International Journal of Information Management 38 (2018) 167–174

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Research Note

Knowledge management in OSS communities: Relationship between dense and sparse network structures

Stefan Kambiz Behfar[a,*], Ekaterina Turkina[b,1], Thierry Burger-Helmchen[a,2]

[a] BETA-CNRS 7522, Bureau d'Economie Théorique et Appliquée, Université de Strasbourg, 61, Avenue de la Forêt Noire, F-67085 Strasbourg Cedex, France
[b] Department of International Business, HEC Montreal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, Québec, H3T 2A7 Canada

ARTICLE INFO

Keywords:
Knowledge transfer
Open source software network
Intergroup diffusion of innovation
Intragroup density
Intergroup coupling

ABSTRACT

Some authors in the literature have addressed knowledge transfer via weak ties between organization's units which are themselves strongly tied inside (e.g. Hansen, 1999). Some others have investigated knowledge management among open-source-software (OSS) developers and discussed factors influencing knowledge transfer within development teams (e.g. Joshi and Sarker, 2006). In the domain of open source software (OSS) communities, more companies are now attempting to establish relationships to benefit from these potential value-creating communities; and project managers could in fact target different goals within software development teams including knowledge transfer within and between teams. We step forward to distinguish knowledge transfer within groups as opposed to knowledge transfer between groups; where relevant projects are bundled into separate strongly intra-connected groups. In knowledge management literature there is a trade-off between sparse network structures (Burt, 2000, 2002) versus dense network structures (Walker et al., 1997; Coleman, 1988). It is argued that the former facilitates the diffusion and generation of ideas among groups, while the latter affects the implementation of idea within each dense group. To our best knowledge, there has been no study to investigate the relationship between dense and sparse network structures. We propose that *knowledge transfer within dense groups has a positive influence on knowledge transfer between sparse groups*, in that intragroup density, group size, developers centrality and betweenness could impact intergroup coupling. To prove our hypothesis, we use a complex network of open source software (OSS) as the domain of interest, where developers represent nodes and two developers contributing to a project task represent a network tie. Developers contributing to tasks in groups other than their own can *explore* novel ideas via sharing knowledge, whereas developers contributing to tasks inside groups *exploit* ideas to improve those projects. We investigate the idea both analytically and empirically within 4 months, 8 months and 1 year lagged time, and finally show that intragroup density has a positive whereas developers' centrality has a negative influence on intergroup coupling.

1. Introduction

Inside organizations, units can learn from each other and knowledge transfer can provide new mutual opportunities for units as well as for the whole organization. New ideas diffuse rapidly when they benefit organizations adopting them, and they vanish, if otherwise (Abrahamson, 1991). Huber (1991) suggested that organizational units transfer knowledge and learn from other units, but not all units have access and capacity to learn knowledge and apply it; they require external access and internal capacity. Internal capacity can be achieved by R & D ability increase, while external access to new knowledge can be improved by networking. In this regard, Hansen (1999) introduced modelling an organization as a complex network with inter-unit links, where knowledge transfer is investigated by analyzing inter-organizational network.

In regards to usage of social network analysis (SNA) in organizations, different authors focused on a wide range of network characteristics from relational (e.g. tie strength) and nodal (e.g. functional background) to positional (e.g. betweenness centrality) and structural (e.g. density), e.g. impact of size of network on innovation (Baer, 2010), relationship strength (Rost, 2011), or weak and strong tie (Nelson, 1989; Tsai, 2000, 2001). Baer, Evan, Oldham, and Boasso (2015) carried out a *meta*-analysis of studies on innovation and social networks and presented insights into the various trade-offs between strength of

* Corresponding author.
E-mail address: kambiz.behfar@gmail.com (S.K. Behfar).
[1] http://www.hec.ca/en/profs/ekaterina.turkina.html.
[2] http://www.beta-umr7522.fr/-BURGER-HELMCHEN-Thierry?lang=fr.

http://dx.doi.org/10.1016/j.ijinfomgt.2017.09.004
Received 28 July 2017; Accepted 26 September 2017
0268-4012/ © 2017 Elsevier Ltd. All rights reserved.

ties and bridging ties among other things. Tsai (2000) suggested that social networks facilitate the creation of new knowledge within organizations. In another study, Tsai (2001) focused on the question "How can an organizational unit gain useful knowledge from other units to enhance its innovation and performance?", and emphasized the role of strong ties in intra-corporate and strategic alliances. Moreover, Ahuja (2000) discussed firm's network relationship impacting the rate of innovation, where network allows for knowledge sharing and information flow. Others have studied the role of networks within the topic of knowledge sharing and innovation adoption where importance was given to the number of firm linkages and geographical proximity (Florida, 1995; Van Oort & Atzema, 2004) impacting rate of adoption.

Apart from the discussion about knowledge management within and between organizations and the discussion about social network analysis in organizations on the topic of innovation, within topic of open source software (OSS) development, researchers have used social network theories to investigate the OSS phenomenon including communication among developers. The positions and relationships among developers in a social network are significant in the efficiency of the network (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 (see Grewal, Lilien, & Mallapragada, 2006; Singh, Tan, & Youn, 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, Siam, and Esfahanipour (2008) demonstrated that the market structure of embedded interpersonal ties enables participants to take advantage of information asymmetry for profit taking Singh, 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, 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. Crowston, Annabi, and Howison (2003) also discussed social structure of open source software development teams based on the analysis of interactions represented in bug reports from 122 large and active projects, and found out that some projects are highly centralized, and others are not.

As above-mentioned, several authors have previously discussed the significance of positions and relationships among developers or so-called community ties in the efficiency of OSS network. In addition, knowledge management among open-source-software (OSS) has been investigated (Joshi and Sarker, 2006), where they discussed factors influencing knowledge transfer within development teams. Ojha (2005) also discussed knowledge sharing between team members based on similarity-attraction paradigm; where he proposed that knowledge sharing more likely happen between same demographic team members. However, developers collaborate not only within the same project group but also across groups, therefore knowledge transfer should be also investigated across groups within sparse network structure. In this regard, 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 impacts on 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. To our best knowledge, there has been no study to investigate relationship between dense and sparse network structures, i.e. impact of dense network on sparse network structure in regards to knowledge transfer. In other words, intragroup density, group size, developers' centrality and betweenness within dense groups could have a positive influence on intergroup coupling between sparse groups. In the theoretical development section, we discuss why we have chosen these independent variables in this causal relationship.

In order to develop our hypotheses, we use a complex network of open source software (OSS) as the domain of interest. In this network, developers represent the nodes and two developers contributing to a project task represent a network tie. Developers contributing to tasks in groups other than their own can *explore* novel ideas for new project creation, whereas developers contributing to project tasks inside their own group *exploit* ideas to improve those existing projects with better inside-group search possibility.

In the theoretical development section, we provide hypotheses and discuss logical and analytical reasoning to prove our hypothesis; then in the empirical section, we alternatively examine the relationship between intragroup density, group size, developers' centrality and betweenness with intergroup coupling, using 4 months, 8 months and 1 year lagged time (to examine robustness), via examining OSS data collected from SourceForge repository.

2. Theory development

In the introduction section, we provided literature and motivation for this paper; here we render the hypotheses and model design to give logic and reasoning to prove the hypothesis.

2.1. 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, as shown in Table 1. We use the word "coupling" between

Table 1
Terminology.

term	definitions	measure
Network tie	two developer working on same project task	frequency of developer contribution in project tasks
Network structure		
	Dense intragroup structure	Densely intraconnected groups, where developers work on relevant project tasks
	Sparse intergroup structure	Sparsely interconnected groups, where developers work on irrelevant project tasks
Intragroup density		Sum of intragroup ties over total possible ties within a group
Intergroup coupling		Sum of intergroup ties (sum of intergroup project tasks)
OSS group	group_project$_{id}$, including project relevant members	Assigned by sourceforge 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

groups, which refers to sum of network ties between network nodes, where tie strength is in fact frequency of developer contribution to project tasks. Intergroup ties are more efficient for acquiring external knowledge, accessing the diversity in contribution in other groups, and facilitating diffusion of new project ideas which leads to new project initiation inside the group, and intragroup ties are more efficient for quick transfer of information, which leads to group growth (Behfar, Turkina, & Burger-Helmchen, 2017).

There are different methods in the literature to calculate intragroup density; alternatively we compute intragroup density by the number of project tasks in each group over total possible ties obtained by the number of nodes represented by developers.

Groups could be connected by a member or more developing projects within different groups; in this case, number of common members would be intergroup coupling. Alternatively here we assume that groups are connected via project tasks by which members from different groups contribute to; in that, intergroup coupling would be measured by number of project tasks connecting developers from the different groups.

2.2. Model design

In OSS network, there are knowledge modules (OSS project), relevant projects are bundled into separate strongly intra-connected groups; whereas irrelevant groups are weakly inter-connected. Ties inside groups resulting from the fact that developers work on the same or relevant project tasks create dense network structure; whereas developers working on other groups' project tasks which are mostly irrelevant create sparse network structure.

Considering a research question which concerns knowledge transfer within dense and sparse network structures, some authors have previously investigated both of them in details, and concluded a trade-off between sparse network structures (Burt, 2000, 2002) versus dense network structure (Walker et al., 1997; Coleman, 1988). It is argued that the former facilitates diffusion and generation of ideas among clusters within the network, while the latter impacts on implementation of idea within each dense cluster. Kogut and Zander (1995) and Tsai (2000) argued that a dense innovative cluster provides quick transfer of information, knowledge sharing, more interactions, and better integration, better coordination. On the other hand, Burt (2000, 2002) argued that strong ties are inefficient to acquire external knowledge as they lack diversity in resources needed for innovation, and at the same time increase communication costs as a result of redundancy of ties. Therefore, weak ties (non-redundant, less-frequent) are more appropriate to communicate, which allow to access variety of knowledge. Despite this trade-off, to our best knowledge, there has been no study to investigate relationship between dense and sparse network structures and whether knowledge transfer in dense network (inside groups) influences on knowledge transfer in sparse network (between groups). At the same time, we propose that *knowledge transfer within dense groups has a positive influence on knowledge transfer between sparse groups.*

Considering that knowledge transfer within dense groups could be proxied by group density, group size, degree centrality and betweenness inside groups; In another word, these measures represent knowledge transfer within groups, we then explore their impact on intergroup coupling which represent intergroup knowledge transfer.

We expect that bigger groups provide developers with more opportunities to contribute to, or increase intragroup density. A developer or user in a larger group has easier access to the right information, knowledge, and resources because there would be a greater number of projects. Therefore it could affect knowledge transfer inside a group. On the other hand, a larger pool of developers leading to a higher level of participation could positively influence developer interaction leading to higher intragroup density, see Fig. 1, however they could indirectly impact intergroup coupling.

Intragroup density refers to the ties and relationships inside a group over total possible number of ties between the developers. In general, these ties represent knowledge transfer inside each group. We claim a positive influence of intragroup density on intergroup coupling. This is based on the fact that developers within a densely-connected group are very likely to create subsequent intergroup connection. The reason for subsequent intergroup connection could include awareness or common neighborhood. Therefore, we propose (see Fig. 1) that

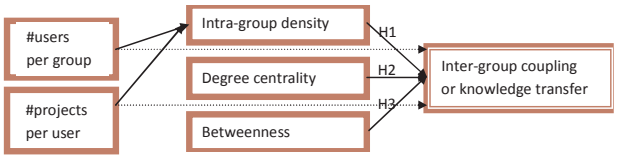

Fig. 1. Illustration of impact of intragroup density, centrality and betweenness on intergroup coupling.

Hypothesis (H1). Intragroup density has a positive influence on intergroup coupling.

Degree centrality for developers is the number of projects in which the developers contribute to. For instance, Developer 1 only links to Project A, and therefore has a degree centrality measure of 1. A high developers' degree centrality for a project implies that developers work on a large number of project tasks simultaneously, resulting in a more structurally diverse team. The variation in structural diversity is obtained through accessing different resources of knowledge or information. The source of knowledge for any given OSS projects can range from developers within and outside the group. Therefore OSS team will have different social networks outside the team; we therefore propose (see Fig. 1) that

Hypothesis (H2). Degree centrality has a positive influence on intergroup coupling.

Betweenness of a developer can be measured by the sum of the probabilities that a developer lies on between pairs of others including both developers and projects. Developers could function as a broker in that they facilitate exchanges between those who are connected through them. Researchers have discovered that people from distant networks have access to distinct knowledge and information which leads to novel knowledge and improve productivity (Granovetter, 1973). If developers' closeness centrality shows developers ability to access other developers' projects through the minimum number of intermediaries, their betweenness implies their ability to control others (Wang, 2007). Developers' betweenness as a matter of fact could function as a broker outside a group, and can supposedly increase intergroup coupling. Therefore, we propose (see Fig. 1) that

Hypothesis (H3). Developers' betweenness has a positive influence on intergroup coupling.

After the 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 use social network dynamics to *explain and predict* our phenomena of interest. The theory components are: the unit of analysis is the group of OSS developers, where the network is made of the node (developer or user) linking with project tasks. Each developer can initiate new projects, but at the same time co-work on project tasks with other developers.

As will be discussed later in the data section, each project initiated by a developer is given a group_project$_{id}$. In fact, it 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 improve those existing projects, which attracts more developers to join the group, but of course

this does not reject possibility of new project creation within the group.

2.3. Theoretical analysis

In information science, clusters or groups could be defined as sum of developers working on related projects. We consider network of open source software (OSS) developers as our domain of interest; where knowledge groups (OSS communities) are distinguished based on relevancy. Links inside each group where projects are relevant represent developers working on the same project tasks. However mechanisms of link formation between two groups with irrelevant projects are different from the ones inside each group. In OSS project collaboration network, link formation mechanisms are as follows:

2.3.1. Visibility/popularity

projects with higher visibility are found first, because search engine gives more weightage of displaying results to those projects with higher in-degree.

2.3.2. Common neighborhood/latent relevancy

projects which have reused similar third project are more visible to the first project; because if project 1 reuses project 3, and project 2 also reuses project 3, project 1 is able to see that project 2 also reuses project 3. This also holds, when different developers could potentially work on the same project tasks.

2.3.3. Awareness

if a project of group 1 reuses a project of group 2, or a developer of a project from group 1 contributes to a project from group 2, then other developers working in similar projects are influenced or become aware of this reuse, then subsequent links will be consequently formed.

In order to investigate how intragroup density influence on intergroup coupling (or subsequent intergroup link formation), we use the link cost/benefit method introduced by Jackson and Wolinsky (1996) using utility function for each project based on benefit and cost of new link formation, e.g. benefit of reuse ρ_E is time saving for rewriting, and cost c_E is time spending for searching the right OSS project, or in case of social communication among individuals, there are values for direct communications as well as indirect communications from their adjacent individuals whose value depends on their distance; of course, communication is also costly, therefore on should compare its benefit against its cost. We compare utility function of individual i when intergroup link is formed to utility function of individual i without new link, where utility function is defined in (1).

$$u_{P_j^i} = \sum_{j \in n_{P^i}} \rho_j - d_j c_E \tag{1}$$

where, d_j denotes degree of individual P_j^i and n_{P^i} denotes number of individuals in groups. Link benefit is represented by $\rho_{External}, \rho_{Internal} \in (0,1)$, where value of i connected to j is proportional to their proximity, and cost of intergroup link formation is represented by c_E, and $c_I = 0$ for connected groups. Figs. 2 and 3 feature the step by step proof. Initially we compare the utility function between two situations, first there is no link existing between groups, and a link between groups 1 and 2 is targeted, as shown in Fig. 2a. Second there is a link existing between groups 1 and 3, and between group 2 and 3, and a link between groups 1 and 2 is expected.

To have the initial link formed between P_1^1 from group 1 and P_1^2 from group 2, the utility function when having a link $l(P_1^1, P_1^2)$ (without a prior intergroup connection) should be less than link creation when there is a prior connection (common neighborhood), as seen in Fig. 2b.

$$u_{P_1^1}(l(P_1^1, P_1^2)) < u_{P_1^1}(l(P_1^1, P_1^3), l(P_1^3, P_1^2), l(P_1^1, P_1^2)) \tag{2}$$

$$(n_{P^1} - 1)\rho_I + \rho_E - c_E < (n_{P^1} - 1)\rho_I + \rho_E + \rho_E^2 - c_E + \rho_E - c_E$$
$$\rightarrow c_E < \rho_E + \rho_E^2 \tag{3}$$

Eq. (3) is not always true; however having a common neighborhood makes link formation $c_E < \rho_E$between groups 1 and 2 more likely in that $c_E < \rho_E + \rho_E^2$ is more beneficial (less costly) than

Therefore, having a common neighborhood makes link formation between groups 1 and 2 more. Now we would like to see whether intragroup density leads to subsequent intergroup link.

In order to have a subsequent link created between P_2^1 and P_1^2, the utility function when having subsequent link $l(P_2^1, P_1^2)$ shown in Fig. 3b, should be greater than the one when just having initial link $l(P_1^1, P_1^2)$ shown in Fig. 3a.

$$u_{P_1^1}(l(P_1^1, P_1^2)) < u_{P_2^1}(l(P_2^1, P_1^1), l(P_1^1, P_1^2)) \tag{4}$$

$$(n_{P^1} - 1)\rho_I + \rho_E - c_E < (n_{P^1} - 1)\rho_I + \rho_E + \rho_E^2 - c_E \rightarrow 0 < \rho_E^2 \tag{5}$$

As given in (4) and (5), when initial link is formed between two groups, subsequent link formation is always cost-wise beneficial to form, shown by $\rho_E^2 > 0$, while formation of initial link on the basis of common collaboration is conditional on link benefit and cost, $\rho_E > c_E$ without a common neighborhood, or $\rho_E + \rho_E^2 > c_E$ when having a common neighborhood in (3). This is based on the condition that groups are densely connected. This indicates that intragroup density leads to subsequent intergroup link (growth in intergroup strength).

We discussed the positive influence of intragroup density on intergroup coupling. This is based on the fact that developers within a group are densely connected, then if any of those intra-connected developers has also intergroup connection, then subsequent intergroup connection is very likely to happen. The reason includes awareness or common neighborhood, as previously discussed, which makes this link formation costwise beneficial.

One can explore this further, and investigate the causal relationship in more detail in that how dense a group should be? Is there a threshold, upon that all intergroup connections are automatically formed; and below that threshold, no extra intergroup connection is formed. Jackson and Wolinsky (1996) discussed unique strongly efficient network where for the case of a complete or fully connected graph, $\rho - \rho^2 > c$ should hold true.

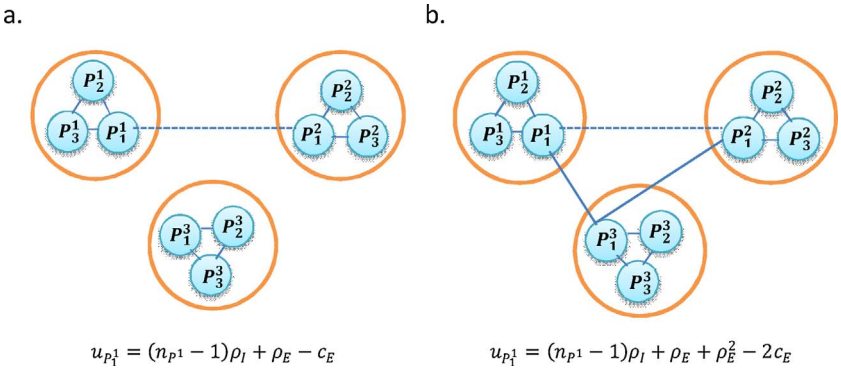

Fig. 2. Illustration of three groups in two cases of a) no intergroup link, b) initial link between groups 1 and 3, and groups 2 and 3.

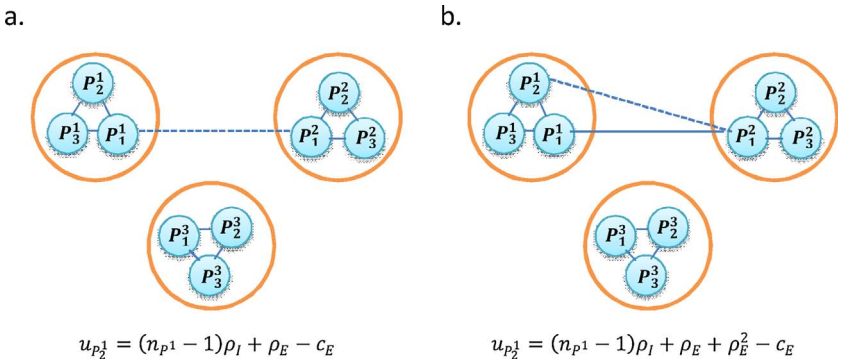

Fig. 3. Illustration of three groups in two cases of a) no initial intergroup link between groups 1 and 2, b) initial and subsequent intergroup link between groups 1 and 2.

However for a condition that all nodes of a network is directly connected (a start encompassing everyone) but is not a complete graph, then $\rho - \rho^2 < c < \rho + ((N-2)/2)\ \rho^2$ should hold true. This is the threshold which shows how dense a group should be, within which all nodes are directly connected.

3. Empirical analysis

In the theoretical development section, we first outlined the literature gap and proposed our hypothesis about the positive influence of intragroup density influence on intergroup coupling, and showed that subsequent intergroup tie is always costwise beneficial when having high intragroup density or fully connected groups. Here, we provide the empirical analysis to validate the relation of intragroup density in intergroup coupling. For this purpose, we use the complex network of open source software (OSS) as the group of interest, and collect OSS project collaboration data, as below.

3.1. Data collection

We collected the data from the website of SourceForge.net, which is the largest repository of OSS projects. At the time it contained more than 150,000 projects and more than 1,600,000 project developers (as indicated by Crowston, et al., 2003). SourceForge.net website has categorized open source software (OSS) 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 website gives $\text{group_project}_{id}$ as an identifier for each project. In fact, sourceforge administration assigns id for a new project; moreover new members/developers are added by the group administrator based on relevancy and of course his or her interest. We downloaded the data ($\text{group_project}_{id}$, project_task_{id}, project_{id} and user_{id}) for 10,000 users for Sep 2013 and Jan 2014, May 2014 and Sep 2014 from SourceForge repository based on multidimensional table shown in Fig. 4. There is no OSS data collected by the University of NotreDame after Sep 2014.

Project_{id} represents just name and id of its initiator, whereas project task shows the number of developers contributing to the particular task. We use task_{id} to find out number of developers contributing to the same task. As already stated, 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.

According to all explanations, different users intra and inter groups are connected by project tasks and are assigned $\text{group_project}_{id}$ by SourceForge Administration.

3.2. Measures

In the complex network of open source software (OSS), a group (team of developers) is represented by $\text{group_project}_{id}$ associated with each project. As previously mentioned, intragroup tie represents developers working on common tasks within a group, and intergroup coupling represents developers contributing to common tasks between two groups (measured by number of intergroup ties); whereas intragroup density is measured by number of intragroup ties over total number of developers, and represent developers contributing to common tasks within a single group. However there are other variables which influence on the output, as we explained in the model design section how group size, degree centrality and developer betweenness could affect intergroup coupling.

In order to show the relationship between intergroup coupling and intragroup density and other explanatory variables, we use regression modelling below, as parameterized in Table 2.

$$Y1 = a_0 + a_1X1 + a_2X2 + a_3X3 + a_4X4 + a_5X5 + \varepsilon_t \quad (6)$$

where $S_t(task_{id}|user_{id}\varepsilon inter - groups)$ represents sum of intergroup ties at time t; whereas $S_{t-1}(task_{id}|user_{id}\varepsilon intra - group)$ represents sum of intragroup ties at tie t-1. Betweenness and degree centrality are already defined. In addition, $S_{t-1}(project_{id}|user_{id})$ represents number of projects associated with a user at time *t-1*, and $S_t(user_{id}|group_{id})$ represents sum of developers within a group at time *t-1*.

The influence of group structure on intergroup coupling may depend on group size in addition to group tie density. As previously stated, we conjecture that more projects associated with a user and more developers within a group could indirectly affect the probability of having more intergroup ties; however our hypothesis and theoretical analysis is based on direct impact of intragroup density, centrality and betweenness on intergroup coupling. A high developers' degree centrality for a project implies that developers of different projects work on a large number of projects simultaneously, resulting in a more structurally diverse team; this could facilitate exchanges between those who are connected could also affect intergroup coupling. In addition, intragroup tie density is measured by sum of intragroup ties over total possible ties within a group.

$$\boldsymbol{totalpossibleties} = \boldsymbol{S_{t-1}(user_{id}|group_{id})}\cdot(\boldsymbol{S_{t-1}(user_{id}|group_{id})} - 1)/2 \quad (7)$$

Finally, we have to control for the group size, $(\boldsymbol{user_{id}|group_{id}})$, in addition to number of projects associated with a user, $S_{t-1}(projects_{id}|user_{id})$, which could also affect the dependent variable.

3.3. Empirical results

We show the results obtained by applying the regression model on the OSS data for Sep 2013, Jan 2014, May 2014 and Sep 2014. In the

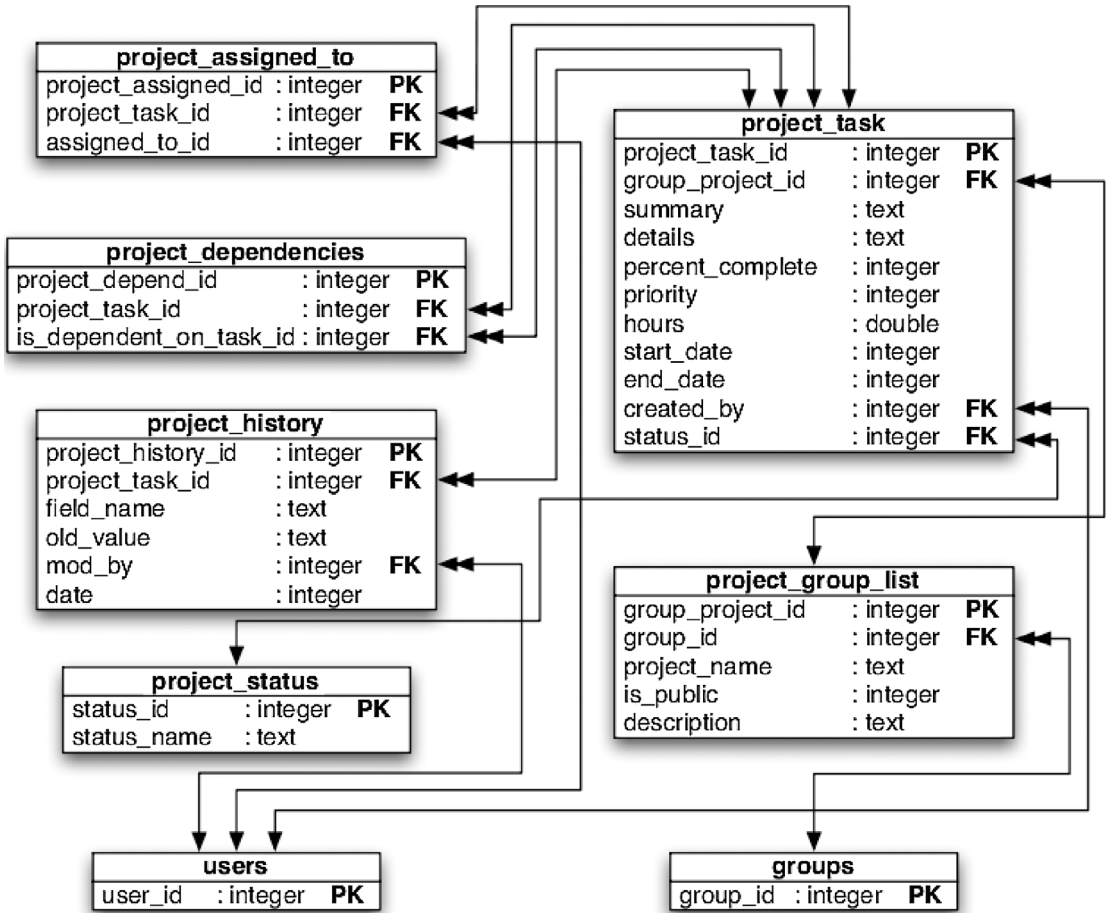

Fig. 4. illustration of relationship between different tables in sourceforge including groups, project_task, project_group_list. This picture is adopted from http://srda.cse.nd.edu/mediawiki/index.php/ER_diagrams.

Table 2
List of variables.

Variable Names	Measures	Notations
Dependent variables		
	$S_t(task_{id}\|user_{id} \varepsilon inter-groups)$	Y_1
Independent variables		
	$S_{t-1}(task_{id}\|user_{id} \varepsilon intra-groups)$	X_1
	$S_{t-1}(betweenness)$	X_2
	$S_{t-1}(degreecentrality)$	X_3
	$S_{t-1}(user_{id}\|group_{id})$	X_4
	$S_{t-1}(projects_{id}\|user_{id})$	X_5

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 are 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. But what is this amount of time? If time *t* is set to Sep 2014, then time *t-1* is 1 month before, 1 year before or 4 months before. As we have checked the data, there is very little change happing during 1 month in independent and dependent variables; however this change is noticeable within 4 months. We then did the analysis for 8 months and 1 year lagged time to examine its robustness.

As shown in Table 3, time lag: 05/2014–09/2014, intragroup density has positive and significant influence on intergroup coupling ($a_1 = 0.107$, and p-value = 0.007) within 4 months lagged time; therefore, the hypothesis is supported. Moreover, betweenness has insignificant influence on intergroup coupling ($a_2 = 0.237$, and p-value = 0.055); degree centrality has significant but negative influence on intergroup coupling ($a_3 = -0.471$, and p-value = 0.010), indicating that users with high degree centrality do not participate in intergroup projects, rather collaborate more with other developers for projects within a group. Number of users has significant influence on the dependent variable ($a_4 = 0.990$, and p-value = 0.113). In addition,

Table 3
Calculation of coefficients 05/2014–09/2014.

Intergroup tie	*Coef.*	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5.409	17.840	0.303	0.762	−29.564	40.383
intragroup	0.107	0.040	2.695	0.007	0.029	0.185
betweenness	0.237	0.052	4.555	0.055	0.135	0.339
centrality	−0.471	0.182	−2.583	0.010	−0.829	−0.114
#users	0.990	0.399	2.481	0.013	0.208	1.772
#projects	−3.201	17.821	−0.180	0.857	−38.136	31.734

Table 4
Calculation of coefficients 01/2014 − 09/2014.

Intergroup tie	*Coef.*	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95.0%*
Intercept	−7.059	17.912	−0.394	0.694	−42.175	28.056
intragroup	0.186	0.040	4.671	0.000	0.108	0.264
betweenness	0.190	0.052	1.734	0.083	−0.012	0.192
centrality	−0.298	0.182	−1.638	0.011	−0.654	0.059
#users	0.938	0.398	2.359	0.018	0.159	1.717
#projects	9.708	17.890	0.543	0.587	−25.363	44.780

number of projects associated with $user_{id}$ has insignificant influence on intergroup coupling ($a_5 = -3.201$, and p-value = 0.857).

This implies that more projects do not have any influence on intergroup coupling.

As shown in Table 4, for lagged time of 01/2014–09/2014, again intragroup density has positive and significant influence on intergroup coupling ($a_1 = 0.186$, and p-value = 0.000) within 8 months lagged time; therefore, the hypothesis is supported. Similar to the case with 4 months lagged time, betweenness has insignificant influence on intergroup coupling ($a_2 = 0.190$, and p-value = 0.083); however degree centrality has significant but negative influence on intergroup coupling ($a_3 = -0.298$, and p-value = 0.011), indicating that users with high degree centrality do not participate in extra (inter) group projects. Again, number of users has significant influence on the dependent variable ($a_4 = 0.938$, and p-value = 0.018), finally number of projects associated with $user_{id}$ has insignificant influence on intergroup coupling ($a_5 = 9.708$, and p-value = 0.587).

As shown in Table 5, for lagged time of 09/2013–09/2014, similar to both lagged time of 4 months and 8 months, the positive influence of intragroup density on intergroup coupling is supported with a close value coefficient ($a_1 = 0.111$, and p-value = 0.005). Degree centrality has negative influence on the dependent variable, whereas number of users has positive influence on the dependent variable. Based on the results within the three tables provided, there is no big difference in the outcome based on the three lagged times of 4 months, 8 months and 1 year, however in the 1 year lagged time, the influence of betweenness on intergroup coupling becomes significant ($a_1 = 0.132$, and p-value = 0.011), which we doubt its precision, and therefore we propose 8 months lagged time as appropriate.

As discussed in the model design section, we expect that the higher the density inside a group is, the more probable it is that intergroup coupling grows.

A summary of the results is that, when users in a group have a lot of intragroup tasks to contribute to, given constant number of users (i.e. intragroup tie density is high), the users would be more likely to contribute to tasks in other groups. In other words, tie density or knowledge transfer within dense groups positively influence on intergroup coupling or knowledge transfer between sparse groups.

In the case of scarcity of resources, e.g. if number of users increases, given same number of tasks to contribute to, the group tie density decreases. This will negatively influence on intergroup coupling, i.e. users would less likely contribute to tasks in other groups.

Some variables such as intragroup density have direct impact on the output, intergroup coupling or knowledge transfer, whereas some others such as group size might have indirect influence on the output. As shown in Fig. 1, the number of developers within a group ($user_{id}|group_{id}$) directly influence on intragroup density, in addition it could influence on intergroup coupling. This result corresponds to the outcome of a statistical paper by Behfar and Behfar (2016), which statistically discusses impact of intragroup tie on intergroup tie strength.

Table 5
Calculation of coefficients 09/2013 − 09/2014.

Intergroup tie	*Coef.*	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	−3.168	12.628	−0.251	0.802	−27.924	21.588
intragroup	0.111	0.040	2.787	0.005	0.033	0.188
betweenness	0.132	0.052	2.549	0.011	0.031	0.234
centrality	−0.242	0.182	−1.331	0.020	−0.598	0.114
#users	1.707	0.397	4.294	0.000	0.928	2.486
#projects	4.588	12.599	0.364	0.716	−20.110	29.286

These results have also 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 more knowledge transfer between groups, i.e. in order to achieve more knowledge transfer between groups, the project manager needs to target number of developers within each group. Moreover, more degree centrality of developers has negative influence on intergroup coupling or knowledge transfer between groups. In addition, the number of developers contributing to particular project tasks implies how popular each project task is; therefore attracting more number of developers contributing to particular tasks which indicates more intragroup coupling, leads to more knowledge transfer between groups.

4. Conclusion

Knowledge management among open-source-software (OSS) teams has been noticed in the literature e.g. Joshi and Sarker (2006) discussed factors influencing knowledge transfer within development teams. Ojha (2005) also discussed knowledge sharing between team members based on similarity-attraction paradigm. In this study, we attempted to distinguish between the factors affecting knowledge transfer within groups as opposed to between groups. To our best knowledge, there has been no study to investigate the relationship between dense and sparse network structures and whether knowledge transfer in dense network (inside groups) has an influence on knowledge transfer in sparse network (between groups). This study demonstrates that knowledge transfer within groups could influence knowledge transfer between groups.

In order to investigate how intragroup density affects intergroup coupling, we used the link cost/benefit method introduced by Jackson and Wolinsky (1996) using utility function for each project based on benefit and cost of new link formation. We showed that when initial link is formed between two groups, subsequent link formation is always cost-wise beneficial to be formed, when there is a higher intragroup density. One can explore this more, and investigate the causal relationship in more detail in that how dense a group should be? Is there a threshold, upon that all intergroup connections are automatically formed; and below that threshold, no extra intergroup connection is formed. Jackson and Wolinsky (1996) discussed unique strongly efficient network where for the case of a complete or fully connected graph, $\rho - \rho^2 > c$ should hold true, however for a condition that all nodes of a network is directly connected (a start encompassing everyone) but is not a complete graph, then $\rho - \rho^2 < c < \rho + ((N-2)/2)\rho^2$ should hold true. This is the threshold which shows how dense a group should be, within which all nodes are directly connected.

Another important contribution of our paper is an empirical analysis validating the relationship between intragroup density and intergroup coupling. For this purpose, we used the complex network of open source software (OSS) as the domain of interest, and collected OSS project data from SourceForge.Net. We showed the results obtained by applying the regression model on the OSS data for Sep 2013, Jan 2014, May 2014 and Sep 2014. In the regression model, we used lagged explanatory variables because of possible existence of simultaneity between dependent and independent as well as rationality that the impacts of group structure on intergroup coupling require a certain time lag before they take place. The results of the regression model showed that intragroup density has a positive and significant influence on

intergroup coupling; therefore, the hypothesis is supported. Moreover, betweenness has insignificant influence on intergroup coupling; and degree centrality has a significant but negative influence on intergroup, indicating that users with high degree centrality do not participate in intergroup projects, rather collaborate more with other developers for projects within a group. Moreover, number of users has a significant influence on the dependent variable, wheras number of projects associated with each group has insignificant influence on intergroup coupling. Our results demonstrate that when users in a group have a lot of in-group tasks to contribute to, given number of users constant, i.e. intragroup tie density is high, the users would be, relatively speaking, more likely to contribute to tasks in other groups. In other words, tie density or knowledge transfer within dense groups positively influence on intergroup coupling or knowledge transfer between sparse groups.

Acknowledgements

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

Abrahamson, E. (1991). Managerial fads and fashions – The diffusion and rejection of innovations. *The Academy of Management Review, 16*(3), 586–612.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly, 45*, 425–455.

Antwerp, M. V., & Madey, G. (2010). The importance of social network structure in the open source software developer community. *Proceeding of the 43rd Hawaii International conference on system sciences*.

Baer, M., Evan, K., Oldham, G. R., & Boasso, A. (2015). The social network side of individual innovation. A meta-analysis and path-analytic integration. *Organizational Psychology Review, 5*, 3191–3223.

Baer, M. (2010). The strength-of-weak-ties perspective on creativity: A comprehensive examination and extension. *Journal of Applied Psychology, 95*(3), 592–601.

Behfar, S. K., & Behfar, Q. (2016). Intragroup density predicting intergroup tie strength within open-source-software collaboration network. *Advances in Intelligent Systems and Computing,* 165–173.

Behfar, S. K., Turkina, E., & Burger-Helmchen, T. (2017). Network tie structure causing OSS group innovation and growth. *Journal of Problems and Perspectives in Management, 15*(1), 7–18.

Burt, R. S. (2000). The network structure of social capital. In I. Robert Sutton, & B. M. Staw Greenwich (Eds.). *Research in organizational behavior* (pp. 345–423). Conn: JAI Press.

Burt, R. S. (2002). The social capital of structural holes. In F. Mauro Guillén, R. Collins, P. England, & M. Meyer (Eds.). *The new economic sociology* (pp. 148–192). New York: Russell Sage Foundation.

Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology, Vol. 94*, S95–S120 Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure.

Crowston, K., Annabi, H., & Howison, J. (2003). *Defining open source software project success paper presented at the international conference on information systems (ICIS),* 1–14.

Florida, R. (1995). Towards the learning regions. *Futures, 27*, 527–536.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*, 1360–1380.

Grewal, R., Lilien, G. L., & Mallapragada, G. (2006). Location, location, location: How network embeddedness affects project success in open source systems. *Management Science, 52*(7), 1043–1056.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization units. *Administrative Science Quarterly, 44*, 82–111.

Hinds, D., & Lee, R. M. (2008). Social network structure as a critical success condition for virtual communities. *HICSS'08: Proceeding of int. conf on system sciences* p. 323.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*, 88–115.

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory, 71*, 44–74.

Jackson, M. O. (2004). A survey of models of network formation: Stability and efficiency. In G. Demange, & M. Wooders (Eds.). *Group formation in economics: Networks, clubs and coalitions.* (pp. 1–62). Cambridge: Cambridge University Press.

Joshi, K. D., & Sarker, S. (2006). Knowledge transfer within information system development teams: Examining the role of knowledge source attributes. *Decision Support Systems, 43*(2), 322–335.

Kogut, B., & Zander, U. (1995). Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organization Science, 6*(1), 76–92 Focused Issue: European Perspective on Organization Theory.

Montazemi, A. R., Siam, J. J., & Esfahanipour (2008). A. Effect of network relations on the adoption of electronic trading systems. *Journal of Management Information Systems, 25*, 233–266.

Nelson, R. E. (1989). The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal, 32*(2), 377–401.

Ojha, A. K. (2005). Impact of team demography on knowledge sharing in software project teams. *South Asian Journal of Management, 12*(3), 67–78.

Rost, K. (2011). The strength of strong ties in the creation of innovation. *Research Policy, 40*, 588–604.

Singh, P. V. (2011). The small world effect: The influence of macro level properties of developer collaboration networks on open source project success. *ACM Transactions on Software Engineering and Methodology, 20*(2).

Singh, P. V., Tan, Y., & Youn, N. (2011). A hidden Markov model of developer learning dynamics in open source software projects. *Information Systems Research, 22*(4), 790–807.

Tsai, W. (2000). Social capital, strategic relatedness and the formation of intraorganizational linkages. *Strategic Management Journal, 21*(9), 925–939.

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), 996–1004.

Van Oort, F. G., & Atzema, O. A. L. C. (2004). On the conceptualization of agglomeration economies: The case of new firm formation in the Dutch ICT sector. *The Annals of Regional Science, 38*(2), 263–290.

Walker, G., Kogut, B., & Shan, W. J. (1997). Social capital, structural holes and the formation of an industry network. *Organization Science, 8*(2), 109–125.

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.