Understanding Collaboration in the Open-Source Arena

The cases of WebKit and OpenStack

Jose Teixeira
Doctoral Candidate
University of Turku, Finland
Turku Centre for Computer Science, Finland
jose.Teixeira@utu.fi

ABSTRACT

In an era of software crisis, the move of firms towards distributed software development teams is being challenged by emerging collaboration issues. On this matter, the open-source phenomenon may shed some light, as successful cases on distributed collaboration in the open-source community have been recurrently reported. In our research we explore collaboration networks in the WebKit and OpenStack open-source projects, by mining their source-code version-control-systems data with Social Network Analysis (SNA). Our approach allows us to observe how key events in the industry affect open-source collaboration networks over time. With our findings, we highlight the explanation power from network visualizations capturing the collaborative dynamics of high-networked software projects over time. Moreover, we depicted that competing companies that market similar products in the same market, can collaborate in the open-source community while publicly manifesting intense rivalry (e.g. Apple vs Samsung patent-wars). After integrating our findings with the current body of theoretical knowledge in management strategy, economics, strategic alliances and coopetition, we propose the novel notion of open-coopetition, where rival firms collaborate with competitors in the open-source community. We argue that classical coopetition management theories do not fully explain the competitive and collaborative issues that are simultaneously present and interconnected in the WebKit and OpenStack open-source communities. We propose the development of the novel open-coopetition theory for a better understanding on how rival-firms collaborate with competitors in an open-source fashion.

Categories and Subject Descriptors

D.9 [**Management**]: Programming teams; K.6.3 [**Software Management**]: Software development,Software process; H.0 [**Information Systems**]: GENERAL

EASE '14 May 13 - 14 2014, London, England, BC, United Kingdom

ACM 978-1-4503-2476-2/14/05 ...$15.00.
http://dx.doi.org/10.1145/2601248.2613086

General Terms

Management,Theory,Measurement

Keywords

Open-Source, OSS, FLOSS, Strategic Alliances, Ecosystems, Collaboration, Competition, Coopetition, Open-Coopetition.

1. INTRODUCTION

In an era of software crisis[1] , the move of firms towards geographically-distributed, and often off-shored, software development teams is being challenged by collaboration issues. On this matter, the open-source phenomenon may shed some light, as successful cases on distributed collaboration in the open-source community have been recurrently reported [3, 12]. While practitioners move with difficulties towards globally distributed software development, there is a lack of research in academia addressing the collaboration dynamics of large-scale distributed software projects [11, 14].

Our research aims to contribute to a better understanding of collaboration in large-scale distributed projects, by mining collaboration networks of open-source projects with social network analysis. While addressing a previous call, from [1] for the advancement of methods and techniques to support the visualization of temporal aspects (e.g. pace, sequence) to represent change and evolution in ecosystems[2], we employed Social Network Analysis over publicly-available and naturally-occurring open-source data that allows us to re-construct and visualize the evolution open-source projects in a sequence of networks.

Our first research unit-of-analysis was the WebKit open-source project and its community. WebKit is an open-source project providing an engine that renders and interprets content from the World Wide Web. Its technology permeates our digital life since it can be found in the most recent computers, tablets and mobile devices sold by Apple, Google, Samsung, Nokia, RIM, HTC, and others. With more than 10 years of history, the WebKit project has brought together volunteers and firm-sponsored software developers that collaborate over the Internet by open and transparent approaches while giving up the traditional intellectual property rights.

[1]A brief discussion on the software-crisis is provided by Fitzgerald, B. "Software Crisis 2.0." Computer 45.4 (2012): 89-91.

[2]Basole, R. employs the ecosystems term as a complex network of companies interacting with each other, directly and indirectly, to provide a broad array of products and services.

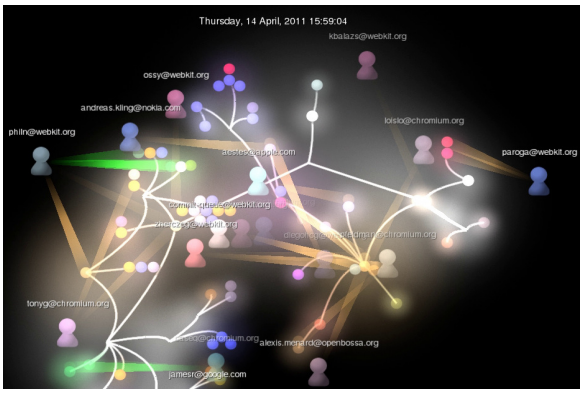

Figure 1: Illustration of the employed Social Network Analysis research approach.

The previous Figure 1, illustrates the power of our research approach [3] that data-mines the WebKit source-code data with Social Network Analysis. From open-source and natural-occurring data, we are able to capture features of collaboration and competition on the WebKit project overtime. On the same Figure 1, we can depict a sample of developers working with each other in specific WebKit source-code artifacts on 14 of April 2013 at 15:59:04.

2. RESEARCH AIM AND QUESTIONS

This proposal aims to contribute with practical and theoretical knowledge regarding collaboration in the open-source arena. More specifically, we seek to explore the notions collaboration, competition and rivalry in large-scale distributed open-source projects.

This research proposal encompasses the following research questions:

RQ1: How rival vendors collaborate in the open-source arena?

RQ2: Can rival corporations collaborate with the same development transparency and sense of community of the open-source software communities?

RQ3: Can rival corporations co-compete in the open-source arena (i.e. collaborate while competing with similar products on the same market)?

RQ4: How can firms develop complex R&D activities by a more open-source fashion? With less space for gate-keepers, lawyers and complex intellectual property arrangements?

RQ5: How key exogenous events in the industry, such as strategic-alliances changes, affect the social-structure of related open-source communities?

3. RESEARCH DESIGN

Previous socio-technological analysis addressing collaboration within large scale open-source software projects tend to adopt either of the two equally unsatisfactory alternatives: (1) providing thick qualitative descriptions of selected cases, thus overlooking the actors, actions and interdependent patterns of the collaborative network [7, 9, 13, 15]; or (2) reducing figurational complexity to a set of quantitative indicators, thus disfiguring all practical purposes of the phenomena under investigation [8, 5, 19].

We worked around this quantitative-qualitative divide by employing an alternative mix methods approach already employed to study collaboration and innovation in antibody reagents in biomedecine [4]. First, we started by gaining a better understanding of the WebKit competitive dynamics by more qualitative and netnographic ways [6]. And then, we used mathematical but non-statistical quantitative features of Social Network Analysis [18] to re-construct and visualize the evolution of the WebKit collaboration in a sequence of networks.

3.1 Unit of analysis

In order to better understand collaboration in the open-source arena, we pointed our lenses to the WebKit open-source project and the firms from the PC and mobile-devices industries joint-developing it. Moreover, we will expand our research to the cloud-computing industry by applying the same approach to understand how companies like HP, Rackspace and Ubuntu collaborate in joint-developing the open-source OpenStack cloud-computing infrastructure software.

3.2 Data collection

Our retrieved data was freely available on the Internet thanks to the public-domain nature of open-source software projects. We made use of the WebKit source-code, its version-control-system and other related web-sites covering developers contributions to the project from 1st September 2006 till 3rd April 2013. All raw-data is natural-occurring and not provoked by the researchers. The initial raw-data, and final-data supporting our research results was archived in our project website at http://users.utu.fi/joante/WebKitSNA/. Data-cleansing efforts were minimal thanks to WebKit very-strict peer-review and code-commit policies.

The data from the OpenStack open-source project was also public available at https://github.com/openstack-dev/. However it's loading, cleansing and analysis depends on funding decisions regarding this research proposal. Initial data screenings suggest that the analysis of the OpenStack project will require more time and labor than the invested in our prior research taking the WebKit-project as unit-of-analysis [16].

3.3 Data analysis

After attaining an initial understanding of the competitive dynamics of the mobile-devices industry with netnography [6] , we extract and analyse the social network of an open-source project levering Social Network Analysis, which is an emergent method widely established across disciplines of social sciences in general [17, 18] and information systems in particular [10]. We focus on the visualization of the collaboration network and sub-community detection using the following established Social Network Analysis methods:

- Visualisation with degree centrality.
- Markov chain clustering.
- Modularity maximization heuristic.
- Hub based community detection with different parameter configurations.

[3] A video illustrating the same research approach, is available at http://users.utu.fi/joante/WebKitSNA/.

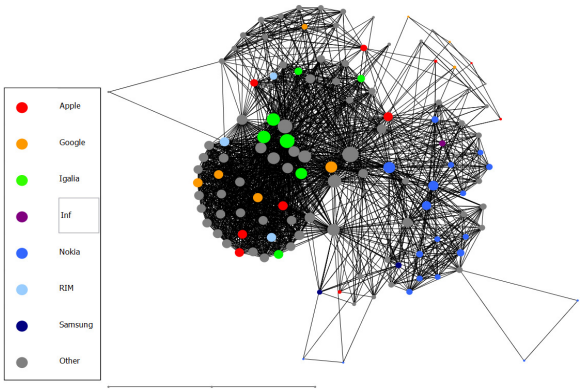

Figure 2: Mapping collaboration in the WebKit project from June 2009 to February 2011: Reflecting the Nokia and Microsoft plans to form a broad strategic partnership that forced Intel to search for new partners for Meego.

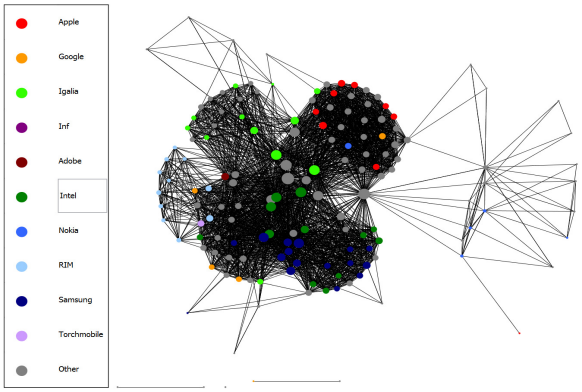

Figure 3: Mapping collaboration in the WebKit project from July 2012 to April 2013: Reflecting patent-wars, trademarking and forking.

More methodological details, data, source-code and visualizations related with this proposal are available on the project website at http://users.utu.fi/joante/WebKitSNA/.

4. PRELIMINARY FINDINGS

Our socio-structural visualizations of collaboration in the WebKit project, such as in Figure 2, Figure 3 and many others visualizations publicly-available on our project website at http://users.utu.fi/joante/WebKitSNA/ lead to a set of interesting findings such as:

- Nokia contributed with a lot of code to the WebKit project, but in a social periphery. Mostly Nokians working with Nokians (forming a sub-community).
- Nokia and Intel breakage of cooperation easily can be easily visualised over time. Nokia marriage with Microsoft caused immediate damage to collaboration in the Webkit project.
- But, even if Samsung and Apple are involved in expensive patent-wars in the courts and stopped collaborating on hardware components, their contributions remained strong and central within the WebKit open-source project.
- Non-affiliated developers, who are often volunteers without firm-sponsorship, together with developers affiliated with smaller firms, were more central within the WebKit collaboration network than developers affiliated with the TOP10 organizations from a previous impirical study from Bitergia [2].
- By forking the project, Google is "recruiting" WebKit developers previously affiliated with Apple and Nokia to its Blink open-source project.

5. SOCIETAL IMPACT

We believe that our research have important contributions for the practice of R&D Management and Software Engineering. Moreover, a better understanding of collaboration in the open-source arena informs the regulatory practice on how rival-firms collaborate and compete in the open-source arena.

5.1 Managerial practice

- Our Social Network Analysis visualizations can help different stakeholders in assessing their inter-firm network positions for better decision-making regarding inter-networked strategic alliances.
- User, adopters and integrators, can depict the project social-structure evolution and dynamics, for thorough assessments of its sustainability when reacting to exogenous events in the industry.
- Investors are provided with a complementary analytical tool for clarifying network dynamics, improving the forecast of product attractiveness and future growth.
- Our proposed theory on open-coopetition [16], derived from previous research efforts, provides guidance on the management of high-networked R&D activities by a more open-source fashion. Minimizing the need for gate-keepers, lawyers and complex intellectual property arrangements, while maximizing development transparency and sense of community as illustrated within Figure 4.

5.2 Software engineering

- Our Social Network Analysis visualizations can also benefit software-developers by providing a better understanding of the socio-structural organization of complex software projects, uncovering possible deficiencies on the development processes.
- Our research also contributes to the area of global-software analytics. The researchers were contacted by commercial companies and open-source developers regarding the development of new Social Network Analysis features for existing analytical tools. The following Figure 5 illustrates how code-driven metrics can be combined with socio-structural visualizations for better decision making in global-software development.

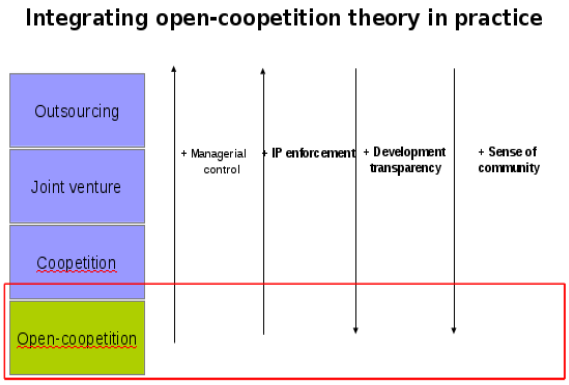

Figure 4: The open-coopetition theory on practice.

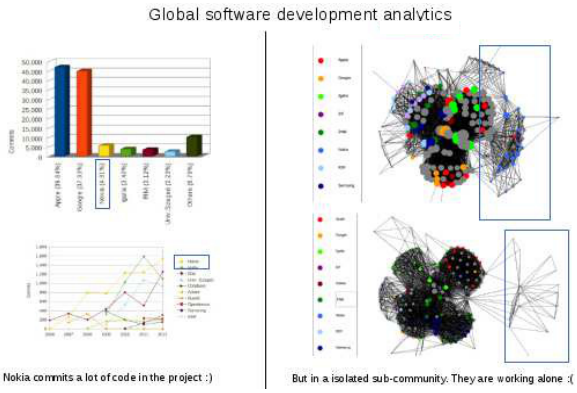

Figure 5: Combining code-driven metrics with Social Network Analysis visualizations.

5.3 Regulatory practice

By better understanding both collaboration and competition in the open-source arena we are better prepared for:

- Fomenting an economic environment with lower entrance-costs in the high-tech industry.
- Enhance industrial competition vis-à-vis with industrial sharing of development costs.
- Minimize problems with the current intellectual property regime currently undermining fair-market competition in the high-tech industry.
- Foment open-source software and consequentially digital-privacy friendly ICT infrastructures.

6. ACKNOWLEDGEMENTS

The author would like to thank Liikesivistysrahasto (grant 32061) and Fundação para a Ciência e a Tecnologia (grant SFRHBD615612009) for funding this doctoral research project. Project was executed under the guidance of Reima Suomi and Hannu Salmela at the University of Turku, Finland. My colleague Lin Tingting introduced me to the world of Social Network Analysis and performed some of the first network visualizations.

More methodological details, data, high-resolution visualizations and source-code are publicly available on the project website at http://users.utu.fi/joante/WebKitSNA/.

7. REFERENCES

[1] R. C. Basole. Visualization of interfirm relations in a converging mobile ecosystem. *Journal of Information Technology*, 24(2):144–159, 2009.

[2] Bitergia. WebKit Report by Bitergia, Jan. 2013.

[3] A. Bonaccorsi and C. Rossi. Why open source software can succeed. *Research policy*, 32(7):1243–1258, 2003.

[4] A. Cambrosio, P. Keating, and A. Mogoutov. Mapping collaborative work and innovation in biomedicine: A computer-assisted analysis of antibody reagent workshops. *Social Studies of Science*, pages 325–364, 2004.

[5] K. Crowston and J. Howison. The social structure of free and open source software development. *First Monday*, 10(2-7), 2005.

[6] R. V. Kozinets. The field behind the screen: using netnography for marketing research in online communities. *Journal of marketing research*, pages 61–72, 2002.

[7] J. Lerner and J. Tirole. Some simple economics of open source. *The journal of industrial economics*, 50(2):197–234, 2002.

[8] L. Lopez-Fernandez, G. Robles, and J. M. Gonzalez-Barahona. Applying social network analysis to the information in CVS repositories. 2004.

[9] S. Q. Mian, J. Teixeira, and E. Koskivaara. Open-Source Software Implications in the Competitive Mobile Platforms Market. In *Building the e-World Ecosystem*, pages 110–128. Springer, 2011.

[10] H. Oinas-Kukkonen, K. Lyytinen, and Y. Yoo. Social networks and information systems: ongoing and future research streams. *Journal of the Association for Information Systems*, 11(2):3, 2010.

[11] M. Paasivaara and C. Lassenius. Collaboration practices in global inter-organizational software development projects. *Software Process: Improvement and Practice*, 8(4):183–199, 2003.

[12] E. Raymond. The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3):23–49, 1999.

[13] J. Salminen and J. Teixeira. Fool's Gold? Developer Dilemmas in a Closed Mobile Application Market Platform. In *Co-created Effective, Agile, and Trusted eServices*, Lecture Notes in Business Information Processing, pages 121–132, Turku, 2013. Springer.

[14] B. Sengupta, S. Chandra, and V. Sinha. A research agenda for distributed software development. In *Proceedings of the 28th international conference on Software engineering*, pages 731–740. ACM, 2006.

[15] J. Teixeira. Open-Source Technologies Realizing Social Networks: A Multiple Descriptive Case-Study. In *Open Source Systems: Long-Term Sustainability*, volume 378 of *IFIP Advances in Information and Communication Technology*, pages 250–255. Springer Berlin Heidelberg, Jan. 2012.

[16] J. Teixeira and T. Lin. Collaboration in the open-source arena: The WebKit case. *CoRR*, abs/1401.5996, 2014.

[17] B. Uzzi. The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American sociological review*, pages 674–698, 1996.

[18] S. Wasserman and K. Faust. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[19] J. Xu, S. Christley, and G. Madey. *Application of social network analysis to the study of open source software*. Elsevier Press, 2006.