Exploring the Ecosystem of Software Developers on GitHub and Other Platforms

Yu Wu
Penn State University
316C IST Building
University Park, PA 16802 USA
yuw132@ist.psu.edu

Patrick C. Shih
Penn State University
313 IST Building
University Park, PA 16802
patshih@ist.psu.edu

Jessica Kropczynski
Penn State University
316A IST Building
University Park, PA 16802
jnk14@ist.psu.edu

John M. Carroll
Penn State University
307H IST Building
University Park, PA 16802
jcarroll@ist.psu.edu

CSCW'14 Companion, February 15-19, 2014, Baltimore, MD, USA.
ACM 978-1-4503-2541-7/14/02.
http://dx.doi.org/10.1145/2556420.2556483

Abstract
GitHub provides various social features for developers to collaborate with others. Those features are important for developers to coordinate their work (Dabbish et al., 2012; Marlow et al., 2013). We hypothesized that the social system of GitHub users was bound by system interactions such that contributing to similar code repositories would lead to users following one another on GitHub or vice versa. Using a quadratic assignment procedure (QAP) correlation, however, only a weak correlation among followship and production activities (code, issue, and wiki contributions) was found. Survey with GitHub users revealed an ecosystem on the Internet for software developers, which includes many platforms, such as Forrst, Twitter, and Hacker News, among others. Developers make social introductions and other interactions on these platforms and engage with one anther on GitHub. Due to these preliminary findings, we describe GitHub as a part of a larger ecosystem of developer interactions.

Author Keywords
GitHub; Ecosystem; Follow; Social connection

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Computer-Supported Cooperative Work

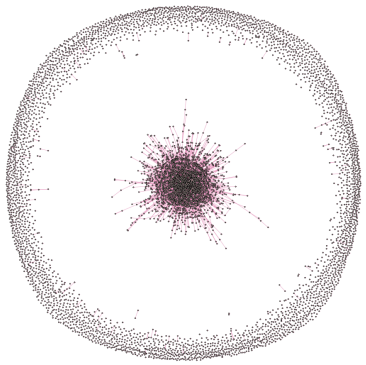

Figure 1. Homebrew Developer Sociogram of Followship

Introduction

GitHub is the primary online work production platform for developers to host their own open source projects and contribute to others. It provides various features, such as user profile, follow function, and activity traces, with which developers can increase awareness of other users' activities. These features are also used to form impressions of others (Marlow et al., 2013), and coordinate work (Dabbish et al., 2012).

The follow feature on social media sites is considered to serve social functions. Previous studies on Facebook and Twitter report that follow or add others as friend express establishment of social ties or social relations (Ellison et al., 2007; Java et al., 2007). Hanna, Rohm, and Crittenden (2011) stated that the power of the social media ecosystem is that "we are all connected." As GitHub also provides follow functions to maintain connections with other developers, how developers follow others and to what extend the follow on GitHub is similar or different to social media functions has not been explored. Do developers make use of GitHub followship to connect to other users that they interact with on the GitHub platform? Or are there alternative explanations for developers' followship and productive connections with each other?

Through large-scale analysis on GitHub timeline data and preliminary survey of GitHub users, we found out that although follow on GitHub indeed correlates with developers' work relations and activities, the effect is small. Developers make use of a diverse set of platforms, such as Twitter and Hacker News, to present their work and make social connections with others, which often result in followship or collaboration on GitHub. Our preliminary results suggest the existence of an ecosystem, which includes many social and production sites, for developers on the Internet.

GitHub Timeline Data Analysis Results

In order to examine the social and work relations among developers on GitHub, we collected one year of timeline data (from 3/11/2012 to 3/10/2013) from GitHub Archive, which contains 51,439,961 activity records. We chose a highly active repository, homebrew, to examine developers' interactions inside and outside of this project. 6,554 developers contributed 59, 270 activities to homebrew, and 1,779,362 activities in total within the time scope. First, we examined the connectedness of users followship relations within this repository. Figure 1 shows the social network of followership within homebrew. It is clear from the network that there is a connected core and a number of unconnected outliers.

While Figure 1 displays a network of followship, other GitHub features can also be thought of as networked relations such as co-coding contributions, co-issue contributions, and co-wiki contributions. A quadratic assignment procedure (QAP) correlation was then used to examine connectedness of GitHub users among relations (followship, code, issue, wiki). QAP correlation generates Pearson correlation between network types of the same nodes in a network.

The Pearson correlations in Table 1 show that relationships between follow and other production activities are extremely low. By moving towards the center of the homebrew community in terms of the activeness of developers, the correlations among these matrices become higher. However, in general, the correlations are not high enough for us to conclude that

"Storing source code, hosting web pages, finding other people's code" – Software Engineer

"I follow other people (on GitHub) who I recognize from my other networks who I converse with… I may appreciate a users feedback on Forrst, and if they have a GitHub, I may just follow them to see what they are up to." – Web Developer

"They only way to I can talk to other dev on GH (GitHub) is through the comment system on GH issue." – Web Developer

"I interact with other developers much more on Twitter than I do on GitHub." – Software Engineer

follow is correlated with production activities, such as code, issue, and wiki. The table shows the highest correlation among code and issue contributions, which is to be expected given the nature of these activities.

	Full repo (n=6554)	**Inner Community (n=1812)**	**Core (n=65)**
Follow and Code	0.010	0.052	0.175
Follow and Issue	0.006	0.041	0.173
Follow and Wiki	0.008	0.018	-0.004
Code and Issue	0.003	0.331	0.530
Code and Wiki	0.010	0.033	0.020
Wiki and Issue	-0.002	0.040	0.086

All values are significant at p < 0.01

Table 1. Matrices QAP correlations for homebrew repository

Developer Preliminary Survey Results

The results from the data analysis of GitHub timeline data that there is a weak correlation between follow and developers' production activities, did not explain followship connections. Therefore, developers follow each other not just because they have working relationships on GitHub. However, as indicated by Figure 1, developers indeed follow each other in the core of a repository. Other than GitHub productive activities, how do users find the developers that they follow, if not through direct co-participation in productive activities?

In order to answer this question, we have begun preliminary survey research with GitHub users (n = 5). We asked GitHub developers about their motivations for using GitHub, whom do they follow on GitHub, and where do they interact with other developers. Most respondents are senior software developers and have years of field experiences. Our preliminary analysis of feedback shows the following insights.

GitHub as a Professional Project Hosting and Collaborating Site

Most respondents regard GitHub primarily as a professional platform where they can host their own projects and find other interesting projects. Developers indeed use GitHub to coordinate their collaborations on software projects and to be aware what their peers are doing. Specifically, they expressed that they consider follow as an awareness tool to be informed of "new, cool" projects. However, users also consider GitHub to be limited in social functions. Developers deem GitHub lack ways for them to interact with other developers and comment on each other's project.

Twitter, Hacker News, … as Social Platforms

Respondents' answers show that they make social connections with others on platforms such as Twitter and Hacker News because they are considered to be superior in terms of social functions. In particular, users mentioned that they use Twitter to learn about ideas, and converse with others. The main function of developers' use of Hacker News is to be aware of the trending projects. Users turn to Quora and StackOverflow to ask and answer technical questions. Also, they use other platforms, like Forrst and JS Weekly, to brand their projects on GitHub and receive feedback, or comment on others' projects. All

developers mentioned that their interactions on these platforms often result in their followship or even collaboration on GitHub.

Discussion

According to our findings from GitHub timeline data analysis and developer survey, most of follow activities on GitHub are not the product of developers' collaborations on GitHub, nor do they lead to work relationships. Rather, follow is the result of their interactions outside GitHub – on other social platforms such as Twitter, Hacker News, StackOverflow, etc.

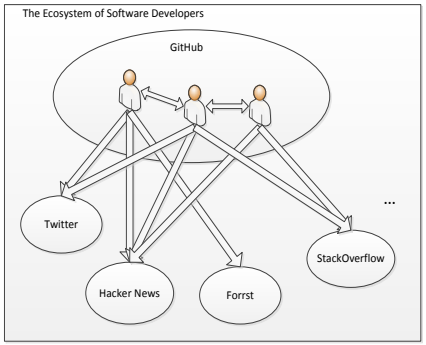

Figure 2. The Ecosystem of Software Developers

GitHub is limited for developers to make social connections. It is more regarded as a legitimated professional site for hosting projects, coordinating and managing software projects, and being aware of what other peers are doing. Social platforms such as Twitter, Hack News, etc., provide rich functions and contexts for developers to learn about new ideas and converse with others, which potentially result in their follow and work relations on GitHub.

Future Work

Based on our current analysis and findings, we believe that there is an ecosystem for software developers on developers to host their own projects and collaborate with others, while other social platforms, such as Twitter, Hacker News, etc., provide context and environment for developers to make social connections (Figure 2).

Our future work includes identifying all platforms that belongs to this large ecosystem and how developers make use of each, and leveraging these platforms to increase developers' performance and effectiveness.

Although on GitHub, follow is not strongly correlated with production activities, our data also suggests that follow is more likely to correlate with production activities for the community center of homebrew. The differences indicate the different functions of follow used by developers on GitHub. Therefore, our next step also includes to expanding this analysis to other software projects on GitHub and understand the how developers use follow differently.

References

[1] Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in github: transparency and collaboration in an open software repository. *Proc. CSCW 2012* (pp. 1277-1286). ACM.

[2] Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *J COMPUT-MEDIAT COMM*, 12(4), 1143-1168.

[3] GitHub Archive. Retrieved May 5, 2013. http://www.githubarchive.org/

[4] Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. *Proc. 9th WebKDD and 1st SNA-KDD 2007* (pp. 56-65). ACM

[5] Hanna, R. Rohm, A., & Crittenden, V. L. (2011). We are all connected: The power of the social media ecosystem. *Business Horizons*, 54(3), 265-273.

[6] Marlow, J., Dabbish, L., & Herbsleb, J. (2013, February). Impression formation in online peer production: activity traces and personal profiles in github. *Proc. CSCW 2013* (pp. 117-128). ACM.