

Tools of the Tool Makers

Joseph A. Cottam
Indiana University
150 S. Woodlawn Ave
Bloomington, IN 47405
01+812-334-9122
jcottam@cs.indiana.edu

Andrew Lumsdaine
Indiana University
150 S. Woodlawn Ave
Bloomington, IN 47405
+01 812-855-6486
lums@cs.indiana.edu

ABSTRACT

In this paper, we describe an analysis of the social network surrounding community developed software as represented on Sourceforge.net. This analysis looks at the factors (network and otherwise) that characterize a successful project both present in its mature state and historically.

Categories and Subject Descriptors

D.2.9 [Software Engineering]: Management – *Programming teams, Life cycle*

K.4.3 [Computers and Society]: Organizational Impacts -- *Computer-supported collaborative work*

General Terms

Measurement, Design, Human Factors.

Keywords

Social Networks, Software Development, Open Source Software, Community Development, Sourceforge

1. INTRODUCTION

Community software development has experienced rapid expansion in the past several years. There are few software markets that do not have viable offerings from a community development project. Unlike traditional software development, community development is self organizing and self regulating. With new mode of organization come the questions: 1) what organization patterns are taken by community projects, and 2) which of those are effective in producing mature software?

By opening up the software development practice to any willing participant, community development groups assert that the social network becomes the most effective tool is the development process [1]. It is argued that the community process empowers the software users to directly motivate development, and provide greater potential to find and fix mistakes. To discover how these philosophies are actually implemented, an analysis of the social networks surrounding software development was undertaken.

2. Methods

A popular community development support site Sourceforge.net

was selected to sample the community development offerings. Sourceforge boasts 120 thousand projects and one million users in its site portfolio. It provides communication, distribution, web presence and source code change control services to the projects it hosts. Further, Sourceforge has taken a *lazie fare* approach to regulation, not placing restrictions on the target audience of the resulting products, design process or organization of the projects hosted there. It represents a strong sample of projects being developed for many different communities and provides easy access to much of its content.

The analysis focused on a snapshot of the projects in the Sourceforge web site in February 2006. The goal was to find a characterization of ‘successful’ projects. This phase examined the network of developers that link projects together and the attributes of the projects themselves. Questions include: 1)How large is the average project, and 2)how is the labor divided up? 3) Is there a process similar to reputation diffusion in academic publication networks at play in the software world [2]? It was performed on a non-directed bi-partite graph where individuals were linked to the projects they worked on.

3. Results

The crawl resulted in a network of 113,000 projects and 160,000 users in the network.. The distribution of degrees roughly follows a power law, with over 75,000 projects having only one developer and the largest project having over 300. These results confirm those reported by Madey [3], but on a much larger network. The network also contained one giant connected component with 90,000 participants (projects and individuals), the remaining nodes divided among 63,000 other components of various sizes. Projects rated as either “Mature” or “Production/Stable” often occur with only one individual, but visual inspection showed the degree distribution decays more gently than in the general graph. This more gentle decay lends some support to the assertion that development communities aid the development process. Preliminary analysis has not uncovered anything analogous to reputation diffusion, instead it suggests that developers come with pre-existing skills and reputations.

4. Future Works and Conclusions

The change control system at Sourceforge provides access to historical records of the projects. The next steps for this project involve analyzing this data to find patterns through time. How does the network grow? Are different roles distinguishable in their change control trail?

It is hoped that the results of this analysis will help inform future community development practices by giving effective reference points for their organizational structures.

5. ACKNOWLEDGMENTS

Thank you to Katy Börner for assisting in analysis direction and selection. Thank you to Chris Mueller and Douglas Gregor for technical assistance.

6. REFERENCES

- [1] Krishnamurthy, S. Cave or Community?: An Empirical Examination of 100 Mature Open Source Projects. First Monday, 2002 Available at SSRN: <http://ssrn.com/abstract=667402>
- [2] Siow, A. Are First Impressions Important in Academia? Journal of Human Resources, Vol. 26, No. 2 (Spring, 1991). pp. 236-255
- [3] Madey, G., Freeh V., Tynan R. The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory. Eighth Americas Conference on Information Systems. 2002.