Understanding Coopetition in the Open-Source Arena: The Cases of WebKit and OpenStack

Research Summary

Jose Teixeira
University of Turku, 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 high-networked 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 explanatory power from network visualizations capturing the collaborative dynamics of high-networked software projects over time. Moreover, we argue that competing companies that sell 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 by open-source manners.

Categories and Subject Descriptors

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

OpenSym '14 , Aug 27-29 2014, Berlin, Germany
ACM 978-1-4503-3016-9/14/08.
http://dx.doi.org/10.1145/2641580.2641627

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 [5, 21]. While practitioners move with difficulty towards globally distributed software development, there is a lack of research in academia addressing the collaborative dynamics of large-scale distributed software projects [20, 23].

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 [3] 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, allowing us to reconstruct and visualize the evolution of 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 means 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.

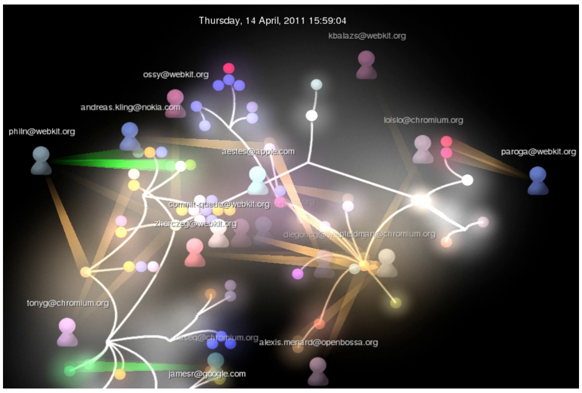

Figure 1: Illustration of the applied Social Network Analysis research approach.

Figure 1 illustrates the power of our research approach [3] which 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 over time. 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 by providing practical and theoretical knowledge regarding collaboration in the open-source arena. More specifically, we seek to explore the notions of collaboration, competition and rivalry in large-scale distributed open-source projects.

This research proposal encompasses the following research questions:

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

RQ2: Can rival corporations collaborate with the same development transparency and sense of community as 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 in an open-source fashion(i.e. with less space for gate-keepers, lawyers and complex intellectual property arrangements)?

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

3. RESEARCH DESIGN

3.1 Unit of analysis

In order to better understand collaboration in the open-source arena, we directed our lenses on 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 for understanding how companies like HP, IBM, Rackspace and Canonical collaborate in joint-developing the open-source OpenStack cloud-computing infrastructure software.

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

3.2 Research method

3.2.1 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 to 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's extremely strict peer-review and code-commit policies.

The data from the OpenStack open-source project was also publicly available at https://github.com/openstack-dev/. However its loading, cleansing and analysis depends on funding decisions that constrain our research project. Initial data investigations [24] suggest that the analysis of the OpenStack project will require more time and labor than that invested in our prior research taking the WebKit-project as unit-of-analysis [16, 27, 25, 28, 26].

3.2.2 Data analysis

After attaining an initial understanding of the competitive dynamics of the mobile-devices industry by ethnographic approaches, we extract and analyse the social network of an open-source project leveraging Social Network Analysis, which is an emergent method widely established across disciplines of social sciences in general [29, 30] and information systems in particular [19]. We focus on the visualization of the collaboration network and sub-community detection, using the following established Social Network Analysis methods:

- Longitudinal visualizations using different geometries and layouts.
- Calculations of nodes and groups centrality measures (classical and eigenvector)
- Sub-community detection with Markov chain clustering, modularity maximization heuristics and hub-based community detection with different parameter configurations.
- Extraction of Simmelian backbones .

3.2.3 Mixed methods and multidisciplinary approach

We engaged in a multidisciplinary approach while employing a multitude of research methods. A lot of Internet-Data was scrutinized using Netnography, an established method for the study of on-line communities in Marketing. The mining of software version-control-systems is also an established method in Software Engineering: the *Working Conference on Mining Software Repositories (MSR)*, currently on its 11th edition, is the main conference in the area. Our most core research approach, Social Network Analysis, is a re-emergent method that is getting widely established across disciplines such Physics, Mathematics, Computer Science,

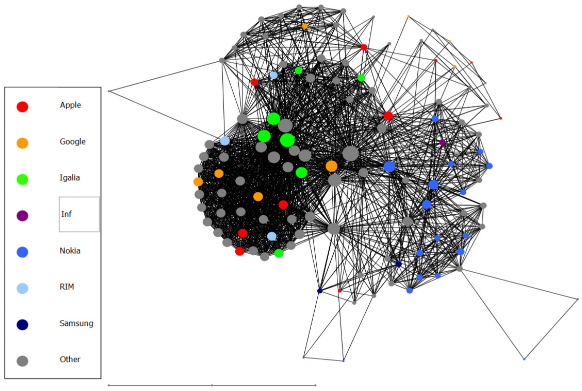

Figure 2: Mapping collaboration in the WebKit project from June 2009 to February 2011: Reflecting the Nokia and Microsoft plans to form a broad strategic partnership that forced Intel to search for new partners for Meego.

Table 1: Multidisciplinary approach

Employed approach	Discipline(s)	Seminal works
Netnography	Marketing	[14] [13]
Mining of software repositories	Software-Engineering	[22] [11]
Network analysis of digital trace data	Information-Systems	[10] [9]
Network analysis with emphasis on the visualization of collaborative activities	Biomedicine Bibliomentrics Innovation-Studies	[6] [15] [8]
Network analysis of massive networked data. Use of clustering and sub-community detection algorithms.	Physics Mathematics Computer-Science Anthropology Neurology Bioinformatics	[7] [31] [12] [18] [1]

Biomedicine, Anthropology, Innovation Studies and many other disciplines [30]. Table 1 captures our cross-disciplinary orientation that celebrates the use and combination of different research methods.

3.2.4 *Validity issues*

Regarding research validity, it is important to mention that we only deal with naturally-occurring data. All data was neither created nor provoked for research purposes. All collected data naturally occurred for the practical purposes of developing WebKit and OpenStack. This minimizes the Hawthorne effect, also referred to as the observer effect, in witch human subjects modify aspects of their behavior in response to the awareness that research is being conducted. In other words: by collecting the data in an unobtrusive way, we did not affect the behavior of WebKit or OpenStack developers during the data-collection process, something that could occur by employing the traditional survey or interview mechanisms.

Moreover, for enhancing rigor and validity, we and took in consideration the following notes:

- Regarding the qualitative research part: We took in consideration key guidelines on now to conduct Netnography [13, 14], such as Kozinets's four criteria for selecting from online data-sources [13]. All research efforts were conducted according a set of established and consistent guidelines on how to conduct ethnographic research [2, 17].

- Regarding the mining software repositories with Social Network Analysis. Besides the following of "standard" methodological guides on how to conduct Social Network Analysis [30], we also took in consideration a set of issues on the use of Social Network Analysis with digital trace data as outlined by Howison et al. [10].

- The triangulation of research results, both from the ethnographic and network analysis research efforts, led to the early test and falsification of preliminary research results, thus increasing the robustness of our findings. Moreover, the access to complementary data from Bitergia [4], a company specialized in software development analytics, allowed us to benchmark our approach on data-mining WebKit's source-code version-control-system.

- Thanks to WebKit's strict policy for committers and reviewers, the collected data set was remarkably clean, facilitating a smooth data extraction ahead of analysis. However, in the OpenStack case, some preliminary data screenings [24] suggest that the data-cleansing and filtering efforts will require greater amounts of time, effort and attention than in the WebKit case [27, 25, 28, 26].

Further methodological details, data, source-code and visualizations related with this proposal are publicly 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 substantial amounts of code to the WebKit project, but in a social periphery, i.e. mostly Nokians working with Nokians (forming a sub-community).

- The Nokia and Intel breakage of cooperation can be easily visualised over time. Nokia's marriage with Microsoft caused immediate damage to collaboration in the Webkit project.

- However, even if Samsung and Apple were 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 outlined in a recent empirical study from Bitergia [4].

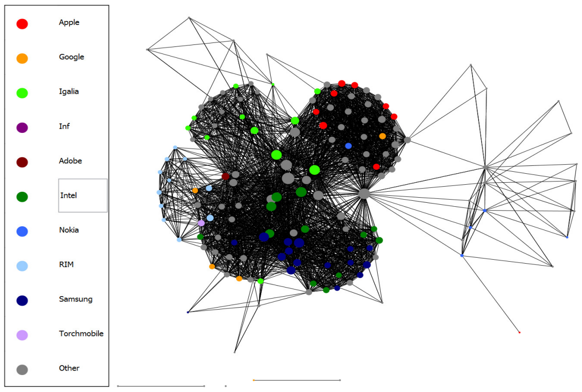

Figure 3: Mapping collaboration in the WebKit project from July 2012 to April 2013: Reflecting patent-wars, trademarking and forking.

- 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 has important implications 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 [25, 24].
- Users, adopters and integrators can better grasp project’s social structure evolution and dynamics. They can then make 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 [28].
- Our proposed theory on open-coopetition [27], derived from our recent research efforts, provides guidance on the management of high-networked R&D activities in a more open-source fashion: i.e minimizing the need for gate-keepers, lawyers and complex intellectual property arrangements, while maximizing development transparency and sense of community as illustrated in 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 in the development processes. Those visualizations are more valuable when combined with the conventional code-driven metrics, as illustrated in Figure 5.

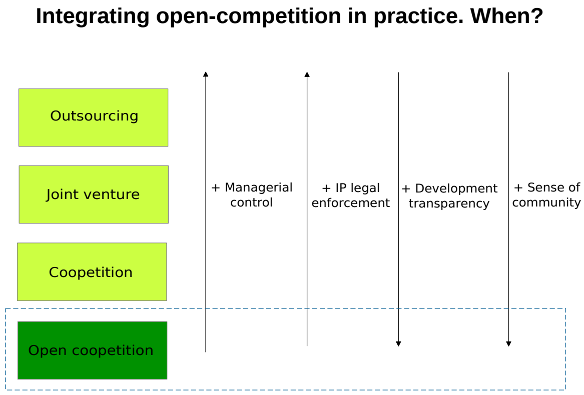

Figure 4: The open-coopetition theory on practice.

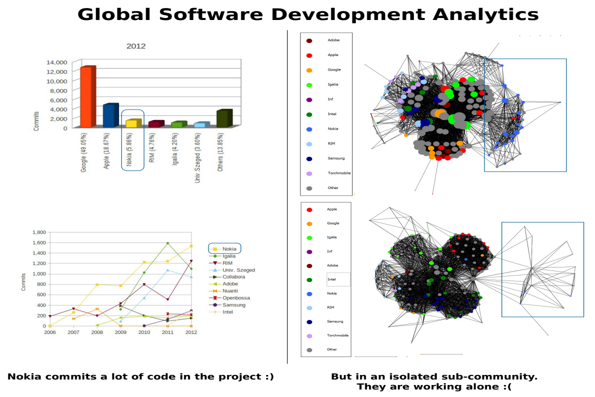

Figure 5: Combining code-driven metrics with social network visualizations.

- Our research also contributes to the area of global-software analytics [26]. 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.

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.
- Enhancing industrial competition vis-à-vis industrial sharing of development costs.
- Minimizing problems with the current intellectual property regime, which currently undermines fair-market competition in the high-tech industry.

- Promote the growth and development of open-source software and consequentially digital-privacy friendly ICT infrastructures.

6. ACKNOWLEDGEMENTS

A word of thanks to the WebKit and OpenStack open-source communities for developing cool, open and research-friendly technologies. Further methodological details, data, high-resolution visualizations and source-code are available on the project websites at http://www.jteixeira.eu/WebKitSNA and http://www.jteixeira.eu/OpenStackSNA.

7. REFERENCES

[1] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi, and T. Vicsek. Cfinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8):1021–1023, 2006.

[2] P. Atkinson and M. Hammersley. *Ethnography: Principles in Practice*. Routledge, New York, 2nd edition, 1995.

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

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

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

[6] 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.

[7] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[8] W. Glänzel and A. Schubert. Analysing scientific networks through co-authorship. In *Handbook of quantitative science and technology research*, pages 257–276. Springer, 2005.

[9] J. Hahn, J. Y. Moon, and C. Zhang. Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Information Systems Research*, 19(3):369–391, 2008.

[10] J. Howison, A. Wiggins, and K. Crowston. Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 2011.

[11] H. Kagdi, M. L. Collard, and J. I. Maletic. A survey and taxonomy of approaches for mining software repositories in the context of software evolution. *Journal of Software Maintenance and Evolution: Research and Practice*, 19(2):77–131, 2007.

[12] J. Kleinberg. An impossibility theorem for clustering. In *NIPS*, volume 15, pages 463–470, 2002.

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

[14] R. V. Kozinets. *Netnography: Doing ethnographic research online*. Sage Publications Limited, 2009.

[15] B.-A. Lundvall. User-producer relationships, national systems of innovation and internationalisation. *National systems of innovation: Towards a theory of innovation and interactive learning*, pages 45–67, 1992.

[16] 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.

[17] M. Myers. Investigating information systems with ethnographic research. *Communications of the AIS*, 2(4es):1, 1999.

[18] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[19] 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.

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

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

[22] G. Robles, J. M. Gonzalez-Barahona, M. Michlmayr, and J. J. Amor. Mining large software compilations over time: another perspective of software evolution. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 3–9. ACM, 2006.

[23] 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.

[24] J. Teixeira. Open-coopetition in the Cloud computing Industry: the OpenStack NOVA case. In *Proceedings of the first European Social Networks Conference*, European Social Networks Conference, 2014.

[25] J. Teixeira. Open-coopetition in the PC and mobile industries: the WebKit case. In *Proceedings of the 6th Workshop on coopetition strategy*, 2014.

[26] J. Teixeira. Understanding collaboration in the open-source arena. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 2014.

[27] J. Teixeira and T. Lin. Collaboration in the open-source arena: The WebKit case. In *Proceedings of the 52th ACM SIGMIS conference on Computers and people research*, Singapore, 2014.

[28] J. Teixeira and T. Lin. Rivalry and collaboration in the open-source arena: The WebKit case. In *Sunbelt XXXIV*, 2014.

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

[30] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*, volume 8. Cambridge University Press, 1994.

[31] W. Zachary. An information flow modelfor conflict and fission in small groups1. *Journal of anthropological research*, 33(4):452–473, 1977.