

# An empirical study of the driving forces behind online communities

Sergio L. Toral, M. Rocío Martínez-Torres, Federico Barrero and  
Francisco Cortés

*University of Seville, Seville, Spain*

## Abstract

**Purpose** – A large variety of online communities have emerged during the last years as a result of the challenges faced by both the business and scientific worlds. This trend has also been promoted by the development of internet and new Web 2.0 technologies. In this context, this paper is focused on the determinants of success of online communities. But, as a difference from other studies, these determinants are analyzed from the social network analysis perspective. Several constructs related to the community organization as a social network are proposed and their interrelations are hypothesized in a general research framework. The obtained results test the proposed model providing the most relevant antecedents of the project success.

**Design/methodology/approach** – A case study based on Linux ports to non-conventional processor and environments is used to test the proposed model. Structural equation modeling analysis is used to validate the structural proposed model.

**Findings** – The main antecedents of online communities' success, quantifying the strength of the relation through the standardized path coefficients.

**Research limitations/implications** – The research is limited to a particular set of online communities engaged with the development of the non-conventional Linux ports. However, they constitute a representative set of communities in the field of the open source projects (OSS) development, which are typically developed using a community of support.

**Originality/value** – This paper fulfils the main antecedents causing the successful development of Internet virtual communities. Instead of using sources of data coming from users' surveys, this study employs community interactions as a source of data. Results have important implications over the development of online communities, like software business models based on virtual communities and open source software.

**Keywords** Communities, Internet, Social networks, Mathematical modelling

**Paper type** Research paper

## Introduction

The notion of community has been at the heart of the internet since its inception (Lesser *et al.*, 2000). Initially, internet was used by scientist to share knowledge, collaborate on research and exchange messages, and today, million of internet users worldwide communicate themselves using electronic tools. The advent of Web 2.0 and social software have propitiated the organization of users around communities of interest. The distinctive feature of these online communities is the intensive use of electronic media for people getting in contact. Although intuitively everyone seems to understand the concept of "online community", there is no agreed definition (De Souza

---

The authors gratefully acknowledge support provided by the Spanish Ministry of Education and Science within the I + D + I national project with reference DPI2007-60128, and the Consejería de Innovación, Ciencia y Empresa (Research Project with reference P07-TIC-02621).

---

and Preece, 2004). This is due to the multidisciplinary nature of this topic, which can be analyzed from different perspectives. The theoretical background behind online communities has been treated by numerous authors. For instance, some authors (Preece, 2001; Amin and Roberts, 2008; Albors *et al.*, 2008) highlight the connection of online communities with the social learning theory and communities of practice developed by Wenger (1998), while others are focused on their relation with knowledge sharing (Kuk, 2006), knowledge creation (Lee and Cole, 2003) and innovation models (Von Hippel and von Krogh, 2003). Another important block of works are devoted to motivation of people participating in online communities (Bonaccorsi and Rossi, 2003; Hertel *et al.*, 2003), the analysis social interactions (Sowe *et al.*, 2006) and participation mechanisms (Shang *et al.*, 2006), and practical applications to successful online communities (Zhao *et al.*, n.d.; Trung and Bieman, 2005). Of particular importance are the online communities supporting open source software (OSS) projects, as they are changing the way in which software is produced. Several case studies can be found on the literature. Mockus *et al.* (2002) raise some questions about OSS development and analyze two case examples based on Apache and Mozilla projects. This study was focused above all on code development and problem resolution, and data were mainly extracted from concurrent versioning systems (CVS) and databases. Lee and Cole (2003) focused on the most well known OSS project: Linux.

Understanding what makes online communities successful is quite complicated. Although the development of such communities requires a specific technology, it is evident that technology alone does not guarantee a successful development (De Souza and Preece, 2004). This is because online communities should be understood as a social phenomenon that establishes social networks of people with common interests. Consequently, success of online communities should be determined by social factors. The purpose of this study consists of going deeply in the social network analysis of online communities. In particular, our work is focused on the discussion space (mailing lists, forums) (Sack *et al.*, 2006), in which most of the design activity takes place (Mockus *et al.*, 2002). A structural and measurement causal model is developed to test several hypotheses about the driving forces behind online communities. The necessity of considering latent variables (not directly observables) and both exogenous and endogenous variables advise the use of structural equation models (SEM) as the appropriate methodology. The findings illustrate which are the causal factors of success of online communities, leading to more effective strategies for their creation and future development.

### **Online communities overview**

The online community can be defined as a social relationship aggregation, facilitated by internet-based technology, in which users communicate and build personal relationships (Rheingold, 1993). They enable the creation of weak structural links between thousands of geographically dispersed individuals. Individuals may quickly and effortlessly access a broad source of expertise through a wide variety of knowledgeable individuals with whom they are not acquainted regardless of their demographic characteristics, organizational setting, or local culture. Individuals engage in knowledge sharing, problem solving, and learning through posting and responding to questions on professional advice, storytelling of personal experiences, and debate on issues relevant to the network. Examples of online communities can be

---

found on fields like education (Barrero *et al.*, 2008; Toral *et al.*, 2005), software development (Barcellini *et al.*, 2008) or consumer behavior (Shang *et al.*, 2006).

Online communities have been frequently connected with communities of practice (Wellman and Gulia, 1995; Lin and Lee, 2006), in the sense that communities develop their own routines, formal and informal “rules”, and practices evolve as a result of learning. One of the basic mechanisms of these communities is participation. The learning process is considered inseparable of the social context in which knowledge is generated (Wenger, 1998; Pan and Leidner, 2003). Expertise does not appear as a result of being taught, but through direct engagement in the social, cultural, and technical practice of the community. At first, newcomers can only peripherally participate in small and easy tasks. Through participation and collaboration with others practitioners, they create their own learning curriculum by developing a global view of the community and what there is to be learned. Gradually, they change their position inside the community gaining knowledge and becoming competent in undertaking more important roles. The described mechanism allows community members to improve their expertise through participation in the community activities. But it also illustrates the structured organization of online communities. Attending to the degree of expertise, different kinds of community members can be distinguished. Several studies of online discussions and online communities (Cassell *et al.*, 2005; Barcellini *et al.*, 2008) have highlighted how participants actively construct their positions and their roles within an online community. These roles reflect the project’s organizational structure, the technical skills and activities exhibited by participants, and the participants’ contributions to the online discussions.

Usually, success of online communities has been studied from the perspective of the information system success models, under the assumption that an online community is a form of the internet-based information system (Wachter *et al.*, 2000). Basically, this model posits that system quality, information quality and service quality affect both user satisfaction and behavioral intention to use IS, which in turn are direct antecedents of IS effectiveness (DeLone and McLean, 2003). The main problem is that this model does not consider the social relationships among community members nor the structure of the community. For this reason, this paper proposes a new approach based on social network analysis (SNA), which is the study of the social relations among a set of actors with the particularity of being focused on the relationships between actors rather than the attributes of these individual actors (Wasserman and Faust, 1994). We argue that network relationships provide the online communities with both opportunities and constraints that can have rich implications for their performance.

### **Theoretical background and hypotheses**

Social network theory uses methods of depicting and analyzing networks of people to help understand and communicate the ways in which they are connected (Wasserman and Faust, 1994). These connections and relationships can be based on online activity, thus allowing social network analysis to be applied to the study of online communities (Garton *et al.*, 1997).

SNA focuses on patterns of relations among people, organizations, states, etc. Researchers have recognized that a broader sense of social network is a self-organized structure of people, information, and communities (Kautz *et al.*, 1997; Raghavan, 2002).

A social network can be modeled by a net structure consisting of nodes and edges. Nodes represent individuals or organizations. The edges connecting nodes are called ties, which represent the relationships between the individuals. Social networks can be represented as matrices; therefore, the properties of the social networks can be analyzed by graph theory (Yang and Chen, 2008). The basic terminology regarding SNA is summarized in Table I.

Grounded on the social network theory, a model of online communities' success is proposed in Figure 1.

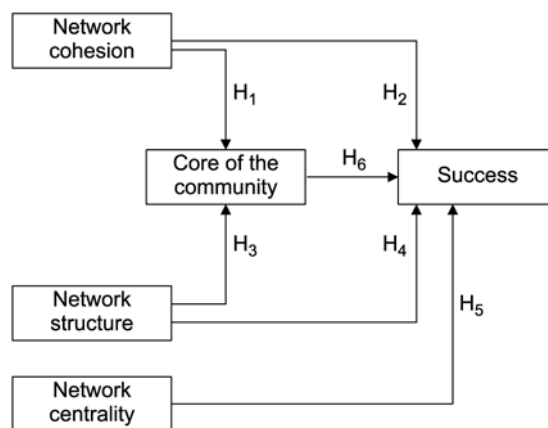
The five considered constructs will be next detailed.

*Network cohesion*

Cohesion describes the overall level of connections in a given network. It is frequently measured using the idea of network density, which is the percentage of all possible lines that are present in a network. However, this definition is not very useful because

Term	Description
Graph	Set of vertices and a set of lines between pairs of vertices
Vertex	Smallest unit in a network
Line	Tie between two vertices in a network
Arc	Directed line
Degree of a vertex	Number of lines incident with it
In-degree of a vertex	Number of arcs it receives
Out-degree of a vertex	Number of arcs it sends
Density of a network	Number of lines in a simple network, expressed as a proportion of the maximum possible number of lines
Betweenness centrality of a vertex	Proportion of all geodesics (shortest path between two vertices) between pairs of other vertices that include this vertex
Betweenness centralization	Variation in the betweenness centrality of vertices divided by the maximum variation in betweenness centrality scores possible in a network of the same size

**Table I.**  
Basic terminology of SNA



**Figure 1.**  
Research model of online communities' success

---

it depends on the size of the network. It is better to look at the number of ties in which each vertex is involved. This is called the degree of a vertex. A higher degree of vertices yields a denser network, because vertices entertain more ties. Therefore, we can use the average degree of all vertices to measure the cohesion of a network. This is a better measure of overall cohesion than density because it does not depend on network size (Nooy *et al.*, 2005).

#### *Network structure*

The challenge of motivating participation revolves around attracting and retaining a critical mass of users. The odds that a community can attract new users are significantly improved if there are already a number of users who contribute to the community. But there is no fixed number of users that defines the threshold for critical mass. Instead, it depends on the number of actively contributing users versus the number of passive users or so-called lurkers (people who just are seeking answers without making any contributions). Usually, this ratio is heavily skewed towards lurkers and online communities can expect 45-90 percent of non-contributing users (Nonnecke and Preece, 2000).

Basically, the following three kinds of participants can be distinguished (Mockus *et al.*, 2002; Xu *et al.*, 2005):

- (1) *Core members*. They are responsible for guiding and coordinating the development of the community. They are usually involved with the community during a long period of time and have made significant contributions to the development and evolution of the system. Moderators and leaders are included in this group.
- (2) *Active developers*. They regularly make contributions to the community.
- (3) *Peripheral developers*. They occasionally contribute with new features to the community. This contribution is irregular, and the period of involvement is short and sporadic. Lurkers are also included in this group.

#### *Core of the community*

The core group is usually responsible of the majority of contributions. But they should be also responsible of ensuring trust and a high response rate (Madanmohan and Navelkar, 2004). That means they are involved in day-to-day interaction and tasks, and they must develop a brokerage role between knowledge seekers and knowledge providers (Sowe *et al.*, 2006).

#### *Network centrality*

A network is highly centralized if there is a clear boundary between the center and the periphery. There are several ways of measuring the centrality of vertices and the centralization of networks. Viewed from a sociocentered perspective, the network as a whole is more or less centralized. But we used an alternative idea of centrality based on the called ego-centered approach. This approach considers that people have better access to information and better opportunities to spread information depending on their position on a network. A person is more central if he or she is more important as an intermediary in the communication network. This idea is included on the concept of betweenness centrality. According to this concept, 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.

#### *Community success*

A challenge for the current literature on the success of online communities is the question of how to measure success. It should be treated as a multi-facet construct. In the literature of information systems (IS), the most widely used model is the one proposed by DeLone and McLean (2003). This model considers that system quality, information quality and service quality affect both user satisfaction and behavioral intention to use IS, which in turn are direct antecedents of IS effectiveness. In the case of internet communities, the widely proposed process outcome measures include level of activities (developer and user contributions), number of developers, and team effectiveness (Preece, 2001; Crowston *et al.*, 2003).

Having defined the conceptual boundary and the constructs of our proposed framework, we will discuss now how these constructs are related to each other.

Cohesive networks facilitate the dissemination of a good reputation and, consequently, make easy new members to join the community (Hagedoorn and Duysters, 2002). As a result, we posit that cohesion would have a positive impact in the success of the virtual community. On the other hand, network cohesion promotes the involvement of developers in the underlying project leading to a bigger core group. Therefore, we propose the following hypotheses:

- H1.* Network cohesion will have a positive impact on the core group of the community.
- H2.* Network cohesion will have a positive impact on the success of the community.

Network structure is due to the different community members profiles. The presence of this kind of structure leads to a core group of developers with the ability of taking decisions. Besides, participation inequality typical of the mentioned network structure has been demonstrated to be beneficial for the development of the community (Kuk, 2006). Therefore, we propose that:

- H3.* Network structure will have a positive impact on the core group of the community.
- H4.* Network structure will have a positive impact on the success of the community.

A project with a higher betweenness centrality is able to spread its various complex and tedious tasks over a larger number of developers, leading to increased technical success and improved efficiency and quality in serving users' needs. A project with a high betweenness centrality has a larger number of developers working closely within the project team and such strong ties benefit the project through the exchange of high-quality information and tacit knowledge (Rowley *et al.*, 2000). Consequently, we propose that:

- H5.* Network centrality will have a positive impact on the success of the community.

---

The core group involvement is critical for building relationships and developing user-created content. They should promote collaboration and trust among community members as well as stimulate participation (Koh *et al.*, 2007). That means they are not only the most frequent contributors but they also should bridge the gap between expert software developers and user communities (Sowe *et al.*, 2006). Therefore, we proposed the following hypothesis:

*H6.* An active core group is positively related to the community success.

### Methodology

The proposed hypotheses were tested using several online communities related to Linux port to non-conventional processors and environments. Linux is perhaps the most well known open source project, and it is developed under the scheme of an online community of support (Lee and Cole, 2003). Just as Windows is the most prominent operating system released under a proprietary software license, Linux is the most prominent operating system released under a free license like GPL (Hertel *et al.*, 2003). Nevertheless, the proposed case study was focused on Linux ports to other processor architectures not intended for desktop or personal computer market (Table II). There are several reasons for this choice. First, Linux is firmly in first place as the operating system of choice for smart gadgets and embedded systems (Henkel, 2006). Second, in contrast to other typical open source projects or even desktop Linux projects, most contributions in this field do not come from volunteers or hobbyists, but from commercial firms, many of which are dedicated embedded Linux firms. Third, there are a lot of communities supporting each one of these Linux ports, and this is an excellent opportunity for analyzing a big group of more or less “homogeneous” communities.

The 11 proposed online communities of Table II were analyzed during the period of time between 2001 and 2007. Each year was considered an independent case study. Consequently, up to 77 cases were considered.

The community interactions are usually structured in 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.

SNA techniques were used to obtain several indicators able to measure the five latent constructs of the research framework. 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, equation (1):

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

In case of a valued graph, real valued weight function  $w(e)$  is defined on the set of edges, i.e.  $w(e) = Ex\Re$ , and the matrix is then defined as given by equation (2).

Project	URL	Description
The ARM Linux Project (ARM)	<a href="http://www.arm.linux.org.uk/">www.arm.linux.org.uk/</a>	ARM Linux is a port of the successful Linux Kernel to ARM processor-based machines
Debian port to ARM (D-ARM)	<a href="http://lists.debian.org/debian-arm/">http://lists.debian.org/debian-arm/</a>	ARM port for Debian GNU/Linux. Debian fully supports a port to little-endian ARM
Linux PPC port (PPC)	<a href="http://penguinppc.org/">http://penguinppc.org/</a>	PowerPC Linux is the Linux kernel running on a PowerPC processor
Debian port to PowerPC (D-PPC)	<a href="http://lists.debian.org/debian-powerpc/">http://lists.debian.org/debian-powerpc/</a>	PowerPC port of Debian GNU/Linux. The PowerPC architecture allows both 64-bit and 32-bit implementations
Debian port to m68k (D-68k)	<a href="http://lists.debian.org/debian-68k/">http://lists.debian.org/debian-68k/</a>	Motorola 68k port of Debian GNU/Linux. Debian currently runs on the 68020, 68030, 68040 and 68060 processors
Debian port to Alpha (D-Alpha)	<a href="http://lists.debian.org/debian-alpha/">http://lists.debian.org/debian-alpha/</a>	The purpose of this project is to assist developers and others interested in the ongoing project to port the Debian distribution of Linux to the Alpha family of processors
Debian port to MIPS (D-MIPS)	<a href="http://lists.debian.org/debian-mips/">http://lists.debian.org/debian-mips/</a>	MIPS port of Debian GNU/Linux, able to run at both endiannesses
Debian port to BSD (D-BSD)	<a href="http://lists.debian.org/debian-bsd/">http://lists.debian.org/debian-bsd/</a>	This is a port of the Debian operating system, complete with apt, dpkg, and GNU userland, to the NetBSD kernel
Debian port to HPPA (D-HPPA)	<a href="http://lists.debian.org/debian-hppa/">http://lists.debian.org/debian-hppa/</a>	This is a port to Hewlett-Packard's PA-RISC architecture
Debian port to Hurd (D-HURD)	<a href="http://lists.debian.org/debian-hurd/">http://lists.debian.org/debian-hurd/</a>	The GNU Hurd is a totally new operating system being put together by the GNU group
Debian port to SPARC (D-SPARC)	<a href="http://lists.debian.org/debian-sparc/">http://lists.debian.org/debian-sparc/</a>	This port runs on the Sun SPARCstation series of workstations, as well as some of their successors in the sun4 architectures

**Table II.**  
Virtual communities considered in the proposed case study

$$m_{i,j} = \begin{cases} w(e) & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

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. 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 (Kuk, 2006). 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 (Knock, 2001). That is the reason why an author posting to a thread was tied to all the authors who have previously posted to the same thread when constructing the social network.

The cohesion of the network was measured using the idea of the degree of a vertex or, more precisely, the idea of outdegree. Outdegree of a vertex is the number of arcs it



sends to other vertices. We are interested in the outdegree because it shows the flow of information between community members. The higher the outdegree of a vertex is, the more this vertex is contributing to the community. Therefore, the outdegree can be used to distinguish the different kind of community members. In particular, we used the community average outdegree as the threshold value to be considered an active member of the community and the community average outdegree plus its standard deviation as the threshold to be considered a core member of the community. Using these thresholds values, the active developers' partition and the core partition can be extracted from the whole community. The average outdegree of these two key partitions was used to measure the cohesion of the network (indicators I1 and I2 of Table III).

The structure of the community is defined by the ratios core/active developers (I3) and active developers/peripheral developers (I4).

The core activity was determined by its size (I5) and the number of core developers performing a brokerage role (I6). This last indicator is essential to promote co-learning experience within the community (Sowe *et al.*, 2006).

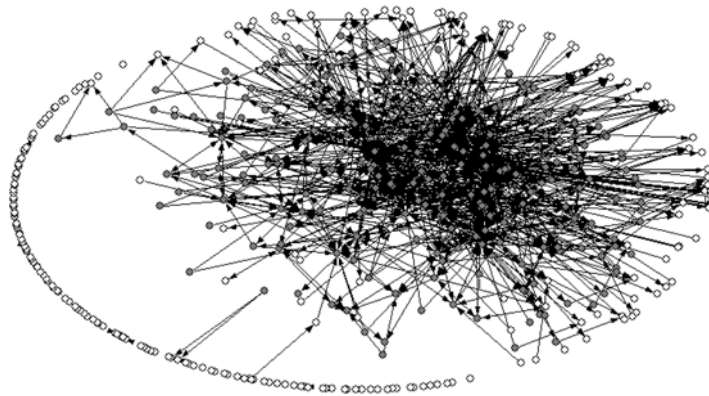
Network centrality was measured using the concept of betweenness centrality. In particular, betweenness centrality of the whole community and active developers' partition (I7 and I8, respectively) were used as indicators of the network centrality construct.

Regarding measurement of the project success, researchers agree that software success is a multi-facet construct. The widely proposed process outcome measures include the size of the community (I9 and I10) and its activity (I11). They are in line with previously proposed measurements in the literature (Crowston *et al.*, 2003).

An example of the resulting social network is illustrated in Figure 2. Vertices represent community members and arcs represent the flow of information through threads of discussion. Arcs are valued with a value showing the number of interactions, although it has been omitted in Figure 2 for clarity purposes. External vertices (filled in white) correspond to peripheral members of the community, characterized by scarce interventions, while inner vertices (filled in grey) correspond to

**Table III.**  
Indicators for measuring the five latent constructs proposed in the research framework

	Indicator	Description
I1	AvCoreOutdegree	Average outdegree of the core partition of the community.
I2	AvActiveOutdegree	Average outdegree of the active developers partition of the community
I3	Core/ActiveDev	Ratio of core developers to active developers
I4	ActiveDev/PeriphDev	Ratio of active developers to peripheral developers
I5	CoreSize	Size of the core group of the community
I6	CoreBrokers	Number of core developers performing a brokerage role
I7	BetwComm	Betweenness centrality of the community
I8	BetwActive	Betweenness centrality of the active developers' partition
I9	ComSize	Community size
I10	ActiveDev	Number of active developers
I11	Threads	Number of threads



**Figure 2.**  
Social network of ARM  
Debian Linux mailing list  
community during 2007

---

active members. The direction of the arc is important because shows the flow of knowledge. That means that a vertex with a lot of inner arcs would be an information receptor while vertices with a lot of outer arcs would be information providers. Vertices without arcs are passive observers. Betweenness centrality is determined by the ability of each vertex to go between two other vertices.

## Results

A structural equation modeling (SEM) analysis has been used to test the research hypotheses. Within SEM techniques, two approaches can be identified. The covariance-based SEM (Bollen and Lennox, 1991; Jöreskog, 1970) is a widely accepted approach, in part due to the development of several computer programs (e.g. LISREL, EQS and AMOS) allowing the estimation of this kind of model. In contrast, partial least squares (PLS) approach has a primary objective to minimize errors in all endogenous constructs (Chin, 1998). Between the two alternative SEM approaches, we have selected PLS mainly because the required assumptions of the observed variables and the sample size are much smaller than in covariance-based SEM (Toral *et al.*, 2007; Martínez-Torres *et al.*, 2008).

Before the testing and assessment of the structural model, we analyzed the reliability of individual reflective items and the corresponding constructs, as well as the convergent validity and discriminant validity of our measures (Table IV). All the reflective item loadings are significant and greater than 0.7, which is the generally recommended threshold (Chin, 1998). We evaluated composite reliability using the internal consistency measure ( $\rho_c$ ) developed by Fornell and Larcker (1981). A construct is considered reliable if  $\rho_c$  is at least 0.70 (Nunnally, 1978). We also examined the average variance extracted (AVE) of each latent construct, which should be greater than 0.5. All the reflective constructs exceed these conditions (see Table IV).

Discriminant validity is established by comparing the square root of AVE (i.e. the diagonals in Table V) with the correlations among latent variables (i.e. the off-diagonal elements in Table V). For each reflective construct, the square root of its AVE should be greater than its correlation with any other construct, which means that it shares more variance with its own measures than with other constructs in the model. This condition is met in all the cases.

Constructs and indicators	Weight	Loading	<i>t</i> -statistic
<i>Success</i> ( $\rho_c = 0.932$ , AVE = 0.821)			
Threads	0.3675	0.9104	55.48
ComSize	0.3294	0.8876	22.39
ActiveDev	0.4056	0.9199	78.18
<i>Cohesion</i> ( $\rho_c = 0.944$ , AVE = 0.894)			
AvActiveOutdegree	0.5755	0.9554	22.18
AvCoreOutdegree	0.4812	0.9355	17.17
<i>Core</i> ( $\rho_c = 0.980$ , AVE = 0.961)			
CoreSize	0.5177	0.9810	245.03
CoreBrokers	0.5022	0.9798	224.69
<i>Structure</i> ( $\rho_c = 0.722$ , AVE = 0.566)			
Core/ActiveDev	-0.5995	-0.6965	9.93
ActiveDev/PeriphDev	0.7241	0.8044	15.00
<i>Centralization</i> ( $\rho_c = 0.903$ , AVE = 0.824)			
BetwComm	1.1113	0.9975	4.05
BetwActive	-0.1343	0.8084	3.46

**Table IV.**  
Reliability and convergent validity of the proposed model

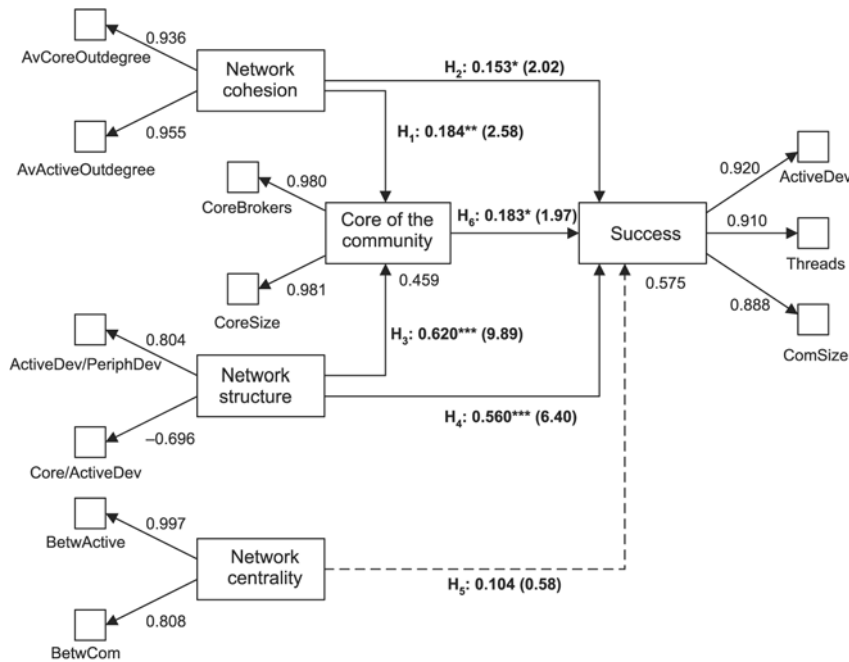
	Success	Cohesion	Core	Structure	Centralization
Success	0.906				
Cohesion	0.329	0.945			
Core	0.603	0.293	0.980		
Structure	0.709	0.176	0.653	0.752	
Centralization	0.165	0.226	0.092	0.018	0.980

**Table V.**  
Correlation matrix

Figure 3 summarizes the results of the PLS analysis performed to test the structural model (Chin, 2003). Standardized path coefficients and their significance level (*t* statistic) between brackets are detailed to test the proposed hypotheses. The value of the  $R^2$  of the two dependent variables is also shown below the corresponding constructs.

All our proposed hypotheses are supported, with the exception of *H5*. We have not found a significant direct relationship between network centrality and project success. The explanation could be that the centrality of the network has been measured based on the concept of betweenness, and the brokerage role is essentially developed by the core group of the community. Consequently, this idea is included in the latent construct referred to the core group.

Network cohesion exhibits a direct and an indirect effect on the project success. The cohesion of the network is supporting the mechanism of participation, necessary for the project development, and participation promotes the success of the underlying project increasing the number of threads and contributions. On the other hand,



**Notes:** \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ;  $t$  statistic values appear in parentheses. Significant relationships appear in bold and with a continuous path

**Figure 3.** Structural model results

cohesion also means a cohesive core group. Consequently, there is an indirect effect on the project success mediated by the activity of the core of the community.

Network structure is the most important antecedent attending to standardized path coefficients. This construct also exhibits a direct and indirect influence on the project success. The structure of the network responds to the participation inequality principle recognized by numerous authors (Kuk, 2006; Sowe *et al.*, 2006). The clear distinction among core, active and peripheral developers promotes the project success. In particular, a good structure promotes a strong core of the community addressing the future development of the underlying project.

Finally, the activity of the core group is essential for the project success. This construct occupies a central position in the proposed model and it is not only causing a direct effect on success, but also mediates an indirect effect of the previous constructs.

### Conclusion

Since two decades ago, internet has promoted the emergence of new kinds of communities. This upward trend has become more pronounced with the advent of Web 2.0 technologies. Communities are basically based on interactions among users, and participation is the basic mechanism promoting its development. This participation has been analyzed using social network analysis techniques. Features like cohesion, structure, centralization and user profiles has been analyzed and defined, and several indicators has been evaluated for a set of online communities related to Linux ports.

---

A research framework considering several antecedents of the project success has been proposed and tested using structural equation modeling analysis. The obtained results reveal that the network structure is the most important antecedent, which is in consonance with the participation inequality typical of online communities. Network cohesion and core group activity also supports the development of the community. These latent constructs have been tested and shown to be the driving forces under which online communities are likely to succeed.

### References

- Albors, J., Ramos, J.C. and Hervás, J.L. (2008), "New learning network paradigms: communities of objectives, crowdsourcing, wikis and open source", *International Journal of Information Management*, Vol. 28 No. 3, pp. 194-202.
- Amin, A. and Roberts, J. (2008), "Knowing in action: beyond communities of practice", *Research Policy*, Vol. 37 No. 2, pp. 353-69.
- Barcellini, F., Détienne, F., Burkhardt, J.-M. and Sack, W. (2008), "A socio-cognitive analysis of online design discussions in an Open Source Software community", *Interacting with Computers*, Vol. 20 No. 1, pp. 141-65.
- Barrero, F., Toral, S.L. and Gallardo, S. (2008), "eDSPLab: remote laboratory for experiments on DSP applications", *Internet Research*, Vol. 18 No. 1, pp. 79-92.
- Bollen, K. and Lennox, R. (1991), "Conventional wisdom on measurement: a structural equation perspective", *Psychological Bulletin*, Vol. 110 No. 2, pp. 305-14.
- Bonaccorsi, A. and Rossi, C. (2003), "Why open source software can succeed", *Research Policy*, Vol. 32, pp. 1243-58.
- Cassell, J., Huffaker, D., Tversky, D. and Ferriman, K. (2005), "How to win a world election: emergent leadership in an international online community", in Van den Besselaar, P., de Michelis, G., Preece, J. and Simone, C. (Eds), *Communities and Technologie*, Springer, Dortmund.
- Chin, W.W. (1998), "The partial least squares approach for structural equation modelling", in Marcoulides, G.A. (Ed.), *Modern Methods for Business Research*, Erlbaum, Mahwah, NJ, pp. 295-336.
- Chin, W.W. (2003), *PLS-Graph*, University of Houston, (Version 3.00, Build 1058), [Computer software].
- Crowston, K., Annabi, H. and Howison, J. (2003), "Defining open source software project success", *ICIS 2003 Proceedings*, available at: <http://aisel.qisnet.org/icis2003/28>
- DeLone, W.H. and McLean, E.R. (2003), "The DeLone and McLean model of information systems success: a ten-year update", *Journal of Management Information Systems*, Vol. 19, pp. 9-30.
- De Souza, C.S. and Preece, J. (2004), "A framework for analyzing and understanding online communities", *Interacting with Computers*, Vol. 16, pp. 579-610.
- Fornell, C. and Larcker, D. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18, pp. 39-50.
- Garton, L., Haythornthwaite, C. and Wellman, B. (1997), "Studying online social networks", *Journal of Computer-Mediated Communication*, Vol. 3 No. 1.
- Hagedoorn, J. and Duysters, G. (2002), "Learning in dynamic inter-firm networks: the efficacy of multiple contacts", *Organization Studies*, Vol. 23 No. 4, pp. 525-48.
- Henkel, J. (2006), "Selective revealing in open innovation processes: the case of embedded Linux", *Research Policy*, Vol. 35 No. 7, pp. 953-69.

- 
- Hertel, G., Niedner, S. and Herrmann, S. (2003), "Motivation of software developers in open source projects: an internet-based survey of contributors to the Linux kernel", *Research Policy*, Vol. 32, pp. 1159-77.
- Jöreskog, K.G. (1970), "A general method for analysis of covariance structures", *Biometrika*, Vol. 57, pp. 239-51.
- Kautz, H., Selman, B. and Shah, M. (1997), "Referral web: combining social networks and collaborative filtering", *Communications of ACM*, Vol. 40 No. 3, pp. 27-36.
- Knock, N. (2001), "Compensatory adaptation to a lean medium: an action research investigation of electronic communication in process involvement groups", *IEEE Trans. on Professional Communication*, Vol. 44 No. 4, pp. 267-85.
- Koh, J., Kim, Y.-G., Butler, B. and Bock, G.-W. (2007), "Encouraging participation in virtual communities", *Communications of the ACM*, Vol. 50 No. 2, pp. 69-73.
- Kuk, G. (2006), "Strategic interaction and knowledge sharing in the KDE developer mailing list", *Management Science*, Vol. 52 No. 7, pp. 1031-42.
- Lee, G.K. and Cole, R.E. (2003), "From a firm-based to a community-based model of knowledge creation: the case of the Linux kernel development", *Organization Science*, Vol. 14 No. 6, pp. 633-49.
- Lesser, E.L., Fontaine, M.A. and Slusher, J.A. (2000), *Knowledge and Communities*, Butterworth-Heinemann, Boston, MA, pp. 86-98.
- Lin, H.-F. and Lee, G.-G. (2006), "Determinants of success for online communities: an empirical study", *Behaviour and Information Technology*, Vol. 25 No. 6, pp. 479-88.
- Madanmohan, T.R. and Navelkar, S. (2004), "Roles and knowledge management in online technology communities: an ethnography study", *International Journal of Web Based Communities*, Vol. 1 No. 1, pp. 71-89.
- Martínez-Torres, M.R., Toral, S.L., Barrero, F., Gallardo, S., Arias Oliva, M. and Torres, T. (2008), "A technological acceptance of e-learning tools used in practical and laboratory teaching, according to the European higher education area", *Behaviour and Information Technology*, Vol. 27 No. 6, pp. 495-505.
- Mockus, A., Fielding, R.T. and Herbsleb, J.D. (2002), "Two case studies of open source software development: Apache and Mozilla", *ACM Transactions on Software Engineering and Methodology*, Vol. 11 No. 3, pp. 309-46.
- Nonnecke, B. and Preece, J. (2000), "Lurker demographics: counting the silent", *Proceedings of Human Factors in Computing Systems (CHI 2000)*, ACM Press, New York, NY, pp. 73-80.
- Nooy, W., Mrvar, A. and Batagelj, V. (2005), *Exploratory Network Analysis with Pajek*, Cambridge University Press, New York, NY.
- Nunnally, J. (1978), *Psychometric Theory*, McGraw-Hill, New York, NY.
- Pan, S.L. and Leidner, D.E. (2003), "Bridging communities of practice with informaton technology in pursuit of global knowledge sharing", *Journal of Strategic Information Systems*, Vol. 12, pp. 71-88.
- Preece, J. (2001), "Sociability and usability: twenty years of chatting online", *Behaviour and Information Technology Journal*, Vol. 20 No. 5, pp. 347-56.
- Raghavan, P. (2002), "Social networks: from the web to the enterprise", *IEEE Internet Computing*, Vol. 6 No. 1, pp. 91-4.
- Rheingold, H. (1993), *The Virtual Community: Homesteading on the Electronic Frontier*, Addison-Wesley, Reading, MA.

- 
- Rowley, T., Behrens, D. and Krackhardt, D. (2000), "Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries", *Strategic Management Journal*, Vol. 21 No. 3, pp. 369-86.
- Sack, W., De'tienne, F., Burkhardt, J.M., Barcellini, F., Ducheneaut, N. and Mahendran, D. (2006), "A methodological framework for socio-cognitive analyses of collaborative design of open source software", *International Journal of Computer Supported Collaborative Work*, Vol. 15 No. 2/3, pp. 229-50.
- Shang, R.-A., Chen, Y.-C. and Liao, H.-J. (2006), "The value of participation in virtual consumer communities on brand loyalty", *Internet Research*, Vol. 16 No. 4, pp. 398-418.
- Sowe, S., Stamelos, I. and Angelis, L. (2006), "Identifying knowledge brokers that yield software engineering knowledge in OSS projects", *Information and Software Technology*, Vol. 48 No. 11, pp. 1025-33.
- Toral, S.L., Barrero, F., Martínez-Torres, M.R., Gallardo, S. and Lillo, J. (2005), "Implementation of a web-based educational tool for digital signal processing teaching using the technological acceptance model", *IEEE Transactions on Education*, Vol. 48 No. 4, pp. 632-41.
- Toral, S.L., Barrero, F. and Martínez-Torres, M.R. (2007), "Analysis of utility and use of a web-based tool for digital signal processing teaching by means of a technological acceptance model", *Computers and Education*, Vol. 49 No. 4, pp. 957-75.
- Trung, T.D.-T. and Bieman, J.M. (2005), "The FreeBSD project: a replication case study of open source development", *IEEE Transactions on Software Engineering*, Vol. 31 No. 6, pp. 481-94.
- Von Hippel, E. and von Krogh, G. (2003), "Open source software and the 'private-collective' innovation model: issues for organization science", *Organization Science*, Vol. 14 No. 2, pp. 209-23.
- Wachter, R.M., Gupta, J.N.D. and Quaddus, M.A. (2000), "IT takes a village: virtual communities in support of education", *International Journal of Information Management*, Vol. 20, pp. 473-89.
- Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, NY.
- Wenger, E. (1998), *Communities of Practice: Learning, Meaning, and Identity*, Cambridge University Press, Cambridge.
- Wellman, B. and Gulia, M. (1995), *Net Surfers Don't Ride Alone: Virtual Communities as Communities*, University of California Press, Berkeley, CA.
- Xu, J., Gao, Y., Christley, S. and Madey, G. (2005), "A topological analysis of the open source software development community", *Proceedings of the 38th Hawaii International Conference on System Sciences, Hawaii*, IEEE Computer Society Press, Los Alamitos, CA.
- Yang, S.J.H. and Chen, I.Y.L. (2008), "A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network", *Int. J. Human-Computer Studies*, Vol. 66, pp. 36-50.
- Zhao, S., Grasmuck, S. and Martin, J. (n.d.), "Identity construction on Facebook: digital empowerment in anchored relationships", *Computers in Human Behavior*, in press.