Social Media and Success in Open Source Projects

Jason Tsay
School of Computer Science and Center on the Future of Work, Heinz College
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
jtsay@cs.cmu.edu

James Herbsleb
School of Computer Science and Center on the Future of Work, Heinz College
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
jdh@cs.cmu.edu

Laura Dabbish
School of Computer Science and Center on the Future of Work, Heinz College
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
dabbish@cs.cmu.edu

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.
ACM 978-1-4503-1051-2/12/02.

Abstract
Social media are being integrated into work environments. They have the potential to provide essential context and awareness, and increase work performance as a result. However, the specific effects of social media that impact productivity are not well understood. We perform a quantitative analysis of project success of over 5,000 open source software projects hosted on GitHub, a website that provides extensive social media functionality. Adapted from the open source literature, we develop two measures of project success, Developer Attention and Work Contribution. We find that projects with highly socially connected developers are not necessarily the most active or popular projects. Oddly, projects with a high level of developer multitasking, i.e., splitting effort equally across multiple projects, tend to receive less Developer Attention, but greater Work Contribution. Success on both measures is strongly positively associated with greater concentration of work among a small number of developers. We discuss the implications of the findings for social media in online production.

Author Keywords
Social media; open source; software development; success

	Developer Attention Model II	*Work Contribu-tion Model IV*
Variable	**IRR**	**IRR**
Followers	1.003***	1.001***
Repositories	1.003	0.984***
Gists	0.999	1.001
Contributions	1.000	-----
Contributed repos.	0.984***	1.001
Task focus	2.136***	0.536***
Work concentration	194.39***	47.644***
Deviance explained	37.56%	23.98%

Table 1. Incidence rate ratios (IRR) and significance of negative binomial models for Developer Attention and Work Contribution. IRR > 1 indicates a positive association, <1 indicates an inverse association. ***P< .001.

ACM Classification Keywords
D.2.6 Programming Environments: Integrated Environments; H.4.3 Communications Applications.

General Terms
Experimentation, Human Factors.

Introduction

Explosively popular social media enable users to create new ties, maintain relationships, and facilitate discussion [1]. Much like email and instant messaging before, enterprises, non-profits and even governments are increasingly trying to find ways to leverage social media to accomplish work [9]. Despite the popularity for personal use [1] and application in a few corporate settings [4], we know very little about social media's impact on how work gets done. Social media in the workplace carries the promise of facilitating communication and awareness, reducing the well-known difficulties of distributed teams [3, 5].

We seek to further our understanding of what leads to software project success in a rich social media environment. We present an analysis of the success of open source software projects as a function of the social and technical characteristics of the developers that contribute to them, as well as the distribution of their effort and attention. We use data from GitHub which is a project hosting site that offers social networking features found in well-known social networking sites such as Facebook and Twitter.

Building on measures used in the published literature, we introduce project success measures that reflect Developer Attention and Work Contribution for a software project. Then we use social characteristics of project contributors and patterns of contribution to predict success of the project along these two dimensions. Our results suggest that project success is not directly related to usage of social media. Rather, projects with contributors who focus on relatively few projects tend to garner more Developer Attention, but less Work Contribution; concentration of work in a small number of individuals is the predictor most strongly related to both types of success.

Method

Development of Work Contribution and Developer Attention, our measures of success, was guided by literature on project success measures in online communities [6], communities of practice [8], and open source software projects [2], and made use of a principal components analysis of six candidate measures. Developer Attention serves as an indicator of the interest software developers have in the project. Work Contribution measures the community's effort toward advancing the functionality of the product.

Using these measures of project success, we use social and technical characteristics of the project's contributors to predict both Developer Attention and Work Contribution of a project. Because of space constraints, the results for control variables are not shown. As seen in Table 1, many of the project contributor characteristics used in the analysis are measures of usage of the different social networking features in GitHub, such as number of Followers, or measures of work performed, such as number of projects owned and projects contributed to.

Some of the characteristics used are aggregate measures with both technical and social components. One of them, Task Focus, measures the extent to which developers on a project focus on a small number of projects. For example, a project with a task focus measure of 1.0 has developers who only contributed to one repository, while a project with a task focus measure of 0.25 means, on average, its developers contribute fairly equally to multiple repositories. Work Concentration, another aggregate measure, is a repository-level measure of how unequally contributions are distributed across developers in a project. For example, a Work Concentration measure of 1 means that one contributor has done all of the work on a project and 0 means that each contributor to the project made the exact same proportion of contributions as the rest of the contributors to the same project.

Results

We created separate negative binomial regression models for the Developer Attention and the Work Contribution dimensions of project. For both measures of success, only the aggregate measures of Task Focus and Work Concentration resulted in substantial and significant associations with project success.

Developer Attention

We found that projects with developers who focus their efforts on a small number of repositories results in a substantial and highly significant association with higher Developer Attention. In other words, Task Focus has a positive relationship with Developer Attention. The strongest association with Developer Attention, however, is Work Concentration. The results show that the more unevenly work is distributed among developers in a project, the more Developer Attention the project receives.

Work Contribution

In direct contrast to the result for Developer Attention, the more developers tended to focus on a small number of projects, the less Work Contribution the project had. However, as with Developer Attention, Work Concentration had by far and away the strongest association with project Work Contribution. Projects where effort was centralized (high Work Concentration) with a few developers rather than spread more equally across developers tended to receive much more Work Contribution than those that distributed the work more evenly.

Discussion and Future Work

We find little evidence that usage of specific social media features in GitHub directly correlates with software project success, however we speculate that these features mediate the observed effects. For example, high Task Focus – reflecting that the developers focus primarily on this project – may indicate a high degree of specialization is on display, which other developers may wish to learn from. Low Task Focus could indicate less specialization, meaning the project is easier for many people to contribute to, hence higher work concentration. Other explanations are possible, of course.

Work Distribution had a very substantial and significant positive relationship with both measures of project success. A similar finding was reported by Kittur and Kraut [7], who found that when work is concentrated in a few editors, the quality of Wikipedia articles was higher. They interpreted this result as an indicator that

high concentration of effort across individuals was a form of implicit coordination, reducing process overhead and the need for explicit coordination through communication media. This may explain why Work Concentration is associated with more Work Contribution, but there is no obvious connection with Developer Attention, since simply attending to a software project requires no coordination overhead. This research raises many questions about the extent to which these findings are mediated by coordination needs (e.g., [7]), or reflect incentives, perceived team boundaries, or other factors. We have much to learn about how developers perceive all of the visible artifacts, activities, and communication, how they draw inferences from them, and how they use what they infer in their daily activities.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant No. IIS1111750.

References

[1] Boyd, D. and Ellison, N. Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication*, 13, 1 (2008), 210-230.

[2] Crowston, K., Howison, J. and Annabi, H. Information systems success in free and open source software development: Theory and measures. *Software Process: Improvement and Practice*, 11, 2 (2006), 123-148.

[3] Cummings, J. N. and Kiesler, S. Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 36, 10 (2007), 1620-1634.

[4] DiMicco, J., Millen, D., Geyer, W., Dugan, C., Brownholtz, B. and Muller, M. (2008). Motivations for social networking at work. In Proceedings*, Computer supported cooperative work,* San Diego, CA, pp. 711-720.

[5] Herbsleb, J. D. and Mockus, A. An Empirical Study of Speed and Communication in Globally-Distributed Software Development. *IEEE Transactions on Software Engineering*, 29, 3 (2003), 1-14.

[6] Iriberri, A. and Leroy, G. A life-cycle perspective on online community success. *ACM Computing Surveys*, 41, 2 (2009), 1-29.

[7] Kittur, A. and Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In Proceedings*, Computer-Supported Cooperative Work,* San Diego, CA, pp. 37-46.

[8] Probst, G. and Borzillo, S. Why communities of practice succeed and why they fail. *European Management Journal*, 26, 5 (2008), 335-347.

[9] Skeels, M. and Grudin, J. (2009). When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In Proceedings*, Conference on Supporting Group Work,* Sanibel Island, FL, pp. 95-10.