Measuring Perceived Trust in Open Source Software Communities

Mahbubul Syeed, Juho Lindman, Imed Hammouda

Measuring Perceived Trust in Open Source Software Communities

Mahbubul Syeed[1], Juho Lindman[2], Imed Hammouda[3]

[1] American International University-Bangladesh, Bangladesh,
mahbubul.syeed@aiub.edu,
[2] Chalmers and University of Gothenburg, Sweden,
imed.hammouda@cse.gu.se,
[3] University of Gothenburg, Sweden,
juho.lindman@ait.gu.se

Abstract. We investigate the different aspects of measuring trust in Open Source Software (OSS) communities. In the theoretical part we review seminal works related to trust in OSS development. This investigation provides background to our empirical part where we measure trust in a community (in terms of *kudo)*. Our efforts provide further avenues to develop trust-based measurement tools. These are helpful for academics and practitioners interesting in quantifiable traits of OSS trust.

1 Introduction

Trust can be perceived as the relationship between people where one person is taking a risk to accept other persons' action [1]. Such trust is one of the fundamental traits of a successful collaborative development environment, e.g., OSS projects [3][4][5][6].

Trust is directly linked to securing a functioning development community, community governance and on-going group work [3][4]. A community with such attributes can attract new developers to join and contribute in the project [4]. Therefore, the sustainability of an OSS project much depends on trust-related questions [2].

Alongside, third party organizations often try to ensure the quality and re-usability of an OSS component before adoption. Such verification is strongly coupled with the trust rating of the developers who developed those modules [10][11][12]. Therefore, assessing trust among the developers is a seminal question in OSS research. Recent study in [5] infer trust among developers based on their contributions. Others measured developers' contributions in Man-Month to investigate trust among them [8].

This study address the following two issues related to trust: first, we investigate the relationship between developers' contribution in OSS projects and their trust rating in the community, which is formulated as follows, *RQ1: How likely a trust rating of a developer changes when contributing to OSS projects?* Second, we investigate how community status affects members endorsing each

others' contribution. Status difference may arise due to: (a) different trust levels in the community, and (b) homophilic elements [9], e.g., same project, country, location, programming language and same community status. Therefore we investigate the following, *RQ2: How likely developers with different community status will endorse each others' contribution?*

2   Methodology

2.1   Data Collection and Presentation

Our data is from OpenHub [13] which records the evaluation information (popularly known as *kudo*) sent or received by the developers over time. Developers are ranked according to their kudo scores, known as *kudo ranks*. Data is collected using the offered open APIs, which returns results in XML. Relevant information is then extracted and stored in database for further analysis.

Developer Account information. This study extracts 563,427 registered developer account information [14] that includes the following: developers account id, name, post count, and kudo rank. *Kudo rank* (a number between 1 and 10) defines ranking of the member based on certain criteria, e.g, kudo received from other members and history of contributions in OSS projects [10].

Project information This study extracted 662,439 OSS project data which holds the following information: project id, name, and total user.

Contributor information A contributor dataset holds project specific contributions made by each developer. A total of 844,012 contribution records were collected, each of which holds, project specific contributor's id and name, personal account id and name, project id in which contribution is made, time of first and last commit, total number of commits and the man month (i.e., the number of months for which a contributor made at least one commit). In the context of OpenHub *Man Month* represents the accumulated contributions made by a developer in OSS projects.

Kudo sent and received history Following information related to kudo sent and received by a member for a given period of time is collected: sender or receiver account id and name, receiver or sender account id and name, project id and name, contributors id, name and date when the kudo was sent or received. A total of 46,926 kudo received records and 57,458 kudo sent records are collected.

Kudo sent and received history data is then combined to generate a uniform dataset that has the following data items: Sender account id and account name, receiver account id and name, project id and name to which the kudo was sent, contributor id and name, and the date when kudo was sent.

2.2   Data Analysis

RQ1:   To determine the extent to which a developer is trusted against his contributions in OSS projects, following approach is adopted. First, contributors (or developers) are grouped based on the first commit date. Then the three

Table 1. Developers count on first commit date

First Commit Date	No of Developers committed
2012-03-26	40
2012-07-05	39
2012-05-10	36

dates in which maximum number of developers count found are taken for further analysis, as shown in Table 1.

Then, the contributions (Section 2.1) and the kudo rank (Section 2.1) of every developer under each of the three dates are measured. This measurement is done from the first commit date till the last. In this study, the *man month* (Section 2.1) is used to measure developer contributions. Alongside, *Kudo rank* is used to represent the trust value of a developer. Additionally, following logical assumptions are made: (a) each of the developers starts with a kudo rank 1 at their first commit date, and (b) change in this kudo rank or trust rating is associated with the amount of contributions made by that developer over a given period of time.

RQ2: Next we examine how often developers of different community status endorse each others contributions. For this, developers are clustered according to their kudo ranks. Example clusters are, kudo rank (9 and 10), (7 and 8),(5 and 6), (3 and 4) and (1 and 2). Then, the log of kudo sent and received (Section 2.1) among developers in different clusters is recorded. This will offer a holistic view on the exchange of kudo among developers of different trust values.

Second, exchange of kudo is examined from two perspectives, namely, sending / receiving kudo (a) directly to developers' personal account and (b) in the contributors' project specific account in which the developer has contributed. This will portray a deeper understanding on how developers of different trust value recognize each others' contributions.

Third, kudo exchange log was generated based on developers who worked in the same project, distinct project, and both. This will provide insight whether working on the same projects stimulate higher kudo exchange.

3   Result and Synthesis

RQ1: How likely that a trust rating of a developer changes when contributing to the OSS projects?
The underlying assumption of this research question is that a developer should start with a kudo rank 1 (or trust rating 1) at the first committing date. And this kudo rank keeps changing along with his contributions in projects over time.

As per the reported results it is observed that 78% to 100% developers who have kudo rank 9 contributed to projects for more than 24 man-months. Whereas, 54% to 79% developers at kudo rank 8 contributed more than 24

man-months. Likewise, the man-months contribution goes down along with low kudo ranks. For instance, none of the developers at kudo rank 5 has contribution record of more than 24 man-months. To summarize our findings, developers trust rating in the community is associated with the contributions they made in the projects over time and the amount of contributions has an impact on the trust value they attain at a given point of time.

RQ2: How likely that developers with different community status will endorse each others' contribution?
We aim to identify the dynamics between the community status of the developers and endorsing each others' contributions. In order to do so, this study analyzed the collected data from three different perspectives as presented in Section 2.2. First perspective is to study the kudo exchange pattern among the developers who belong to different kudo rank clusters. Reported results show that almost all the kudos are sent to the developers who belong to the kudo rank between 7 and 10. For instance, developers having kudo ranks between 9 and 10 sent almost all of their kudos to the developer having kudo rank between 7 and 10. This pattern holds for all other kudo rank clusters as well.

This outcome leads to several observations: First, developers living at higher kudo ranks are often the ones who commit the most significant contributions to the projects, hence, the majority of kudos are attributed to them as a token of appreciation. Second, these group of developers are the ones who are trusted by all the community members irrespective of their ranking in the community. Third, developers residing at lower kudo ranks (e.g., ranking between 1 and 6) rarely receive kudos for their contributions.

Investigation on the second perspective reveals that developers with higher kudo ranking or trust rating most often receive their kudos directly to their personal account. In a very few occasions, kudos are attributed to their project specific accounts. For instance, about 83% to 93% of the kudos received by developers having rank between 7 and 10 are attributed to their personal accounts. This highlights how high ranked developers are often appreciated irrespective of the projects they have contributed to. However, Low kudo ranked (e.g., rank between 1 and 5) developers are not investigated due to lack of sample data.

Finally, the study on one of the homophilic factors, e.g. the effect of working on same or different projects, on exchanging kudo rank reveals inconclusive results. For instance, developers who work on the same and distinct projects at the same time, share kudo more frequently than developers working in either distinct or same projects. Therefore, this study does not conclusively support the earlier research on homophilic factors [9] which claims that such factors have positive impact on endorsing each others contributions.

Acknowledgments

References

1. Golbeck, J.: Analyzing the Social Web, ISBN-13: 978-0124055315, Chapter 6, Page 76.
2. Sirkkala, P., Hammouda, I., Aaltonen, T.: From proprietary to open source: Building a network of trust. Proceedings of Second International Workshop on Building Sustainable Open Source Communities, OSCOMM 2010, Pages 26-30, 2010.
3. Stewart, K.J., Gosain, S.: The impact of ideology on effectiveness in open source software development teams. Mis Quarterly, pages 291314, 2006.
4. Lane, M.S., Vyver, G., Basnet, P., Howard, S.: Inter-preventative insights into interpersonal trust and effectiveness of virtual communities of open source software (OSS) developers. 2004.
5. Laat, P.B.De.: How can contributors to open-source communities be trusted? on the assumption, inference, and substitution of trust. Ethics and Information Technology, 12(4), Pages 327341, 2010.
6. Dabbish, L., Stuart, C., Tsay, J., Herbsleb, J.: Social coding in Github: transparency and collaboration in an open software repository. pages 1277-1286, 2012.
7. Zhang, H., Wang, Y., Zhang, X.: A trust vector approach to trans- action context-aware trust evaluation in e-commerce and e-service environments. Pages 1-8, 2012.
8. Arafat, O., Riehle, D.: The commit size distribution of open source software. Pages 1-8, 2009.
9. Hu, D., Zhao, J.L., Cheng, J.: Reputation management in an open source developer social network: An empirical study on determinants of positive evaluations. Decision Support Systems, 53(3), Pages 526-533, 2012.
10. Gallardo-Valencia, R.E., Tantikul, P., Elliott Sim, S.: Searching for reputable source code on the web. Proceedings of the 16th ACM international conference on Supporting group work, Pages 183-186, 2010.
11. Orsila, H., Geldenhuys, J., Ruokonen, A., Hammouda, I.: Trust issues in open source software development. Proceedings of the Warm Up Workshop for ACM/IEEE ICSE, Pages 912, 2010.
12. Gysin, F.S., Kuhn, A.: A trustability metric for code search based on developer karma. Proceedings of 2010 ICSE Workshop on Search-driven Development: Users, Infrastructure, Tools and Evaluation. Pages 41-44, 2010.
13. OpenHub data repository, http:// www.openhub.net/.
14. OpenHub API documentation, https://github.com/blackducksw/ohloh_api.
15. Massa, P., Avesani, P.: Controversial users demand local trust metrics: An experimental study on epinions.com community. Proceedings of the 20th national conference on Artificial intelligence. Volume 1, Pages 121-126, 2005.
16. Golbeck, J., Kuter, U.: The ripple effect: change in trust and its impact over a social network. In book: Computing with Social Trust, pp.169-181, 2009.
17. Golbeck, A.J.: Computing and applying trust in web based social networks. Doctoral Dissertation. University of Maryland at College Park College Park, MD, USA, 2005.