Open Source Foundations as Social Innovators in Emerging Economies: The case study in Brazil

Bert-Erboul, Clément (1); Ferreira dos santos, Jean Carlos (2); Vonortas, Nicholas (3)

1: UNICAMP INSYSPO, Brazil; 2: GEICT - Grupo de Estudos Interdisciplinares em Ciência e Tecnologia; 3: The George Washington University

Abstract

This paper highlights the social innovators' organisation through an open source community. Our goal is to understand if the institutional origins matter in terms of involvement? With social network analysis we study the institutional and structural heterogeneity in a community. The case study of Brazilian social innovators in the Open Knowledge Foundation gives us the very opportunity to study over time a full community with contributors' real identity consistent with their offline occupation. Our analysis is based on data from a mailing list archive to rebuild each year the community social network. We contribute to the open source field by studding interactions in an open community's contributors from different organisations such as academics, companies, non-profit institutions and individual involvement. Our results show the importance of groups' centrality degree. We distinguish two types of institutional contributors: insiders and outsiders. Insiders represent a community of belonging deeply involved in the online activities. The outsiders contribute in a more silent way. Even if their demography is important, their community of belonging do not have a strong collective structural online presence. The implication of these results completes the point of view concerning the core/periphery management in distant communities.

Key words: Social networks, heterogeneity, open innovation, social innovation, institution.

1 Introduction:

The economic models of Internet activities are shaped by two main questions. First, how the big platforms such as Google, Apple, Facebook, Amazon or Netflix produce and provide innovation to users (Barrett et al., 2016; Bosch-Sijtsema and Bosch, 2015; Gawer and Cusumano, 2014)? Second, how users innovate when the big platforms are not providing the service they want, considering political goals they want to reach and the alternative property rules needed (Bauer et al., 2016)? We focus our work on the second question related to the uneven development strategy. We use the theoretical framework of the economics of innovation and social network analysis. We study the institutional and structural heterogeneity

of contributors in the Brazilian Open knowledge community to have a better understanding of the open source communities' social embedment (Grewal et al., 2006).

Our goal is to understand if the institutional origins matter in terms of involvement? Do the central contributors have the same base and the peripheral people have another? The paper concentrates on the case of social innovators in Brazil creating a new activity by remixing the open source communities' model. We use this case to develop our main contribution in terms of analysing a type of actors that are often mentioned but rarely studied: the open source foundations (Liu et al., 2016; O'Mahony, 2005). These organisations contribute to the open source ecosystem by writing and defending licences, and by advocating and proselytising people and organisations. Furthermore, the foundations contribute to renewing the open community management model and the organisational types involved in open source projects (O'Mahony, 2007).

The open source institutional ecosystems have already evolved since the birth of the movement at the end of the 1980s (Gehring, 2006). Open source had originally been designed by informatics engineers. Now open source innovation has crossed to different sectors (Bonaccorsi et al., 2016) involving collaboration between academia and the private sector (Spanjol et al., 2014). It has also become a valuable asset for developing countries in strategic social sectors such as health (Chavez and Kovarik, 2017) and public administration (Ghosh and Kumar Das, 2007).

Even though the literature on social innovation is not unified, open innovation is an important field of social innovation (van der Have and Rubalcaba, 2016). The core contributors are using the sharing economy as a viable business. Furthermore their activities are rooted in the social context they are embedded in, trying to improve social needs such as education, the health system, or political organisation (Mulgan, 2006). They are not the inventors of what they use, but they remix previous concepts in original ways (Stanko, 2016). These people play the role of "knowledge activists" in organisations (Von Krogh et al., 2000). They stimulate creativity and find economical ways to maintain activity. They are brokers in knowledge circulation (Edquist, 2010) between different areas.

The people merging from different fields is highly important in the innovation process (Etzkowitz and Leydesdorff, 2000). We analyse this convergence by observing and analysing one of the 55 Open Knowledge Foundation communities: its Brazilian chapter. The Open Knowledge Foundation (OKFN) was founded in 2004 by the economist Rufus Pollock(Pollock, 2008) as a pro-openness organisation (Molloy, 2011) and its Brazilian chapter (OKFNBr) became quickly the most active in the world. The analysis of the OKFNBr

shows the Brazilian open source community institutionalisation process in specific areas such as public administration. This group contribute to creating new tools to public administration like data visualisation and ergonomic digital framework (Fig. 1).

Figure 1 Data visualisation in mosaic of Public budget on the Diretoria de Análise de Políticas Públicas website[1]

We are studding the main public mailing of the Brazilian Open Knowledge Foundation where these tools are discussed and conceived. This community of activists gives us the very specific opportunity to analyse the contributors with their real identity and refine their institutional attachments such as universities, non-profit organisation, companies or individual enrolment. We compare structural importance of these categories in terms of the brokerage and contributions. During the early years of the community there is an institutional heterogeneity and different groups contribute collectively to the discussions. Over time this collective heterogeneity decrease but some leaders from different institutions constitute the core and the community is led by the OKFN members. The last result concerns the role of academics. They are numerous in the social network but not collectively well connected. This result informs on a contribution to the community outside of the OKFN organisation. To summarise our point of view, we distinguish two types of open source projects contributors. A first one is an insider. He dedicated to an organisation and collective projects. The second one is an outsider. He is acting in more traditional and general organisations and tends to implement isolated activity related to different organisations.

The paper has five sections. The first part presents the conceptual background from heterogeneity and centrality studies in open source research field and frames our hypothesis.

---

[1] http://dapp.fgv.br/mosaico/mosaic_f/2014

The second part is dedicated to the field work presentation. Third we present the methodology and data. Finally, in a fourth part we resume our results. The end of the paper is a discussion of results, limits and perspectives.

2 Conceptual background and hypotheses: institutional and structural heterogeneity in open communities.

The open source story is framed by its opposition to the proprietary model especially in the software industry. Nevertheless, the free/open source movement is a global social movement with specific ethics and aesthetic (Coleman, 2013). The spread of these values is in part due to the foundations situated between users and makers and contribute to the open source communities' institutionalisation. Our work is focused on the institutionalisation as Lynne Zucker defined such as a transmission of practices with a possible reinterpretation of meanings (Zucker, 1977). In other words, all the groups who are using free/open licences do not have the same ideology as the first free software developers in the 1980th. For this reason the literature about open source communities' institutionalisation is shared between the homogeneous and the heterogeneous approach (Markus, 2007).

The homogenous approach focuses on projects that succeed economically and technically. Steven Weber (Weber, 2000) or Eric Raymond (Raymond, 1999) had early analysed institutional imitation and open source model spread close to the homomorphism process observed by Paul Di Maggio and Walter Powell (DiMaggio and Powell, 1991). This model is pointed on economic efficiency in a specific institutional environment.

The heterogeneous approach is focusing on the broad free software movement history. From this point of view there are not only economic incentives, production, and use in open movement. The first social innovators commitments are usually motivated by moral concerns. Then, the second step concerning the activism institutionalisation and economic issues (Roberts and Woods, 2005). The research in the open source studies confirm this activist root (Alleyne, 2011; Ghosh, 2005). This process is documented in a dedicated literature about the heterogeneity of open source communities' institutional shapes(O'mahony and Ferraro, 2007) and their change (Fitzgerald, 2006; Schweik, 2003).

Below, we first present how the institutional open source project ecosystem can be shaped by inter-organisational agreements, by the national context, and by individual behaviour. We then present how social network analysis can capture these dynamics by focusing on structural dimension.

Institutional heterogeneity in the open community: Organisation, national context and individuals

Innovation in informatics had challenged organisation theories early on. The first open innovation process questioned the merging of heterogeneous needs, ideas and work habits. In informatics, the intermediaries between organisations are historically important. In the fifties the Society to Help Avoid Redundant Effort (SHARE) was the first world wild source code-sharing community before the birth of the free software movement in the eighties (Armer, 1980; Mounier-Kuhn, 2011). This non-profit organisation created technical norms, with the support of aeronautic, energy and telecommunication companies. Some common engineering problems linked these organisations in the use of calculators. The existence of SHARE was due to the antitrust law forbidding alliances between firms in charge of strategic businesses such as energy and telecommunications after the economic crisis in 1930.

Since the 1990s, the institutionalisation of globalised free/open source foundations has created a new context. The open source universe is not homogenous any longer (if it ever were) even if the users share some similar mental representation about technology policy created by the "voices from the open source revolution" such as Richard Stallman, Linus Torvald, Eric Raymon, or Tim O'Reilly (DiBona and Ockman, 1999).

Open source sympathisers, have numerous motivations and profiles (Bitzer et al., 2007; Bonaccorsi and Rossi, 2003; David and Shapiro, 2008). Furthermore these differences are observable between individual and organisation strategies (Bonaccorsi and Rossi Lamastra, 2004) and between organisations themselves (Henkel, 2006).

Structural Heterogeneity in open communities: involvement, hierarchy and socialization

Inter organisation innovation and open source process create large network involving industry, university and government (Gustafsson and Jarvenpaa, 2017). This complex network structure is a necessary ecosystem for communities. The multi commitment overlapping in projects improve the communities' survival (Zhu et al., 2014).However the community social network is often reduced to the one of the project's members and not the entire ecosystem.

Different strategies exist to study online collaboration in open source communities. The learning process of the new comers, gives the opportunity to analyse how the language, and the skills are enriched once he is embedded (Steinmacher et al., 2015; Von Krogh et al.,

2003). This approach is seeking the epistemic organisation theory where "the involved parties contribute to, and learn from the community" (Edwards, 2001).

In addition, a part of the literature is focusing on the most central contributors and the commitment diversity(Daniel et al., 2013). This approach observe the importance of the founders' social capital (Mallapragada et al., 2012). The leader point of view insists on the equilibrium between technical and social skills that central contributors have to attract and to maintain the community's activity (Fleming and Waguespack, 2007).

The diversity between newcomers and central contributors is observed by the degree of participation and sub communities (Zhu et al., 2016). The studies analyse the core/periphery dynamics that structure the online open source communities (Crowston and Howison, 2006). Hierarchy between heterogeneous contributors can be related to their technical knowledge (Singh and Tan, 2010), or their social status considering their prestige (Stewart, 2005). The social network structure on mailing list or the source code data base, show specific socialisation in open source communities between distant contributors (Ducheneaut, 2005). One of the characteristics is the technical skills concentration and knowledge about the community history in the core contributors sub network (Crowston and Scozzi, 2008).In other words, the core contributors look like a college of experts useful to solve problems in specific fields (Faraj and Sproull, 2000)and take strategic decisions (Lazega, 1992, 2001).

Expertise centralisation is not the only centrality source. The brokers play a strategic role considering their expertise and their capability to communicate with different social environment(Tan et al., 2007). The literature shows that the core contributors have to improve the social network for themselves but also for the community to maintain the collective activity (Ganley and Lampe, 2009). In that perspective, it seems important to focus both on the contributors' origin and their relationships' intensity with their initial group and with the entire community.

Hypothesis

We focused our conceptual background on the institutional heterogeneity and centrality in open source communities because behave in a distant organisation requires a specific sociability that is not equally distributed (Wei et al., 2016). Our review shows that it exists different motivations and ways to use the open source licences and ideology. Furthermore the literature concerning the open social network highlights the ordeal of

multiple influences on the communities. However there is no clear insight concerning the impact of institutional origin of contributors and their involvement in the community.

The wide use of open source innovation practice has changed the problematic of the communities' governance previously focused on software quality. Now, inter-organisational agencies (Benkler, 2006), the externalities (Dalle and Jullien, 2003) and ecosystem (Chesbrough, 2003) shape the open source innovation academic field in more broad way. Open innovation is a process in technological activities but also in social and development policy. Our question is not to evaluate how the open source movement has changed the offline policies (Rushkoff, 2003)?

The free/open source practices' institutionalisation creates new innovation dynamics. Initially build by community implementing bottom/up process, now some institution more traditional such as governmental agencies are leading open policies (Fuggetta, 2003; Lewis, 2007). Some new open concepts are already mainstream, such as open government. This context should create a new innovation flow through top/down dynamics. For these reasons we frame our hypothesis around the question concerning institutional heterogeneity. The first hypothesis has a collective point of view concerning the contribution in the community. The second hypothesis highlights the difference at the individual level.

H1 The open source project is built by contributors from heterogeneous institutions

The open source ecosystem innovation is usually presented as a direct meeting between producers and users (Lerner and Tirole, 2002). This system does not need big majors or publishers to select the good and bad contributions (Stallman, 2001). The peer communities are efficient curators to provide competitive content front of the proprietary concurrence in diverse highly innovative sectors such as operation system or database management.

However, the new actor in the open activity creates a heterogeneity that implied a new regulation way. In that perspective open production seems to be an answer to bureaucracy malfunctions (Kreiss et al., 2011). We do network centralities analyse by taking into account the institutional contributors' origins to observe the community heterogeneity. This approach contributes to know if online relationship reproduce some offline social structure or hierarchy or if they have their own rules (Wellman and Haythornthwaite, 2008).

H2 The open source project counts structural heterogeneous contributors from different institutions.

The presence of heterogeneous contributors is one of the biggest ordeals for distant innovative community. In the software development case, a common mental representation is needed to avoid accidents because the software is intangible goods. Strong communication and organisation should compensate the impossibility to represent the product of their action (Brooks Jr, 1956). For that reason lots of studies had already focus on the developers' mailing list and their repository where are stored their contributions (Mockus et al., 2002). These media are strategic in organisation and contain a part of the developers' everyday life activities.

Recent research (Faraj et al., 2016) insists on the sociability in the open source communities. The tacit knowledge seems an important factor in coordination (Von Krogh et al., 2000). Classic tacit knowledge definition in economics is framed by Michael Polanyi's approach resumed by the sentence "*we can know more than we can tell*"(Polanyi and Sen, 2009). This approach points the capability to verbalise information and implied more communication between heterogeneous people from further institutions. The observation of the structural network between the main open source contributors from different institutions should reveal if it exists communication needs.

3   Field work

The Brazilian OKFN community counts two levels: one online, and a second offline. The online community establish internal rules, and common projects. The offline level is marked by local product digital implementation and conferences. The online discussions shape the offline project. The short history and the project's resume presented below illustrate this organisation.

The first Brazilians people who participate to OKFN in 2009 subscribed to the world OKFN open government mailing list in the context of political scandals during national elections and a period of important spending in public infrastructure. Openness has been seen as a means to improve state bureaucracy and the public funds management. These first volunteers are not public choice theory advocates. They look at how political elite use public money. This monitoring is motivated by a wish of control of the connections between state bureaucracy and private sector.

The OKFN institutionalisation in Brazil is concomitant with a national legislative change. In 2011 the Federal Information Access Law (Lei de Acesso a Informação) established openness principle in the state administration digital data. This data give the opportunity to citizens to investigate stratospheric amount of data concerning the cities

hardware infrastructure and bureaucratic institutions. In the Brazilian cities the open source is deeply embedded in the state bureaucracy, transportation, bank, school, security and animated by the rise of local elites with academics, developers, artists, and politicians. After few years of social and nonprofit experimentation, some companies start viable economic activities and extend the first movement.

The main project of the OKFN Brazil is named "Gastos Abertos[2]" and started in 2012. This project is built by the strategic alliance between the OKFN crew, a Brazilian startup "AppCivico" located in São Paulo and Google as the economic funder of the experience. The objective of OKFN non-profit organisation is to promote information as public commons and promote open tools. The AppCivico provides data visualisation of public expenditures to city halls such in São Paulo city or some NGOs. Google develops partnerships with emerging technological partners in philanthropic view and feed an internal policy of innovation by hosting heterogeneous projects from all over the world.

In these projects the innovation concerned design and ergonomic gain to manage the avalanche of data produce by the digitised bureaucracy and public services in the biggest cities of South America located in Brazil. This alliance looks like a win/win agreement. Public servants had seen evolved their tools, and citizen have a new view on their institutions. Furthermore companies found a business model by providing nonexistent skills in public administration because of the recent emergence of the data mining.

After these experiments other projects are led by the AppCivico with the State of São Paulo such as the "Observatório de Indicadores da Cidade de São Paulo[3]". This product is a complex monitoring of 280 economic and social indicators (the number of inhabitants per mortality rate due to diseases, the proportion of adolescent pregnant women) with automatic layouts (Fig. 2) . The company continues to develop free framework influenced by the activist experience with OKFN.
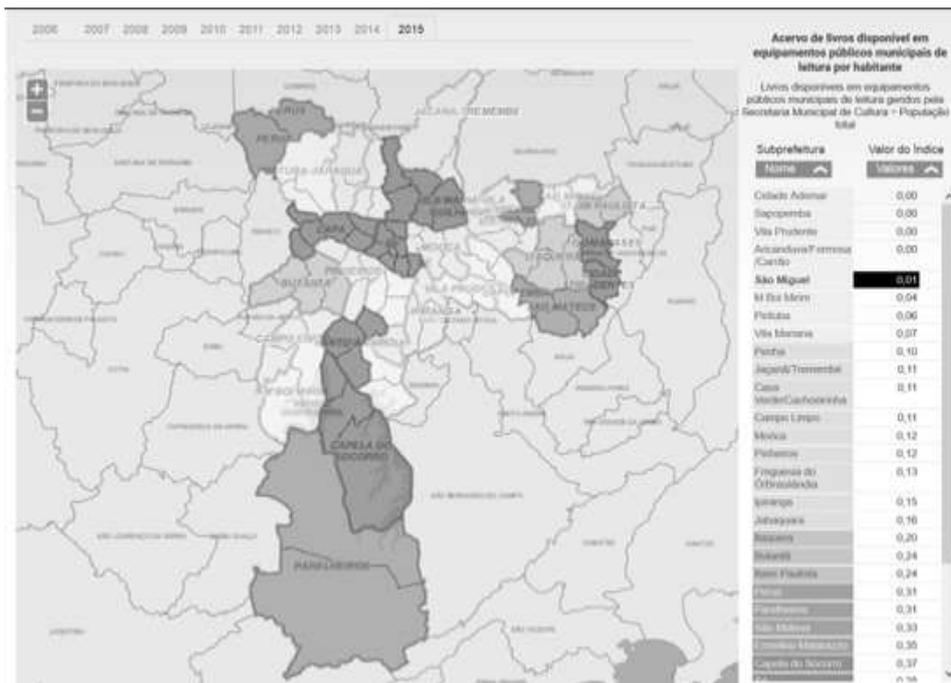

Figure 2 : Books available in public municipal reading equipment per inhabitant in the Subprefeitura of Sao Miguel in the State of San Paolo.[4]

This case shows that different ideologies feed the social innovation. In a top down dynamic we see the Brazilian state process with a bureaucratic transparency policy since the end of the dictatorship with budget divulgation or public procurement offer. In parallel an international corporation led by a liberal and technophile ideology marked by the free software culture contribute to the local ecosystem and funding initiatives. In another way, a bottom/up dynamic comes from nonprofit organisations and small local corporations to provide digital services in niches to the local government. OKFN hosts discussions from a wide community interest to contribute to these experiences or adapt them in other contexts. Inside these processes, academics contribute to events or discussion concerning community regulation or local experiments.

The mix from state to local hacker influences tends to scramble the process of community organisation. In the next part we will study how these contributors from groups are interacting online.

4 Methodology& data

Our study is a descriptive statistical analysis of 8118 messages from the main Brazilian OKFN mailing list. It contains 853 names registered from September 2011 to

---

[4] http://observasampa.prefeitura.sp.gov.br/index.php/indicadores/indicadores-por-regiao/

December 2016 (Tab. 1)[5]. We are using the variables of institutionalisation process and the structural indicators. The table behind summarises the indicator we are defining below in this methodological part (Fig. 3).

Tab 1 Effectives on the Brazilian OKFN mailing list

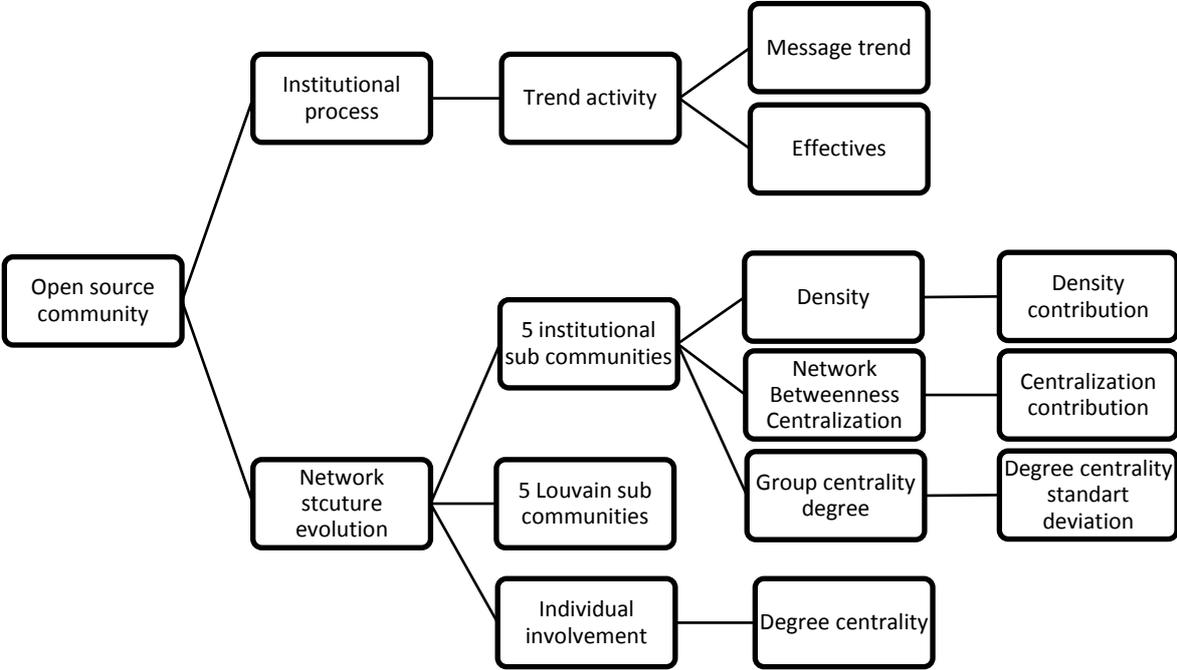|	2011		2012		2013		2014		2015		2016	
Open knowledge	7	18%	16	14%	26	10%	47	13%	35	12%	21	24%
Companies	2	6%	4	3%	13	5%	28	8%	24	8%	4	4%
Individual commitment	15	40%	63	49%	133	46%	177	43%	131	43%	27	30%
Universities & Governmental agencies	7	18%	14	11%	46	18%	65	17%	56	18%	20	23%
Nonprofit organisations	7	18%	27	23%	54	21%	75	19%	57	19%	16	19%
TOTAL	38	100	124	100	272	100	392	100	303	100	88	100

Figure 3 Analytical framework

Most of research look at latent relationships between contributors from structural indicators measured in the discussion dataset (Bird et al., 2008). This approach implies to create block model and sub communities to evaluate the structural equivalence and the group's membership of the contributors. We change from the classic method. We build a priori contributor categories to check if they have regular structural content. The updating of profiles in the social network is inspired by affiliation networks studies focused on interlocks with common members in two or more parallel networks (Carrington et al., 2005). For each

---

[5]Public archive available at : https://lists.OKFN.org/pipermail/OKFN-br/

name and each year we assign an institutional affiliation. The categories are Open knowledge foundation, Foundation & nonprofit organisations[6], Companies[7], Universities &Governmental agencies[8] and individual commitment[9].

We affiliate the names by taking into account the e-mail addressee extension, the information available on the Internet on social network websites or blogs. Real identity data are sometimes missing in traditional software communities studies. The contributors maintain their anonymity by using pseudo for reasons of patent infractions or company agreements. The political issues of the OKFN community push the contributors to use their real identity. The democratic context of the commitment gives the opportunity to access to the real contributors' names, clearly expressed on the public mailing list. This material gives us the opportunity to analyse the open community social embedment.

Between these categories we create a hierarchy, to organise data. The strongest affiliation is the one with OKFN. When some affiliations are overlapping we keep the one from whom the contributor received his/her wage if it's not OKFN. To evaluate individuals and institutional hierarchies in the networks we compare their demographic importance with their structural weight.

We compare these a priori categories with the categories built by the Louvain Community Detection algorithm and its coarsening and refinement options (Tab. 2). The Louvain algorithm creates sub communities with stronger intra-density than the enter community one. In other words, these communities are the most cohesive group in the global network. This calculation gives us the opportunity to control our institutional categorisation. The Louvain method shows that the four most active sub communities represent a majority of the population in the network list[10]. The others sub communities count one or two nodes and do not represent important components. The large communities are not institutionally nor demographically homogenous over time. Each year they are recomposed and show the complexity of strong individual involvement from different origins in the specific OKFN projects. This confirms at least two points of our research strategy. First, the division of the global network in five sub communities looks consequent with the structural organisation in five principal sub communities. Second, the importance to distinguish collective and

---

[6] It can be organization from the free software ecosystem or social foundation or hurban activist.
[7] It can be new paper corporation or software services or data mining firms.
[8] It can be Brazilian or foreign universities, Brazilian state or international institutions.
[9] It can be unemployed people and/or people with personal project.
[10] More than 60% of the entire population each year.

individual involvement is important to understand the difference between structural contribution and institutional contribution.

Tableau 2 Population in the most important Louvain method sub communities in the main Brazilian OKFN mailing list

	2011		2012		2013		2014		2015		2016	
G1	14	37%	32	26%	62	23%	43	11%	26	8%	24	26%
G2	2	5%	23	18%	29	11%	62	16%	66	21%	11	12%
G3	4	10%	22	18%	19	7%	54	14%	71	23%	26	28%
G4	11	29%	17	14%	28	10%	49	12%	23	7%	1	1%
G5	2	5%	2	2%	33	12%	37	9%	21	6%	1	1%
TOTAL	33	86%	96	78%	171	63%	245	62%	207	65%	63	68%

The analysis in our categories is mainly focus on the centrality and density calculation to observe the hierarchy and structure in the network (Freeman, 1979). Degree centrality and betweenness are the simplest and most efficient indicator to observe individual involvement and structural roles in a large media communication network. We carry out this calculation by transforming the relationship between an individual and topics by links between individuals. In other words, if A, B and C contributes to a topic we build the links: A/B, A/C, and C/B. These edges create one mode and undirected network. Each time a contributor gain a new contact his/her degree centrality grows by 1. Stronger is the topic diversity contributed, stronger is the expertise and stronger is the contributor's centrality.

The betweenness centrality is calculated in the same way with one mode dataset (Freeman, 1977). If a contributor participates in different topics, that underline his expertise and his capacity to speak with different people. We calculate the betweenness indicator by taking into how many times a node is on the shortest path between two contributors (the geodesic distance). We do not take into account some variable such as the time spent to answer or repetition of interaction in the e-mail flow (Borgatti, 2005). In our study, the expertise is more important than availability. To observe the betweenness effect at the collective level we analyse the general network betweenness centralisation (Freeman, 1979). Freeman defined this index as the "average difference between the relative centrality of the most central point, and that of all other points". For each sub institutional group we analyse the contribution to that index (see annexe).

To limit the sub communities demographic size impact on their structural centrality, we calculate group degree centrality in each institutional sub networks to obtain the group's weight (Everett and Borgatti, 1999). Everett and Borgatti define the group degree centrality as

"the number of non-group nodes that are connected to group members". This method tends to advantage the large sub communities, but the open community organisation with very central core and less committed periphery rebalancing this effect[11]. Furthermore as Everett and Borgatti recommend we normalise the degree by "dividing the group degree by the number of non-group actors".

We evaluate the network structure through the density. We calculate it, for each institutional subnetworks. The network density is evaluated by the number of ties established, reported to the potential number of links if all nodes were connected. This cohesion indicator gives the opportunity to check if some subgroup's contributors participate in the same topics and are linked to each other.

We analyse the network data with the Pajek software and the partition shapes layout to visualise the inter institutional relationships (Fig. 4). This multilevel approach gives visual information about the contribution centralisation around few contributors. The centralisation is due to the strong communication between a limited contributors number in each sub communities.

---
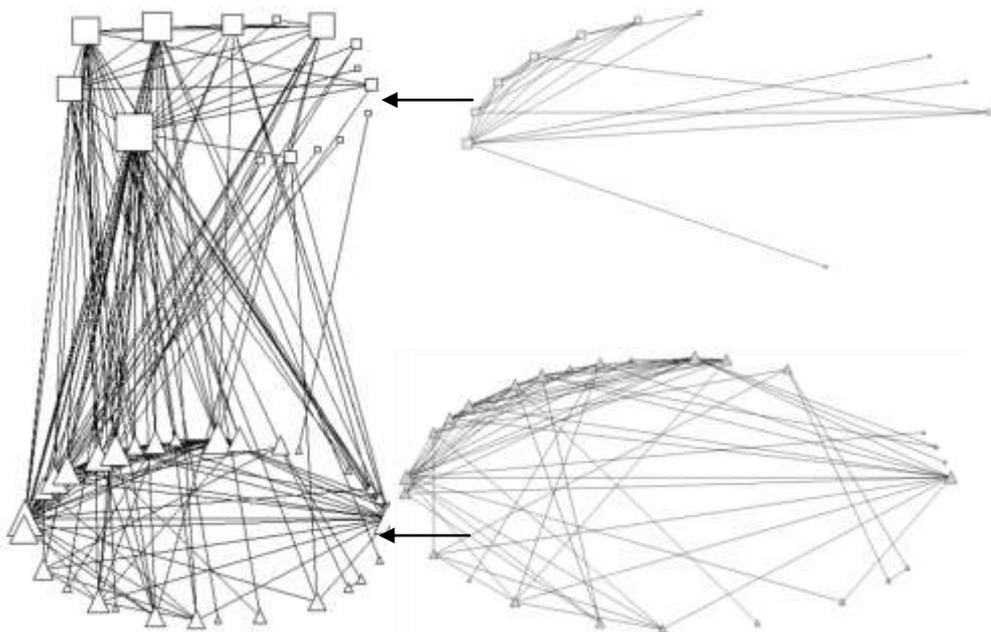
[11]This method seems limited for the group of Companies, because of a low effective and low centrality.

Figure 4: Social network between OKFN members and Nonprofits organisation in 2012 on the OKFN mailing list (n = 80)

We do a network robustness test to evaluate the contribution for each sub communities in terms of degree of centrality and density per years. Each year we reduce the entire network by one sub network (see the methodological appendix). We compare the structural indicators from the entire network with the reduced network. This test contributes to analysing how the sub network internal density impacts the entire reflexivity. We made an annual analyse of these indicators between 2011 and 2016.

The mailing list trend shows a specific period with specific mail number (Fig. 5), close to the process already observed in the literature (de Laat, 2007). We observe a progressive increase in contributions and contextual events such as "flaming" due to strong discussion on sensitive topics. During June 2014 and the August 2015 the most important discussions concerned about the ways of political action and definition of collective rules. The main theme of the collective debate is to know if the OKFN has to work with the governmental institution or beside them.

In the global mailing list the dynamic is characterised by an increase phase during the first years (between 2011 and 2014). Then there is a contributor's increase on the list and topics number increases too (between 2014 and 2015). The 2016 is marked by a strong diminution of discussion implied by the decrease of the contributors' number, the diversity of communication media, the community professionalization and the centralisation around specific contributors. These time frames create homogenous periods that we reuse to create temporal networks.
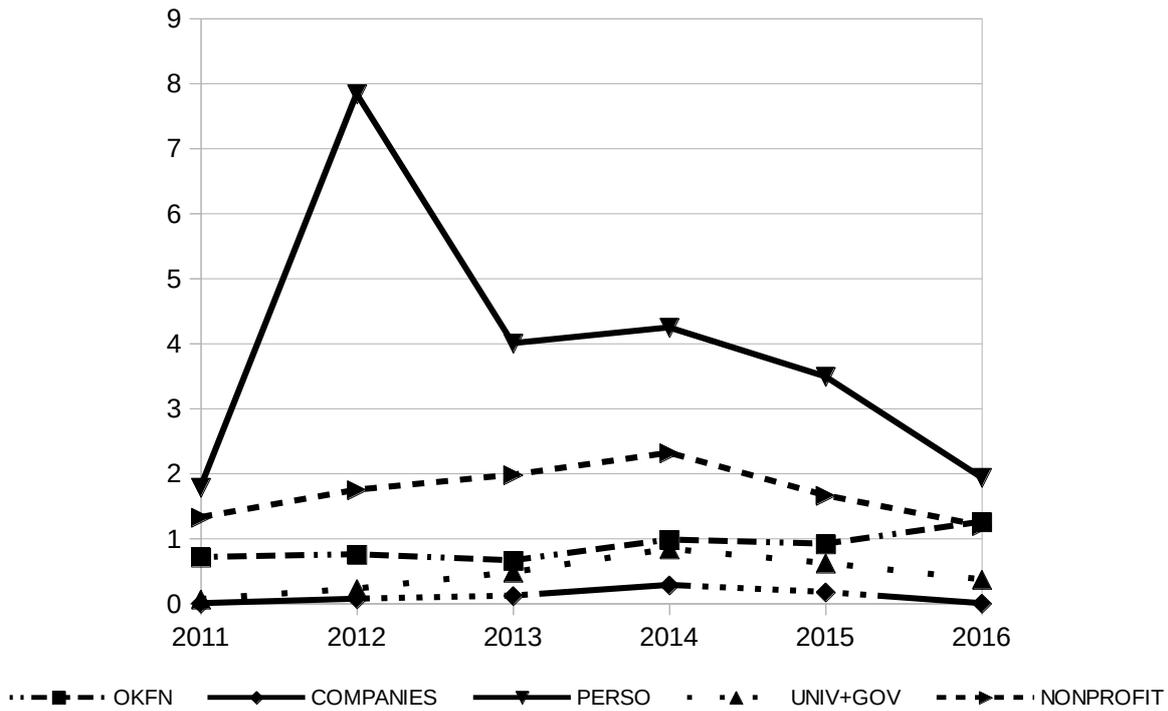
Figure 5 Brazilian OKFN mailing list trend

5   Results

H1 The open source project is built by contributors from heterogeneous institutions.
The first hypothesis contains two main possible answers. The heterogeneity is validating if the groups have similar centrality and betweenness. However, as often, centrality and betweenness are not related to the same groups.

The group centrality index shows the hierarchy of the collectives the most active in the network in terms of external contact (Fig. 6)[12]. The individual contributors and the non-profit organisation members are the most extroverted groups. They feed on the list and received answers from different categories. The OKFN members and the academics do not have a strong group centrality score even if at least the OKFN should be more central because they host the community. That reveals a specialisation in terms of topics followed in these two communities. This first result tends to underline an institutional heterogeneity of relationships but not shared by all categories.

---

[12]Excepted one value in 2016

Figure 6 Group centrality degree index

The general betweenness centralisation index gives another point of view concerning the sub community clustering hierarchy and highlights evolution (Fig. 7). The main result is the broker position of the OKFN members sub community is the mailing list. That means this group concentrate resources such as technical skills, social network and an interest in a wide open social innovation types. This involvement can concern the social use of technology or technology as a means to renew classic social and political activity. They are writing about topics requiring a deep knowledge of Portuguese language, and Brazilian institutions' specificities. The OKFBR looks like as a project incubator. The community make in connects people and originations with framed objectives (open data, accountability).

However the OKFN members did not lead the first years of the mailing list. The individual commitment seems the more important to explain the first coordination movements. The decline of individual brokers is partly due to their enlistment in the OKFN organisation and their global demographic decline. What we also observed is the low brokerage level from academics and non-profit organisations during the first years. That result is surprised by taking into account their important demographic shares. That highlight the outsider profiles, widely involved but not strongly contributory. They stay focused on their organisational projects and do not transfer them from an institution to another. The use of open source looks like more as a window of opportunity rather than a total package implementation with it ideology and specific techniques.
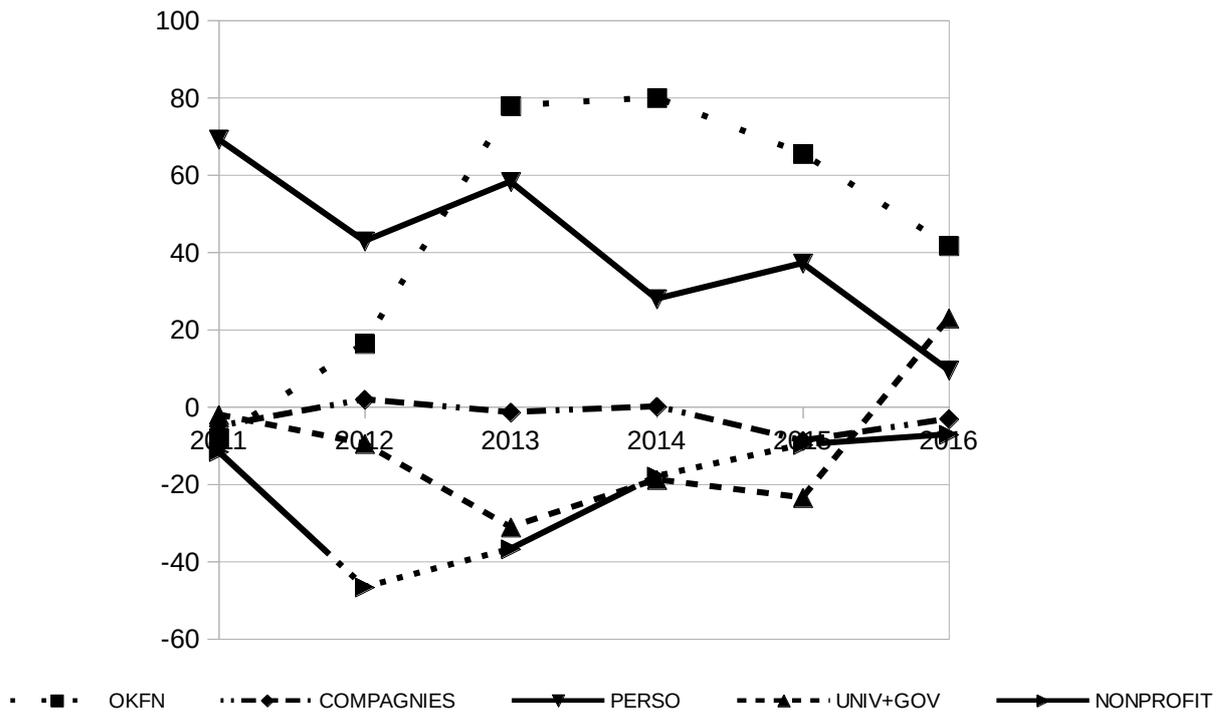
17

Figure 7 General centralisation betweenness contribution of sub networks in the Brazilian OKFN mailing list

The structural hierarchy of institutional commitment shows a dynamic hierarchy related to demographic evolution and individual transfer from an organisation to another. Furthermore, the interaction concerning the individual committers and OKFN members tend to crush other institutional involvement in terms of topic diversity and expertise.

H2 The open source project counts structural heterogeneous contributors from different institutions.

The second hypothesis counts two main possible answers. The heterogeneity is validating if we observe strong diversity between contributors in subgroups in terms of centrality. During the first hypothesis tests we have seen a limited institutional heterogeneity in the network. However, we observe involvement in the contributors' core community from different institutional origins with a high centrality degree in a classic core/periphery structure (Fig. 8).
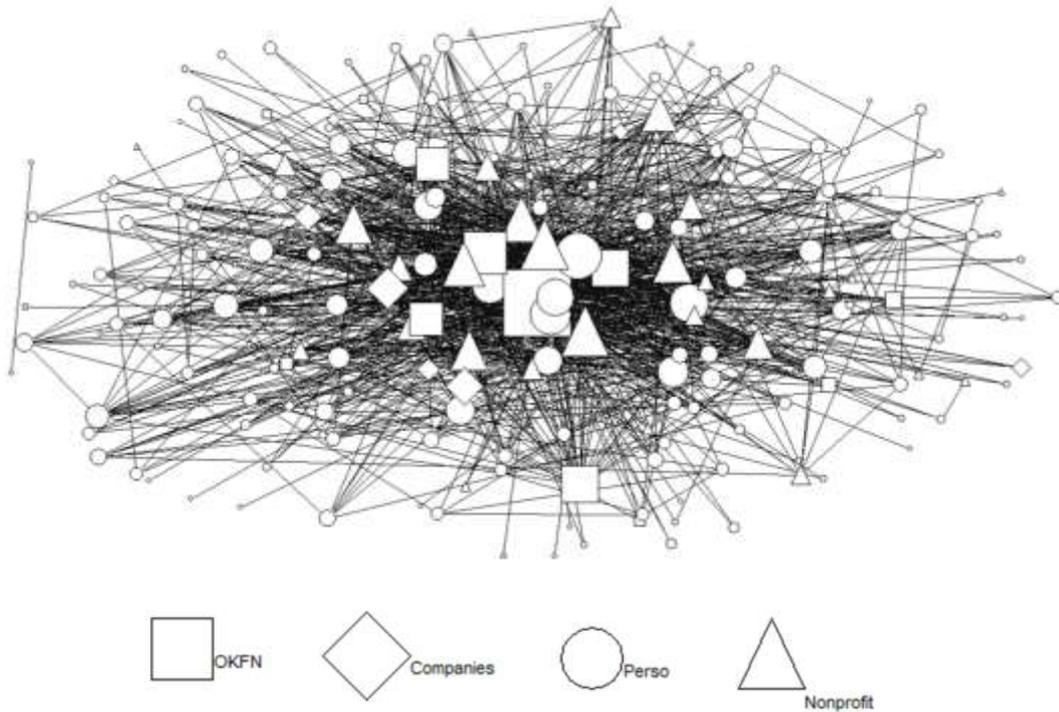
Figure 8Network of the Brazilian OKFN mailing list in 2013 with OKFN, Companies, Perso and Nonprofit categories (layout Fruchterman and Reingold, vector size: centrality degree)

The academics and the other categories such as some companies' members play an important role in the core mailing list. Even during the last period where the OKFN institutionalisation is strong, the nonprofit members' organisations are the main contributors to the mailing list topics.

As the global hierarchy evolves, the core contributors' category hierarchies also evolve. The centrality degree standard deviation underlines a horizontality in each group and a very heterogeneous starting point (Fig. 9). In a second time diversity in the contribution intensity rises in each group. The trends follow the message numbers, but the group reacts differently. The highest involvement inequality level is in the OKFN group. This result reveals an internal hierarchy, or at least a work division related to the community management on the mailing list. In comparison, the academics and nonprofit organisation groups have a similar trend with a low standard deviation. It reveals the organisation heterogeneity involved with some leaders. The individual group has a stable trend, insensible to the increasing messages numbers on the list. This reveals a profile of occasional users more than a community contributor. In 2014 the companies' trend reveals an internal diversity and highlights the strongest involvement of some companies' members in a specific project such as data-driven journalism.
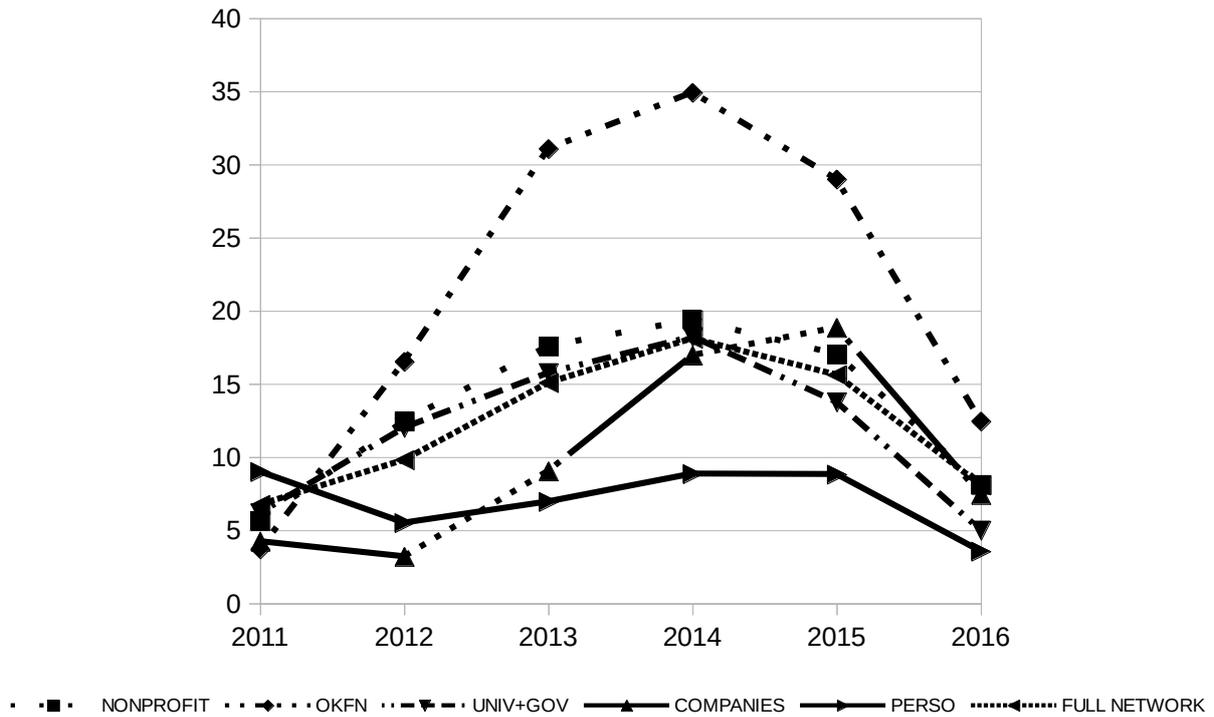
Figure 9 Degree centrality standard deviation on the Brazilian OKFN mailing list

The robustness density test shows a stronger reflexivity in the OKFN community over time (Fig. 10). At the opposite, the group of individual committers shows an increasing negative contribution to the global density. In other words, the community institutionalisation is flowing isolated contributions and do not suffer free rider behaviours. The density robustness test density also gives a new point of view on the contribution from academics with changes since 2014. Between 2011 and 2014, the trend is close to the average and does not inform a specific comportment. Since 2014 the academics' density contribution is decreasing without this category losing demographic shares. This change indicates a certain marginality of topics from these institutions because they do not generate reactions and discussion flow.
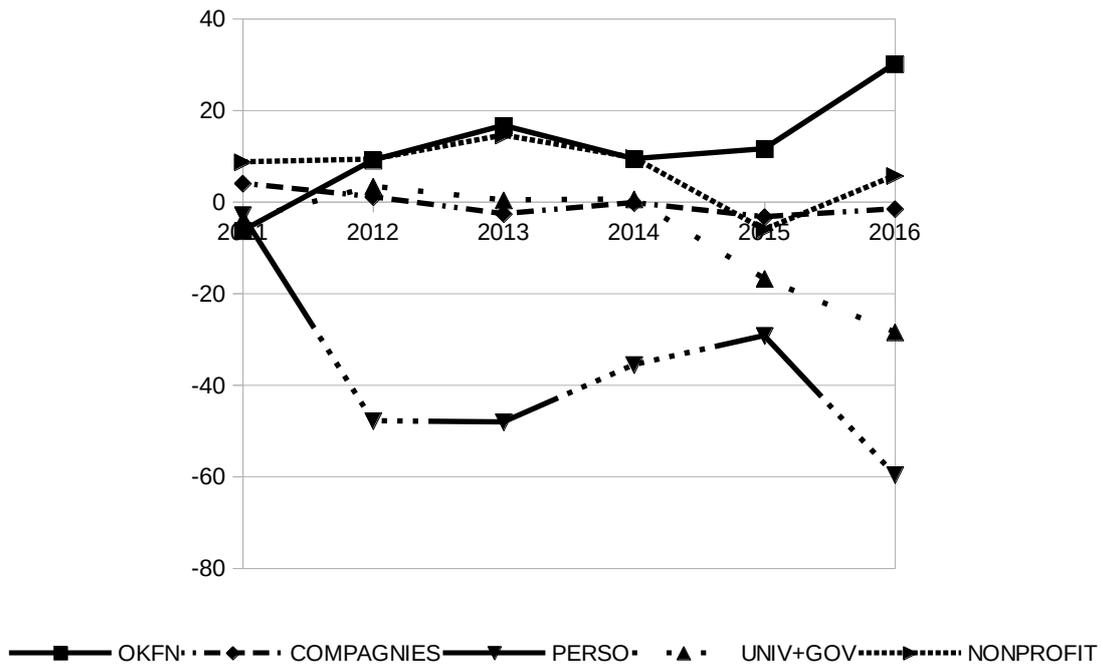
Figure 10 Density robustness test

The question of free riders (Olson, 1965) in collective action is particularly sensitive in heterogeneous context. In the Brazilian OKFN mailing list we observe a turnover more than half of the contributors each year. However, a control on distant, unpaid and voluntary commitments is complicated to setup. To structure its projects the Open Knowledge Foundation Brazil has dedicated employees and budgets. The organisation hosts discussion from nonprofit organisations, academics and individual volunteers. Most of these communities count some leaders inside the discussion network with different intensities. However the representatively of these most active contributors is questionable. Their activity does not systemically trigger feedback inside their groups of belongings. However, the contribution from structured institutions are more followed and discussed than the individual participation. Finally, our results tend to confirm the individual heterogeneity from different institutions.

Discussion, limits and perspectives

We can conclude that the institutional matters in terms of online involvement. Far from the hacker culture and the academic labs, the OKFN community show another way to practise openness as social innovation. We studied the both institutional and structural heterogeneity in the Brazilian OKFN community with individual and collective centrality indicators. The different trends we studied highlight specific structural positions after the first

year. Our method proposes to cross instructional contributors' affiliation with structural position in a network mailing list. The foundation appears as an intermediary in the free/open source ecosystem between different institutions beyond the engineering field. This is confirmed by the observation of offline implementation of digital products and framework such as automatic layout for official website.

The principal finding matches with other studies concerning the structural hierarchy in open source projects (Shaw and Hill, 2014). The community institutionalisation tends to down the diversity. The most involved contributors are part of nonprofit organisations related to Brazilian social movement and free/open source environment. This innovative elite underlines the progressive building of a viable economic model in the community management. The second findings show that the open ecosystem is built by strong individual commitments from different organisations. Even if we observe the hegemony management from one institution, the institutional heterogeneity is maintained over time at the collective and individual level with others non profit organizations and individual involvements.

In parallel to these strongly involved categories, the academics and companies and have their own dynamics. Their low structural involvement is quite surprising in comparison to the free/open source Brazilian institutional history (Schoonmaker, 2007; Shaw, 2011; Takhteyev, 2012). These results can signify at least two things not contradictory. First the academic and companies represents specific independent social innovators in comparison to the non-profit organisations involved in specific cities. Second the academic and companies' communities are very wide and scattered at the Brazilian level but also at the international scale. For future investigation, the evaluation of these groups cannot be resumed in one monograph but implied a multi fieldwork investigation.

To summarise our results we can say that it exists two types of social innovators in the open source: insiders and outsiders. Firstly, the insiders, they are involved in a bottom-up process and using a dedicated institution to perform their ideas in collaboration with a structured organisation in a restricted geographic area. Secondly, the outsiders, they are participating from afar with different organisations and working in institutions on the periphery of the open source fields such as the universities. They have the possibility to engage top/down processes of innovation without strong participation from local online open source communities. These categories can be identified in a social network, and their specific management should improve innovation by targeting key actors, usually ignored because of their a priori low involvement in a core/periphery layout.

Between these two categories, the foundations, such as the OKFN play a role of intermediary to share ideas, and enlist social innovators who share latent ties through tacit knowledge. We observe that product management creates alliances with companies even if they are lowered involved in the community discussion. The study concerning this balance between directive product management and heterogeneous contributions should be deepened by future studies. The online leadership and the merging of the production process are hot topics in the free software matters, and should be continued in the social innovation process issues.

The structural study of communication media in a community such as a mailing list has management implication for innovative products. Considering the needs to express tacit knowledge, these kinds of datasets give an opportunity to observe collective innovative process. However the social network analyses contain bias. This approach creates a positive representation in the social relationships (Lazega 1998). The negative or implicit links between two people or institutions are uneasy to observe because of matrices are coding only the absence or presence of relations.

Our comparison between different institutional involvements in the OKFN community should be crossed with other case studies. The relationship contextualisation is a methodological ordeal in most of the community, because the online identity does not always overlap the offline ones. Our work shows that it's possible to update the profiles. A possible improvement could enlarge the heterogeneity perceptive with other variables such as gender, ages, or geographic location. This approach could complete analyse concerning the context effects on innovative online communities' structures.

References

Alleyne, B., 2011. Challenging Code: A sociological reading of the KDE Free Software project. Sociology 45, 496-511.

Armer, P., 1980. SHARE-A Eulogy to Cooperative Effort. Annals of the History of Computing 2, 122-129.

Barrett, M., Oborn, E., Orlikowski, W., 2016. Creating value in online communities: the sociomaterial configuring of strategy, platform, and stakeholder engagement. Information Systems Research 27, 704-723.

Bauer, J., Franke, N., Tuertscher, P., 2016. Intellectual property norms in online communities: How user-organized intellectual property regulation supports innovation. Information Systems Research 27, 724-750.

Benkler, Y., 2006. The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press.

Bird, C., Pattison, D., D'Souza, R., Filkov, V., Devanbu, P., 2008. Latent social structure in open source projects, Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering. ACM, pp. 24-35.

Bitzer, J., Schrettl, W., Schröder, P.J., 2007. Intrinsic motivation in open source software development. Journal of Comparative Economics 35, 160-169.

Bonaccorsi, A., Lorenzi, D., Merito, M., Rossi, C., 2016. Firms' Participation in Free/Open Source Projects: Theory and Preliminary Evidence. Successful OSS Project Design and Implementation: Requirements, Tools, Social Designs and Reward Structures, 27.

Bonaccorsi, A., Rossi, C., 2003. Why open source software can succeed. Research Policy 32, 1243-1258.

Bonaccorsi, A., Rossi Lamastra, C., 2004. Altruistic individuals, selfish firms? The structure of motivation in Open Source software.

Borgatti, S.P., 2005. Centrality and network flow. Social networks 27, 55-71.

Bosch-Sijtsema, P., Bosch, J., 2015. User Involvement throughout the Innovation Process in High-Tech Industries. Journal of Product Innovation Management 32, 793-807.

Brooks Jr, F.P., 1956. No silver bullet essence and accidents of software engineering.

Carrington, P.J., Scott, J., Wasserman, S., 2005. Models and methods in social network analysis. Cambridge university press.

Chavez, A., Kovarik, C., 2017. Open Source Technology for Medical Practice in Developing Countries, Health Information Systems and the Advancement of Medical Practice in Developing Countries. IGI Global, pp. 33-59.

Chesbrough, H.W., 2003. Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

Coleman, E.G., 2013. Coding freedom: The ethics and aesthetics of hacking. Princeton University Press.

Crowston, K., Howison, J., 2006. Hierarchy and centralization in free and open source software team communications. Knowledge, Technology & Policy 18, 65-85.

Crowston, K., Scozzi, B., 2008. Bug fixing practices within free/libre open source software development teams.

Dalle, J.-M., Jullien, N., 2003. 'Libre'software: turning fads into institutions? Research Policy 32, 1-11.

Daniel, S., Agarwal, R., Stewart, K.J., 2013. The effects of diversity in global, distributed collectives: A study of open source project success. Information Systems Research 24, 312-333.

David, P.A., Shapiro, J.S., 2008. Community-based production of open-source software: What do we know about the developers who participate? Information Economics and Policy 20, 364-398.

de Laat, P.B., 2007. Governance of open source software: state of the art. Journal of Management & Governance 11, 165-177.

DiBona, C., Ockman, S., 1999. Open sources: Voices from the open source revolution. O'Reilly Media, Inc.

DiMaggio, P.J., Powell, W.W., 1991. The new institutionalism in organizational analysis. University of Chicago Press Chicago.

Ducheneaut, N., 2005. Socialization in an open source software community: A socio-technical analysis. Computer Supported Cooperative Work (CSCW) 14, 323-368.

Edquist, C., 2010. Systems of innovation perspectives and challenges. African Journal of Science, Technology, Innovation and Development 2, 14-45.

Edwards, K., 2001. Epistemic communities, situated learning and open source software development. Epistemic Cultures and the Practice of Interdisciplinarity.

Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Research Policy 29, 109-123.

Everett, M.G., Borgatti, S.P., 1999. The centrality of groups and classes. The Journal of mathematical sociology 23, 181-201.

Faraj, S., Sproull, L., 2000. Coordinating expertise in software development teams. Management science 46, 1554-1568.

Faraj, S., von Krogh, G., Monteiro, E., Lakhani, K.R., 2016. Special Section Introduction—Online Community as Space for Knowledge Flows. Information Systems Research 27, 668-684.

Fitzgerald, B., 2006. The transformation of open source software. Mis Quarterly, 587-598.

Fleming, L., Waguespack, D.M., 2007. Brokerage, boundary spanning, and leadership in open innovation communities. Organization science 18, 165-180.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry, 35-41.

Freeman, L.C., 1979. Centrality in social networks conceptual clarification. Social networks 1, 215-239.

Fuggetta, A., 2003. Open source software—an evaluation. Journal of Systems and Software 66, 77-90.

Ganley, D., Lampe, C., 2009. The ties that bind: Social network principles in online communities. Decision Support Systems 47, 266-274.

Gawer, A., Cusumano, M.A., 2014. Industry platforms and ecosystem innovation. Journal of Product Innovation Management 31, 417-433.

Gehring, R.A., 2006. The institutionalization of open source. Poiesis & Praxis 4, 54-73.

Ghosh, R.A., 2005. Understanding free software developers: Findings from the FLOSS study. Perspectives on free and open source software, 23-46.

Ghosh, S., Kumar Das, A., 2007. Open access and institutional repositories—a developing country perspective: a case study of India. IFLA journal 33, 229-250.

Grewal, R., Lilien, G.L., Mallapragada, G., 2006. Location, location, location: How network embeddedness affects project success in open source systems. Management science 52, 1043-1056.

Gustafsson, R., Jarvenpaa, S., 2017. Extending community management to industry-university-government organizations. R&D Management.

Henkel, J., 2006. Selective revealing in open innovation processes: The case of embedded Linux. Research Policy 35, 953-969.

Kreiss, D., Finn, M., Turner, F., 2011. The limits of peer production: Some reminders from Max Weber for the network society. New Media & Society 13, 243-259.

Lazega, E., 1992. Analyse de réseaux d'une organisation collégiale : les avocats d'affaires. Revue française de sociologie, 559-589.

Lazega, E., 2001. The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership. Oxford University Press on Demand.

Lerner, J., Tirole, J., 2002. The Simple Economics of Open Source. The journal of industrial economics, 50, 197-234.

Lewis, J.A., 2007. Government open source policies. Center for Strategic and International Studies.

Liu, M., Hull, C.E., Hung, Y.T.C., 2016. Starting open source collaborative innovation: the antecedents of network formation in community source. Information Systems Journal.

Mallapragada, G., Grewal, R., Lilien, G., 2012. User-generated open source products: Founder's social capital and time to product release. Marketing Science 31, 474-492.

Markus, M.L., 2007. The governance of free/open source software projects: monolithic, multidimensional, or configurational? Journal of Management & Governance 11, 151-163.

Mockus, A., Fielding, R.T., Herbsleb, J.D., 2002. Two case studies of open source software development: Apache and Mozilla. ACM Transactions on Software Engineering and Methodology (TOSEM) 11, 309-346.

Molloy, J.C., 2011. The open knowledge foundation: open data means better science. PLoS Biol 9, e1001195.

Mounier-Kuhn, P.-E., 2011. Les clubs d'utilisateurs: entre syndicats de clients, outils marketing et «logiciel libre» avant la lettre. Entreprises et histoire, 158-169.

Mulgan, G., 2006. The process of social innovation. innovations 1, 145-162.

O'mahony, S., Ferraro, F., 2007. The emergence of governance in an open source community. Academy of Management Journal 50, 1079-1106.

O'Mahony, S., 2005. Nonprofit Foundations and Their Role in Community-Firm Software Collaboration.

O'Mahony, S., 2007. The governance of open source initiatives: what does it mean to be community managed? Journal of Management & Governance 11, 139-150.

Olson, M., 1965. The logic of collective action: public goods and the theory of groups. Harvard University Press.

Polanyi, M., Sen, A., 2009. The tacit dimension. University of Chicago press.

Pollock, R., 2008. Innovation, imitation and open source.

Raymond, E., 1999. The cathedral and the bazaar. Knowledge, Technology & Policy 12, 23-49.

Roberts, D., Woods, C., 2005. Changing the world on a shoestring: The concept of social entrepreneurship. University of Auckland Business Review 7, 45-51.

Rushkoff, D., 2003. Open source democracy: How online communication is changing offline politics. Demos.

Schoonmaker, S., 2007. Globalization from below: Free software and alternatives to neoliberalism. Development and Change 38, 999-1020.

Schweik, C., 2003. The institutional design of open source programming: implications for addressing complex public policy and management problems. First Monday 8.

Shaw, A., 2011. Insurgent expertise: The politics of free/livre and open source software in Brazil. Journal of Information Technology & Politics 8, 253-272.

Shaw, A., Hill, B.M., 2014. Laboratories of oligarchy? How the iron law extends to peer production. Journal of Communication 64, 215-238.

Singh, P.V., Tan, Y., 2010. Developer heterogeneity and formation of communication networks in open source software projects. Journal of Management Information Systems 27, 179-210.

Spanjol, J., Scott, M.J., Melamed, S., Page, A.L., Bergh, D., Pfanner, P., 2014. Collaborative Innovation Across Industry-Academy and Functional Boundaries: How Companies Innovate with Interdisciplinary Faculty and Student Teams. Open Innovation: New Product Development Essentials from the PDMA, 175-223.

Stallman, R., 2001. Copyright and globalisation in the age of computer networks. CODE: Collaborative Ownership in the Digital Economy, 331.

Stanko, M.A., 2016. Toward a theory of remixing in online innovation communities. Information Systems Research 27, 773-791.

Steinmacher, I., Graciotto Silva, M.A., Gerosa, M.A., Redmiles, D.F., 2015. A systematic literature review on the barriers faced by newcomers to open source software projects. Information and Software Technology 59, 67-85.

Stewart, D., 2005. Social status in an open-source community. American sociological review 70, 823-842.

Takhteyev, Y., 2012. Coding places: Software practice in a South American city. MIT Press.

Tan, Y., Mookerjee, V., Singh, P., 2007. Social capital, structural holes and team composition: Collaborative networks of the open source software community. ICIS 2007 Proceedings, 155.

van der Have, R.P., Rubalcaba, L., 2016. Social innovation research: An emerging area of innovation studies? Research Policy 45, 1923-1935.

Von Krogh, G., Ichijo, K., Nonaka, I., 2000. Enabling knowledge creation: How to unlock the mystery of tacit knowledge and release the power of innovation. Oxford University Press on Demand.

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

Weber, S., 2000. The political economy of open source software, E-conomy Project Working Paper 15, BRIE, University of California, Berkeley.

Wei, K., Crowston, K., Eseryel, U.Y., Heckman, R., 2016. Roles and politeness behavior in community-based free/libre open source software development. Information & Management.

Wellman, B., Haythornthwaite, C., 2008. The Internet in everyday life. John Wiley & Sons.

Zhu, H., Kraut, R.E., Kittur, A., 2014. The impact of membership overlap on the survival of online communities, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 281-290.

Zhu, Y., Zhang, X., Hu, Y., 2016. Cluster analysis of participants of open source design community, 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 68-73.

Zucker, L.G., 1977. The role of institutionalization in cultural persistence. American sociological review, 726-743.

Methodological appendix

General betweenness centrality index

The general centrality is defines by Freeman as (Freeman 1979)

$$C_D = \frac{\sum_{i=1}^{n} [C_D(p^*) - C_D(p_i)]}{\max \sum_{i=1}^{n} [C_D(p^*) - C_D(p_i)]}$$

The general betweenness centrality is defines "average difference between the relative centrality of the most central point, and that of all other points". The model developed by Freeman and implemented in Pajek is the below:

$$C_B = \frac{\sum_{i=1}^{n} [C_B(p^*) - C_B(p_i)]}{n-1}$$

$$= \frac{\sum_{i=1}^{n} \left| \frac{C_B(p^*)}{n^2 - 3n + 2} - \frac{C_B(p_i)}{n^2 - 3n + 2} \right|}{n-1}$$

$$= \frac{\sum_{i=1}^{n} [C_B(p^*) - C_B(p_i)]}{(n-1)(n^2 - 3n + 2)}$$

$$= \frac{\sum_{i=1}^{n} [C_B(p^*) - C_B(p_i)]}{n^3 - 4n^2 + 5n - 2}$$

To evaluate the sub communities's general betweenness centrality impact (Ct) we subtract the community one (Cn) from the general betweenness centrality (C_B).

$$Ct = \frac{C_B - Cn}{C_B} \times 100$$

Density robustness calculation

To evaluate the sub density robustness (G) we subtract the community one (n) from the general betweenness centrality (N).

N=Global density

n=sub network density

G=contribution to the global density

$$G = \frac{N-n}{N} \text{x}100$$