Analysing Collaboration in OPAALS' Wiki: A Comparative Study among Collaboration Networks

Fernando A. Basile Colugnati and Lia Carrari R. Lopes

Instituto de Pesquisa em Tecnologia e Inovação,
Av. São Luis, 86 cj 192, São Paulo, SP, Brazil
{fernando,lia.carrari}@ipti.org.br

Abstract. This work aims to analyse the wiki tool from OPAALS as a collaborative environment. To achieve that, methods from social network analysis and statistics are employed. The analysis is compared with other collaboration networks. The results obtained here show the evolution of the tool and that the adoption was successful.

Keywords: Social Network Analysis, collaborative process, Virtual Communities, wiki.

1 Introduction

OPAALS Network of Excellence (NoE) "aims to work toward an integrated theory of digital ecosystems, to create new disciplines from the merging of old disciplines, and to do so by leveraging community values and principles." (Phase II Workplan, pp. 5). In this NoE, the comprehension of how networks behave and communities emerge can be considered a self reflection for the consortium governance. Another significant consideration is that this is an important research framework for the Social and Computer Sciences domain since SMEs scenarios in the adoption and use of Information and Communication Technologies (ICTs) for collaborative and democratic business is one of the main objectives of Digital Business Ecosystems (DBEs).

Information and communication technologies (ICTs) have an effect on coordination, communication and control in all societal networks and communities. OPAALS developed the concept of Open Knowledge Space (OKS) - as part of a broad definition - that is technically a set of computational tools for collaboration and knowledge construction by means of virtual communities. Created to be the first tool of the OKS, the wiki tool was until 2008 the most used environment for collaboration in OPAALS. The main knowledge repository for the consortium is formed by: Management issues, events calendar, dissemination material, Work Packages descriptions, among others.

F.A. Basile Colugnati et al. (Eds.): OPAALS 2010, LNICST 67, pp. 109–117, 2010.

So far there are only few studies analyzing in detail how virtual knowledge networks grow, how the process of communication and articulation between their nodes (local actors) happens, and the dynamics that are behind those aspects. More research is needed to generate knowledge about applicable strategies and mechanisms that are essential in order to make a newly established network successful and sustainable.

From the analytical point of view, the community emergence processes are nowadays understood as complex emergent systems [1-4], driving quantitative methods to look at them as physical and dynamic structures. The development of Social Network Analysis (SNA) methods, that in the beginning was static and purely deterministic, is now directed to dynamic stochastic processes where the whole of the actors are the central point, modelling the network emergence form a bottom-up perspective that will lead to a certain topology. The dynamic process led by actors by means of selection and social influence, and the topology of the network both provide interpretable parameters and metrics that quantify the intensity and velocity of the emergence and formation of communities or sub-communities.

This paper explores the evolution of collaborations in this repository using methods for Social Network Analysis over the four years of wiki existence. Also, the results are compared with other collaboration networks that used different ways for communication.

2 Methods

To verify the collaboration process over tasks defined in work packages, only the WP pages are considered, from the first publication in 2006 until June of 2009. Data used in this work were extracted with the toolkit Wille2 [2], developed by TUT.

In terms of SNA, two approaches are used. The classical approach using the methods based on degree, centrality and clustering [6] using these measures in time series and the final degree distributions was estimated to explore a possible scale-free behaviour [7]. Also, the dynamic of the network is modelled using methods presented by Snijders [8], and already presented for OPAALS consortium in [9] and [10].

3 Social Network Analysis for OPAALS' Wiki

Table 1 describes the OPAALS wiki network, which is not a significantly large network. It had a major growth between 2007 and 2008, with 37 collaborating researchers. Fig. 1 evidences that it does not follow a power-law, but show that there are two sets of actors in terms of degrees, above and below the horizontal line formed by the dots ("circled dots"). The set above the line are actors with few degrees, from 2 to 4. The "circled dots" comprises a few actors with degrees around from 10 to 15, and then actors with more than 20 ties.

Table 1. Number of actors and network degree for OPAALS Wiki

Year	Active Users	Network Degree
2006	15	46
2007	23	80
2008	34	222
2009	37	224

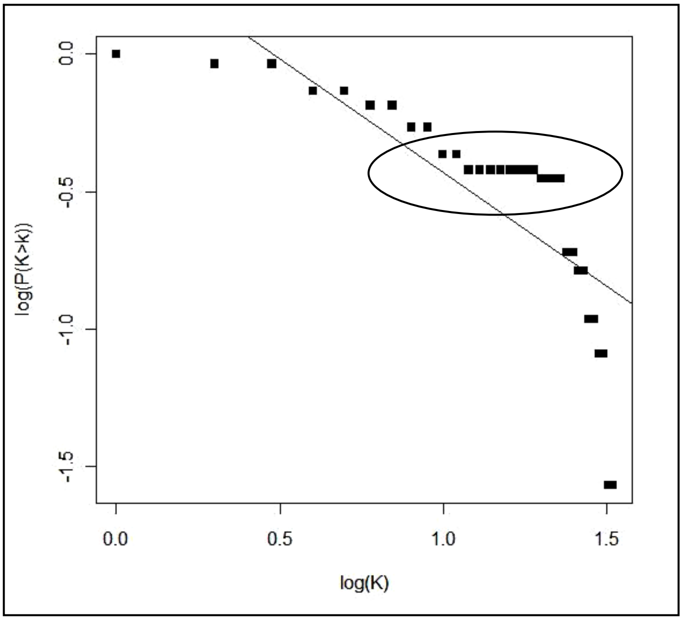

Fig. 1. Cumulative degree distribution for OPAALS Wiki collaboration network

The social graphs or sociograms confirm the described structure (Fig. 2). In 2009 in the accumulated network, a main core is formed by 12 actors (bottom part) with high degree values and many others spread out around this core. It is noticeable also the presence of many articulators, making bridges between smaller structures with the main one. Also, the interdisciplinary characteristic is a remarkable aspect; in Fig. 2 each colour represent a research domain and the blue one are consultants or administrative persons.

In terms of evolution, the increasing transitivity (Fig. 3d) claims attention, since this effect normally decreases in this type of network. The presence of many articulators can explain this high transitivity, increasing the probability of new ties mediated by them.

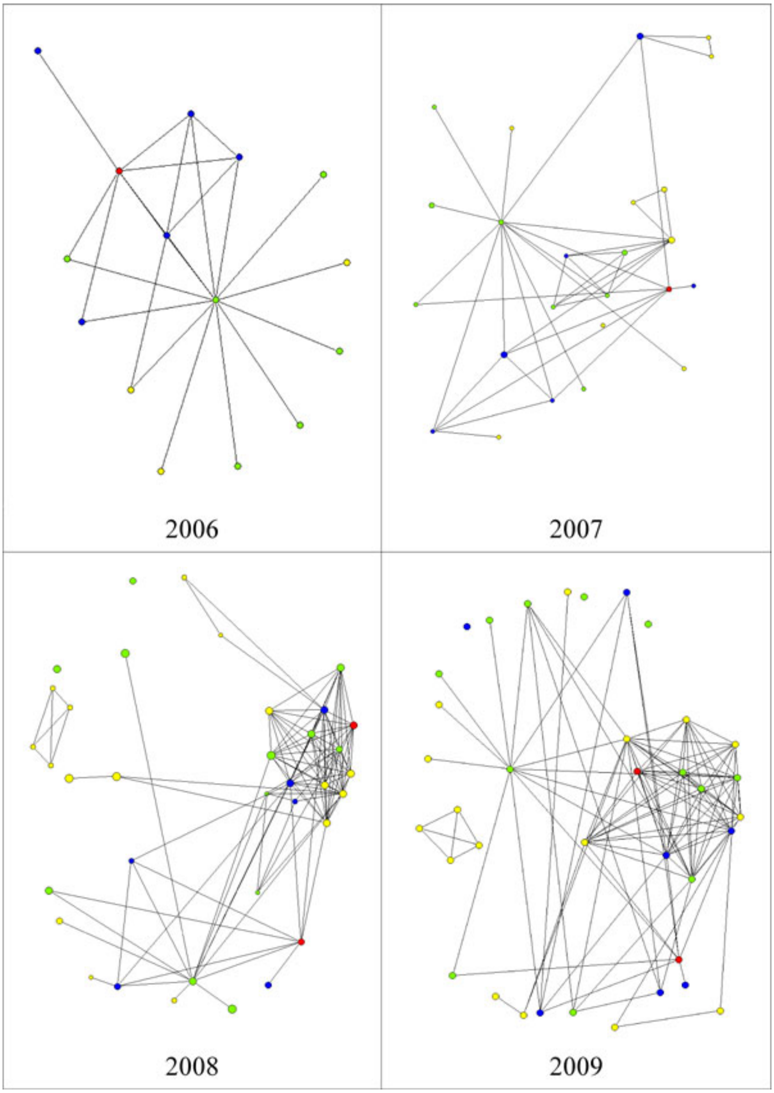

Fig. 2. Sociograms for the OPAALS wiki collaboration evolution. Yellow = Social Sciences, Green = Computer Science, Red = Natural Sciences, Blue = Consultants/Adm.

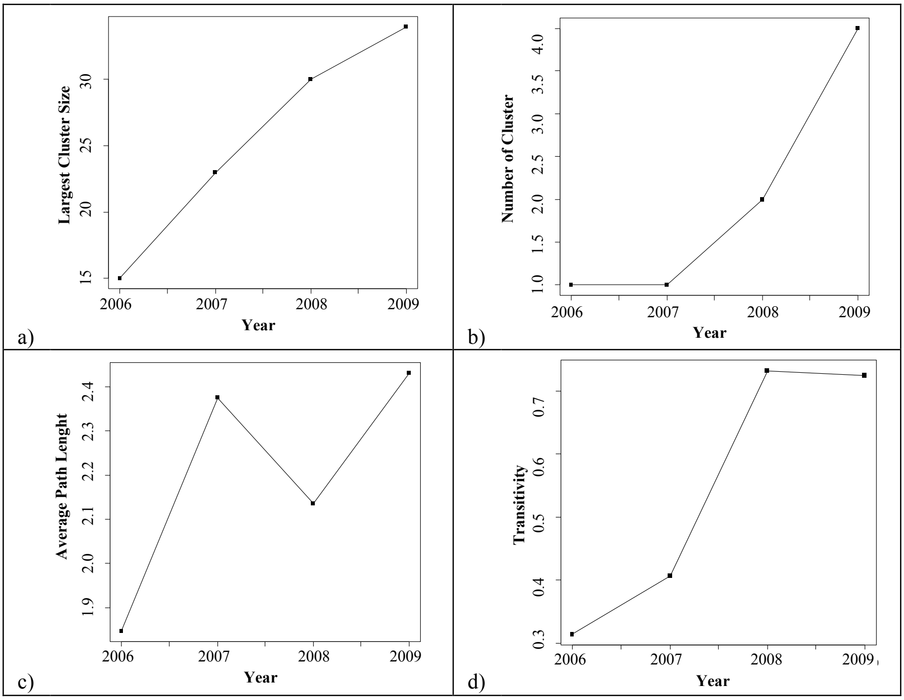

Fig. 3. a) Larger cluster size, b) Number of Clusters, c) Average path length and d) Clustering Coefficient evolutions for OPAALS wiki

4 Stochastic Dynamic Modeling

OPAALS wiki is modelled as a non-directed network (meaning that the relationships are reciprocal), adequate objective functions are introduced, as well as their effects interpretation. A covariate was introduced representing the domain of actors, as represented in the sociograms. Despite the fact that these domains have been coded as ordinal numbers, this was not taken in account in the analysis, being considered a categorical (nominal) variable. The rate parameters presented here have the same meaning and interpretation as in the directed case. Four objective functions were tested for this model, firstly without any covariate effect:

1. *Transitive ties*: is similar to the transitive triplets effect, but instead of considering for each other actor j how many two-paths i - h - j are there, it is only considered whether there is at least one indirect connection. Thus, one indirect tie suffices for the network embeddedness. It can be interpreted as a measure of clustering for the network, or even network closure.
2. *Betweenness*: represents brokerage: the tendency for actors to position themselves between not directly connected others, i.e., a preference of i for ties i - j to those j for which there are many h with h - i and h !-j ("!-"means not tied).

3. *Number of actors at distance 2*: expresses network closure inversely: stronger network closure (when the total number of ties is fixed) will lead to smaller geodesic distances equal to 2. When this effect has a negative parameter, actors will have a preference for having few others at a geodesic distance of 2 (given their degree, which is the number of others at distance 1); this is one of the ways for expressing network closure.
4. *Assortativity*: reflects tendencies to actors with high degrees to preferably be tied to other actors with high degree. If this parameter is positive and significant, preferential attachment is not likely.

One covariate effect is tested, the *Identity*, in this case the Domain identity. Positive values for this effect means that actors tend to choose similes in domain, a homophily representation. Due to the small number of actors in 2006, which means that too many structural zeros are present, the analysis will comprises 2007, 2008 and 2009 only. Table 2 present the parameters and different models adjusted. A parameter is considered statistically significant if the *t-statistic* is bigger than | 2 |.

Table 2. Dynamic Modeling for the Wiki collaboration

Effect	Parameter	Std. Error	t-statistic
Rate Parameters			
2007 - 2008	2.77	0.33	8.34
2008 - 2009	0.02	0.02	1
Objective functions (simple models)			
Transitive triads	0.63	0.54	1.16
Betweeness	-0.56	-	-
Assortativity	0.34	0.12	2.81
Multieffects Models			
Assortativity	0.34	0.12	2.81
Domain	-0.70	0.24	2.91

The estimated rates shows that in the first period, each actor had at least 3 chances to establish a tie, or at least 3 actors could collaborate with other, while in the second period this is almost 0, which is coherent with the data. Looking at the sociograms and the network evolution, few new actors entered and few new contributions were made from 2008 to 2009. So, the great part of the effects presented bellow happened between 2007 and 2008.

The only effects on objective functions statistically significant were the Nad2 and the *Assortativity*. Transitive ties effect was not significant and *Betweenness* did not achieve numerical convergence, so that standard errors could not be computed. This situation means a lack of stability for the function used, given the dataset, i.e., it is a numerical problem rather than a network topology problem. The negative value for Nad2 express the tendency to the network closure, such that few actors are separated at a distance two.

This is can be complementary to the positive *Assortativity*, where the selection of actors is lead by the degree centrality, such that, high degree actors are more likely to tie with other high degree actors. When adjusting for Domain, that is negative and significant, Nad2 loose statistical importance, but *Assortativity* remains the same. The negative sign means that the selection of ties happens among different domains preferably, and this selection does not depend on the *Assortativity* since this parameter is the same in the simple and multivariate model.

To interpret this composition of effects is not a simple task, and the definition of these effects above need to be verified. Nad2 expresses the effect of distance 2 actors, given the degree of that actor (number of actors at distance 1). Considering *Assortativity* a degree effect, where actors tend to choose others with same degree, the chance so that distance 2 links vanishes. The interesting conclusion is that it does not depend on the Domain homophily; actors collaborate with others with the same intensity of collaborative activity, and not by the research domain. By having the NoE as the interdisciplinary characteristic, this is an expected effect. Next, the collaboration process will be compared with other analysis.

5 Comparing to Other Collaboration Processess

In terms of information and communication technology, the OKS is a set of online collaboration tools such as chats, wikis, document editing tools, forums among others. The understanding of how the process of collaboration behaves in each of those virtual environments, in terms of the network formed by the researchers, can help one to follow the actual OKS dynamics. It also allows for intervention planning and strategies to keep this a sustainable community. Workshops, conferences, meetings, web-conferences are all ways to intervene in, foster or change direction in the use of the OKS.

In D 10.8 [11], twelve different collaboration networks were analysed, such as Forums, Wikis and co-authoring for scientific papers. All those networks could then be compared. The idea is to assess differences in terms of topology and clustering, having this last network property as an indicative for communities emergency.

It is clear from Table 3 that collaboration process differs basically in the environment in which it happens. Behaviour and properties are similar for both forums and co-authoring networks. All networks that follow approximately the power law distribution, present coefficients between 2 and 3, as found in many other studies about collaboration networks [12]. Transitivity is lower for Forums than for any other kind of text elaboration, as is suggested by the imposed and fixed treelike structure, forcing in a certain way the preferential attachment.

Then, the probability of new links among any 2 step actors is low, while for a structure free environment like a Wiki this tends to be high, as those actors have more chances to collaborate with any other actor. For the co-authoring network, the only one where ICT is not the environment, although email was and still is quite useful to write in a collaborative way, the behaviour is a little different. They are truly small- worlds but not entirely scale-free networks in the examples studied. This result does not match with other findings, like Barabasi *et al.* [13], who identified this behavior. Here, degree distributions tended to present heavier tales due to an effect sometimes called "finite population effect", in this case that dataset is not complete in terms of all collaborations or authors. The dataset studied just had ISI indexed papers, so journals not indexed by this database were not contemplated.

Table 3. Scale-free and small-wordness assessment parameters for networks in the last panel

Discipline	*α*	*l*	*R (l)*	*CC*	*R(CC)*
Mathematics	3.23*	3.17	0.48	0.55	**151.67**
Applied Mathematics	3.30	6.22	1.43	0.58	**237.95**
Interdisciplinary Applications	2.84	4.91	1.27	0.38	**115.23**
Computation Science	3.22*	2.95	0.54	0.70	**303.76**
Artificial Intelligence	2.76	2.66	0.59	0.63	**60.00**
Interdisciplinary Applications	2.70	1.38	0,29	0.85	**99.67**
Psychology	3.18*	3.36	0.73	0.76	**115.35**
Treatment and Prevention	3.10	2.49	0.66	0.78	**63.85**
Multidisciplinary	2.76	1.38	0.29	0.84	16.68
Supera	1.98	3.06	0.77	0.06	4.57
Converse	2.43	2.88	0.88	0.18	10.24
OPAALS Wiki	1.94	2.43	1.12	0.72	4.05

α: Adjusted exponent for the Power-law, **L**: Average path lenght, **R(L)**: ratio **L / $\mathbf{L_{random}}$**, **CC:** Clustering coefficient, **R(CC):** ratio **CC / $\mathbf{CC_{random}}$**
* Do not follows power-law distribution

6 Conclusion

From this project it can be concluded that the wiki tool adoption was successful over time. Most active users found on the e-mail lists were using the wiki tool collaboratively by the end of the analysis. It was also observed that collaboration on the wiki tool does not follow a preferential attachment. The fact of the wiki tool usage was a migration from e-mail lists, were subgroups migrate together might suggest this result.

Another consideration for this includes the multidisciplinarity of the group as well as the level of homogeneity from most collaborators. The experiment results show different domains collaborating with each other in the wiki environment. Some actors that shower low collaboration and group integration could be temporary researchers that left the project or temporary consultants.

Finally, comparing the usage of the wiki tool with other collaboration networks and tools, one can conclude that media changes the network evolution. This can be suggested by the fact that different types of groups may obtain fluency to the media that is more adequate to their profile, and this is essential for the adoption success of the system [14].

References

1. Buckley, W.: Sociology and modern systems theory. Prentice Hall, New Jersey (1967)
2. Monteiro, L.H.A.: Sistemas Dinâmicos. Editora da Física, São Paulo (2006)
3. Mitchell, M.: Complex systems: network thinking. Artificial Intelligence 170(18), 1194–1212 (2006)

4. Salazar, J.: Complex systems theory, virtual worlds and MMORPGs: complexities embodied, http://salazarjavier.mindspages.net/complex.pdf
5. Tampere University of Technology. Wille (2009), http://matriisi.ee.tut.fi/hypermedia/en/publications/software/wille/
6. Wasserman, S., Faust, K.: Social network analysis: methods and applications. Cambridge University Press, Cambridge (1994)
7. Barabási, A.L., Albert, R., Jeong, H.: Scale-free characteristics of random networks: The topology of the world wide web. Physica A 281, 69–77 (2000)
8. Snijders, T.A.B.: The Statistical Evaluation of Social Network Dynamics. In: Sobel, M.E., Becker, M.P. (eds.) Sociological Methodology, pp. 361–395. Basil Blackwell, Boston (2001)
9. Colugnati, F.A.B.: Dynamic social network modeling and perspectives in OPAALS frameworks. In: 2nd OPAALS Conference, Tampere, Finland (October 2008)
10. Kurz, T., Heistracher, T.J., Colugnati, F., Razavi, A.R., Moschoyiannis, S.: Deliverable 3.7: Biological and evolutionary approaches for P2P systems and their impact on SME networks. In: OPAALS (August 2009), http://files.opaals.org/OPAALS/Year_3_Deliverables/WP03/D3.7.pdf
11. Colugnati, F., Lopes, L., Kurz, T., English, A.: Deliverable 10.8: Report on cross-domain networks. In: OPAALS (August 2009), http://files.opaals.org/OPAALS/Year_3_Deliverables/WP10/D10.8.pdf
12. Barabási, A.L.: Linked. Plume, Cambridge (2003)
13. Barabási, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Physica A 311, pp. 590–614 (2002)
14. Koh, J., Kim, Y., Butler, B., Bock, G.: Encouraging participation in virtual communities. Communications of the ACM 50(2), 68–73 (2007)