

# Flat for the few, steep for the many: Structural cohesion as a measure of hierarchy in FLOSS communities

Guido Conaldi  
Institute of Management,  
University of Lugano,  
CH-6904 Lugano, CH.  
guido.conaldi@lu.unisi.ch

## ABSTRACT

A discrepancy exists between the emphasis posed by practitioners on decentralized and non-hierarchical communication in Free/Libre Open Source Software (FLOSS) communities and empirical evidence of their hierarchical structure. In order to explain this apparent paradox it is here hypothesized that in FLOSS communities local sub-groups exist and are less hierarchical, more decentralized than the whole social network to which they belong. A measure of structural cohesion based on network node connectivity is proposed as an effective method to test whether FLOSS communication networks can be decomposed in nested hierarchies of progressively less centralized sub-groups. Preliminary results from a case study that are consistent with the hypothesis are presented and discussed.

## Categories and Subject Descriptors

D.2.8 [Software Engineering]: Metrics—*Empirical, Open Source*

## General Terms

Open Source, Social Networks, Structural Cohesion

## 1. INTRODUCTION

Although practitioners and advocates emphasized decentralized and non-hierarchical communication as one essential component of FLOSS success [13, 3], a growing body of empirical evidence supports the claim that FLOSS communities manifest forms of internal hierarchy [4, 1, 9, 10]. How can this discrepancy be explained? One possible explanation is that perception of practitioners is influenced by the position they occupy as developers inside the social networks of FLOSS communities.

Indeed, it is known in the organizational literature [15, 8] that direct social relationships and social network structure surrounding individuals in an organization influence their perceptions. Therefore, the hypothesis here advanced is that in FLOSS communities local sub-groups exist and are less hierarchical, more decentralized than the whole social network to which they belong. The position relative to these sub-groups occupied by developers could contribute to explain the above-mentioned discrepancy.

In the following sections a measure of structural cohesion based on network node connectivity [12] is proposed in order to investigate the existence of decentralized sub-groups within hierarchical FLOSS communities. The measure is

firstly discussed with reference to the existing literature on hierarchy in FLOSS projects. Preliminary results from a case study are then presented and discussed.

## 1.1 Related Work

Several studies presented evidence of hierarchy in FLOSS communities. A first indication of widely varying levels of participation to FLOSS communities is represented by the Pareto distribution followed by the number of messages sent and received in FLOSS communities' mailing lists [18, 1]. These results focus on the individual level of analysis: FLOSS communication networks show small-world characteristics as few individuals account for the bulk of the communication flow.

Other studies [7, 11] demonstrated that well-established and large FLOSS communities manifest hierarchical structures. Also smaller projects revealed a non-flat network structure. Specifically, several studies [4, 14, 10, 17] suggested that different levels of involvement manifested by developers and different statuses achieved by them in the recognition of their peers shape the social networks of FLOSS communities in a hierarchical, stratified structure.

The main social network analysis measures applied by this body of research on hierarchization are Krackhardt's graph theoretical dimensions [4], the core-periphery index [10] and centralization [17]. All these measures have been successfully used to summarize in singular values different dimensions of hierarchy exhibited by FLOSS communication networks. However, these measures cannot capture the existence of local sub-groups with distinguishable structural characteristics. On the contrary, Bird et al. [2] demonstrated the existence of latent sub-communities in FLOSS communication networks. However they identified sub-communities by looking at the difference in density within and between them and this approach does not investigate their potential hierarchical nesting.

## 2. RESEARCH DESIGN

The body of literature discussed in the previous section does not directly test whether FLOSS communication networks can be decomposed in nested hierarchies of progressively less centralized sub-groups. In order to test this hypothesis it is here proposed to use a measure of structural cohesion. More specifically, structural cohesion is here measured following Moody and White [12] in order to analyze simultaneously the tendencies toward hierarchy and centralization in the communication network generated by FLOSS developers in a development mailing list. Nonetheless, the

same procedure could be applied to bug-fixing interactions occurring in FLOSS bug trackers.

The structural cohesion of a network as devised by Moody and White [12] can be defined as the minimum number of actors who, if removed from that network, would disconnect it (i.e. the  $k$ -connectivity of that network), and equivalently as the minimum number of independent paths linking each pair of actors in that network. Indeed, these two minimum numbers can be proven to be equal by virtue of Menger’s Theorem. Once the sets of actors (i.e. the  $k$ -cutsets, one or more than one) holding a given network in one component are identified, they can be removed from it and the structural cohesion of the resulting sub-network be determined. This recursive procedure, defined as cohesive blocking, identifies progressively more (structurally) cohesive sub-sets of actors within the original network, until no further cutting can be done without leaving only isolates.

When applied to an email communication network, cohesive blocking recursively removes the individuals communicating the least with the other members of the community, thus finding sub-groups of the original network progressively more interconnected by the exchange of emails. Progressively smaller sub-groups of individuals hierarchically nested are then found, until either one or more ‘core’ sub-groups are identified that cannot be further reduced without dissolving all the connections existing among the remaining individuals. If multiple equally cohesive sub-groups are found, inter-group communication is assessed by looking at whether hierarchically equivalent groups share some of their members. Therefore a horizontal partial overlapping of sub-groups can be detected together with the hierarchical decomposition of the original network.

The FLOSS project selected as case study is the GNOME web browser Epiphany. The communication network was generated tracing backwards mail threads in the development mailing list. A tie was registered from actor  $i$  to actor  $j$  if  $i$  replied to  $j$ . Similarly to Wiggins et al. [17], a subsequent reply from  $j$  to  $i$  was required for reciprocation to be registered. The number of email exchanged among individuals was recorded, therefore the generated networks were valued di-graphs and the outdegree centrality of the actors was computed in order to complement the analysis of structural cohesion conducted on the underlying graphs. In order to overcome the problem of email aliasing a semi-automatic procedure similar to the algorithm described by Bird et al. [1] was implemented and 587 active individuals were identified over the entire life of the project.

The communication network was measured at 10 points in time. Each time period represents one GNOME six-month release cycle in order to capture potential cyclicalities in the communication pattern. The time period covered by the analysis starts in September 2003, when Epiphany became part of GNOME, and ends in September 2008. The software used to harvest the mailing list is *MailingListStats* developed by the FLOSSMetrics project. Analyses were conducted using the *igraph* package [5] for the R environment.

### 3. PRELIMINARY RESULTS

Epiphany’s communication network possesses a degree distribution typical for FLOSS communities [18, 1]. For example, the 10 most active community members account for respectively 54% and 39% of the entire outgoing and ingoing email exchanges registered over the entire life of the project.

The analysis reveals the existence of a nested hierarchy of overlapping sub-groups generated by the distribution of email exchanges among developers (see Figures 1 and 2). The hierarchy on the right of the networks represents this feature by placing the progressively smaller, more cohesive sub-groups towards the bottom of the graph. The nodes in the communication networks depicted on the left are colored according to the most cohesive sub-group to which they belong. Furthermore, the size of the nodes represents their outdegree centrality, i.e. the number of emails sent as replies during that release cycle.

Due to space limitations, the results for only two time periods are presented<sup>1</sup>. Nonetheless, the two networks shown are representative of the two types of cohesive blocking configuration assumed by Epiphany’s mailing list over time. In the first configuration (see Figure 1) communication exchanges connect developers in one single hierarchy inside which progressively smaller sub-groups of more strongly connected developers can be distinguished. In the second configuration (see Figure 2) the developers are still connected in one single hierarchy. However, at the most strongly connected level of communication two groups form and, thus, the communication tends to cluster in two groups of equally strongly connected developers that overlap only partially. Further analyses are being conducted in order to investi-

<sup>1</sup>High-resolution figures for all time periods and all the generated matrices are available upon request.

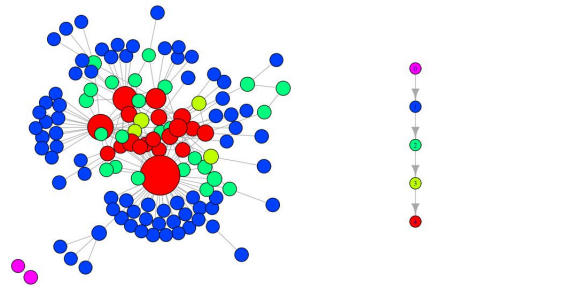


Figure 1: Cohesive Blocking: 09-2003–03-2004

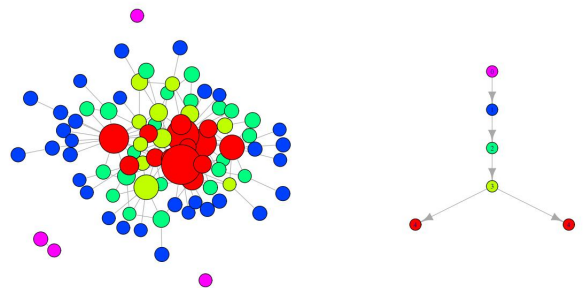


Figure 2: Cohesive Blocking: 09-2006–03-2007

gate whether these different configurations at the deepest levels of the cohesive structure can be related with different development phases of Epiphany [6]. Finally, in both configurations the actors with the highest outdegree centrality are included in the most strongly connected sub-groups.

#### 4. DISCUSSION AND CONCLUSIONS

Empirical evidence gathered from Epiphany remark that sub-groups of intensely communicating developers manifest themselves not simply at the center, but at the structurally cohesive core of the FLOSS communication network. Looking at the entire network, these cohesive sub-groups become ‘central’, in the sense that they concentrate the communication flow by redirecting it towards developers that are inside them. Indeed this is confirmed by the presence of the most active actors in the deepest levels of the cohesive structure of the communication network. Nonetheless, the same maximally cohesive sub-groups are internally highly decentralized. Up to the point that, by definition, developers inside them cannot be removed without dissolving them entirely.

Consequently, in Epiphany an egalitarian local structure appears to exist at the core of the communication network, to which the most active developers are dragged. Nonetheless, Epiphany’s overall structure remains hierarchical as it structures itself in progressive layers of involvement in the communication exchange process.

Provided that similar structural characteristics could be found to characterize other FLOSS communication networks, the hypothesis could be supported that in the FLOSS world the most active developers perceive as flat and egalitarian a network configuration that overall remains highly hierarchical because they implicitly take into consideration only the most cohesive layers of the communication network.

Several opportunities subsist for further research. Firstly the results of this study have to be replicated with respect to different projects and organizational settings. This would also serve to test the external validity of the preliminary findings here presented that rely on a single case study. Secondly, a more extensive research should take into account the interaction between structural properties and individual attributes. A study in this direction is being conducted [6] in order to investigate the relationship between the position occupied by the developers in the cohesive structure of a FLOSS communication network and their level of contribution to the codebase.

Finally, empirical evidence show, for example, a significant effect of structural cohesion of the sub-groups of friends to which students belong on the level of attachment to their school students manifest [12]. Moreover, structurally cohesive sub-groups have been proven to predict effectively the faction individuals will choose in case of conflicts that divide a community [16]. Further research could draw on these results in order to investigate the relationship linking the cohesive structure of FLOSS communities and the likelihood of internal conflict, as well as the final outcomes to which conflict may lead, i.e. for example the death of the project or forking.

#### 5. REFERENCES

- [1] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 137–143, 2006.
- [2] C. Bird, D. Pattison, R. D’Souza, V. Filkov, and Devanbu. P. Chapels in the bazaar? latent social structure in OSS. In *Proceedings of International Symposium in Foundations of Software Engineering*, 2008.
- [3] A. Cox. Cathedrals, bazaars and the town council. *Slashdot*, 13 Oct. 1998.
- [4] K. Crowston and J. Howison. Hierarchy and centralization in free and open source software team communications. *Knowledge Technology & Policy*, 18:65–85, 2006.
- [5] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*:1695, 2006.
- [6] R. De Vita and G. Conaldi. Knowledge networks change during radical and incremental innovation episodes: A case study on software development. Conference Paper Presented at Sunbelt Conference XXIX, San Diego, CA, 2009.
- [7] C. Gacek and B. Arief. The many meanings of open source. *IEEE Software*, 21(1):34–40, 2004.
- [8] H. Ibarra and S. B. Andrews. Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly*, 38(2):277–303, 1993.
- [9] C. Jensen and W. Scacchi. Role migration and advancement processes in OSSD projects: A comparative case study. In *Proceedings of the 29th international conference on Software Engineering*, pages 364–374, 2007.
- [10] Y. Long and K. Siau. Social network structures in open source software development teams. *Journal of Database Management*, 18(2):25–40, 2007.
- [11] A. Mockus, R. Fielding, and J. D. T. Herbsleb. Two case studies of open source software development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, 11(3):309–346, 2002.
- [12] J. Moody and D. R. White. Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 68(1):103–127, 2003.
- [13] E. S. Raymond. The cathedral and the bazaar. *First Monday*, 3(3), 1998.
- [14] G. Robles and J. M. González-Barahona. Contributor turnover in libre software projects. In *Proceedings of the Second International Conference on Open Source Systems*, 2006.
- [15] G. R. Salancik and J. Pfeffer. A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23(2):224–253, 1978.
- [16] D. R. White and F. Harary. The cohesiveness of blocks in social networks: Node connectivity and conditional density. *Sociological Methodology*, 31:305–359, 2001.
- [17] A. Wiggins, J. Howison, and K. Crowston. Social dynamics of FLOSS team communication across channels. In *Proceedings of the Fourth International Conference on Open Source Software*, 2008.
- [18] J. Xu, Y. Gao, S. Christley, and G. Madey. A topological analysis of the open source software development community. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005.