What Makes a Good Developer? An Empirical Study of Developers' Technical and Social Competencies

Cheng Zhou
Tandy School of Computer Science
University of Tulsa
Tulsa, Oklahoma
cheng-zhou-73@utulsa.edu

Sandeep Kaur Kuttal
Tandy School of Computer Science
University of Tulsa
Tulsa, Oklahoma
sandeep-kuttal@utulsa.edu

Iftekhar Ahmed
School of Elect. Eng. & Computer Science
Oregon State University
Corvallis, Oregon
ahmedi@oregonstate.edu

***Abstract*—Technical and social competencies are highly desirable for a protean developer. Managers make hiring decisions based on developer's contributions to online peer production sites like GitHub and Stack Overflow. These sites provide ample history regarding developers' technical and social skills. Although these histories are utilized by hiring tools to help managers make their hiring decisions, little is known empirically how developers' social skills affect their technical skills and vice versa. Without such knowledge, tools, research, and training might be flawed.**

We present an in-depth empirical study investigating the correlation between the technical and social skills of developers. Our quantitative analysis of factors influencing the social skills of developers compared with factors affecting their technical skills indicates that better collaboration competency skills are associated with enhanced coding abilities as well as the quality of code.

I. Introduction

Technical and social competencies are vital for a successful developer. Developers are using online peer production sites like GitHub for software development and Stack Overflow for learning, which provide ample histories regarding developers' social and technical activities. These are used as a proxy for measuring their social and technical competencies [3], [4], [8]. Managers are using these proxies to assess a potential candidates for hiring in their teams or companies [4]–[9].

Technical Skills are vital for writing code. The two most important skills revealed in the literature are coding competency and quality of work. *Coding ability*: How proficient is an individual's knowledge and ability to code? On a global platform, these skills can be measured by the log activities, number of projects owned or forked, number and frequency of commits/issues/comments and number of languages the professional is proficient in [1], [7]. *Quality of work*: How good is the code that an individual produces? It can be measured by number of accepted commits and inclusion of test cases [5], [6], [9], [10].

Social Skills are soft skills that measure the ability to work as an individual and in teams. Three important skills are collaboration proficiency, project management ability, and motivation. *Collaboration proficiency*: How well can an individual work with other team members? This is measured by communication activity through the number of comments/answers/questions and reputation. Good team players are vital for the success and timely release of large projects [2]. *Project management ability*: How well can an individual manage the project? This can be measured by the number of projects owned by an individual [6]. *Motivation*: How passionate is an individual about the project? These can be measured by number of the commits/issues/comments of the contributions, non-related side projects, and diversity of languages known [9].

There exists little knowledge about which of the technical or social skills are important and what correlation exists between them. This is the basis of our study.

978-1-5386-4235-1/18/$31.00 © 2018 IEEE

II. Methodology

We used data dumps, GHTorrent [11]–[13] and Stack Exchange [14]. To find common active users, we selected users who provide their GitHub link on their Stack Overflow profiles, and we filtered out those who were not active contributors on GitHub by established criterias [15]. First, we removed the projects from our analysis that didn't have language information in the GHTorrent database. Declaring the languages used in a project is a part of the initial setup of a project in GitHub, and missing information in this field raises concerns about the validity of the project. In order to make sure that our results are free from such noise, we filtered out those projects. Secondly, to avoid personal projects, we set a standard that projects should have at least five committers. We found 467,770 GitHub projects from 12,831 common users (on GitHub and Stack Overflow), and after implementing criterias, we were left with 3,266 projects and 1,749 users. We retrieved the data from Stack Exchange for all 1,749 users in Stack Overflow and had 221,219 comments, 19,635 questions, and 90,795 answers.

III. Results

The goal of this paper is to investigate which technical skills or social skills are important when it comes to measuring

TABLE I: Linear model with coefficients

	Dependent Variable	Quality Inferred	# of Answers	Reputation score	# of Questions	# of Contributed projects	McFadden Pseudo R-squared value
RQ1	# of Project owned	Coding ability	3.93e-04	-8.44e-07	1.25e-03	1.34e-01	0.13
	# of Commits	Coding ability	7.78e-04	-2.57e-06	8.72e-04	1.20e-01	0.26
	# of Issues	Coding ability	2.30e-03	-2.95e-06	2.30e-03	8.08e-02	0.27
	# of Comments	Coding ability	1.12e-03	3.46e-06	-1.62e-03	1.50e-01	0.20
	Languages used	Coding ability	8.68e-04	-1.52e-06	null	5.61e-02	0.01
RQ2	# of Accepted commits	Quality of work	1.28e-03	1.52e-06	-2.95e-03	1.49e-01	0.17
	Test case inclusion	Quality of work	3.59e-03	-1.52e-05	3.61e-03	1.36e-01	0.12

the competency of a developer. To answer this overarching question, we analyzed the correlation between various competency measures, and built models using various factors to understand how effective these factors are in explaining technical and social skills. Hence, we targeted two research questions presented here.

RQ1: Which is the most important factor among social skills in relation to Coding ability - a technical skill?

We attempted to identify whether *motivation*, *project management ability*, or *collaboration proficiency* was the most effective factor for coding ability. In order to answer this question, we first computed the pearson correlation coefficients for all the factors. As visible in Figure 1, none of the factors are highly associated with each other.

Next, we built Poisson regression models using all of the coding ability indicators, such as # of Project owned, # of commits etc. with a log linking function and filtered the factors with VIF>5. The significant contributors towards individual's *coding ability* are shown in RQ1 section of Table I. The McFadden Pseudo R-squared [16] for the models are shown in RQ1 section of Table I. We used McFadden's Pseudo R-squared as a quality indicator of the model because there is no direct equivalent of R-squared for Poisson regression. The ordinary least square (OLS) regression approach to goodness-of-fit does not apply for Poisson regression. Moreover, pseudo R-squared values like McFadden's cannot be interpreted as one would interpret OLS R-squared values. McFadden's Pseudo R-squared values tend to be considerably lower than those of the R-squared and values of 0.2 to 0.4 represent an excellent fit.

Next, we wanted to check the kind of quality inferred (discussed in the introduction) by these factors. From RQ1 section of Table II, we can see that factors associated with *collaboration proficiency* are most frequently identified as significant when we try to build models to predict coding abilities of a contributor.

RQ2: Which is the most important factor among social skills for Quality of work - a technical skill?

Our second research question attempted to identify whether *motivation*, *project management ability*, or *collaboration proficiency* was the most important factor in determining the *quality of work*. We followed the same procedure of building Poisson regression models using all of the *quality of work* indicators with a log linking function, shown in Table I.

Then we looked into the category of the factors based on the quality inferred to discover the most frequent category as shown in Table II. *Collaboration proficiency* is the most common factors that is associated with the quality of work.

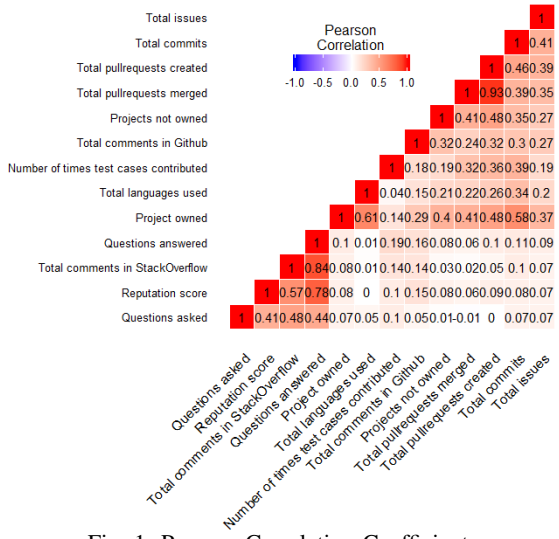

Fig. 1: Pearson Correlation Coefficients

TABLE II: Factor wise frequency along with their categories

	Factors	Freq.	Quality Inferred
RQ1	# of Answers	5	Collaboration proficiency
	Reputation score	5	Collaboration proficiency
	# of Contributed projects	6	Motivation
	# of Questions	4	Collaboration proficiency
	# of Languages	3	Motivation
	# of Projects owned	2	Project management ability
	# of Accepted commits	2	Quality of work
	# of comments	2	Collaboration proficiency
RQ2	# of Answers	2	Collaboration proficiency
	Reputation score	2	Collaboration proficiency
	# of Contributed Projects	2	Motivation
	# of Questions	2	Collaboration proficiency
	# of Languages	2	Motivation
	# of Projects owned	2	Project management ability
	# of Comments	2	Collaboration proficiency

IV. Conclusion

In our large scale study, we find that collaboration proficiency is the most frequently identified competency category, and there is a lack of strong association between technical and social skills. The results reaffirm that collaboration is an important factor while developing large software, but there is a lack of strong association between technical and social competency. This opens up an opportunity to identify the reason behind such lack of association and also instigates the need for longitudinal studies to investigate the association over time.

REFERENCES

[1] Al-Ani, Ban, Matthew J. Bietz, Yi Wang, Erik Trainer, Benjamin Koehne, Sabrina Marczak, David Redmiles, and Rafael Prikladnicki. "Globally distributed system developers: their trust expectations and processes." In conference on Computer supported cooperative work. ACM, 2013.

[2] Al-Ani, Ban, and David Redmiles. "In strangers we trust? Findings of an empirical study of distributed teams." In International Conference on Global Software Engineering. IEEE, 2009.

[3] Kristof-Brown, Amy, Murray R. Barrick, and Melinda Franke. "Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity." In Journal of Management 28.1 (2002): 27-46.

[4] Long, Ju. "Open Source Software Development Experiences on the Students' Resumes: Do They Count?-Insights from the Employers' Perspectives." In Journal of Information Technology Education: Research 8 (2009): 229-242.

[5] Movshovitz-Attias, Dana, et al. "Analysis of the reputation system and user contributions on a question answering website: Stackoverflow." In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, 2013.

[6] Marlow, Jennifer, and Laura Dabbish. "Activity traces and signals in software developer recruitment and hiring." In conference on Computer supported cooperative work. ACM, 2013.

[7] Marlow, Jennifer, Laura Dabbish, and Jim Herbsleb. "Impression formation in online peer production: activity traces and personal profiles in github." In conference on Computer supported cooperative work. ACM, 2013.

[8] Sarma, Anita, et al. "Hiring in the global stage: Profiles of online contributions." In Conference on Global Software Engineering. IEEE, 2016.

[9] Singer, Leif, et al. "Mutual assessment in the social programmer ecosystem: an empirical investigation of developer profile aggregators." In conference on Computer supported cooperative work. ACM, 2013.

[10] Tsay, Jason, Laura Dabbish, and James Herbsleb. "Influence of social and technical factors for evaluating contribution in GitHub." In international conference on Software engineering. ACM, 2014.

[11] Ghtorrent, http://ghtorrent.org, accessed: Nov 2016.

[12] Gousios, Georgios, and Diomidis Spinellis. "GHTorrent: GitHub's data from a firehose." In conference on Mining software repositories. IEEE, 2012.

[13] Gousios, Georgios. "The GHTorent dataset and tool suite." In conference on mining software repositories. IEEE, 2013.

[14] Stack exchange data explorer, https://data.stackexchange.com, accessed: Feb 2018.

[15] Kalliamvakou, Eirini, Georgios Gousios, Kelly Blincoe, Leif Singer, Daniel M. German, and Daniela Damian. "The promises and perils of mining GitHub." In conference on mining software repositories. ACM, 2014.

[16] Hensher, David A., and Peter R. Stopher, eds. Behavioural travel modelling. Taylor & Francis, 1979.