ABSTRACT
How can we understand FOSS collaboration better? Can social issues that emerge be identified and addressed before it is too late? Can the community heal itself, become more transparent and inclusive, and promote diversity? We propose a technique to address these issues by quantitative analysis of social dynamics in FOSS communities. We propose using social network analysis metrics to identify growth patterns and unhealthy dynamics; giving the community a heads-up when they can still take action to ensure the sustainability of the project.

Categories and Subject Descriptors
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

General Terms
Measurement, Reliability, Human Factors.

Keywords

1. INTRODUCTION
Social networks are a ubiquitous part of our social lives, and the creation of online social communities has been a natural extension of this phenomena. Free/Open Source Software (FOSS) development efforts are prime examples of how community can be leveraged in software development, as efforts are formed around volunteer-based communities of interest, and depend on continued interest and involvement in order to stay alive [2].

Though the bulk of collaboration and communication in FOSS communities occurs online and is publicly accessible, there are many open questions about the social dynamics in FOSS communities. Projects might go through a metamorphosis when faced with an influx of new developers or the involvement of an outside organization. Conflicts between developers raised as the result of divergent opinions about the future of the project might lead to an ensuing fork of project and the dilution of the community. Forking, either as a violent split when there is a conflict or as a friendly divide when new features are experimentally added both affect the community [1].

Most recent studies of FOSS communities have tended to suffer from two important limitations. First, they treat community as a static structure rather than a dynamic process. Second, many social dynamics in FOSS have been studied using a case-study methodology, focusing on a selected subset of the available data.

In this paper, we propose to use social network analysis to study the evolution and social dynamics of FOSS communities. With these techniques we hope to identify measures associated with unhealthy group dynamics, e.g. a simmering conflict, as well as early indicators of major events in the lifespan of an online community. One dynamic we are especially interested in are those of forked FOSS projects. We will seek to validate this technique by comparing the results of our analysis to the results of a study of forked FOSS projects by Robles et al. [5]. The goal is to demonstrate that this quantitative approach can be applied to commonly available FOSS archives to get a better understanding of the evolution of these communities.

This paper is organized as follows: We present related literature on online social communities, recounting their focus and the findings. We then present the gap in the literature, and what further study needs to be done. Next, we discuss why the issue needs to be addressed and who benefits from it, in the motivation section. Following that, we present three research questions framed as hypotheses that we are going to test. After that, we propose a methodology as to how to test the validity of the hypotheses, which includes gathering data, doing the analysis, and the visualization of the findings. At the end, we present future work and challenges.

2. RESEARCH QUESTIONS
We argue that the social interactions data reflects the changes the community goes through, and will be able to describe the context surrounding a forking event. Robles et al. [5] classify the main reasons for forking into six classes, listed in Table 2. Three of the six listed reasons are socially related, and so should arguably be reflected somehow in the social interaction data. As an example, if a fork occurs because of a desire for “more community-driven development”, we expect to see an interaction pattern in the collaboration data showing, a strongly-connected core community getting more isolated from the rest of the community, and the formation of another very active and well-connected core.

More specifically, our research hypotheses are:

Hypothesis #1: Social interactions data can be used to predict an imminent fork.

Hypothesis #2: “Personal differences” and “technical differences” are associated with distinctly different interaction patterns. It is possible to tell the difference between when the reason for forking was personal vs. technical.

Hypothesis #3: It is possible to tell through collaboration patterns between when the reason for forking is “more community-driven development” and “differences among the developer team”.

In this paper, we propose to use social network analysis to study the evolution and social dynamics of FOSS communities. With these techniques we hope to identify measures associated with unhealthy group dynamics, e.g. a simmering conflict, as well as early indicators of major events in the lifespan of an online community. One dynamic we are especially interested in are those of forked FOSS projects. We will seek to validate this technique by comparing the results of our analysis to the results of a study of forked FOSS projects by Robles et al. [5]. The goal is to demonstrate that this quantitative approach can be applied to commonly available FOSS archives to get a better understanding of the evolution of these communities.

This paper is organized as follows: We present related literature on online social communities, recounting their focus and the findings. We then present the gap in the literature, and what further study needs to be done. Next, we discuss why the issue needs to be addressed and who benefits from it, in the motivation section. Following that, we present three research questions framed as hypotheses that we are going to test. After that, we propose a methodology as to how to test the validity of the hypotheses, which includes gathering data, doing the analysis, and the visualization of the findings. At the end, we present future work and challenges.

2. RESEARCH QUESTIONS

We argue that the social interactions data reflects the changes the community goes through, and will be able to describe the context surrounding a forking event. Robles et al. [5] classify the main reasons for forking into six classes, listed in Table 2. Three of the six listed reasons are socially related, and so should arguably be reflected somehow in the social interaction data. As an example, if a fork occurs because of a desire for “more community-driven development”, we expect to see an interaction pattern in the collaboration data showing, a strongly-connected core community getting more isolated from the rest of the community, and the formation of another very active and well-connected core.

More specifically, our research hypotheses are:

Hypothesis #1: Social interactions data can be used to predict an imminent fork.

Hypothesis #2: “Personal differences” and “technical differences” are associated with distinctly different interaction patterns. It is possible to tell the difference between when the reason for forking was personal vs. technical.

Hypothesis #3: It is possible to tell through collaboration patterns between when the reason for forking is “more community-driven development” and “differences among the developer team”.

In this paper, we propose to use social network analysis to study the evolution and social dynamics of FOSS communities. With these techniques we hope to identify measures associated with unhealthy group dynamics, e.g. a simmering conflict, as well as early indicators of major events in the lifespan of an online community. One dynamic we are especially interested in are those of forked FOSS projects. We will seek to validate this technique by comparing the results of our analysis to the results of a study of forked FOSS projects by Robles et al. [5]. The goal is to demonstrate that this quantitative approach can be applied to commonly available FOSS archives to get a better understanding of the evolution of these communities.
3. RESEARCH PROGRESS
Some initial work is done, namely scraping mailing list data off several FOSS projects, as well as their bug repositories. We have also done a literature review of the existing work on social network analysis on FOSS communities, and have experienced working with large graphs. There are a myriad of possible applications for the possible outcomes of this line of research, and the gap in the existing body of research shows a need for further study of FOSS social network analysis.

4. CONTRIBUTIONS
To better understand and measure the evolution and social dynamics of FOSS projects, integral components to understanding their evolution and direction, we need new and better tools. With this knowledge and these tools, we could help projects reflect on their actions, and help community leaders make informed decisions about possible changes or interventions. We want to map the dynamics of communities to real world phenomena. Identification is the first step to rectify an undesired dynamic before the damage is done. A community that does not manage growing pains may end up stagnating or dissolving.

Managing growing pains is especially important in case of FOSS, where near half the project contributors are volunteers [6]. Oh et al. [7] have argued that openness in FOSS is “[…] generally perceived as having a positive connotation, however, the term can also be interpreted as referring to some unconstructive characteristics, such as unobstructed exit, susceptible, vulnerable, fragile, lacking effective regulation, and so on. The unobstructed exit and lack of regulatory force inherent in the OSS community can result in a community’s susceptibility and vulnerability to herded exits by its participants. Commercial vendor intervention, an alternative project becoming available, and licensing issues can result in some original core members ceasing to provide their loyal service for the community, which can prompt their coworkers to leave as well” [7].

Recipes for success or stagnation, sustainability or fragmentation could be identifiable, leading to a set of best practices and pitfalls.

---

Table 1. The main reasons for forking as classified by Robles et al. [5] as well as the proposed associated metrics & collaboration data sources.

<table>
<thead>
<tr>
<th>Reason for Forking</th>
<th>Example Forks</th>
<th>Frequency</th>
<th>Associated Collaboration data</th>
<th>Associated Metrics</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical (Addition of functionality)</td>
<td>Xpdf &amp; Poppler</td>
<td>60 (27.3%)</td>
<td>Commits logs, Bug repositories, Mailing list</td>
<td>Effective graph diameter, Resilience of the entire network, Clustering coefficient of nodes</td>
<td>6 months before and 3 months after the forking date</td>
</tr>
<tr>
<td>Discontinuation of the original project</td>
<td>Apache web server</td>
<td>44 (20.0%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>More community-driven development</td>
<td>EGCS &amp; GCC</td>
<td>29 (13.2%)</td>
<td>Commits logs, Bug repositories</td>
<td>Clustering coefficient of nodes, Closeness centrality, Degree distribution of the entire network, Effective graph diameter, Eigenvector centrality, Resilience of the entire network</td>
<td>6 months before and 3 months after the forking date</td>
</tr>
<tr>
<td>Legal issues</td>
<td>X.Org &amp; XFree</td>
<td>24 (10.9%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Commercial strategy forks</td>
<td>LibreOffice &amp; OpenOffice.org</td>
<td>20 (9.1%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Differences among developer team</td>
<td>OpenBSD &amp; NetBSD</td>
<td>16 (7.3%)</td>
<td>Commits logs, Bug repositories, Mailing list (1. who emailed who, 2. what they are saying (content))</td>
<td>Effective graph diameter, Edge betweenness centrality, Eigenvector centrality, Node betweenness centrality, Clustering coefficient of nodes, Edge betweenness centrality, Degree distribution of the entire network, Resilience of the entire network</td>
<td>6 months before and 3 months after the forking date</td>
</tr>
<tr>
<td>Experimental</td>
<td>Blackbox &amp; Openbox</td>
<td>5 (2.3%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Not Found</td>
<td>SMPlayer &amp; UMPlayer</td>
<td>22 (10.0%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
5. EXPECTED BENEFITS OF PARTICIPATION IN OPENSYM 2013 DOCTORAL SYMPOSIA

Having my research work evaluated and provided feedback from the experts in the community is going to be an invaluable opportunity that would help chisel out my research. Meeting other doctoral symposium participants will also help broaden my network of researchers with similar interest, which will facilitate possible future collaboration.

6. ACKNOWLEDGMENTS

My thanks to Carlos Jensen, my advisor, and the peers at the Human-Computer Interaction Lab at Oregon State University for their feedback.

7. REFERENCES


Biographical Sketch

Amir Azarbakhht
PhD Student
Oregon State University
1148 Kelley Engineering Center
Corvallis, OR 97331
U.S.A.
Phone: +1 (347) 276-0790
Email: azarbaam@eecs.oregonstate.edu
Web Page: http://eecs.oregonstate.edu/azarbaam

Amir Azarbakhht is a PhD student in computer science at Oregon State University. His research focuses on Social Network Analysis (SNA) on online communities, especially Free/Libre Open Source Software (FLOSS) communities. He earned his master’s degree in Human-Computer Interaction and Artificial Intelligence at Chalmers University of Technology in Gothenburg, Sweden, and holds a bachelor’s degree in computer engineering from Azad University of Central Tehran, Iran. He is currently working on a project that focuses on analyzing online communications of online FLOSS communities, particularly the ones that have forked, to identify common measures in such complex networks. Temporal visualization of these complex communication networks over time is another focus of the project which aims to array disparate information to make sense of the underlying structure and dynamics of such social networks. He is trilingual, a GNU/Linux user and advocate, and an avid traveler who has visited 9 countries and has lived on three continents and is aiming to work as a data scientist.

Major Advisor:

Carlos Jensen
Associate Professor
Oregon State University
3061 Kelley Engineering Center
Corvallis, OR 97331-5501
U.S.A.
Phone: +1 (541) 737-2555
Fax: +1 (541) 737-1300
Email: cjensen@eecs.oregonstate.edu
Web Page: http://eecs.oregonstate.edu/people/jensen

Carlos Jensen is an associate professor in the School of Electrical Engineering and Computer Science (EECS) at Oregon State University (OSU). He received his BS. degree in Computer Science from the State University of New York (SUNY) Brockport, and a Ph.D. in Computer Science from the Georgia Institute of Technology in 2005, where he was a member of the Graphics, Visualization and Usability Center (GVU). His research Interests are Human-Computer Interaction, Usability Engineering, End-user software engineering, Usable Privacy and Security. He leads the HCI research group. He lead OSU’s Computer Science Platform for Learning efforts, looking to give OSU’s students more hands-on and engaging experiences in CS. As part of these effort, they developed a custom community of code called Beaversource, the OSWALD ultra-mobile PC for students, both of which were part of a NSF sponsored CPATH project under his leadership. The HCI research group are working on getting a better understanding, and addressing the needs of users in a number of different communities, including Open Source developers, sports broadcasters at the Olympics, overloaded information workers at Intel, and web-surfers making privacy and security decisions.
Amir Azarbakht

**Contact Information**
3212 NW Foxtail Pl.
Corvallis, OR 97330
U.S.A.

Mobile: (347) 276-0790
E-mail: azarbaam@eecs.oregonstate.edu
Webpage: http://eecs.oregonstate.edu/~azarbaam
LinkedIn: http://www.linkedin.com/profile/view?id=63719178

**Areas of Expertise**
- User Experience (UX)
- Mobile Computing (Android Development)
- Social Network Analysis (SNA)
- Human-Computer Interaction (HCI)

**Publication**

**Education**

**Ph.D. in Computer Science**
*Oregon State University* 2011-present
- Analyzing Online Communities of FLOSS & their Social Dynamics with Temporal Social Networks

**M.S. in Intelligent Systems Design**
*Chalmers University of Technology* 2009-2011
- Human-Computer Interaction (HCI) & Artificial Intelligence (AI)

**B.S. in Computer Engineering**
*Azad University of Central Tehran* 2004-2008
- Completed the 4-year program in 3.5 years, as an exceptional student

**Graduate Coursework**
Mobile & Cloud Software Development (CS569); Information Visualization (CS519); HCI meets Software Development: The Case Study (CS569); Qualitative & Quantitative research methods; Open Source Software Development (CS519); Theory of Statistics I (ST561); Theory of Statistics II (ST562); Statistical Methods of Data Analysis (ST511); Unix Internals: FreeBSD Operating System; Artificial Intelligence I; Artificial Intelligence II (Stochastic Optimization); Ubiquitous Computing; Physical Computing; Algorithms & Data Structures (CS515); Computer Vision (CS556)

**Skills**

**Academic Appointment**
- Teaching Assistant for Operating Systems (CS 311) Winter 2013
- Teaching Assistant for Parallel Programming (CS 575) Spring 2012
- Teaching Assistant for Data Structures (CS 261) Winter 2012, Fall/Spring 2013
  - Lecturer for weekly recitations
  - Grader for assignments on C programming
  - Held Office hours providing programming & debugging help to students

**Languages**
- English (Expert)
- Persian (Expert)
• FRENCH (Intermediate)
• SWEDISH (Elementary)

Scholarships
The Adlerbert Scholarship Foundation  
Gothenburg, Sweden, 2010

Volunteer Work
• Orientation leader for 20+ exchange & international students.  
  Chalmers University’s International Reception Committee. Sweden, 2011
• Representative of computer engineering students in the Student Union  
  2005-2006
• Elected Head of the association of computer engineering students.  
  2006-2007
• Elected Steward for the Coalition of Graduate Employees (CGE).  
  2012-present

Memberships
• ASSOCIATION FOR COMPUTING MACHINERY (ACM)
• AMERICAN FEDERATION OF TEACHERS (AFT)
• National Iranian-American Council
• AMNESTY INTERNATIONAL
• MÉDECINS SANS FRONTIÈRES
• UNITED NATIONS CHILDREN’S FUND (UNICEF)
• Chalmers Dance Society (I dance salsa)

Further Info
Available on my webpage: www.azarbakht.info