

Analysis of virtual communities supporting OSS projects using social network analysis

S.L. Toral^{a,*}, M.R. Martínez-Torres^b, F. Barrero^a

^aE.S. Ingenieros, University of Seville, Camino de los Descubrimientos s/n, 41092 Seville, Spain

^bEscuela Universitaria de Estudios Empresariales, University of Seville, Avda. San Francisco Javier s/n, 41018 Seville, Spain

A B S T R A C T

This paper analyses the behaviour of virtual communities for Open Source Software (OSS) projects. The development of OSS projects relies on virtual communities, which are built on relationships among members, being their final objective sharing knowledge and improving the underlying project. This study addresses the interactive collaboration in these kinds of communities applying social network analysis (SNA). In particular, SNA techniques will be used to identify those members playing a middle-man role among other community members. Results will illustrate the importance of this role to achieve successful virtual communities.

Keywords:

Virtual communities
Knowledge brokers
Social network analysis
Open Source Software

1. Introduction

VIRTUAL communities can be defined as a social relationship aggregation, facilitated by Internet-based technology, in which users communicate and build personal relationships [37]. They allow the creation of weak links among geographically dispersed individual who regularly participates in the community. Examples of virtual communities can be found on fields like education [3,41], software development [2] or consumer behaviour [38].

The theoretical background behind virtual communities has been treated by numerous authors. For instance, some authors [34,1] highlight the connection of virtual communities with the social learning theory and communities of practice developed by Wenger [46], while others are focused on their relation with knowledge sharing [24], knowledge creation [26] and innovation models [45]. Another important block of works are devoted to motivation of people participating in online communities [6], the analysis social interactions [39] and participation mechanisms [38,42,43], and practical applications to successful online communities [44].

The social dimension of virtual communities can be justified by thousands of people who freely decide to take part in developing new structures and content [23]. This trend is also confirmed by those organisations which are moving away from traditional structures based on hierarchies, discrete groups and teams towards no-

vel structures based on more fluid and emergent organisational forms such as these networks and communities [9]. When people join together in a common enterprise sharing a common expertise, they led to a novel organization form known as *community of practice* (CoP). CoPs are organized around circumscribed sets of activities and their members develop their own routines, formal and informal “rules”, and practices evolve as a result of learning [46]. They do not need to be in direct contact, as nowadays people are in contact using information and communication technologies. Brown and Duguid [7] denominate this kind of communities “networks of practice”. This is the case of Open Source Software (OSS) projects, developed under the scheme of a virtual community of support. The user community of an OSS consists of people that, enabled by Internet, use the software and participate in its development in some way. The result is geographically distributed individuals who voluntarily contribute to a project by means of the Bazaar model [36,39]. The community of an OSS project can be considered a resource: it is the community that does most of the testing and provides quality feedback, instead of investing huge financial resources to put the software through extensive testing and quality assurance, like a proprietary vendor would do. The major productive assets of a OSS project are the developers, who interact with other developers by forming a small but well organized structure to intensify the learning interactions. It has been demonstrated that much of the OSS development is realized by a small percentage of individuals despite the fact that there are tens of thousands of available developers. Such concentration is called “participation inequality” [24], and it can be explained by the different user profiles of open source communities. Moreover, OSS

* Corresponding author. Tel.: +34 954 48 12 93; fax: +34 954 48 73 73.

E-mail addresses: toral@esi.us.es (S.L. Toral), rmtorres@us.es (M.R. Martínez-Torres), fbarrero@esi.us.es (F. Barrero).

virtual communities are typically sustained by a small group of core developers responsible for the majority of contributions [39]. Consequently, they always exhibit a high degree of centralization. Although the presence of such group of core developer is necessary to promote the participation process, it has been proved that an excessive participation of this particular group may lead to situations in which very few questions are debated among a reduced group of core participants [42,43]. This situation can be explained because core participants are not developing a brokerage role respect to other members of the community, reducing their possibilities of becoming an expert. In this context, this paper is focused on the identification of this particular group of users among the core group of participants in order to take some conclusions about their incidence in the successful development of the community. In particular, the middle-man role developed by several users will be highlighted as one of the most important strengths of a virtual community. For this purpose, Social Network Analysis (SNA) will be used to analyze the interactions among participants. This is an appropriate technique for this kind of communities, as one of the mechanisms for sharing information is based on participation [28].

The rest of the paper is organized as follows: The next section is a review of related work. After that, the methodology is presented, and SNA is applied to a particular case study. The one after discusses the implications of the empirical results and the final section discusses conclusions for practice and research.

2. Related work

Networks of practice refers to the overall set of various types of informal, emergent social networks that facilitate learning and knowledge sharing between individuals conducting practice-related tasks. They use networked technologies, especially the Internet, to establish collaboration across geographical barriers and time zones [17]. In contrast to traditional communities, virtual communities require less formal borders, and norms do not dominate as much as in traditional communities because members cannot see each other [18]. Although the idea of communities and networks of practice emerges from the social learning theory of Wenger [46], their importance in terms of knowledge management has been described by numerous authors [16,27,33]. One of the processes underlying the construction and nurturing of knowledge in networks of practice is called Legitimate Peripheral Participation (LPP) [25,11]. LPP describes the process by which a newcomer is integrated into the community. In this process, new members learn how to function as a community member through participation, and acquire the language, values, and norms of the community. Learning is gradually achieved as an individual moves from being a novice, gaining access to community practices to complete socialization and therefore becoming an insider or full member of the community. This process is similar to that described by Ducheneau [10], who explains how people enter an OSS. He states that successful participants progressively construct identities as software craftsmen, and this process is punctuated by specific rites of passage. Successful participants also understand the political nature of software development and progressively enroll a network of human and material allies to support their efforts [10]. For instance, OSS projects web sites provide forums and mailing lists where participants and contributors can report software improvements, needs or bugs, and share and discuss solutions to posted messages. For new members, bug fixes are usually the way to start contributing. If somebody has found a problem, it becomes accessible to the whole community. As members of the community browse through the OSS project tools, it is very likely that someone will consider the problem, jump right into action and concrete experience [13]. One of the most interesting features

of forums and mailing lists consists of enabling re-experience by collective reflection and virtual experimentation [13]. As a difference to bug reporting databases or Concurrent Version Systems, both forums and mailing lists details the sequence of discussion as they allow the possibility of being organized through threads of discussion. Threads are groups of messages sharing the same subject. A thread is initiated by someone who posts a message asking for help, suggesting some improvements, or just considering some new idea. Then people start answering this initial message, posting possible solutions, sources of information or just extending posted considerations. Some members of the community become engaged in a process of conceptualization, leading to some collective innovation and new knowledge. The result is a list of related messages where the sequence of reflections is detailed, so newcomers can follow expert reasoning step by step. Consequently, LPP and typical participation inequality of OSS projects lead to participation patterns driving the behaviour of virtual communities. These patterns should be analyzed using SNA techniques to visualize the performance of virtual communities.

Social networks are self-organized structures of people, information, and communities [20,35], which can be modelled by a net structure consisting of vertices and edges. Vertices represent individuals or organizations and the edges connecting nodes are called ties, which represent the relationships between the individuals and organizations. The strength of a tie indicates how strong the relationship is [49]. The main goal of social network analysis is detecting and interpreting patterns of social ties among actors [31]. In the case of OSS projects, several actors can be distinguished. We refer not only to the distinction between peripheral contributors and community members, but also to the different user profiles among the members themselves, like moderator and knowledge brokers [39].

3. Methodology

This section will be focused on the methodology used for the analysis of the participation mechanism in virtual communities. First, participation inequality will be measured to test the existence of a core/periphery structure. Then the structure of the community will be analyzed using SNA techniques.

Participation is on the basis of the development of virtual communities, and participation inequality is the typical taxonomy in OSS projects. The Gini coefficient is frequently used to provide a measure of the level of participation based on the numbers of postings made by individual developers within a community [12]. This coefficient has a value ranging from 0 to 1: with "too little" concentration (i.e., all developers posted fairly evenly), it approaches 0; and with "too much" concentration (i.e., a very few developers actively engaged in posting), it approaches to 1. Mathematically, the Gini coefficient (or Gini ratio) is a summary statistic of the Lorenz curve, and the Lorenz curve is a function of the cumulative proportion of ordered developers mapped onto the corresponding cumulative proportion of their contributions. Given a sample of n ordered individuals with x_i the size of individual i and $x_1 < x_2 < \dots < x_n$, then the sample Lorenz curve is the polygon joining the points $(h/n, L_h/L_n)$, where $h = 0, 1, 2, \dots, n$, $L_0 = 0$, and $L_h = \sum_{i=1}^h x_i$. Alternatively, the Lorenz curve can be expressed as:

$$L(y) = \frac{\int_0^y x dF(x)}{\mu} \quad (1)$$

where $F(x)$ is the cumulative distribution function of ordered individuals and μ is the average size. If all individual contributions are the same, the Lorenz curve is a straight diagonal line, called the line of equality. Otherwise, the Lorenz curve falls below the line of equality. The Gini coefficient is the ratio between the area en-

closed by the line of equality and the Lorenz curve, and the total triangular area under the line of equality (see Fig. 1). Operationally, the Gini coefficient is most easily calculated from unordered size data as the “relative mean difference”, i.e., as the mean of the difference between every possible pair of individuals, divided by the mean size μ , Eq. (2)

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2 \mu} \quad (2)$$

Once the participation inequality is checked, the behaviour of the community will be analyzed using SNA techniques. Basically, OSS project community participation is structured as threads of discussion. The simplest way to classify threads is using their length, i.e. the total number of posts they contain. Posts per thread – how densely packed posts are in a collection of threads – turns out to be a reliable metric to determine the degree of “conversational concentration” of an author in a given group [5]. Nevertheless, this kind of data does not provide any information about the social structure of the community, or about the relationships among authors. In this paper, social networks will be extracted from threads of discussion and SNA techniques will be applied to provide new insights in the community organization. A (social) network can be represented as a graph $G = (V, E)$ where V denotes a finite set of vertices and E denotes a finite set of edges such that $E \subseteq V \times V$. Some network analysis methods are easier to understand when graphs are conceptualized as matrices, Eq. (3)

$$M = (m_{ij})_{n \times n} \quad \text{where } n = |V| \quad m_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

In case of a valued graph, real valued weight function $w(e)$ is defined on the set of edges, i.e. $w(e) = Ex\mathfrak{R}$, and the matrix is then defined as given by Eq. (4).

$$m_{ij} = \begin{cases} w(e) & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In the context of threads of discussion, V is given by all the authors posting messages and E is given by the successive answers among authors inside a thread, which is the basic unit considered [19]. The use of discussion threads as the basic unit of analysis is highly valid, considering that the epistemic interactions in support of OSS development often take place in discussion threads where individual postings provide the context to encourage participation

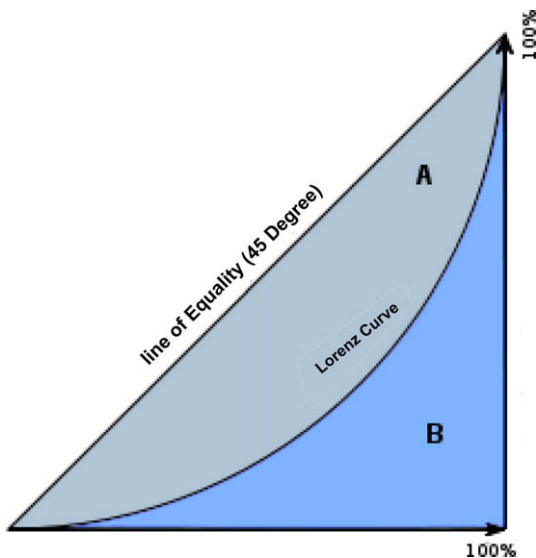


Fig. 1. Gini coefficient calculation.

[24]. In contrast to a reply to a single message, it is more cognitively complex to reply to a threaded discussion, because the ebb and flow of earlier postings must be taken into account to develop a coherent answer [21]. That is the reason why an author posting to a thread will be tied to all the authors who have previously posted to the same thread when constructing the social network. The resulting graph will exhibit the following features:

- It will be a directed graph. Usually, the word edge is reserved for undirected lines, while arc is the notation used for directed lines. The direction of the arc is given by the flow of information between two authors. That means that a sender (the tail of the arc) is answering a receiver (the head of the arc) inside a thread of discussion.
- It will be a valued graph. An author is able to participate several times inside a thread or can answer to the same authors in different threads. Consequently, the function $w(e)$ is a measure of the strength of the relationship among two authors.

Given the network, we are interested in studying both the macro-structure and the micro-structure of the social network.

- From a macro-structure point of view, networks can be partitioned using some discrete characteristics of vertices. For instance, several classes of vertices can be obtained using the function $w(e)$, that is, the strength of arcs. In the case of OSS projects, these kinds of partitions should highlight the core/periphery (C/P) structure of the community. A C/P structure divides vertices in two distinct subgroups: vertices in the core, densely connected with each other, and vertices on the periphery, not connected with each other, only nodes in the core. In network analysis, density is a measure of the cohesion of the network. More ties between people yield a tighter structure, which is, presumably, more cohesive. Density can be defined as the number of lines in a simple network, expressed as a proportion of the maximum possible number of lines. Consequently, maximum density is found in a network where all pairs of vertices are linked by two arcs, one in each direction. Sometimes, network density is not very useful because it depends on the size of the network. In this case, it is better to look at the number of ties in which each vertex is involved. This is called the degree of a vertex. As we are involved with a directed network, we will actually use the concept of out-degree, of a vertex, that is, the number of arcs it sends. Therefore, the average out-degree of all vertices could be used to measure the structural cohesion of a network independently of the network size.
- From a micro-structural point of view, we will concentrate on the different roles played by people belonging to the core of the community. One of the most pre-eminent roles is the one developed by the moderator of the community. But some other key people participate as a mediator among contributors, or among contributors and newcomers. The purpose of this local analysis is the identification of the group of people representing the heart of the community.

4. Case study

A support community for the most well known OSS project, that is, Linux, has been chosen as a case study. Particularly, we have focused on a Linux port for a particular class of processor like ARM. ARM, which stands for Advanced RISC Machine, is a family of processors maintained and promoted by ARM Holdings Ltd. Contrary to other chip manufacturers such as IBM, Motorola, and Intel, ARM Holdings does not manufacture its own processors. Instead,

ARM designs the CPU cores for its customers based on the ARM core, charges customers licensing fees on the design, and lets them manufacture the chip wherever they see fit. All ARM processors share the same ARM instruction set, which makes all their variants fully software compatible. Presently, ARM Linux port has been implemented in more than 1200 related boards, and it is supported by a project web site at <http://www.arm.linux.org.uk/>. Particularly, the “lists.arm.linux.org.uk” mailing list is one of the highest quality and most effective forum for finding answers to problems when working with ARM embedded Linux [48]. In general, the mailing list data of OSS projects are widely available, providing an excellent infrastructure to study community participation in an OSS project. Data from ARM Linux mailing list located at “lists.arm.linux.org.uk” during the years 2001–2006 have been used in this research. A total of 12,010 messages posted by 2086 authors have been considered. When accounting authors, it is necessary to consider the fact that they are identified using aliases. They usually correspond to a unique e-mail address, but this is not always truth. As a consequence, header of messages should be processed to check there is no duplicity of aliases or e-mails [4].

The Gini coefficient was calculated for the case study during the period 2001–2006 to check the typical core/periphery structure of the virtual community. The obtained result of 0.684 (see Fig. 2) clearly proves the existence of a core of regular contributors inside the community. They are responsible of the majority of contributions. Quite the opposite, free riders can be defined as “members who enjoy the benefits of the collective good without contributing to its establishment” [29]. They usually post questions but they are nor involved in the development of the community. They are tolerated because they are an important resource. As time goes by, some users begin to take interest in the project and voice their opinion, becoming contributors.

In the following subsections we will study the evolution of the virtual community year by year using SNA. The analysis has been divided in two parts: a macro-structural analysis to obtain a global view of the community evolution, and a micro-structural analysis to zoom in on particular and interesting areas of the network.

4.1. Macro-structural analysis of the virtual community

Table 1 details the general data of the ARM Linux mailing list for the considered period 2001–2006. In addition to the absolute values of messages, threads and authors, the cohesion of the network has also been evaluated. Cohesion is calculated as the average out-

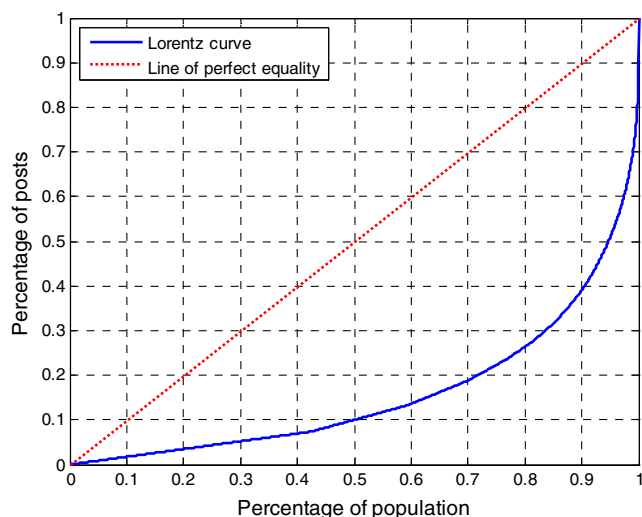


Fig. 2. Gini coefficient for ARM Linux port mailing list (0.684).

degree of all vertices, as it was pointed out in the previous section. The average out-degree has been used as the criterion to distinguish between peripheral members and regular contributors of the virtual community. Their average values and their standard deviation (between brackets) are included in the fifth column of Table 1. The sixth column of this table details the obtained regular contributors using the average out-degree as the threshold participation value (rounded to the next integer value). The value in brackets of the same column represents the percentage of regular contributors over the full community size. Notice that the percentage of regular contributors oscillates between 28% and 35% of the community. This is a logical value taking into account the typical participation inequality of OSS projects. Out-degree concept is based on the idea of centrality of a person within a community, but understanding centrality in the sense of a minimum distance to the rest of the members of the community. A second approach to centrality rests on the idea that a person is more central if he or she is more important as an intermediary in the communication network [31]. This approach is based on the concept of betweenness. The centrality of a person depends on the extent to which he or she is needed as a link in the chains of contacts that facilitate the spread of information within the network. The more a person is a go-between, the more central his or her position in the network. If we consider that the shortest path between two vertices (geodesic) is the most likely channel for transporting information between actors, an actor who is situated on the geodesics between many pairs of vertices is very important to the flow of information within the network. The betweenness centrality of a vertex is the proportion of all geodesics between pairs of other vertices that include this vertex and betweenness centralization of the network is the variation in the betweenness centrality of vertices divided by the maximum variation in betweenness centrality scores possible in a network of the same size. [31]. The betweenness centrality of the network is also detailed in Table 1 for each year.

Fig. 3 illustrates the networks graphs corresponding to the regular contributors of the ARM Linux community (peripheral contributors have been removed in accordance with the criterion of Table 1). The area of each vertex is proportional to its out-degree value. Consequently, the most participative contributors exhibit a higher vertex area (they have been drawn in the central part of each network graph of Fig. 3).

The out-degree value can be used as a discriminant criterion to extract the core of the community. The biggest vertex area corresponds to the moderator of the community, who is logically the most active contributor. However, some other members of the community are also developing an outstanding role attending to their numerous contributions. Following this general criterion, we have considered as the core of the community those members with an out-degree value higher than the mean value plus the standard deviation detailed in Table 1. The obtained results are shown in Table 2. The second and third columns are the size and the percentage of the regular contributors owning to the core of the community. The fourth and fifth columns are the average out-degree value of members owning to the core and the percentage of the out-degree values accumulated by them.

Looking together at Table 1 and Table 2, the following considerations can be highlighted:

- The size of the core of the community is changing over the years. Although four of the members are present along the studied years (including the moderator), the rest of members are changing from one year to another. This is due to the fact that most of people attending the ARM Linux is coming from commercial companies and universities or research institutes, instead of volunteers or hobbyists like usually happens in desktop Linux

Table 1
General data of the studied mailing list during the period 2001–2006.

Year	Messages	Threads	Authors	Average out-degree	Regular contributors	Betweenness centrality
2001	2483	1160	499	6.076 (23.1)	117 (30.6%)	0.55
2002	1770	827	388	5.474 (23.1)	86 (28.3%)	0.73
2003	1988	910	466	4.577 (13.8)	111 (31.1)	0.55
2004	2284	1084	494	5.168 (14.1)	117 (31.0%)	0.29
2005	2244	1025	453	6.086 (15.8)	119 (35.6%)	0.37
2006	1249	594	314	4.812 (13.3)	78 (33.0%)	0.59

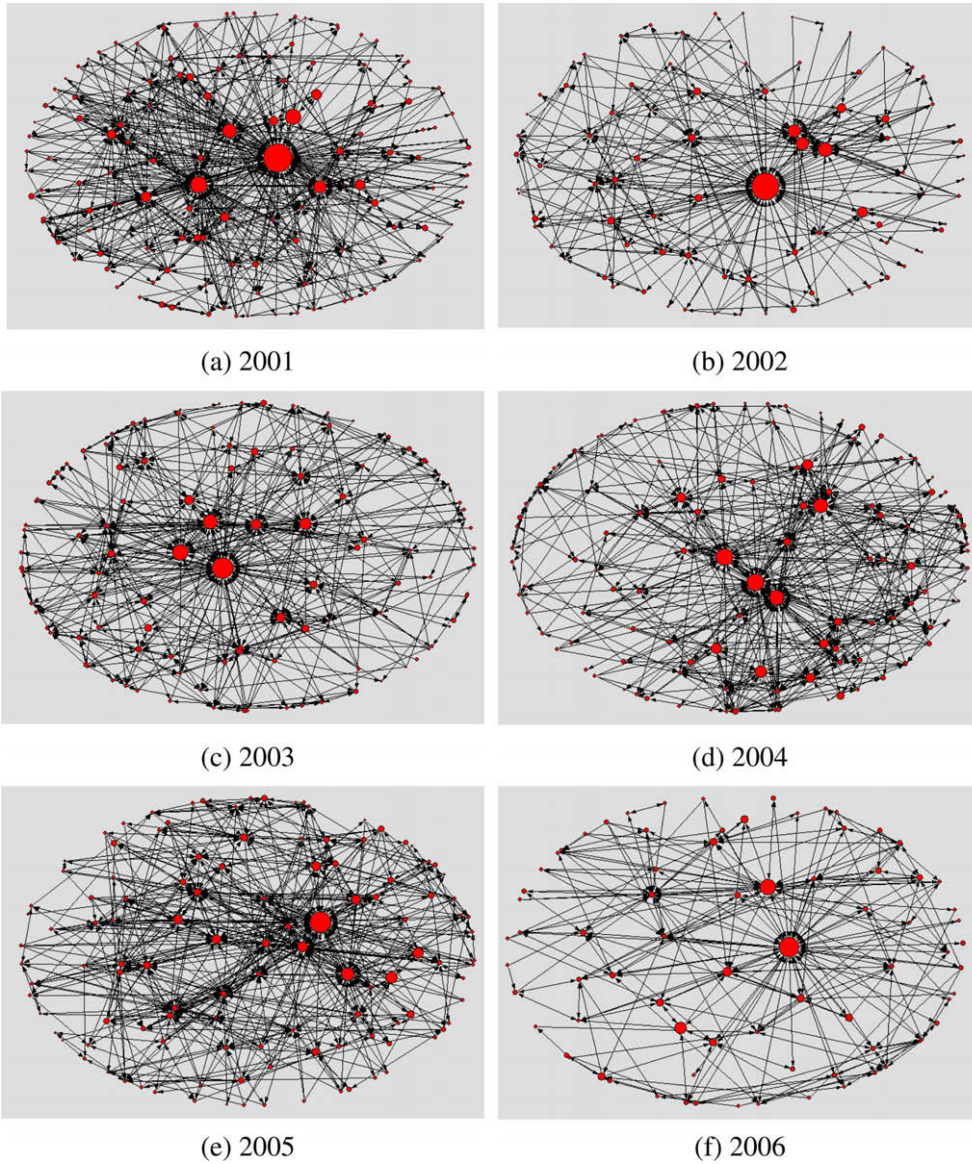


Fig. 3. Regular contributors network graphs of the ARM Linux community during the years 2001–2006 (a–f).

Table 2
Core of the community during the years 2001–2006.

Year	Size	Percentage over regular contributors (%)	Average out-degree	Out-degree percentage over regular contributors (%)
2001	15	12.82	86.67	78.64
2002	11	2.82	87.09	74.61
2003	23	4.91	37.91	85.07
2004	25	5.06	46.40	87.15
2005	34	7.50	32.71	83.99
2006	15	4.77	38.60	80.75

community [14]. They join the community during the period of time in which they are involved in a particular project, and after that, they leave the community.

- The topology of the network is centralized around the core of the community. The degree of centralization can be obtained as the betweenness centrality, and it has been observed that this value is highly dependent of how the core of the community is organized. The biggest value in this study corresponds to year 2002. Notice that 2002 exhibits a very small core but very active, as it can be deduced by its high value of the average out-degree (see Table 2). The smallest values correspond to years 2004 and 2005. Those cases are just the opposite, a bigger core but without a very active participation.
- The core of the community is responsible of the majority of interactions inside the community. The last column of Table 2 clearly illustrates this property. The core of the community is responsible of around a 75–87% of the total out-degree value of regular contributors. Again, this fact emphasizes the important role developed by the core of the community as well as the typical participation inequality of OSS projects.

4.2. Micro-structural analysis of the virtual community

In the previous section, the structure of the entire network was analyzed, following a socio-centred approach. Now we turn to the ego-network and ego-centred approach: we focus on the position of one person in the network and his or her opportunities to broker or mediate between other people. Consequently, the micro-structural analysis refers to the analysis of particular actors of a network who performs a specific brokerage role. They bridge the gap between expert software developers and peripheral users, helping OSS projects to engage in a discourse and co-learning experience with their user communities [39].

A broker is a middle node in a directed triad (a set of three vertices and the lines among them). We have considered brokers to those members of the community performing such role more than four times. The results are illustrated in Fig. 4. Brokers are highlighted in white colour while the other vertices of the network remain in grey colour.

The size of each vertex is proportional to the number of times this vertex is performing a broker role. In 2001, 2002, 2003 and 2006, the biggest vertex corresponds to the moderator of the list,

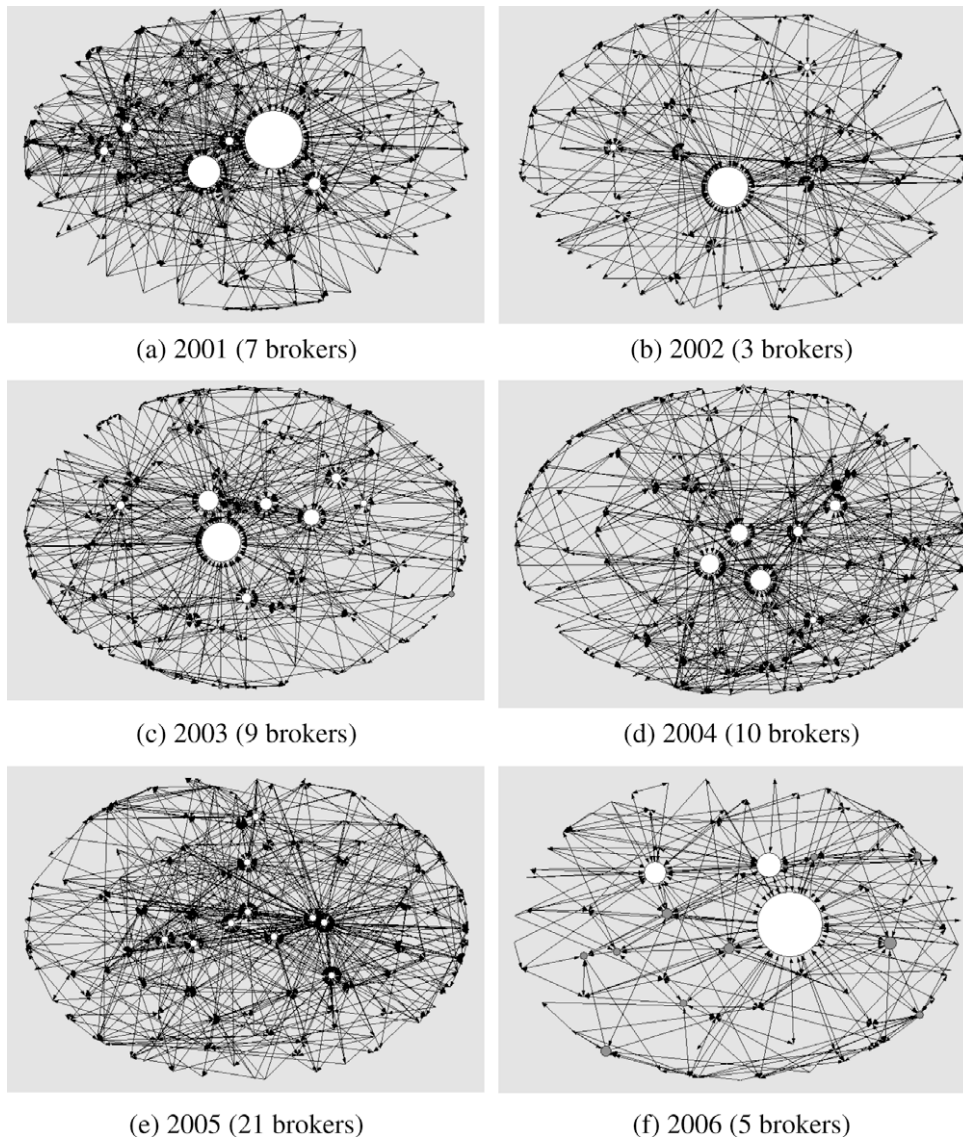


Fig. 4. Brokers' identification in the ARM Linux community during the years 2001–2006 (a–f).

who is consequently performing a preponderant role in the network activity. Although there are some other brokers, the difference in size respect to the moderator is clear. As a difference, there are several brokers with the same size in 2004 and 2005, including the moderator. They are precisely the years in which, according to Table 1, the network is less centralized. This fact suggests that the presence of enough brokers with an active mediation leads to less centralized topologies and more dynamic communities.

5. Discussion and implications

OSS projects success relies essentially on the virtual community of support, who is responsible not only of the efficient development of the project, but also of tasks like solving bugs, improvements and doubts of users and potential users. People tend to choose software not only because its features, but also because of the provided support and this support emerges from the virtual community [8,32]. Consequently, it is important the organization of the community to achieve an efficient support. It must be taken into account that participation inequality is always present, so the structure of a core versus a periphery will be typically found on OSS projects. The core must be constituted by people with an active participation, not only in the sense of number of times they are contributing, but also of the number of times they are participating as brokers or mediators among other users. Brokers develop a dual role as knowledge brokers and knowledge providers, performing an important role in the project. Their presence and their level of activity as brokers is a key parameter for the degree of centralization of the network. In general, OSS communities exhibit a high degree of centralization, and this fact is in harmony with the participation inequality idea and with the existence of a core inside the community. However, excessively centralized network with small cores will negatively affect the overall performance in the sense that newcomers will find more difficulties and barriers for improving their knowledge and becoming a potential expert or even a core member. In this case, the community is highly dependent on a few people, and their role as brokers is also lower, decreasing the flow of knowledge between knowledge seekers and providers.

The organization of virtual communities has several implications. The first one is related to OSS project success. OSS projects compete mainly to provide good services to consumers. As long as they achieve to bring closer a higher quantity of users to their communities, they will obtain a great amount of potential users and a continuous feedback of bugs and improvements that in turn will attract new users. In this virtuous circle, the efficiency of support is extremely important. The second one is related to the development of open source in general. Commercial firms can obtain potential benefits by publishing the source code [15], so they can be interested in transmitting their software widely, promoting a virtual community around the project [22].

6. Conclusion

OSS projects constitute today a clear example of social software and collective innovation with a notorious impact on the society. OSS projects are based on a virtual community of users who freely participate in the development of the underlying project, under the principle of participation inequality. This concentration on a small number of developers leads to network topologies that have been studied in this paper using SNA techniques. A macro-structural analysis and a micro-structural analysis have been developed in order to study social networks.

The macro-structural analysis of the network allowed the identification of three different groups: periphery users, regular contributors, and the core of the community, which is a subset of the regular contributors group. This result agrees with those from other authors who have also analyzed social networks structure (see, for example [30,47], rejecting the idea of flat structure claimed by the bazaar model of full participation [36].

The micro-structural analysis allowed the identification of brokerage as the key role to be performed by the core of the community. Particularly, the presence of brokers contributes to decrease the traditional high level of centralization of such communities as these brokers behave as intermediary between expert software developers and peripheral users, helping OSS projects to engage in a discourse and co-learning experience with their user communities. In accordance to Timmers [40], "these brokers prescribe additional strategic opportunities either as third party value-added suppliers or as information brokers providing a virtual structure around specific business information services".

As a future expansion of this work, the outcomes of the studied virtual community could be incorporated to the analysis using text categorizations tools such as latent semantic analysis or latent Dirichlet allocation. The idea would be to identify the number of topics the mailing list is addressing in order to determine if the presence of brokers is contributing to improve the tangible and quantifiable results of the community.

The comparison of the macro- and micro-structural analysis has provided new insights about the organization of virtual communities that can be useful for those companies and firms interested in developing communities of supports of their particular projects.

Acknowledgements

This work has been supported by the Spanish Ministry of Education and Science (Research Project with reference DPI2007-60128) and the Consejería de Innovación, Ciencia y Empresa (Research Project with reference P07-TIC-02621).

References

- [1] A. Amin, J. Roberts, Knowing in action: beyond communities of practice, *Research Policy* 37 (2) (2008) 353–369.
- [2] F. Barcellini, F. Détienne, J.-M. Burkhardt, W. Sack, A socio-cognitive analysis of online design discussions in an open source software community, *Interacting with Computers* 20 (1) (2008) 141–165.
- [3] F. Barrero, S.L. Toral, S. Gallardo, EDSPLab: remote laboratory for experiments on DSP applications, *Internet Research* 18 (1) (2008) 79–92.
- [4] C. Bird, A. Gourley, P. Devanbu, M. Gertz, A. Swaminathan, Mining Email Social Networks, in: *Proceedings of the International Workshop on Mining Software Repositories, MSR'06*, May 22–23, Shanghai, China, 2006, pp. 137–143.
- [5] D. Bonacci, Towards quantitative tools for analysing qualitative properties of virtual communities, *Interdisciplinary Description of Complex Systems* 2 (2) (2004) 126–135.
- [6] A. Bonaccorsi, C. Rossi, Why open source software can succeed, *Research Policy* 32 (2003) 1243–1258.
- [7] J.S. Brown, P. Duguid, *The Social Life of Information*, Harvard Business School Press, MA, 2000.
- [8] L. Carotenuto, W. Etienne, M. Fontaine, J. Friedman, M. Muller, H. Newberg, M. Simpson, J. Slusher, K. Stevenson, *CommunitySpace: Toward Flexible Support for Voluntary Knowledge Communities*, Changing Places workshop, London, April 1999.
- [9] R. Cross, A. Parker, *The hidden power of social networks. Understanding how work really gets done in organizations*, Harvard Business School Publishing, Boston, 2004.
- [10] N. Ducheneau, Socialization in an open source software community: a socio-technical analysis, *Computer Supported Cooperative Work (CSCW)* 14 (4) (2005) 323–368.
- [11] S. Fox, Communities of practice, Foucault and actor-network theory, *Journal of Management Studies* 37 (6) (2000) 853–868.
- [12] C. Gini, *On the Measure of Concentration with Espacial Reference to Income and Wealth*. Cowles Commission, 1936.
- [13] A. Hemetsberger, C. Reinhardt, Learning and knowledge-building in open-source communities a social-experiential approach, *Management Learning* 37 (2) (2006) 1350–5076.

- [14] J. Henkel, Software development in embedded Linux – informal collaboration of competing firms, in: Proceedings der 6. International en Tagung Wirtschaftsinformatik, Dresden, 2003, pp. 1–20.
- [15] J. Henkel, Selective revealing in open innovation processes: the case of embedded Linux, *Research Policy* 35 (2006) 953–969.
- [16] P. Hildreth, C. Kimble, P. Wright, Communities of practice in the distributed international environment, *Journal of Knowledge Management* 4 (1) (2000) 27–38.
- [17] C.M. Johnson, A survey of current research on online communities of practice, *Internet and Higher Education* 4 (2001) 45–60.
- [18] J.P. Johnson, Collaboration, peer review and open source software, *Information Economics and Policy* 18 (2006) 477–497.
- [19] G. Jones, G. Ravid, S. Rafaela, Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration, *Information Systems Research* 15 (2004) 194–210.
- [20] H. Kautz, B. Selman, M. Shah, Referral Web: combining social networks and collaborative filtering, *Communications of ACM* 40 (3) (1997) 27–36.
- [21] N. Knock, Compensatory adaptation to a lean medium: an action research investigation of electronic communication in process involvement groups, *IEEE Transactions on Professional Communication* 44 (4) (2001) 267–285.
- [22] B. Kogut, A. Metiu, The emergence of E-innovation: insights from open source software development. Reginald H. Jones Center Working Paper, Philadelphia, PA, 2000.
- [23] J. Kolbitsch, H. Maurer, The transformation of the web: How emerging communities shape the information we consume, *Journal of Universal Computer Science* 12 (2) (2006) 187–213.
- [24] G. Kuk, Strategic interaction and knowledge sharing in the KDE developer mailing list, *Management Science* 52 (7) (2006) 1031–1042.
- [25] J. Lave, E. Wenger, R. Pea, *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, New York, USA, 1991.
- [26] G.K. Lee, R.E. Cole, From a firm-based to a community-based model of knowledge creation: the case of the Linux Kernel development, *Organization Science* 14 (6) (2003) 633–649.
- [27] M.R. Martínez-Torres, A procedure to design a structural and measurement model of intellectual capital: an exploratory study, *Information and Management* 43 (5) (2006) 617–626.
- [28] M.R. Martínez-Torres, S.L. Toral, F. Barrero, Analysis of online participation in open source communities, in: Proceedings del IADIS International Conference on Web based Communities (IADIS-09), 2009, pp. 96–101.
- [29] D. Millen, J. Feinberg, B. Kerr, Social bookmarking in the enterprise, *ACM Queue* 3 (9) (2005) 28–35.
- [30] A. Mockus, T. Fielding, D. Herbsleb, Two case studies of open source software development: Apache and Mozilla, *ACM Transactions on Software Engineering and Methodology* 11 (3) (2002) 309–346.
- [31] W. Nooy, A. Mrvar, V. Batagelj, *Exploratory Network Analysis with Pajek*, Cambridge University Press, New York, 2005.
- [32] L. Palen, J. Grudin, Discretionary adoption of group support software: lessons from calendar applications, in: B.E. Munkvold (Ed.), *Organizational*, 2002.
- [33] S.L. Pan, D.E. Leidner, Bridging communities of practice with information technology in pursuit of global knowledge sharing, *Journal of Strategic Information Systems* 12 (2003) 71–88.
- [34] J. Preece, Sociability and usability: twenty years of chatting online, *Behaviour and Information Technology Journal* 20 (5) (2001) 347–356.
- [35] P. Raghavan, Social networks: from the web to the enterprise, *IEEE Internet Computing* 6 (1) (2002) 91–94.
- [36] E.S. Raymond, *The cathedral and the bazaar*, 1999. <<http://www.tuxedo.org/~esr/writings/cathedral-bazaar/>>.
- [37] H. Rheingold, *The Virtual Community: Homesteading on the Electronic Frontier*, Addison-Wesley, Reading, MA, 1993.
- [38] R.-A. Shang, Y.-C. Chen, H.-J. Liao, The value of participation in virtual consumer communities on brand loyalty, *Internet Research* 16 (4) (2006) 398–418.
- [39] S. Sowe, I. Stamelos, L. Angelis, Identifying knowledge brokers that yield software engineering knowledge in OSS projects, *Information and Software Technology* 48 (11) (2006) 1025–1033.
- [40] P. Timmers, Business models for electronic markets, *Electronic Markets* 8 (2) (1998) 3–8.
- [41] S.L. Toral, F. Barrero, M.R. Martínez-Torres, S. Gallardo, J. Lillo, Implementation of a web-based educational tool for digital signal processing teaching using the technological acceptance model, *IEEE Transactions on Education* 48 (4) (2005) 632–641.
- [42] S.L. Toral, M.R. Martínez-Torres, F. Barrero, Modelling mailing list behaviour in open source projects: the case of ARM embedded Linux, *Journal of Universal Computer Science* 15 (3) (2009) 648–664.
- [43] S.L. Toral, M.R. Martínez-Torres, F. Barrero, Virtual communities as a resource for the development of OSS projects: the case of Linux ports to embedded processors, *Behavior and Information Technology* 28 (5) (2009) 405–419.
- [44] T. Trung, Dinh-Trong, J.M. Bieman, The FreeBSD project: a replication case study of open source development, *IEEE Transactions on Software Engineering* 31 (6) (2005) 481–494.
- [45] E. von Hippel, G. von Krogh, Open source software and the “private-collective” innovation model: issues for organization science, *Organization Science* 14 (2) (2003) 209–223.
- [46] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity*, Cambridge University Press, Cambridge, 1998.
- [47] J. Xu, Y. Gao, S. Christley, G. Madey, A topological analysis of the open source software development community, in: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. HICSS '05, 2005, pp. 188–198.
- [48] K. Yaghmour, *Building Embedded Linux Systems*, O'Reilly, 2003.
- [49] S.J.H. Yang, I.Y.L. Chen, A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network, *International Journal of Human-Computer Studies* 66 (1) (2008) 36–50.