The Impact of Gamification on Socio-technical Communities: A Case for Network Analysis

Antti Knutas
Lero, the Irish Software Research Centre
Dublin, Ireland
antti.knutas@dcu.ie

Timo Hynninen
Lappeenranta University of Technology
Lappeenranta, Finland
timo.hynninen@lut.fi

ABSTRACT
Designing for motivating and engaging experiences is at the core of gamification. The results of gamification are often evaluated with user experience testing involving recordings, surveys, and interviews. However, in multi-user socio-technical environments the benefits of gamification are often realized in interactions between users. We propose that social network analysis should be used more to analyze the impact of gamification at community level. To demonstrate the approach, we present a study where a gamified computer-supported collaborative learning system was introduced to a course, and compare the course to a previous instance. Furthermore, we present several examples of how social network analysis can be used with hypothesis testing and discuss the benefits of the approach.

ACM Classification Keywords
H.1.2. User/Machine Systems: Human Factors; I.2.1. Applications and Expert Systems: Games; K.3.1. Computer Uses in Education: Collaborative Learning

Author Keywords
gamification, social network analysis, hypothesis testing, computer-supported collaborative learning, socio-technical systems

INTRODUCTION
Collaborative learning is a learning method where students have a symmetry of action, knowledge and status, and have a low division of labor [15]. Computer-supported collaborative learning (CSCL) facilitates the interaction with software tools and increases potential for creative activities and social interaction [30]. Collaborative learning is a commonly used method in software engineering education as it prepares students to work in software teams as independent experts. In recent studies, it has been shown that students can be guided towards educational goals like collaboration by using gamification [25], which is the application of game-like elements to non-game environments [13].

GHItaly18: 2^{nd} Workshop on Games-Human Interaction, May 29th, 2018, Castiglione della Pescaia, Grosseto (Italy).

The gamification of learning in software engineering education has been studied earlier [25, 14, 3], mostly at the level of individual students. However, software engineering education classes and software engineering are communities, or *socio-technical entities* [17], which are interconnected by personal and technical communication channels. In this context, we define a socio-technical system as a complex system which involves both physical–technical elements and networks of interdependent actors [10]. These technological structures are not only neutral instruments, but also shape users' perceptions, behavioural patterns, and activities [17]. Computer-supported collaborative learning tools as software systems always include certain inbuilt assumptions, interaction methods, and rules. An issue tracker such as GitHub enables collaboration in a different way than for example a forum software. When gamification is added into the mix, it adds explicit interaction rules. Embedding a technical element into the social environment changes not only the experience of an individual users, but has the potential to impact interaction patterns and rules of social interaction.

We propose that social network analysis (SNA) is a fitting research approach for investigating the impact of new technologies in socio-technical systems, such as in the case of applying gamification. Social network analysis can be summarized to be an interdisciplinary research field about who communicates with whom, in which social relationships are viewed in terms of the network theory [26]. In social network analysis, communication between individual or social units are mapped into a communication matrix and then modeled as graphs that are composed of nodes and edges. Nodes represent individuals and edges the connections or communications between them. These graphs can be used to visualize communication patterns in socio-technical systems systems. Additionally, in the graph theory there are different mathematical tools available, which can be used for example to estimate the relative influence of nodes in the graph or analyze the graph by the connection patterns of the nodes [2, 4, 29].

Currently many social network analysis -based studies rely only on descriptive statistics and omit hypothesis formulation and evaluation. We make a case that hypothesis testing is an essential part of empirical studies that use quantitative methods, such as SNA. Wohlin et al. [32, p. 12] summarize the the importance of hypothesis testing for generating new knowledge as follows. "In science, physical phenomena are addressed by putting forward hypotheses. The phenomenon is observed and if the observations are in line with the hypothesis, this becomes evidence for the hypothesis. Experiments are important to test the hypothesis and in particular the predictive ability of the hypothesis. If the new experiments support the hypothesis, then we have more evidence in favor of the hypothesis." *If the study remains at the level of descriptive statistics, no evidence is created to support the presented new knowledge.* [1]

To illustrate the social network analysis as a research method and how it can be used in hypothesis testing, we present a study of how embedding technical elements into a collaborative learning course creates a new socio-technical system. The goal is to evaluate how the new elements affect the structure of collaboration in and between student teams. First, we present descriptive statistics generated by social network analysis and then test several hypotheses about the impact based on that data. We conclude the paper with discussion of the study results and what kind of insight social network analysis can provide for other studies.

RELATED WORK ON ANALYSIS OF GAMIFIED COMMUNITIES

Social network analysis has been utilized by social scientists to explain different phenomena. Borgatti et al. [7] divide the various themes that can be examined by SNA into the similarities of actors, their social relations, interactions and flows (transferring physical or intellectual property between actors).

A number of authors have previously conducted research in group and peer learning activities in software engineering education using social network analysis [21]. This is because the effects of social networking in an educational setting can have a substantial impact on learning performance [11]. For example, the study by De Laat et al. [24] presents an overview of how social network analysis can be applied in the field of CSCL. The authors build a case for utilizing SNA to investigate the group dynamics in a CSCL environment. The study emphasizes that in order to fully understand student participation in CSCL environments, we must analyze who are the actors in the collaborative learning task, and if actors are participating actively or peripherally. It is also important to understand how the different methods of participation change over time.

In terms of social network analysis in gamified CSCL context, de-Marcos et al. [12] used SNA to investigate a gamified e-learning course. In the study different network metrics were assessed in how well they can predict academic performance. The investigation found potential in the use of network metrics to predict learning achievement but also found limitations to their applicability beyond central nodes in the social network. This study is one of the few studies which have evaluated the impact of gamification on socio-technical communities from a network analysis perspective.

[1] The authors acknowledge that this is a very positivist view and other philosophies of science or qualitative approaches are just as valid. However, in positivist-quantitative approaches hypothesis testing is essential.

More recently SNA has been utilized by Wise & Cui [31], for example, to distinguish social relationships in MOOC discussion forums. Their study showed connections between network structure and discussion practices. The study also pointed out that interactions involving course content related discussion and other discussion should be examined separately.

SOCIAL NETWORK ANALYSIS AS A RESEARCH METHOD

Social network analysis is based on graph theory, which is a concept related to mathematics and information sciences [18]. Central concepts in graph theory are network, node, and edge. A network is composed of actors, or nodes, between which are connections, or edges. A common method in social network analysis is to store graphs as adjacency matrices. From these matrices one can calculate different statistics and characteristics with graph theory and matrix algebra. Several applications have been developed to automate the analysis, such as UCINET [6] and Gephi [4]. A common method to visualize the graph matrix is a sociogram, which is is a visual representation of the nodes and and their edges. Most often the nodes are represented as circles and edges as connecting lines.

Descriptive Statistics for Graphs and Nodes

The degree is most essential of node characteristics. The node degree indicates the number of edges connected to a node, or the number of connections the node has to other nodes. The weighted degree also takes into account the different values the edges might have. Additionally, the most simple measure of centrality, the degree centrality, is derived from node degrees. Degree centrality is calculated from the sum of edges a node has.

Graph density is derived from node degrees. It is a value between zero and one, and it indicates how many connections have been established of all possible connections. If all nodes are connected to all other nodes, it is 1. If the graph has no edges, then network density is 0.

A social network can be composed of one or several components. A component is an element of a social network where a path can be traced from one node to another. The number of connected components can be evaluated how fragmented the network is. In well-functioning socio-technical systems the desired number of connected components is just one.

Centrality is a method to measure nodes' relative importance in a graph [29]. Application areas for centrality include finding influencers in social media or finding the most active collaborators in learning. Degree centrality is the easiest to calculate and was covered earlier in this section. Betweenness centrality is the measure of all the shortest paths that go through the node. Other, more advanced methods such as eigenvector centrality are beyond the scope of this article.

Regression Analysis and Comparisons in Social Network Analysis

Traditional hypothesis testing, such as regression analysis, is difficult to use on networks, because edges exist as dyads between two nodes. The Multiple Regression Quadratic Assignment Procedure (MRQAP) method was developed to address this issue and it allows analyzing the effect of external variables on network structure [23]. Correlation between networks can be analyzed with a related method, the Quadratic Assignment Procedure (QAP) [22].

CASE STUDY: ANALYZING THE IMPACT OF GAMIFICATION ON STUDENT COLLABORATION

In this section, we present a case study where we perform comparative social network analysis of two subsequent teamwork-based computer-supported collaborative courses in software engineering at master's level. The research data has been collected in earlier studies [19, 20]. The two courses were arranged in subsequent years with the main difference being that in the second year a gamified, computer-supported collaboration platform was introduced to the course. In this paper, we take the analysis further with hypothesis testing, with an emphasis on evaluating the impact of the artifact introduced in the second study.

In the first study [20], we observed 17 students over a five day long intensive format and collaborative software engineering course, arranged in the Code Camp format [27]. The course had 10 hours of lecturing and 40 to 64 hours (depending on the student team) of collaborative teamwork around a set task. The topic for the course was to develop a new mobile or tablet application before the deadline on the fifth day. The teams were free to choose their own path to a solution within the theme and how to achieve it. The students had no other courses during the week. The students spent their time in the same shared computer classroom, with each student team sitting at their own table group. After the kickoff event and a technology tutorial, the teachers were available to advise when requested during the rest of the week and to facilitate inter-team collaboration. Most of the course participants were master's level students with previous experience of other programming projects. All recorded collaborative communications occurred offline in the classroom in this study.

All student interactions that occurred in the classroom were recorded during the first study. The video and audio recordings were combined into multi-angle and surround sound videos that allowed the researchers reviewing the video to analyze several concurrent interactions. This resulted in 40 hours of video, from which 3366 interactions were coded for analysis. The data analysis presented in this paper is sourced from the quantitative analysis results of the video recordings.

The second study [19], with the computer-supported collaborative learning system, was similar to the first one and also arranged in the Code Camp format [27]. The only differences were the introduction of the computer-supported collaborative learning system, the theme of the course, and the number of participants. This time there were 22 participants divided into six teams. The topic of the course was using open data for sustainability. In this study, all recorded collaborative communications occurred online in the computer-supported collaboration system. The data analysis presented in this paper was sourced from the system log files.

Software Artifact

The software artifact for computer-supported collaboration in software engineering student teamwork, originally presented in [19], was created following along the following principles: Increased team collaboration, extended collaborative communication between students, and explicit goal communication that supports shared goal setting and goal achievement.

These tenets were realized as a collaboration platform that concentrated around setting team goals and issues, viewing the status of other teams and using a chat-based tool, Slack, to communicate and share information regarding the goals and issues. The information and views concentrated around three main views, with a sidebar and notifications showing additional information when needed. The guidelines of responsive design were followed in the implementation of the views, with the system working equally well on desktops or mobile phones. The selected chat system also had a mobile client application for iOS and Android, enabling continuous collaborative communication for the community.

Social Network Analysis Results

We performed descriptive social network analysis with the Gephi software [4] and hypothesis testing with the R statistical programming language [28]. First, we produced descriptive statistics for the two networks for comparison purposes, such as the number of components in the network, various indicators for network size, and the average degree for the nodes. Then we tested hypotheses of whether the distribution of node attributes, such as degree and centrality, were different in the networks, and how well being in the same group predicts two students having a collaborative connection.

Descriptive Statistics

We analyzed the social networks of two studies, classroom and online communication, and produced descriptive statistics of the networks as presented in Table 1. The online system had a little more nodes and a lot more recorded interactions. Both networks had only one connected component, which means that there was a path from each node to any other node. The online network was a little less connected, which is signified by an increased network diameter and lower network density.

Despite the lower network density, on average a node had more connections to other nodes in the online network as signified by the average degree. However, the average weighed degree was much higher in the classroom, implying that students in the online system were slightly better connected and classroom-based students established stronger and more even ties.

Hypothesis testing

Descriptive statistics are not enough to evaluate proof of evidence when for example making claims about the differences in social networks. For this, we need statistical hypothesis testing which can give us the probability of the evidence supporting our hypothesis being significant, and the effect size of the evidence [9]. An obtained p-value represents the probability of observing the current data due to chance when the null hypothesis is true. Effect size allows us to evaluate how substantial the evaluated phenomenon is, after it has been shown that the phenomenon probably exists.

Graph	No. of nodes	Recorded interactions	No. of connected components	Network diameter	Average degree	Average degree (weighted)	Network density
Classroom	18	703	1	3	9.39	374	0.55
Online, gamified	23	3216	1	4	10.5	54.5	0.46

Table 1. Descriptive social network analysis statistics

We first used the Mann–Whitney U test [32] to test the difference in distributions between the datasets because of its suitability for non-normal data. A continuity correction was enabled to compensate for non-continuous variables [5]. The Bonferroni correction was used to adjust the p-value to compensate for the family-wise error rate in multiple comparisons [1]. We calculated effect size r using guidelines by Fritz et al. [16] for the Mann-Whitney U test. We evaluate the effect size as proposed by Cohen [9] that in r a large effect is .5, a medium effect is .3, and a small effect is .1.

Our null hypotheses is that there is no difference in the distributions of node degrees or betweenness centralities between the networks. Node degree and two node centrality variables were tested to establish differences in the distribution of connections and whether there was overall differences in how the relative importance of nodes were distributed. The Mann-Whitney U test results are summarized in Table 2, with the sum of ranks denoted by the U-value, the probability by the p-value, and effect size by r. When evaluating node degrees and betweenness centrality we fail to reject the null hypothesis, with there being no statistically significant difference between the networks in these aspects. When evaluating weighted degrees we can reject the null hypothesis and accept an alternative hypothesis that there is a difference in distributions between the two networks in this aspect.

Finally, we want to evaluate the impact of external factors on the network. For this purpose we use MRQAP to apply regression analysis to evaluate the effect of external variables on dyads, or the connections between nodes. For the MRQAP analysis we used the R sna library [8] and its network multiple regression function.

In our first regression analysis we evaluate whether being in the same team predicts a connection in the classroom. The regression model using MRQAP and ordinal least squares regression statistically significantly predicts collaborative connections, $F(1, 304) = 38.38$, $p < .001$, adj. $R^2 = .11$. Regression coefficient estimate of 0.41 on the same team variable means that being in the same team increases the possibility of a connection by 41%. Cohen's f^2 for effect size is 0.13, which means a medium effect [9].

In the second regression analysis we evaluate whether being in the same team predicts a connection in the online platform. The regression model using MRQAP and ordinal least squares regression statistically significantly predicts collaborative connections, $F(1, 504) = 72.28$, $p < .001$, adj. $R^2 = .12$. Regression coefficient estimate of 0.56 on the same team variable means that being in the same team increases the possibility of a connection by 56%. Cohen's f^2 for effect size is 0.14 which means a medium effect [9].

Despite the medium effect size and statistical significance, in both cases the R^2 value is low. The same-team variable predicts only 11 to 12% of the variance, which means that a social connection is mainly predicted by other factors.

Discussing the Analysis Results

When evaluating the differences in the networks, we made the following findings: The online network from the second study was slightly larger than the classroom collaboration network in the first study. The online network had vastly more interactions, but the nodes had stronger ties in the classroom network in a statistically significant manner. In both networks, there was the same distribution of betweenness centralities, which means there were the same distribution of central nodes mediating communication. According to the MRQAP results, being in the same team was a stronger predictor for collaboration in the online network, which means that the classroom network had stronger inter-team collaboration. A summary of the tested hypotheses and outcomes are presented in Table 3.

At a first look the gamified online communication platform appears to increase collaborative communication between students. The three essential goals of the platform were to increase communication, enable stronger inter-team communication, and to reduce the hierarchy in communication. However, after testing our hypotheses with social network analysis data, we found that compared to plain classroom communication, the online platform failed to enable stronger social ties, increased inter-team communication, or more diverse patterns of collaboration. As the CSCL platform had an increased number of interactions, these outcomes might not have been discovered without in-depth social network analysis and statistical evaluation.

These discoveries were enabled by statistical hypothesis testing, which allowed us to evaluate the strength of the evidence in regard to statistical significance and effect size. For example, the beneficial differences in betweenness centralities might seem plausible when presented side by side as descriptive statistics. However, by statistically testing our hypotheses, we were able to conclude that the difference is not significant and the effect size was trivial.

DISCUSSION AND CONCLUSION

In this paper, we demonstrated how social network analysis can be used in hypothesis testing -based evaluation of socio-technical communities. In the demonstration we evaluated the impact of a new software artifact. The artifact at a glance appeared to increase communication between participants of the community, but when evaluating social ties, the artifact's impact was more complex than reflected by simple usage statistics, and not fully beneficial.

Variable	U-value	Mdn (Classroom)	Mdn (Online)	p-value	Adjusted p (Bonferroni) / significance	Effect size (r)
Degree	251.5	20.00	21.08	0.37	1 / no	0.13 (small)
Weighted degree (centrality)	20	748.00	108.97	0.00000067	0.000002019891 / yes	0.72 (large)
Betweenness centrality	218	7.24	6.35	0.97	1 / no	0.007 (none)

Table 2. Mann-Whitney U test results

Test	Hypothesis	Outcomes
Mann-Whitney U	The distribution of node attributes is different in the two datasets.	Node degrees and betweenness centralities have no significant differences. Weighed degrees have.
MRQAP	Being in the same team is a statistically significant predictor for the strength of connection between the nodes.	Statistically significant relationship exists between the variables. In the online case the relationship is somewhat stronger.

Table 3. Summary of tested hypotheses and outcomes

Social network analysis is a research method worth considering when studying the impact of technological artifacts on socio-technical communities, such as gamified systems. It is a powerful technique, especially when combined with hypothesis testing. It can reveal patterns of actions in complex, interlinked systems that involve multiple interdependent actors. The number of these socio-technical systems will only increase with the ongoing digital transformation in our society.

Some studies already use social network analysis to evaluate the impact of gamification on the educational communities, such as the line of research by de-Marcos et al. [11, 12]. However, in gamification, computer-supported collaboration, and socio-technical research this is still in the minority [3, 14, 21, 25]. Additionally, many social network analysis studies do not use hypothesis testing or regression analysis and rely just on descriptive statistics. We recommend publishing tutorials, reviews, and methodological demonstrations of the benefits of more in-depth and rigorous research approaches in social network analysis. Social network analysis has potential as a research method in the field of gamification and beyond. As future work, aspects of social network analysis not covered by this paper, such as community detection and more advanced approaches to centrality, should be demonstrated in regard to compatibility with hypothesis testing.

ACKNOWLEDGMENTS

The work of the first author was supported by the Ulla Tuominen Foundation. This study was partially funded by the European Union Regional Development Fund grant number A70554, “Kyberturvallisuusosaamisen ja liiketoiminnan kehittäminen”, administrated by the council of Kymenlaakso. This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

REFERENCES

1. Hervé Abdi. 2007. Bonferroni and Šidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics* 3 (2007), 103–107.
2. Ajith Abraham and Aboul Ella Hassanien. 2010. *Computational Social Network Analysis: Trends, Tools and Research Advances*. Springer.
3. Manal M Alhammad and Ana M Moreno. 2018. Gamification in software engineering education: a systematic mapping. *Journal of Systems and Software* (2018).
4. Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. In *International AAAI conference on weblogs and social media*, Vol. 2.
5. Reinhard Bergmann, John Ludbrook, and Will PJM Spooren. 2000. Different outcomes of the Wilcoxon—Mann—Whitney test from different statistics packages. *The American Statistician* 54, 1 (2000), 72–77.
6. Stephen P Borgatti, Martin G Everett, and Linton C Freeman. 2002. Ucinet for Windows: Software for social network analysis. (2002).
7. Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. 2009. Network analysis in the social sciences. *science* 323, 5916 (2009), 892–895.
8. Carter T. Butts. 2016. *sna: Tools for Social Network Analysis*. https://CRAN.R-project.org/package=sna R package version 2.4.
9. Jacob Cohen. 1992. A power primer. *Psychological bulletin* 112, 1 (1992), 155.
10. Hans De Bruijn and Paulien M Herder. 2009. System and actor perspectives on sociotechnical systems. *IEEE Transactions on systems, man, and cybernetics-part A: Systems and Humans* 39, 5 (2009), 981–992.

11. Luis de Marcos, Eva Garcia-Lopez, and Antonio Garcia-Cabot. 2016a. On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers & Education* 95 (2016), 99–113.

12. Luis de Marcos, Eva García-López, Antonio García-Cabot, José-Amelio Medina-Merodio, Adrián Domínguez, José-Javier Martínez-Herráiz, and Teresa Diez-Folledo. 2016b. Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance. *Computers in Human Behavior* 60 (2016), 312–321.

13. Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. ACM, New York, NY, USA, 9–15.

14. Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. Gamification in education: a systematic mapping study. *Journal of Educational Technology & Society* 18, 3 (2015), 75.

15. Pierre Dillenbourg. 1999. What do you mean by collaborative learning? *Collaborative-learning: Cognitive and computational approaches* 1 (1999), 1–15.

16. Catherine O Fritz, Peter E Morris, and Jennifer J Richler. 2012. Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General* 141, 1 (2012), 2.

17. Frank W Geels. 2004. From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research policy* 33, 6-7 (2004), 897–920.

18. Frank Harary. 1969. *Graph theory*. Addison-Wesley, Reading, MA.

19. Antti Knutas. 2016. *Increasing Beneficial Interactions in a Computer-Supported Collaborative Environment*. Number 718 in Acta Universitatis Lappeenrantaensis. Lappeenranta University of Technology.

20. Antti Knutas, Jouni Ikonen, Dario Maggiorini, Laura Ripamonti, and Jari Porras. 2016. Creating Student Interaction Profiles for Adaptive Collaboration Gamification Design. *International Journal of Human Capital and Information Technology Professionals (IJHCITP)* 7, 3 (2016), 47–62.

21. Antti Knutas, Jouni Ikonen, and Jari Porras. 2015. Computer-Supported Collaborative Learing in Software Engineering Education: A Systematic Mapping Study. *Journal on Information Technologies & Security* 7 (2015), 4.

22. David Krackardt. 1987. QAP partialling as a test of spuriousness. *Social networks* 9, 2 (1987), 171–186.

23. David Krackhardt. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks* 10, 4 (1988), 359–381.

24. Maarten de Laat, Vic Lally, Lasse Lipponen, and Robert-Jan Simons. 2007. Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning* 2, 1 (March 2007), 87–103.

25. Fiona Fui-Hoon Nah, Qing Zeng, Venkata Rajasekhar Telaprolu, Abhishek Padmanabhuni Ayyappa, and Brenda Eschenbrenner. 2014. Gamification of education: a review of literature. In *International Conference on HCI in Business*. Springer, 401–409.

26. Evelien Otte and Ronald Rousseau. 2002. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science* 28, 6 (Dec. 2002), 441–453.

27. J. Porras, K. Heikkinen, and J. Ikonen. 2007. Code camp: a setting for collaborative learning of programming. *Adv. Technol. Learn.* 4, 1 (Jan. 2007), 43–52.

28. R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

29. John Scott. 2012. *Social network analysis*. SAGE Publications Limited.

30. Gerry Stahl, Timothy Koschmann, and Dan Suthers. 2006. Computer-supported collaborative learning: An historical perspective. *Cambridge handbook of the learning sciences* 2006 (2006), 409–426.

31. Alyssa Friend Wise and Yi Cui. 2018. Learning communities in the crowd: Characteristics of content related interactions and social relationships in MOOC discussion forums. *Computers & Education* (2018).

32. Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in software engineering*. Springer Science & Business Media.