causing OSS group innovation and growth. *Problems and Perspectives in Management*, *15*(1), 7-18. doi:10.21511/ppm.15(1).2017.01
DOI	http://dx.doi.org/10.21511/ppm.15(1).2017.01
RELEASED ON	Tuesday, 28 March 2017

LICENSE	This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

JOURNAL	"Problems and Perspectives in Management"
ISSN PRINT	1727-7051
ISSN ONLINE	1810-5467
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
34	7	6

© The author(s) 2018. This publication is an open access article.

Stefan Kambiz Behfar (France), Ekaterina Turkina (Canada), Thierry Burger-Helmchen (France)

Network tie structure causing OSS group innovation and growth

Abstract

Open source software (OSS) development as an inexpensive process to develop software threatens proprietary software business strategies. Providing business strategy to benefit from volunteer developers for the purpose of contributing to existing projects, as well as initiating new OSS projects is of utmost significance for companies in that industry. Therefore, it is important to figure out how groups of volunteer developers are formed as new developers join existing projects, and it is even more important to investigate what causes these developers to initiate new projects. The authors investigate network structure as a causal factor for both new project initiation within a group (representing group innovation) as well as new developers joining existing projects within a group (representing group growth). The authors develop four hypotheses:

4. Intra-group coupling has a positive impact on group growth,
5. Inter-group coupling has a positive impact on group innovation,
6. Inter-group structural hole has a positive impact on group innovation,
7. There is a trade-off between the effects of inter-group structural hole and inter-group coupling on group innovation.

The authors test these four hypotheses using data from OSS. Developers contributing to project tasks in groups other than their own can explore novel ideas for new project creation, because they can benefit from sharing knowledge, whereas developers contributing to project tasks inside their own group *exploit* ideas to improve those existing projects with better inside-group search possibility; and this demands more developers to join those group projects.

Keywords: open source software, cluster management, network management.
JEL Classification: D85, L14, O31.

Introduction

Open source software (OSS) project collaboration has been analyzed from various perspectives within different disciplines from computer science to business and economics, as well as multidisciplinary network theory. This collaboration constitutes the means of producing goods and services by self-organizing groups within worldwide networks, and represents a form of partnership between businesses and customers.

While there were skeptics over the quality of OSS products, and software industry was struggling to find innovative methods of developing quality software products, Linux and the Apache server achieved a big success, which led to the potential of new approach to produce reliable and high quality products that are also produced inexpensively (von Hippel, 2001). Due to these advantages, they claim that OSS development has the potential to compete with traditionally produced software, and even replace traditional development methods (Mockus et al., 2002). Software developers are now facing new labor market, where participation in OSS projects could lead to increased salaries and improved job security. Three forms of competitive advantage have emerged: verifiable technical skills, peer-certified competencies and positional power, as stated by Riehle (2015).

Researchers have widely used social network theories to investigate the OSS phenomenon. They showed that the positions and relationships among developers in a social network are significant in the efficiency of the network (Jackson and Wolinsky, 1996; Jackson, 2004) using different techniques and tools such as social network analysis (SNA). Success of many OSS projects is closely related with the communication structure (Grewal et al., 2006; Singh et al., 2011). One distinguished feature of the OSS development model is the cooperation and collaboration among the members, which will cause various social networks to emerge (Grewal et al., 2006). To some extent, the OSS community is a more networked world than the traditional organizational communities, where programmers can join, participate, and leave a project at any time and developers collaborate not only within the same project team, but also across teams. It has also been shown that the structure of an interproject network affects knowledge sharing within and across open source projects. Montazemi et al. (2008) demonstrated that the market structure of embedded interpersonal ties enables participants to take advantage of information asymmetry for profit taking (Singh et al., 2011).

Hinds and Lee (2008) discussed costs and benefits of community ties, and concluded that *social network structure of open source software has no important effect on community structure*. On the other hand,

© Stefan Kambiz Behfar, Ekaterina Turkina, Thierry Burger-Helmchen, 2017.
Stefan Kambiz Behfar, Researcher, Faculty of Management and Economic Science, University of Strasbourg, France.
Ekaterina Turkina, Associate Professor, Department of International Business, HEC Montreal, Canada.
Thierry Burger-Helmchen, Professor of Management, University of Strasbourg, France.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International license, which permits re-use, distribution, and reproduction, provided the materials aren't used for commercial purposes and the original work is properly cited.

Antwerp and Madey (2010) investigated social network structure of open source software, and used long term popularity as the metric developer-developer tie and concluded that *previous ties are generally an indicator of past success and usually lead to future success*. Hahn et al. (2008) also investigated the *personal factors* causing a new developer to join a project, and concluded that prior collaboration between a new developer and the project initiator or previous experience of the group members are determining factors. Rather than discussing which *personal factors* make developers initiate new projects, in this study, we focus on *network structural factors* that influence developers to join existing projects or initiate new projects within a group, as shown grey in Table 1.

Table 1. Factors influencing developers to join projects or initiate new projects

Factors	*Developers join existing projects in a group*	*Developers initiate new projects in a group*
Personal	Prior collaboration of new developer with project initiator and prior experience of actual project members	
Structural	Intra-group coupling and structural hole have positive effect on new developers joining projects inside a group	Inter-group coupling and inter-structural hole have positive effect on developers initiating new projects.

Innovation results from interactions among different bodies or sources of knowledge, where these sources of knowledge aggregate into groups interacting within (intra) and between (inter) groups. In information science, groups could be defined as the sum of developers working on related projects. Intra- and inter-group coupling have been investigated in the literature within sociological systems in terms of tie strength by Granovetter (1973), in social and biological systems (Newman, 2004) represented by community structure, in organizational systems (Simon, 1962; Weick, 1976) by loose versus tight couplings. In addition, various authors have investigated the impact on innovation by tie strength (weak versus strong) (Granovetter, 1973; Hansen, 1999), and by network structure (sparse versus dense) (Burt, 1992; Walker et al., 1997). At the same time, there is ambiguity and conflicting theories linking network and innovation. Ahuja (2000) investigated the impact of direct and indirect ties on firm innovation, and reported that a) "the more direct ties a firm maintains, the greater the firm's subsequent innovation output", b) "the greater a firm's number of indirect ties, the greater the subsequent innovation output of the firm", c) "the impact of indirect ties on a firm's innovation output will be moderated by the level of the firm's direct ties".

There is also ambiguity in the benefit to networks from structural holes, where innovation generation is moderated by type of innovations and type of firms. For some types of new technology diffusion, trust and cooperation between firms is required, which demands fewer structural holes, whereas for firms where brokerage of information is the primary business more structural holes are required (see Burt, 1992; Ahuja, 2000).

Tedeschi et al. (2014) studied the dynamic of innovation networks, where they introduced an agent-based model, where heterogeneous firms compare and modify their innovation strategies. Kogut (2000) proposed that part of the value of a firm comes from its participation in a network.

Lastly, there are conflicting explanations concerning the impact of sparse and dense network structure for the purpose of innovation. Walker et al. (1997) and Coleman (1988) stressed that dense network structure has a positive impact on the implementation of idea within each group, and argued that strong ties are required for exchange of complex knowledge, whereas Burt (2000, 2002) emphasized that a sparse network structure facilitates diffusion of ideas and argued that strong ties within dense network are inefficient for acquiring external knowledge, as they do not promote diversity in resources.

In this study, we place our major contributions within the afore mentioned literature gap, to the best of our knowledge, there has been no study investigating the managerial and economic *impact of network group structure on group innovation*. We focus on network group rather than individual for both network structure as input and innovation and growth as output, because a) group represents the collective impact on network output rather than the individuals' impact, b) intra- and inter-group couplings both represent group structure, but impact differently on group innovation or growth, c) trade-offs among dense and sparse network cluster structures are different from those associated with networks of individuals. Moreover, we focus on network structural factors, and attempt to apply the concept of "*the impact of network structure on innovation*" from organizational science to information system. We make two assumptions: 1) new project initiation within a group represents group innovation and 2) new developers joining existing projects represent group growth.

The paper is structured as follows: in the first section, we present our theoretical framework and hypotheses. We review network structural perspectives on innovation output from the two network structure aspects: intra- and inter-group couplings, and structural hole. This section is followed by the method section, where we will discuss data collection and measures. In the next section, we provide empirical analysis to validate the three hypotheses. We discuss data collection, and propose the method, which includes data collection,

measurement and results. Our analysis is based on the data collected from the website of SourceForge.net, which is the largest repository of OSS projects.

1. Theoretical framework and hypotheses

1.1. Network group structure. As discussed by Burt (2000), groups are inter-connected via both strong and weak ties, where weak ties are far more numerous. Groups are also intra-connected via both strong and weak ties, where strong ties are far more numerous, while intergroup coupling is used between groups. Inter-group coupling should not be confused with tie strength (weak-strong) between network nodes which accounts for frequency of developer contribution in project tasks, as shown in Figure 1.

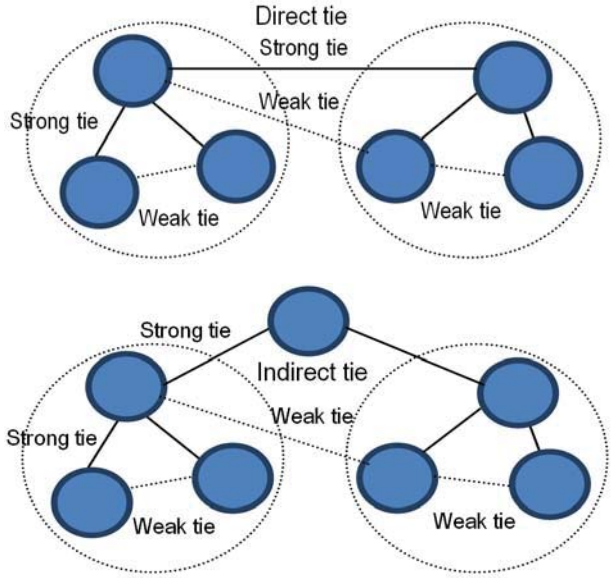

Fig. 1. Illustration of both weak and strong ties within and between two groups

We use the word "coupling" between groups, which is different from concept of tie strength (weak-strong) between network nodes. Tie strength is, in fact, frequency of developer contribution to project tasks, as shown in

Table . We do not measure ties by their weight, rather, a developer contributing to one project task within a group or among different groups forms one network tie.

Table 2. Terminology

Term	Definitions	Measure
Network tie	Weak/strong tie	Frequency of developer contribution in project tasks
Network structure		
	Dense	Densely intraconnected groups, where developers work on relevant project tasks
	Sparse	Sparsely interconnected groups, where developers work on irrelevant project tasks
Intergroup coupling		Sum of intergroup ties (sum of intergroup project tasks)
OSS group	$Group_{id}$	Assigned by source forge administration for any new project; moreover, new members/developers are added by the group administrator based on relevancy and of course his or her interest
OSS project	$Project_{id}$,	Anyone can initiate a project by signing up at source forge, or alternatively create a subproject, in case he or she is already member of a project and a group
Project task	$Task_{id}$	Project related tasks, where developers within a group or across groups work on (here it is considered as network tie)

We use the word "coupling" between clusters ranging from loose to tight (Simon, 1962). After a description of network group structure, we present what the complex network components *node and tie* are. In our network of OSS project collaboration, each developer represents a node, whereas two developers contributing to the same project task represent a tie. We still need to define the constructs: *intra-group coupling* measured by the number of project tasks in each group, *inter-group coupling* measured by the number of project tasks between any two groups, and *inter-group structural hole* measured by the number of opportunities contributed by project tasks between any two groups. The structural hole concept (relationship of non-redundancy between two contacts) was initially introduced by Burt (1992), and implies a brokerage opportunity (creating competitive advantage for an individual whose relationships span the holes). In fact, structural holes shown in Fig. 2 are gaps in information flow between alter linked to the same actor, but not linked to each other (Ahuja, 2000).

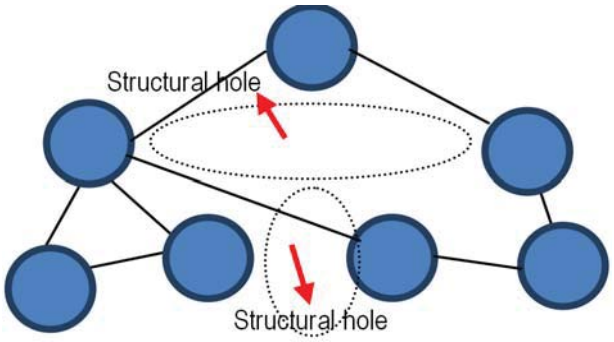

Fig. 2. Illustration of structural holes

Group innovation is defined as new project initiation within each group, whereas group growth is defined by the number of new developers joining existing groups. We use social network dynamics to *explain and predict* our phenomena of interest "OSS group innovation".

The theory components are: the unit of analysis is the group of OSS developers, where the network is composed of nodes (developers) linking by project tasks. Each developer can initiate new projects, but, at the same time, co-work on project tasks with other developers. Network outputs are group growth and innova-

tion, where group innovation is measured by the number of new projects created within each group and group growth is measured by the number of new developers joining existing projects within each group. There are some conceptual and contextual assumptions underlying our proposed theory:

- Innovation usually results from interactions among different sources of knowledge. Here we assume that new project initiation representing innovation is solely caused by these interactions; however, sometimes it could happen due to different reasons such as an OSS project being large in size (e.g., Apache subdivides into some smaller projects).
- New innovative projects contain either just one initiator or additional members, where we assume that the initiator has been influenced by prior intra- or inter-group interactions.
- Studies of the factors that cause a new developer to join a project (Hahn et al., 2008) conclude that prior collaboration between a new developer and a project initiator and the experience of actual project members cause a new developer to join the project. Here, the reputation of developers (represented by the number of projects he or she has initiated), popularity of project tasks (shown by the number of developers contributing to the project tasks) and other factors (such as the number of project tasks that one developer contributes to). Although they might influence the developer's decision to initiate a new project or join an existing project, here we consider them as control variables.
- We use the word "coupling" within and between groups, which should not be confused with tie strength (weak-strong) between network nodes. Tie strength is, in fact, frequency of developer contribution to project tasks, used when measuring intra- and inter-group coupling, however, we do not include frequency of developer contribution in our modeling.

1.2. Impact of group coupling on innovation output. Management and economics literature provides different perspectives on clusters in different contexts such as knowledge sharing, governance and transaction cost economics (Williamson, 1975; Lefebvre et al., 2016), network as solution to exploration (discovery, development of idea) and exploitation (implementation of idea) dilemma (Stadler et al., 2014). In the context of information systems within the domain of open source software (OSS) project collaboration network, similar to the concept of cluster in an organizational science, projects are grouped based on their characteristics. Developers could contribute to one or several project tasks within a group or among different groups. Bridging between groups allows exploration or access to novel ideas, whereas dense groups promote exploitation of those ideas. Therefore, developers working on project tasks across different groups can access novel ideas for new project creation, whereas developers working on projects inside one group exploit those ideas to improve those existing projects; and this requires more developers to join the group.

Inter-group coupling leads to both group innovation and growth, but with greater impact on group innovation. This is because inter-group ties are more efficient for acquiring external knowledge, accessing the diversity in projects in other groups, and facilitating diffusion of new project ideas, which leads to new project initiation inside the group (so-called group innovation). On the other hand, intra-group coupling leads to both group innovation and growth, but with greater impact on group growth. This is because intra-group ties are more efficient for quick transfer of information via group factors ($group_{id}$), which leads to group growth (Tsai, 2000, 2001).

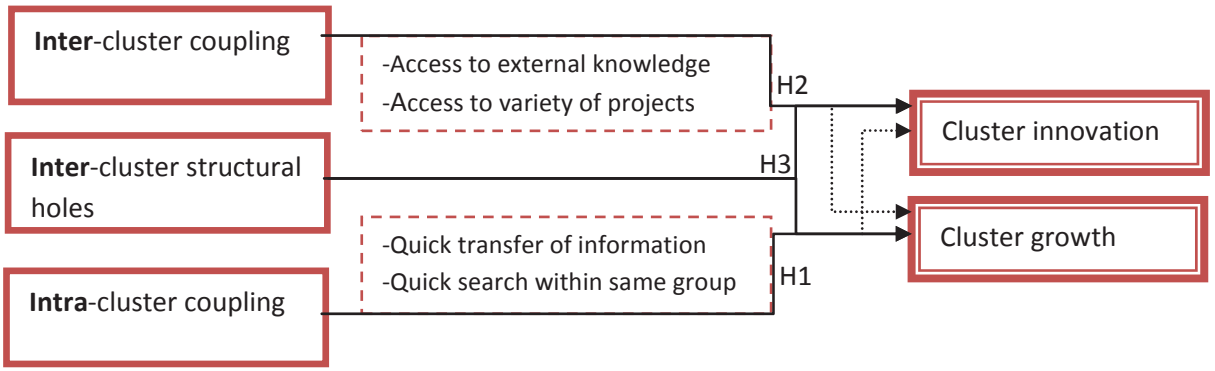

Fig. 3. Illustration of theory design on the impact of the three constructs (intra-group coupling and inter-group coupling, as well as inter-group structural holes) on group growth and innovation

We use three constructs "intra-group coupling", "inter-group coupling" and "inter-group structural hole" shown in the model diagram in Fig. 3. First, we intend to investigate the impact of intra/inter coupling on group growth and, therefore, answer the question "Does intra/inter group coupling have a

positive impact OSS project group innovation?" If yes, is it due to quicker search and transfer of information and better accessibility inside group?" As will be discussed later in the data section, each project initiated by a developer is given a $group_{id}$, which contains both projects and developers. In fact, $group_{id}$ benefits developers allowing them to search related projects faster, as well as benefiting other developers working on similar project tasks. In this way, developers within each group have quicker transfer of information and contribute to the same project tasks. This helps to improve those existing projects, which attracts more developers to join the group, but of course this does not reject the possibility of new project creation within the group. Therefore, we propose the following hypothesis:

Hypothesis 1 (H1): Intra-group coupling has a positive impact on group growth.

Second, we intend to investigate the impact of intra/inter coupling on group innovation. Therefore, we answer the question "Does intra/inter group coupling have an impact on OSS project group innovation?" As mentioned earlier, developers can explore a variety of projects in other groups by contributing to the same project tasks as other group members (inter-coupling). This leads to access to other various projects, and this facilitates diffusion of ideas between the two groups, which leads to new project creation. Of course this does not reject the possibility of new developers joining existing projects within the group. Therefore, we propose the following hypothesis:

Hypothesis 2 (H2): Inter-group coupling has positive impact on group innovation.

There is a trade-off between the effects of sparse and dense network structures on innovation. As mentioned above, Ahuja (2000) investigated the effect of structural holes on firm innovation, and reported a trade-off between dense and sparse network structures. Inter-group structural holes are defined as the number of opportunities for developers to contribute between two connected groups. This leads to a positive impact to group innovation; however, it is predicted that similar to Ahuja's conclusion on a trade-off between dense and sparse network structures, there is a trade-off between impact of inter-group coupling and inter-group structural holes on group innovation, as more inter-group coupling means more communication channels and, therefore, fewer opportunities for developer contribution. Therefore, we propose the hypothesis that:

Hypothesis 3 (H3): Inter-group structural hole has a positive impact on group innovation.

Hypothesis 4 (H4): There is a trade-off between the impacts of inter-group structural hole and inter-group coupling on group innovation.

2. Method

We aim to determine separately the impacts of intra-group coupling, inter-group coupling and inter-group structural hole on group innovation in the domain of OSS projects. We use the complex network of open source software (OSS) as the domain of interest for this purpose, and collect OSS project collaboration data, as will be explained in the data subsection. We, then, use regression method and define dependent, independent and control variables, as will be explained in the measurement subsection.

2.1 Data. We collected the data from the website of SourceForge.net, which is the largest repository of OSS projects. Crowston et al. (2004) indicated that at the time it contained more than 150,000 projects and more than 1,600,000 project developers. SourceForge currently hosts over 430,000 open source software projects, which are categorized into several categories such as Audio and Video, Business and Enterprise, Communications, Development, Home and Education, Games, Graphics, Science and Engineering, Security and Utilities, System Administration. SourceForge gives $group_{id}$ as an identifier for each project. In fact, SourceForge administration assigns id for any new project; moreover, new members/developers are added by the group administrator based on relevancy and of course his or her interest. As a project administrator, one can control over who is a member of his or her project and what they can do. We downloaded data ($group_{id}$, $task_{id}$, $project_{id}$ and $user_{id}$) for January 2013 and January 2014 from SourceForge repository based on multidimensional table. Downloaded data include 10.000 users.

1. They are in random, because those users belong to random projects or group.
2. Any additional group user is added to the group by the administrator.
3. Each user selects what project to initiate, or decide which project task to contribute to.

In order to find out the relationship between the fields: $group_{id}$, $task_{id}$, $project_{id}$ and $user_{id}$, as seen in Figure 4, we organize the graphs based on differences of *shared users, shared projects and shared tasks*, where $group_{id}$ is represented by g, $user_{id}$ by u, $project_{id}$ by p, and $task_{id}$ by t.

As shown in Figure 4.a (*non-zero values for shared users w.r.t. $group_{id}$*), one user (indicated by u1=u2) can contribute to the same task (shown by t1=t2) and different projects (denoted by p1≠p2), intra-groups (shown by g1=g2) or inter-groups (shown by g1≠g2).

Figure 4.b (g1=g2, u1=u2, t1≠t2, p1≠p2) implies that one user can contribute to different tasks, as well as different projects intra or inter groups.

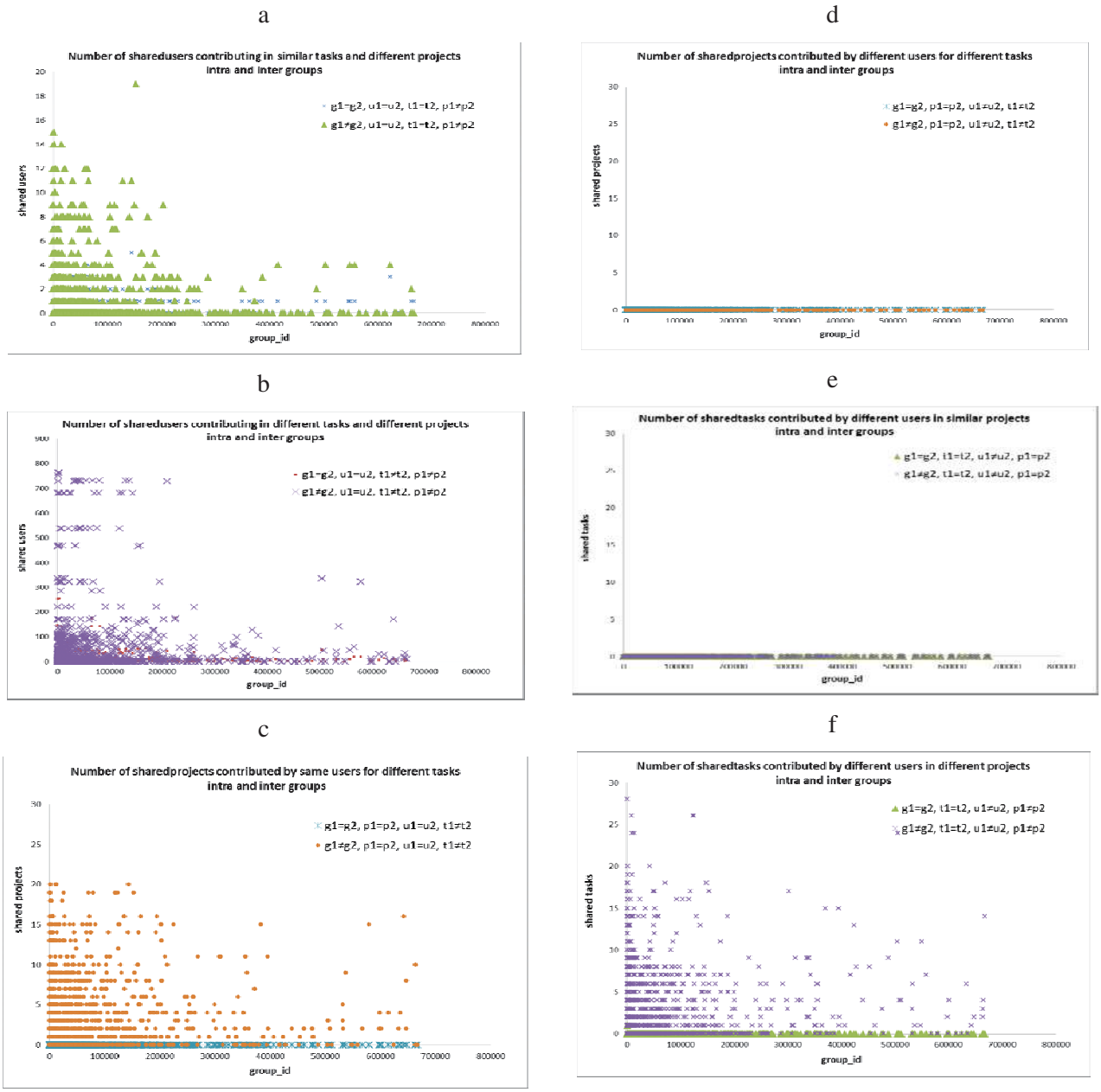

Fig. 4. Illustration of a,b) shared users implying one developer contributing to similar or different tasks and different projects intra or inter groups, c,d) shared projects implying one project contributed by the same or different developers and different projects intra or inter-groups, e,f) shared tasks implying one project task contributed by different developers in similar or different projects intra or inter groups.

As seen in Figure 4.c (g1=g2, p1=p2, u1=u2, t1≠t2), one project can be contributed by one user for different tasks; whereas Figure 4.d (g1=g2, p1=p2, u1≠u2, t1≠t2) implies that one project cannot be contributed by different users intra or inter groups.

As seen in Figure 4.e (g1=g2, t1=t2, p1=p2, u1≠u2) and Figure 4.f (g1=g2, t1=t2, p1≠p2, u1≠u2), different users contribute in different tasks for different projects, but not the same projects (*shown by zero values*).

Figure 5 illustrates different projects, users, tasks and groups, where each project could be related to diferent users, and that $project_{id}$ represents just name and id of its initiator.

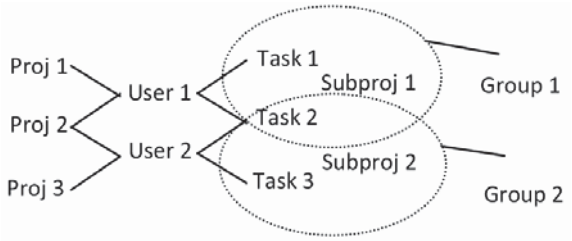

Fig. 5. Illustration of group of OSS developers co-working on some project tasks from the same group or two connected groups.

Project task shows the number of developers contributing to a particular task. We use $task_{id}$ to find out number of developers contributing to the same task. At the same time, each user could create a new

subproject; therefore, each group contains number of users, tasks and projects.

2.2. Analysis. We conduct an empirical analysis to validate the hypotheses; for this purpose, we use a complex network of open source software (OSS). As afore mentioned, we use three constructs "intragroup coupling", "inter-group coupling" and "intergroup structural hole", however, there are other variables which could influence on the output (group innovation and growth) such as how number of developers contributing to a particular task, and number of tasks that one developer contributes to, as well as number of projects that one developer has initiated, however, we have to control all these variables. Moreover, group size might also affect the dependent variable in that group size has a positive effect on its member projects' performance, because bigger groups provide the members with more opportunities. A developer or user in a larger cohesive group has easier access to the right information, knowledge, and resources, because there would be a greater number of developers. On the other hand, a larger cohesive network has a larger number of developers who are familiar with each other. A larger network also guarantees the availability of a larger pool of developers or users, leading to a higher level of user participation. However, we include this factor by its outcome, which is number of tasks that one developer contributes to.

As previously mentioned, *Inter-group coupling* is the developers (denoted by $user_{id}$) contributing to project tasks between two groups (measured by number of intergroup links); whereas *Intra-group coupling* is the developers (denoted by $user_{id}$) contributing to project tasks within a single group (measured by number of intra-group links). Moreover, structural holes are measured by clustering coefficients among users.

We use regression modelling to prove the three hypotheses. In the regression model, we use lagged explanatory variables, first, because there is possible existence of simultaneity between dependent and independent variables. The simultaneity problem stems from possible confusion in the direction of causality between dependent and independent variables. For example, network structures may influence project performance, but, meanwhile, performance is likely to influence network structures. Second, the specification of lagged structural variables is also based on rationality that the impacts of group structure on intergroup coupling require a certain time lag before they take place.

$$Y1 = a_0 + a_1 C1 + a_2 C2 + a_3 C3 + \alpha_4 X1 + \\ + \alpha_5 X2 + \alpha_6 X3 + \alpha_7 X4 \quad (1)$$

$$Y2 = \beta_0 + \beta_1 C1 + \beta_2 C2 + \beta_3 C3 + \beta_4 X1 + \\ + \beta_5 X2 + \beta_6 X3 + \beta_7 X4 \quad (2)$$

Table 3. List of variables

Variable names	Measures	Notations
Dependent variables		
	$\Delta_t\ (user_{id} \in groups)$	Y_1
	$\Delta_t\ (project_{id} \in groups)$	Y_2
Independent variables		
	$\Delta_t\ (task_{id}\ \|user_{id} \in intra-groups)$	X_1
	$\Delta_t\ (task_{id}\|user_{id} \in inter-groups)$	X_2
	$\Delta_t\ (structural\ holes\|user_{id})$	X_3
	$\Delta_t\ (structural\ holes\|user_{id})$ $\cdot\ (task_{id}\|user_{id} \in inter-groups)$	X_4
Control variables		
	$\Delta_t\ (task_{id}\|user_{id})$	C_1
	$\Delta_t\ (user_{id}\|task_{id})$	C_2
	$\Delta_t\ (project_{id}\|user_{id}, task_{id})$	C_3

Dependent variables. We use an approach to define the outputs (network innovation and growth), where new OSS project initiation within each group represents innovation $(project_{id} \in groups)$ and new developers joining existing OSS projects within each group represents group growth $(user_{id} \in groups)$. Innovation usually results from interactions among different sources of knowledge (in this study, it refers to OSS project tasks among developers from same group or different groups).

Independent variables. Inter-group coupling represents the developers (denoted by $user_{id}$) contributing to separate tasks between two groups (measured by number of inter-group links), whereas intra-group coupling represents the developers (denoted by $user_{id}$) contributing to separate tasks within a single group (measured by number of intra-group links). Inter-group structural holes are the number of opportunities for developer contribution between two connected groups.

Control variables. While our study focuses on examining the impact of intra and inter-group coupling and structural hole on group innovation, other factors might also have an influence on group innovation. Hence, we control for three factors:

- The number of developers contributing to a particular task $(user_{id}|task_{id})$ implies how popular each project task is. The more popular each task is, the higher the number of developers contributing to the task and the more infor-

mation is exchanged, and the more possibility of new developers joining the existing task.

- The higher the number of tasks $(task_{id}|user_{id})$ one developer contributes to, the less time the developer has to spend, and more time developer has in order to initiate a new project.
- Developers who have initiated some projects (represented by $project_{id}$) are more probable to initiate another new project than those who have just contributed to project tasks, and have not initiated a project (they are not as innovative). Therefore, we control for number of projects $(projekt_{id}|user_{id}, task_{id})$ that one developer has initiated.

3. Results

The source of knowledge and information for OSS projects can range from collaboration within and outside the group, wherein OSS team members have different social networks outside the team and may exchange information and collaborate with particular groups of developers. In this section, we illustrate how this collaboration will influence on groups' innovation. The results will be examined both graphically and statistically using the regression model.

3.1. Graphical representation. In the following graphs, we show the change in number of developers and projects for each group from January 2013 to January 2014, and attempt to see whether the hypotheses are supported.

As shown in Figure 6.a, number of users belonging to two connected groups w.r.t number of intra-group coupling have increased from January 2013 and to January 2014, so this is an indication of growth as a result of intra-group coupling. As shown in Figure 6.c, number of users belonging to two connected groups w.r.t number of inter-group coupling have increased, so this is an indication of growth as a result of inter-group coupling. However, the raise in number of users due to intra-group coupling is far more than its raise in number of users due to inter-group coupling; therefore, H1 is supported, implying positive impact of intra-group on number of users.

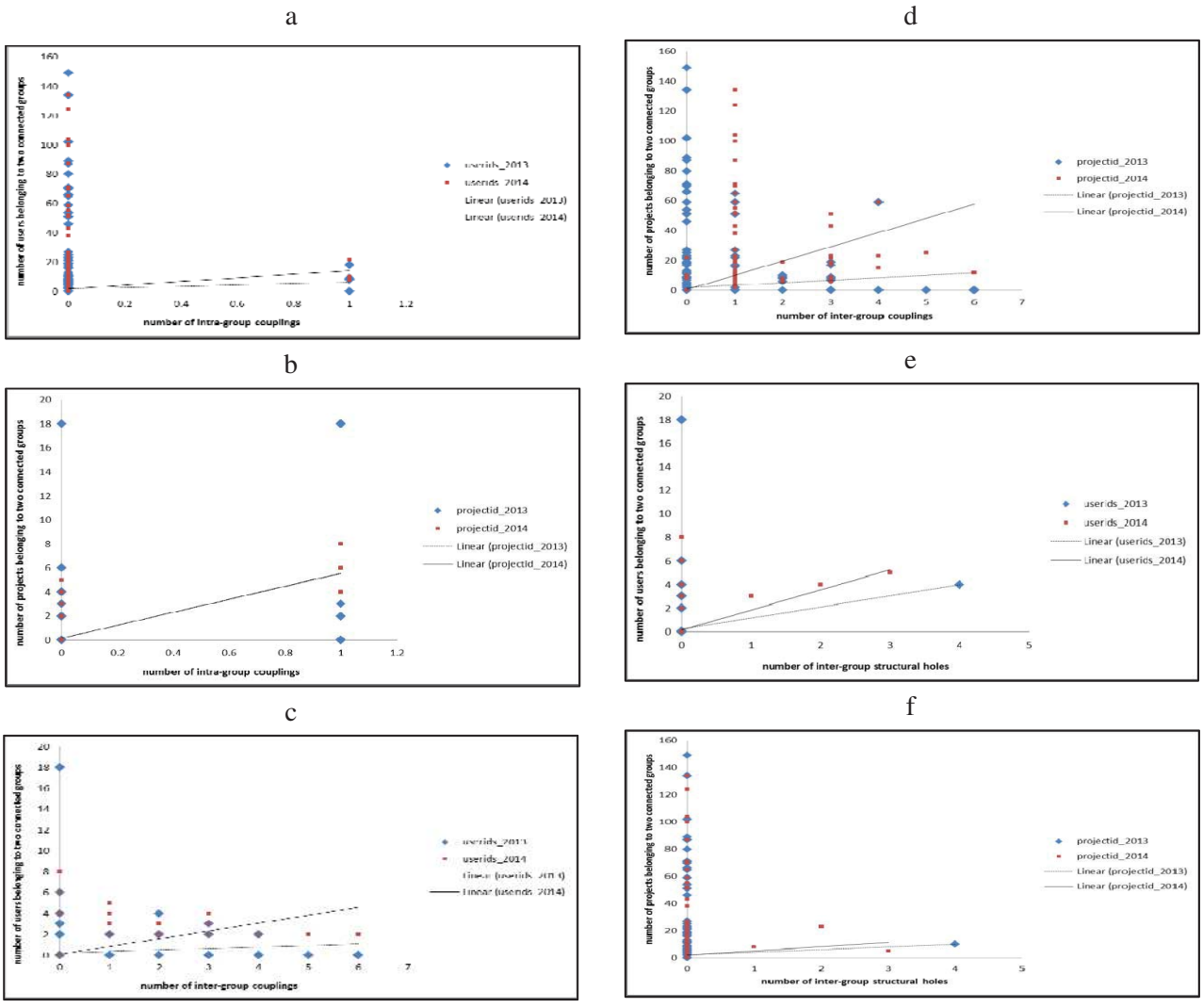

Fig. 6. Illustration of growth and innovation as results of intra-, inter-group couplings and structural holes

As shown in Figure 6.b, number of projects belonging to two connected groups w.r.t number of intra-group coupling has not increased; therefore, there is no indication of innovation as a result of intra-group coupling. As shown in Figure 6.d, number of projects belonging to two connected groups w.r.t number of inter-group coupling have increased, so this is an indication of innovation as a result of inter-group coupling; therefore, H2 is supported, implying positive impact of inter-group coupling on number of projects.

As shown in Figure 6.e, number of users belonging to two connected groups w.r.t number of inter-group structural holes have increased over the period, so this is an indication of growth as a result of structural holes. As shown in Figure 6.f, number of projects belonging to two connected groups w.r.t number of inter-group structural holes have increased, so this is an indication of innovation as a result of structural holes; therefore, H3 is supported.

3.2. Statistical representation. We attempt to test the three hypotheses using regression modeling and determine if structural variables are significant predictors of OSS group innovation and growth. As observed inTable 4, intra-coupling is positive and significant (a4=5.556, p<0.01) implying that ties within groups has positive influence on group growth. However, inter-coupling is positive, but non-significant (a5=1.637, p>0.1) implying that H1 is supported. In addition, number of structural holes is negative and insignificant (a6=-0.473, p>0.1) implying that structural holes within groups has no influence on group growth. However, structural holes * inter-coupling is negative and significant (a7=-0.287, p<0.01) implying that number of structural holes negatively influences the impact of inter-coupling on group growth. In other words, there is a trade-off between impact of inter-group structural holes and inter-coupling.

Finally, among the control variables, number of developers for a particular task $(user_{id}|task_{id})$ is positive and significant implying that how popular each project task is. The more popular each task is, the higher number of developers will contribute to the task. In addition, neither the number of projects nor the number of tasks is statistically significant.

Table 4. Number of new developers (dependent variable) as a function of independent and control variables

	Coefficients	*Std Dev*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.188	0.017	0.000	-0.222	-0.154
#tasks (C1)	0.001	0.000	0.056	0.000	0.002
#developers (C2)	0.080	0.024	0.001	0.034	0.127
#projects (C3)	0.000	0.001	0.893	-0.003	0.002
#intra-coupling (X1)	5.556	0.377	0.000	4.816	6.295
#inter-coupling (X2)	1.638	0.052	0.150	1.536	1.739
#struct. Holes (X3)	-0.473	1.324	0.721	-3.070	2.124
#struct. holes* #inter-coupling (X4)	-0.287	0.021	0.000	-0.327	-0.246

As observed inTable 4, intra-coupling is positive but not significant (b4=30.069, p>0.1) implying that intra-group coupling does not influence on group innovation. However, inter-coupling is positive and significant (b5=35.611, p<0.01) implying that H2 is supported. In addition, number of inter-group structural holes is not significant (b6=0.233, p<0.01) implying that structural holes between groups has no influence on group innovation. However, inter-group structural holes * inter-group coupling is negative but significant (b7=-8.407, p<0.01) implying that number of structural holes negatively influences the impact of inter-coupling on group growth; therefore, H3 is supported.

Finally, among the control variables, number of developers for a particular task $(user_{id}|task_{id})$ is positive and significant implying how popular each project task is to attract higher number of developers in order to contribute to the task. In addition, neither the number of projects nor the number of tasks is statistically significant. As observed in the following correlation matrix, Table 6, the relationship between explanatory variables is fairly low, therefore, multicollinarity is not a problem.

Table 5. Number of new projects (dependent variable) as a function of independent and control variables

	Coefficients	*Std. Dev.*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.031	0.378	0.006	-1.772	-0.290
#tasks (C1)	-0.011	0.010	0.275	-0.031	0.009
#developers (C2)	-5.059	0.519	0.000	-6.076	-4.042
#projects (C3)	0.017	0.029	0.564	-0.040	0.074
#intra-coupling (X1)	30.069	8.220	0.258	13.953	46.185
#inter-coupling (X2)	35.611	1.127	0.000	33.401	37.821
#struct. Holes (X3)	0.233	0.061	0.000	0.114	0.353
#struct. holes* #inter-coupling (X4)	-8.407	0.448	0.000	-9.285	-7.530

Table 6. Correlation matrix

	#tasks	#developers	#projects	#struct. holes	#intra-coupling	#inter-coupling	#struct. holes* #inter-coupling
#tasks	1.0000						
#developers	0.0175	1.0000					
#projects	0.2999	0.0203	1.0000				
#struct. holes	0.0003	-0.0001	0.0003	1.0000			
#intra-coupling	-0.0257	0.0063	-0.0254	-0.2278	1.0000		
#inter-coupling	0.0074	0.0141	0.0076	0.1203	-0.2578	1.0000	
#struct. holes* #inter-coupling	-0.0018	0.0004	-0.0018	0.2506	-0.1839	0.1133	1.0000

3.3. Strategy options. As well demonstrated, strategy on OSS group innovation and growth depends on OSS network group structure. Some variables such as group coupling have direct impact on group innovation, and some others such as number of tasks per user might have indirect influence on group innovation and growth. Other researchers have investigated OSS project success in terms of commercial and technical success. The most widely indicators of OSS success include number of downloads, number of releases, task completion rate, number of concurrent versioning system (CVS) commits (Crowston et al., 2003; Grewal et al., 2006). Download rate as a proxy for acceptance rate could reflect commercial success (Wang, 2007).

As shown in Figure 7, the number of tasks one developer contributes to $(task_{id}|user_{id})$ is associated with the number of intra-group coupling and inter-group coupling. As a result, this could impact both group innovation and growth. In another study, we have shown that intra-group coupling could lead to inter-group coupling (Behfar and Behfar, 2016). Therefore, in order to achieve more group innovation, one needs to target task contribution between groups or inter-group structural hole, whereas to achieve more group growth, one needs to target a number of task contributions inside a group or a number of users per task.

The number of developers contributing to a particular task $(user_{id}|task_{id})$ implies how popular each project task is. The more popular each task is, the higher the number of developers contributing to the task, which indicates group coupling. This could lead to group innovation. This has implications for project managers in open source environment, such as IBM and Sun Microsystems actively working in open source projects with decision to sponsor project tasks for the purpose of group innovation or group growth.

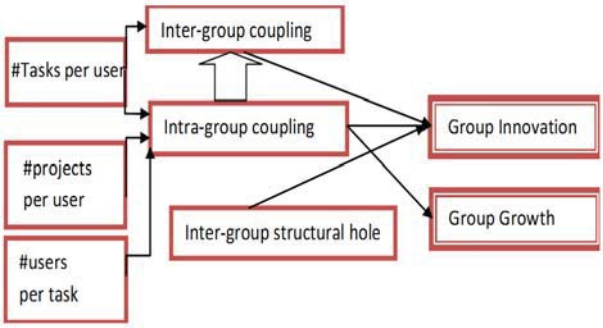

Fig. 7. Illustration of strategy options to augment the group innovation.

Conclusion

Open source software (OSS) projects could be launched by both commercial and non-commercial sectors, however, as opposed to conventional organizational software development, where projects are assigned by managers to skilled individuals, OSS collaboration teams are based on voluntary groups composed of individuals with different ranges of skills. The success of OSS projects relies on the extent to which they attract individuals to join projects. Recently organizations have shown much interest in OSS collaboration teams both as pools of projects for reuse, as well as volunteer groups for contributing to existing OSS projects and initiation of new projects.

Although it is significant to identify the factors that attract developers to join projects, it would be interesting to find the factors that make developers initiate a new project. These factors could be personal reasons or network structural factors. Hahn et al. (2008) have investigated personal reasons as the factors causing a new developer to join a project, concluding that prior collaboration between a new developer and the project initiator or previous experience of group members are relevant factors. However, in this study, we are only concerned about network structural factors that influence developers to join existing projects or initiate new projects within a group. We focused on the network group rather than the individual for network structure (input), as well as innovation and growth (output). In this regard, we discussed three aspects of network

structure with regard to innovation: 1) tie strength (weak versus strong), 2) network structure (structural hole) and 3) network structure (group coupling). We defined three constructs: 1) *intra-group coupling* measured by the number of project tasks in each group, 2) *inter-group coupling* measured by the number of project tasks between any two groups, and 3) *structural hole* measured by the number of developer contribution opportunities between any two groups. Group innovation was defined by new project initiation within each group, whereas group growth was defined by the number of new developers joining existing groups.

We showed that ties within groups have a positive influence on group growth. Moreover, we demonstrated that ties between groups have positive influence on group innovation. In addition, structural holes between groups have positive influence on group innovation. However, the number of structural holes negatively influences the impact of inter-coupling on group innovation. In other words, there is a trade-off between the impact of inter-group structural holes and inter-coupling on group innovation.

In order to achieve more group innovation, one needs to target task contribution between groups or inter-group structural hole, whereas to aim more group growth, one needs to target number of task contributions inside a group or number of users per task. From a managerial and economic point of view, several researchers have pinpointed the need to manage and organize adequately such groups (Benkler, 2006; Borgatti and Foster, 2003). As implications for project managers in open source environment, e.g., an IBM executive to make a financial or human resource allocation decision to which project tasks programmers should work on, his/her focus could be more group innovation or growth, then, he or she has to target task contribution between groups, number of users per tasks or more.

The practical significance of these contributions to the literature is to benefit business strategy by the use of volunteer groups for the purpose of both contributing to existing OSS projects and initiating new OSS projects. This has made it worthwhile to the factors (personal and structural) influencing developers to join projects or initiate new projects. Future research could examine relative activity of users as group members, or look at the application of our proposed definitions of innovation and growth to other domains.

Acknowledgement

We collected data from SourceForge.net, which is the largest repository of OSS projects. We appreciate access to this repository given by prof. Greg Madey at the department of Computer Science & Engineering, University of Notre Dame.

References

1. Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study, *Administrative Science Quarterly*, 45, pp. 425-455.
2. Antwerp, M.V. and Madey, G. (2010). The Importance of Social Network Structure in the Open Source Software Developer Community, *The 43 Hawaii International Conference on System Science*, Big Island, Hawaii.
3. Behfar, S.K. and Behfar, Q. (2016). Intragroup density predicting intergroup tie strength within open-source-software collaboration network, *Distributed Computing and Artificial Intelligence, 13th International Conference*, pp. 165-173.
4. Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*, Yale university press, New Haven, CT.
5. Borgatti and Foster. (2003). The network paradigm in organizational research: a review and typology, *Journal of Management*, 29 (6), pp. 991-1013.
6. Burt, R.S. (1992). *Structural Holes*, Cambridge, Mass.: Harvard University Press.
7. Burt, R.S. (2000). "The network structure of social capital", in R.I. Sutton and B.M. Staw (eds). *Research in Organizational Behavior*, Greenwich, Conn.: JAI Press, pp. 345-423.
8. Burt, R.S. (2002). The social capital of structural holes", in M.F. Guillén, R.Collins, P.England, M.Meyer (eds); *The New Economic Sociology*, New York: Russell Sage Foundation. pp. 148-192.
9. Coleman, J.S. (1988). Social capital in the creation of human capital, *The American Journal of Sociology*, 94, pp. S95-S120.
10. Crowston, K., Annabi, H. and Howison, J. (2003). Defining open source software project success, Paper presented at *the 24th International Conference on Information Systems* (ICIS), Seattle, pp. 1-14
11. Granovetter, M. (1973). The strength of weak ties, *American Journal of Sociology*, 78, pp. 1360-1380.
12. Grewal, R., Lilien, G.L. and Mallapragada, G. (2006). Location, location, location: how network embeddedness affects project success in open source systems, *Management Science*, 52 (7), pp. 1043-1056.
13. Hahn, J., Moonm, J.Y. and Zhang, C. (2008). Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties, *Information System Research*, 19, pp. 369-391.
14. Hansen, M.T. (1999). The search-transfer problem: the role of weak ties in sharing knowledge across organization units, *Administrative Science Quarterly*, 44, pp. 82-111.

15. Hinds, D. and Lee, R.M. (2008). Social network structure as a critical success condition for virtual communities, *The 41 Hawaii International Conference on System Science*, Big Island, Hawaii.
16. Jackson, M.O. and Wolinsky, A. (1996). A strategic model of social and economic networks, *Journal of Economic Theory*, 71, pp. 44-74.
17. Jackson, M.O. (2004). A survey of models of network formation: stability and efficiency, In G.Demange and M.Wooders (eds.), G*roup Formation in Economics: Networks, Clubs and Coalitions*. Cambridge: Cambridge University Press, pp. 1-62.
18. Kogut, B. (2000). The network as knowledge: generative rules and the emergence of structure, *Strategic Management Journal*, 21 (3), pp. 405-425.
19. Lefebvre, V.M., Sorenson, D., Henchion, M. and Gellynck, X. (2016). Social capital and knowledge sharing performance of learning networks, *International Journal of Information Management*, 36 (4), pp. 570-579.
20. Montazemi, A.R., Siam, J.J. and Esfahanipour. (2008). Effect of network relations on the adoption of electronic trading systems, *Journal of Management Information Systems*, 25, pp.233-266.
21. Mockus, A., Fielding, R.T. and Herbsleb, J.S. (2002). Two case studies of open source software development: Apache and Mozilla, *ACM Transactions on Software Engineering and Methodology*, 11 (3) pp. 309-346.
22. Newman, M.E.J. (2004). Detecting community structure in networks, *The European Physical Journal B*, 38 (2), pp. 321-330.
23. Riehle, D. (2015). How open source is changing the software developer's career. *IEEE Computer*, 48 (5), pp. 51-57.
24. Simon, H.A. (1962). The architecture of complexity, *Proceedings of American Philosophical Society*, 106 (6), pp. 462-482.
25. Singh, P.V., Tan, Y. and Mookerjee, V. (2011). Network effects: The influence of social capital on open source project success, *MIS Quarterly*, 35 (4), pp. 813-829.
26. Stadler, C., Rajwani, T. and Karaba, F. (2014). Solutions to the exploration/exploitation dilemma: networks as a new level of analysis, *International Journal of Management Reviews*, 16, pp.172-193.
27. Tedeschi, G., Vitali, S. and Gallegati, M. (2014). The dynamic of innovation networks: a switching model on technological change, *Journal of Evolutionary Economics*, 24 (4), pp. 817-834.
28. Tsai, W. (2000). Social capital, strategic relatedness and the formation of intraorganizational linkages, *Strategic Management Journal*, 21 (9), pp. 925-939.
29. Tsai, W. (2001). Knowledge transfer in intra-organizational networks: effects of network position and absorptive capacity on business unit innovation and performance, *Academy of Management Journal*, 44 (5), pp. 996-1004.
30. Von Hippel, E. (2001). Innovation by user communities: learning from open-source software, *Sloan Review*, 42 (4), pp. 82-86.
31. Walker, G., Kogut, B. and Shan, W.J. (1997). Social capital, structural holes and the formation of an industry network, *Organization Science*, 8 (2), pp. 109-125.
32. Wang, J. (2007). The role of social networks in the success of open-source-software systems: a theoretical framework and an empirical investigation, *PhD Thesis at Kent State University Graduate School of Management.*
33. Weick, K. (1976). Educational organizations as loosely coupled systems, *Administrative Science Quarterly*, 21, pp.1-18.
34. Williamson, O.E. (1975). *Markets and hierarchies: Analysis and antitrust implications*, New York: The Free Press.