Master Thesis

Collaboration networks in open-source software development

Jozsef Csepanyi

Date of Birth: 06.09.1996
Student ID: 11927479

Subject Area: Information Systems

Studienkennzahl: h11927479

Supervisor: Johannes Wachs

Date of Submission: 31.March 2021

Department of Information Systems and Operations, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

Contents

List of Figures

List of Tables

Abstract

Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus....

1 Introduction

In recent years open source software solutions have become widely popular and frequently used in both scientific and enterprise use, which can be attributed to a number of factors, most importantly the ease of development and deployment of IT projects, improved cybersecurity and enhanced scalability [1]. This increases the contribution to open source projects from enterprises and individuals alike. Due to its nature, open source software projects are driven by community contributions, and depend heavily on active participation in all phases of the project. Because there is a high dependency on the community in open source software projects, by understanding how contributions are included and what patterns emerge we can gain valuable insight into the project's current state and its trajectory.

...

1.1 Literature review

- Open-source software development properties

    - centralized vs decentralized
    - No collocation
    - Enterprise support
    - Version control, issue tracking

- Relevant social aspects of OS projects

- State of the art

    - Collaboration by coediting files
    - Contributors form dynamic social networks
    - Problem of analysing changes over time in a network
    - Other studies in this field...

- Preliminary analysis results (pandas, networkx, ...)

1.2 Motivation of research problem and research question

- Importance of OS project analysis based on lit rew

- Analysing effects of large events within the lifecycle of the OS project in order to improve them or adapt

  - Planned, foreseeable changes (e.g. upcoming major release)
  - Unforeseeable changes (e.g. end of support, pandemic)

- Research questions

  - What social patterns emerge within large-scale open-source software projects?
    * Are there smaller "core" collaborator networks connected with weak links or do they form one large interconnected network?
    * Are there usually key contributors, who are central to the project and collaborate with most contributors, or is it completely decentralized?
    * How does the size of the project change these properties?

  - How does the structure of OS software development collaboration change over time?
    * Are there any major changes over the natural project lifecycle? Are they visible in the collaboration network? (e.g. planning, developing, bugfixing, sunset?)
    * How does a sudden major event change the participation and development?

2 Proposed research method

- Developing a tool, that can extract the collaboration information from any OS project (from GitHub/git repository)

- Data cleaning - method to merge authors, excluding common folders, etc...

- Qualitative research

  - Observing collaboration statistics and networks in order to discover patterns: connected components, centrality, changes over time

- Quantitative research

  - Composing a large set of repositories (different sizes, properties)
  - Detecting past changes automatically based on changes in measured statistics

3 Outline of thesis

- Literature review

  - Network analysis, relevant metrics
  - Properties of social collaboration networks

- Used repositories, selection criteria

- Data cleaning - files, authors, max modifications

- Implementation

  - ...

- Qualitative analysis

- Quantitative analysis

- Conclusion

3.1 (Preliminary literature list - in references)

3.2 Work plan including milestones

- Data cleaning - files, authors, max modifications

- Implementation

  - ...

- Qualitative analysis

- Quantitative analysis

- Conclusion

References

[1] PwC. Leading benefits of open-source software among enterprises world-wide as of 2016. *Statista*, 2016.