Social network analysis in a software crowdsourcing perspective

Leticia S. Machado, Ricardo Rodrigo M. Melo, Cleidson R. B. de Souza

Universidade Federal do Pará (UFPA)
{leticia.smachado, ricardorodrigomm}@gmail.com,
cleidson.desouza@acm.org

Abstract. *Social network analysis (SNA) has been an important research tool in the Software Engineering community in the last years. The goal of this work is to bring together researcher and practitioners who have studied SNA from software crowdsourcing (SW CS) perspective, and to refine our ways of thinking about issues on communication and collaboration in SW CS projects.*

1. Introduction

Software Crowdsourcing (SW CS) is an alternative software development strategy where a large number of online users is engaged to contribute in several development activities. Such strategy, based on the crowd, has been used for companies who are seeking to increase the speed of their software development efforts [Stol and Fitzgerald, 2014]; [Yang et al., 2016]. In SW CS projects, particularly in competitive activities, there are potentially misunderstandings, restricted communication, and lack of social interactions during the development software activities and this often impacts on the success of the projects [Stol and Fitzgerald, 2014; LaToza et al., 2015; Yang et al., 2016]. The lack of appropriate communication can result in difficulties to understand the task requirements reducing the productivity and affecting the quality of the solutions.

Software engineering is an inherently sociotechnical endeavor [Whitehead, 2007]. Some studies demonstrate the importance of social interaction in SW CS projects [Nag et al., 2012; Gray et al., 2016]. For instance, in SW CS projects, it was found that the effort involved in collaboration readiness was an essential challenge among crowd workers and the crowdsourcing platform impact the work. During each development challenge communication among crowd and platform members is strongly focused on the task (goals, requirements, restrictions, technology and so on) in terms of request for help, help answers, problems identified, and responses to problems.

Communication and collaboration play an important role in supporting the quantity and quality of crowdsourced task submissions in competitive SW CS projects [Yang et al., 2016; LaToza et al., 2015; Tausczik and Wang, 2017]. The interactions among crowd participants during the tasks configure implicit social networks. Making these network connections explicit might help researchers and platforms designers to better understand and support communication and collaboration tools. A way to make these connections explicit is through social networks analysis [Wasserman and Faust, 1994].

The term “social network” refers to the graph structure made up of social actors (individuals, teams, or organizations) and connections between these actors, reflecting the relational information about the actors [Wasserman and Faust, 1994]. Social network analysis (SNA) is a part of this field, and many aspects from SNA support the

understanding of the individual roles and interaction patterns and anti-patterns in a social network. From a SW CS perspective, social networks can be extracted by analyzing the post (messages) from communication forums. Usually, these messages are exchanged by crowd participants during competitive SW CS tasks.

This study is based on the preliminary results from [Machado, 2019] that show that crowd workers who communicate with platforms moderators are more productive in their work: they submit solutions more often and consequently, are more likely to win a SW CS challenge. Therefore, we plan to explore SNA and cluster analysis in the SW CS context to examine the relationships between crowd participants and platform moderators to better understand the influence of these social interactions (communication) in terms of task completion. Moreover, we plan to investigate how social network analysis can be used to support SW CS tasks and understand how platform moderators and crowd participants collaborate and compete, how platforms can reduce the amount of crowd members who withdraw from the competitions, and how a social network sites can support a professional development community

To illustrate a possible example of the social network in SW CS competitions, Figure 1 illustrates a social network that capture the collaborative network over a specific competition. Different metrics can be used to analyze this particular social network including number of nodes, number of edges, network density, all measures of community-size, collaborative behavior and community cohesion respectively. Different metrics of centrality can also be used to explore the role of particular actors in this network [Wasserman and Faust, 1994].

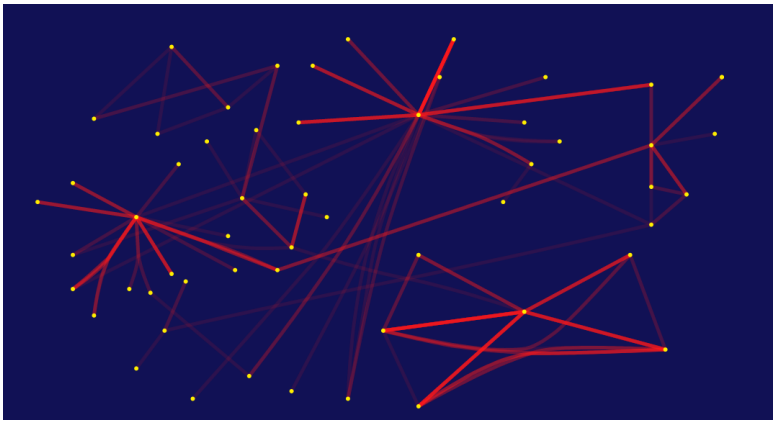

Figure 1. Example of a social network created from SW CS forum messages.

2. Research Design

We plan to combine qualitative analysis of archival data (QA), mining software repositories (MSR) and Social Network Analysis (SNA) on available communication forums from SW CS platforms, specifically the TopCoder platform. Our goal is to re-construct and visualize the social interactions as a sequence of networks. We plan in our study to evaluate, for instance, whether crowd members who are winners of challenges have more connection ties with the other members or not.

Dissanayake et al. [2014] points out that team performance in crowdsourcing competitions is higher when members with high task-related skills are not centralized in the network. They reveal that teams benefit if their members have higher task-related skills or more connection ties with other members.

2.1 Data Collection

We will start in a qualitative way, by screening online forum's messages hosted on the TopCoder platform [Topcoder, 2019]. We will read the content of each message and classify in categories and topics using grounded theory coding procedures [Strauss and Corbin, 2007] to identify meanings associated with the messages. Topics will be defined at a level of detail that could allow us to understand which subjects crowd members communicate with each other about during the competitions. Examples of topics include requirements, libraries, output, access, etc.

Thus, based on the interpretation of the content of each message, we will visualize which topics the crowd and platform moderators communicate about in the competitions. Furthermore, the quantitative data from the number of messages sent by involved parties (crowd members and platform moderators) may detect the distinct user roles and demonstrate the associations between these roles and the number of messages associated with each role. After the data is collected, we will construct the social networks, which will allow us to analyze the overall interaction behavior and the communication dynamics among different crowd participants (including the platform moderators) in SW CS contests.

We believe the social network analysis might help identify empirical groupings of persons, which can then serve as a basis for further analysis in order to identify different user roles (crowd participants as set of quitters, submitters and winners). From this data we will extract: date and time of the forum's messages (i), thread (ii), sender (iii), recipient (iv), and finally the message itself (v). The nodes of the social network will be the crowd participants who registered on the same SW CS competition and the platform moderator – who works for the SW CS platforms to mediate the communication between the customer and the crowd. Meanwhile, the relationship between the nodes will be extracted from the messages exchanged among the participants.

2.2 Data analysis

Our empirical materials to be analyzed will be the SW CS forums associated with challenges from the busiest Topcoder months registered in the coding category competitions from 2017 and 2018 [Dubey et al., 2016]. Thus, we will plan to take a longitudinal approach and to construct SNA visualizations that depict collaborative behaviors competition by competition. As mentioned, we plan to use different metrics to evaluate the social networks that will be created.

Users on the Topcoder platform must be registered in the competitions they want to participate. After registering, they are able to either write a post or receive posts via the online forum, therefore, two different measures can be used to capture commenting behavior, namely in-degree and out-degree [Wasserman and Faust, 1994]. With the help of in-degree and out-degree centrality, popularity or activeness of a user can be determined [Kratzer et al., 2008].

3. Expected Outcome

We expect with the help of SNA to investigated social structures of crowdsourcing communities and identify different user roles in SW CS projects. SNA represents a valuable method to identify user roles as the derived social network methods can be used as a practical diagnostic and monitoring tool for community behavior. Specifically, we hope that SNA will refine our understanding of the communication and collaboration taking place in competitive software crowdsourcing projects, an understanding that is currently limited.

Acknowledgments

The authors thank the Brazilian Council for Research (CNPq) for the financial support through processes 420801/2016-2 and 311256/2018-0.

References

Dissanayake, I., Zhang, J., Gu, B. "Virtual team performances in crowdsourcing contests: A social network perspective completed research paper". In 35th International Conference on Information Systems: Building a Better World Through Information Systems, ICIS 2014. Association for Information Systems.

Dubey, A.; Abhinav, K.; Taneja, S.; Virdi, G.; Dwarakanath, A.; Kass, A.; Kuriakose, M.S. Dynamics of software development crowdsourcing. ICGSE'16: Proceedings of the IEEE International Conference on Global Software Engineering, Orange County, California, 2 August – 5 August 2016, pp. 49-58.

Gray, M.L.; Suri, S.; Ali, S.S.; Kulkarni, D. "The crowd is a collaborative network". In: 19th ACM conference on computer-supported cooperative work & social computing, 2016, ACM, pp. 134-147.

Kratzer, J., Lettl, C. "A social network perspective of lead users and creativity: An empirical study among children", 2008. Creativity and Innovation Management, 17(1), 26–36.

LaToza, T.D.; Chen, M.; Jiang, L.; Zhao, M.; Van Der Hoek, A. "Borrowing from the crowd: A study of recombination in software design competitions". In: 37th International Conference on Software Engineering, v1, 2015, pp. 551-562.

Machado, L.S; Melo, R.; Prikladnicki, R.; de Souza, C.R.B. "Collaboration in Competitive Software Crowdsourcing: An Empirical Study of TopCoder. Submitted to the CSCW Journal, 2019.

Strauss A.; Corbin, J. M. "Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory". SAGE Publications, 2007, 312p.

Stol, K.-J.; Fitzgerald, B. "Two's company, three's a crowd: a case study of crowdsourcing software development". In: 36th International Conference on Software Engineering (ICSE), 2014, ACM Press, p.p.187.

Tausczik, Y.; Wang. P. "To Share, or Not to Share? Community-Level Collaboration in Open Innovation Contests". CSCW, 2017.

Topcoder, 2019. Access in: https://www.topcoder.com. July 2019.

Yang, Y.; Karim, M.R.; Saremi, R.; Ruhe, G. "Who should take this task? Dynamic decision support for crowd workers". In: Proceedings of the 10th ESEM, 2016.

Wasserman S, Faust K. Social network analysis: methods and applications. 1st ed. Cambridge, New York: Cambridge University Press; 1994.

Whitehead, J. "Collaboration in software engineering: A roadmap". In: Proceedings of Future of Software Engineering, 2007, pp. 214-225.