

Need for Tweet: How Open Source Developers Talk About Their GitHub Work on Twitter

Hongbo Fang, Daniel Klug, Hemank Lamba, James Herbsleb, Bogdan Vasilescu
Carnegie Mellon University

ABSTRACT

Social media, especially Twitter, has always been a part of the professional lives of software developers, with prior work reporting on a diversity of usage scenarios, including sharing information, staying current, and promoting one's work. However, previous studies of Twitter use by software developers typically lack information about activities of the study subjects (and their outcomes) on other platforms. To enable such future research, in this paper we propose a computational approach to cross-link users across Twitter and GitHub, revealing (at least) 70,427 users active on both. As a preliminary analysis of this dataset, we report on a case study of 786 tweets by open-source developers about GitHub work, combining automatic characterization of tweet authors in terms of their relationship to the GitHub items linked in their tweets with qualitative analysis of the tweet contents. We find that different developer roles tend to have different tweeting behaviors, with repository owners being perhaps the most distinctive group compared to other project contributors and followers. We also note a sizeable group of people who follow others on GitHub and tweet about these people's work, but do not otherwise contribute to those open-source projects. Our results and public dataset open up multiple future research directions.

ACM Reference Format:

Hongbo Fang, Daniel Klug, Hemank Lamba, James Herbsleb, Bogdan Vasilescu. 2020. Need for Tweet: How Open Source Developers Talk About Their GitHub Work on Twitter. In *17th International Conference on Mining Software Repositories (MSR '20), October 5–6, 2020, Seoul, Republic of Korea*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3379597.3387466>

1 INTRODUCTION

General social media platforms like Twitter, Facebook, and LinkedIn are impacting the professional lives of many software developers, facilitating communication and coordination, learning and knowledge sharing, or recruiting and hiring [3, 4, 16, 17], just name a few.

Among these platforms, Twitter is especially popular with software developers [4] and actively studied by software engineering researchers (Sec. 2). Prior work has been exploring, e.g., how developers use Twitter and what they talk about [15, 18], and has been theorizing about Twitter's effects on individual developers or software development practices and outcomes.

Still missing, however, are studies testing, refining, extending,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MSR '20, October 5–6, 2020, Seoul, Republic of Korea

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7517-7/20/05...\$15.00

<https://doi.org/10.1145/3379597.3387466>

and validating such theory, especially when it comes to cross-platform effects such as the impact of Twitter activity on software project outcomes. Despite a widespread recognition that activities on open source platforms like GitHub are only a part of the overall software development ecosystem [15, 22] and that open source developers are often active on multiple platforms simultaneously [16, 20], there is a paucity of research studying the same populations of developers *across* platforms.

To enable and encourage such future research, in this paper we propose a computational approach (Sec. 3) to cross-link users on Twitter and GitHub, the dominant platform for hosting open-source development, revealing (at least) 70,427 users with accounts on both. Using this dataset, we then report on a case study of 786 tweets by open-source developers about GitHub work (Sec. 4). Specifically, we analyze tweets containing links to GitHub repositories, combining automatic characterization of tweet authors in terms of their relationship to the GitHub items linked in their tweets with qualitative analysis of the tweet contents.

We find (Sec. 5) that different developer roles tend to have different tweeting patterns when including GitHub links in their tweets. For example, repository owners seem to engage less in work discussions or answering others' questions, but do share information and updates, and promote their GitHub projects. Interestingly, people who follow others on GitHub also tweet about these people's work, but do not otherwise contribute to those open-source projects. We discuss implications of these results in Sec. 6.

2 RELATED WORK

Social Media and Software Engineering. The emergence of social media has generated considerable attention from software engineering researchers, who studied its broad impact on software development practices. As early as 2010, a survey by Black et al. [4] found that almost all respondents use social media to communicate with their colleagues. Further studies have shown that different social media are used for different purposes and have various impacts, from improving the quality of communication, to staying aware of the status of other developers [3, 5, 17].

The use of micro-blog services like Twitter in software development has also drawn some attention from researchers. Early empirical results show that developers differ from the general public in terms of their use of URLs, @-mentions, and other features while tweeting [7, 19]; and those characteristics vary across software communities [7]. Sharma et al. [12] found that information sharing and online conversation are the two major categories of tweets by developers [12, 18]. Yasir et al. [24] found that software-related topics like discussions on features, and community-related topics like informing about an upcoming event, are the two major tweet topics among Python core contributors. Singer et al. [15] found that Twitter can help build communities and help individual

developers maintain professional relationships and build trust.

However, studies linking activity on one platform to outcomes on another are rare. Yang et al. [23] found that the average number of followers of micro-bloggers in a project and the project developers' access to network structural holes are positively related to the number of project commits and downloads. Borges and Valente [6] found that popular open-source projects are more likely to use Twitter as a promotion channel, and hypothesized that such promotion may increase project adoption and popularity. None of these studies considered possible connections between software project role and social media posting patterns, which we address in this work.

Merging Identities. Our work also relates to the literature on linking multiple accounts associated with a single person, a known problem in the mining software repositories community [1, 2, 9, 13, 14]. Prior work used a diversity of approaches, ranging from heuristics based on names, emails, and other features [14, 21], to machine learning [1] and information retrieval [9]. The approach we propose here draws inspiration from all of these.

3 CROSS-LINKING GITHUB AND TWITTER

This section presents our heuristic approach to cross-link GitHub and Twitter user accounts, based on mining GitHub user profile pages and personal blogs for explicit URLs pointing to Twitter.

Heuristics. Typically, identity merging (aka. de-aliasing) approaches for user accounts in software repositories are based on heuristics and involve a precision-recall trade-off. For example, linking two accounts if they share the same *email address* tends to be a highly precise heuristic (valid email addresses are typically individual), but with relatively low recall (many people have multiple email addresses and they would not be linked this way). Matching people based on *names* is expected to have lower precision (common names are shared by many different people), but relatively higher recall (more accounts would get merged).

Facing the same precision-recall trade-off, we developed two heuristics expected to have relatively high precision: (1) mining explicit links to Twitter accounts from GitHub user profile pages; (2) crawling personal websites linked from GitHub user profile pages and mining links to Twitter accounts therein. The closest prior approach to ours is that of Silvestri et al. [14]. The two are complementary and could be combined to increase coverage: we crawl personal websites listed on GitHub profiles, while Silvestri et al. don't; in contrast, Silvestri et al. use heuristics based on full names to identify additional links, while we don't. The goal of this short paper is not to be exhaustive, nor to systematically compare many different heuristics, though we believe these to be worthwhile directions for future work. Rather, we aim to provide a large initial dataset to catalyze research in this promising area.

Implementation. Using a GHTORRENT [8] MySQL dump, we compile a list of all GitHub user logins (excluding organizational and deleted accounts), then use the GitHub API to mine, for each user, their profile page *blog* URL field (which is not part of MySQL GHTORRENT dump).

If the target of the URL is a Twitter profile, then we assume the linked Twitter account and the GitHub account belong to the same person and we record this account pair. If the URL points

to a non-Twitter page, we crawl the content of that website for URLs pointing to Twitter; if the page only contains one such URL pointing to a Twitter profile, we assume the linked Twitter account and the GitHub account belong to the same person and we record this pair; otherwise, we don't record any link.

This way we identify 72,668 GitHub-Twitter user account pairs.

Validation. We take several steps to filter out obvious false positives, and to estimate and increase the validity of our data.

First, we filter out those pairs where multiple GitHub accounts are linked to the same Twitter account, leaving 70,427 pairs total. Our manual inspection of a sample of such pairs reveals two common scenarios: (i) people linking to organizational or business Twitter accounts (e.g., multiple developers working for the same company) that are not explicitly recorded as such on Twitter; (ii) people linking to celebrity accounts.

Second, we randomly select 50 GitHub-Twitter pairs and manually review them, relying on profile images and public profile information on both platforms to determine the validity of the link. Among the 50 pairs, three of them have either their Twitter or GitHub accounts currently inaccessible and could not be verified. Of the 47 remaining pairs, 40 of them are identified as valid (85 % accuracy). Of the remaining seven, six users (13 %) link to an organizational Twitter account that is not explicitly recorded as such on Twitter, and the other one (2 %) links to a Twitter account which likely belongs to another individual and has been incorrectly linked. We conclude that the overall accuracy of our approach is high.

Dataset. Our final dataset, consisting of 70,427 GitHub-Twitter user ID pairs, is available online [DOI 10.5281/zenodo.3711629](https://doi.org/10.5281/zenodo.3711629).

4 CASE STUDY: HOW GITHUB DEVELOPERS TALK ABOUT THEIR WORK ON TWITTER

As a preliminary analysis of our dataset, we perform an exploratory case study of the content of a sample of tweets by GitHub users in our dataset, that contain links to GitHub artifacts. As a key contribution compared to prior work, we use the cross-links between the two platforms to *automatically characterize the relationship of the tweet authors to the GitHub artifacts their tweets point to*, and use this information as part of our analysis.

We purposefully focus on tweets containing links to GitHub since these are more likely to be work-related as opposed to social and, therefore, among the most relevant content on which to study the impact of Twitter activity on software development outcomes.

Sample Selection. To create our case study sample, we first collected all the tweets returned by the Twitter API for each of the 70,427 users in our dataset (max 3,200 most recent per person), and kept the recent ones, posted between January 1, 2018 and July 1, 2019. Among these tweets, we then identified those with URLs containing the substring `/github.com/`, for a total of 131,217 tweets. For each GitHub URL (repositories, issues, pull requests, comments, etc), we recorded the repository the URL points to. After we filtered out tweets containing two or more distinct GitHub URLs, to simplify the subsequent qualitative analysis, 129,843 tweets remained.

Using GHTORRENT, the GitHub API, and the GitHub-Twitter user account map we created earlier, we then characterized the relationship of the tweet author to the GitHub repository referenced

in each tweet. Suspecting different tweeting behavior by different types of stakeholders, we distinguished between the following five mutually-exclusive groups, sorted roughly by the level of authority and involvement with the project (if a user is part of multiple groups, we assign them to the topmost one):

- *Owner* (36,480 tweets, or 28 %): The tweet author owns the repo. Note that by construction (see §3), our sample contains only repositories owned by individual user accounts, not organizations; our results should only be interpreted within this context.
- *Collaborator* (14,512; 11 %): The tweet author has write access to the repo, *i.e.*, closed others' issues or pull requests.
- *Contributor* (20,316; 16 %): The tweet author contributed to the repo pull requests, commits, or issues.
- *Follower* (7,697; 6 %): The tweet author follows on GitHub any of the repo owner, collaborators, or contributors.
- *Other* (50,838; 39 %): None of the above.

Finally, we randomly sampled 200 tweets from each of the five strata for our qualitative analysis (details next). A random sample of 1,000 tweets is representative of the population of 129,843 tweets we started from at 95 % confidence with $+/-3\%$ margin of error; within each stratum, 200 random tweets are representative of the toplevel group at 95 % confidence with $+/-7\%$ margin of error.

Qualitative Analysis. Two researchers qualitatively coded the content of the tweets, using the contextual information in the tweet conversations (responses, etc) whenever available, to explore the motivation and purpose of the tweets.

The two researchers proceeded iteratively. First, each independently coded a random sample of the same 50 tweets, then compared results. Multiple rounds of discussion followed to resolve disagreements, after which a unified and fully agreed upon coding scheme was developed and applied to those 50 tweets. Finally, one researcher coded all the remaining tweets using this coding scheme. We only assigned one code per tweet. During coding, we discovered multiple tweets not written in English and we removed these from further analysis, leaving a total of 786 valid tweets.

5 RESULTS

While we did not perform a formal statistical analysis, our results seem to suggest that GitHub project role and Twitter posting patterns are related. We discuss how this can inform future research.

Emerging Themes. The following six major themes emerged from our qualitative analysis.

- *Question* (20 tweets, or 3 %). These tweets ask about technical details of a repository, or about an open issue. Most tweets in this category link to a GitHub issue.
- *Answer* (94 tweets, or 12 %). These tweets respond to questions by other Twitter users. They may explain technical details of the linked repository, or point others to resources on GitHub. Most tweets here link to a repository homepage, a file, or an issue.
- *Call for Action* (32 tweets, or 4 %). The authors of these tweets actively seek or request actions from others. This includes collective actions like starring a repository, helping to solve an issue, or requesting someone to merge a pull request. Most of those tweets link to either an issue or a repository homepage.
- *Repository Advertisement* (366 tweets, or 47 %). Tweets advertising

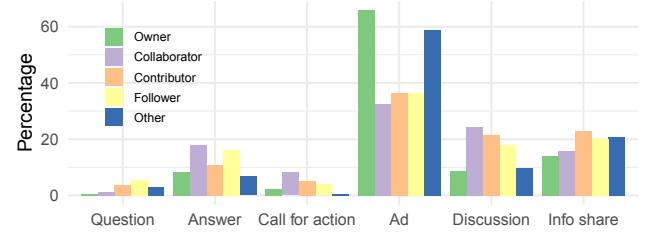


Figure 1: Distribution of emerging themes across the strata.

or advocating for the use of the linked repository. Most such tweets link to a repository homepage.

- *Work Discussion* (129 tweets, or 16 %). Tweets which are part of a discussion on Twitter, but are not a suggestion to use a certain repository. This usually includes discussing the solution of a technical problem, or how to implement certain features. Most of those tweets link to issues, repository homepages, or files.
- *Information Sharing* (145 tweets, or 18 %). Tweets which inform the public or specific individuals about the existence of a repository, a file, an issue discussion, or other work-related news, but with no explicit intention to advocate for the use. Tweets in this category are mostly not part of a Twitter conversation. Most link to a repository, release page, or a specific commit or pull request.

Our qualitative coding results align well with prior analyses of different samples of tweets [12, 18], except that our sample does not include topics on job openings and personal life, which seem generally less likely for GitHub links.

Relationship to GitHub Repository. Figure 1 shows how the different emerging themes vary across the five strata. While we leave more formal statistical analysis for future work, we make some informal observations here, interpreting the figure.

First, when Twitter users tweet about their own repositories, in most cases they intend to advertise (66 %) or inform others about related information (14 %). Tweeting about the repositories they own rarely happens during work discussion (9 %) or answering others' questions (8 %), and almost never appears as a question to ask ($<1\%$), or a call for action (2 %).

The patterns for tweets with links to a repository the author contributes to, has write access to, or follows people in, are similar. They are mostly advertisements for a repository or intended to share related information. Comparing to the previous group of tweets from repository owners, the tweets in this group seem to appear more frequently as work discussion and answers to questions, but less frequently as advertisement or information sharing.

Within these 3 tweeting pattern categories, tweeting for repositories which the author has write access to seem more likely to appear as an answer to a question, or in a work-related discussion, but less likely appear as a repository advertisement. Contributors appear to answer others' questions less than collaborators, but they do more advertisement and information sharing. Contributors and followers have similar tweeting patterns, with contributors appearing to link to GitHub pages more during work discussion, but less as answers to others' questions.

People who tweet a GitHub link and have no identifiable relationship to the linked repository use their tweets mostly as advertisement for the repository (59 %) or to share work-related information (21 %). They seem to rarely tweet a link in work discussions (10 %), when answering questions (7 %), or when calling for actions (1 %).

Our results are partially aligned with Yasir et al. [24], who found that Python core contributors mostly tweet about software-related (corresponding to *Work Discussion*, *Question*, and *Answer*) and community-related topics (*Sharing Information*); the main difference between our results and theirs is the bulk of tweets on repository advertisement, which may be explained by the fact that Python is a well-known language which needs less advertisement.

6 DISCUSSION

Implications. We found that among the tweets whose authors have an identifiable relationship to the repository (*i.e.*, non-*Other*), most are posted by the repository owner and only a few are posted by followers. This suggests that people engaged in the development of the repository post most of the tweets linking back to GitHub. On the other hand, around 40 % of all tweets with a GitHub link are authored by people with no identifiable relationship to the linked repository. We hypothesize that people knew about those repositories from other places, and posted about them on Twitter to further inform others. This suggests that there are channels where people can learn about repositories without engaging in their development or actively following their developers.

Repository owners do not seem to engage a lot in work discussion or answering others' questions. Instead, they mostly share work-related information and updates, and advertise their work. Collaborators seem to participate more in work discussion (*e.g.*, whether to add new features) and answer others' questions (*e.g.*, how to use this repository for a particular job); those user-developer interactions are important to improve the project usability and user satisfaction [10, 11], and we expect they may help attract more adopters and contributors. Contributors to a repository seem less likely to answer others' questions on Twitter, but do more advertisement than collaborators. It might be that most of the advertisements are about the commits or pull requests the contributors made. Contributors also seem to provide less developer-user interaction compared to collaborators, but they may help expand the influence of the project by making more people aware of it.

Interestingly, people who follow others on GitHub tweet a lot about these people's work as advertising or information sharing. However, we do not have timestamped data about GitHub followers, thus we do not know whether the GitHub follow event happens before the tweet or after. It might be that people find the repository first and then start to follow developers therein.

While our case study deepens our understanding of how different types of open-source developers tweet about GitHub work, we have barely began to understand how developers' Twitter activity and social standing may impact the success of their open-source projects. To this end, our dataset mapping 70,427 open-source developer GitHub and Twitter accounts, available online, can be the starting point for several future research directions. For example, one can begin analyzing the impact of Twitter use on outcomes

relating to project sustainability, such as increased project popularity and attracting new contributors—both can now be studied quantitatively. Similarly, one can begin analyzing how open-source developers form and evolve over time (social) networks of collaborators, mentors, and peers, across the two platforms.

Limitations. A notable threat to the validity of our analysis is the accuracy of tweet author-repository relationship inference. For computational reasons, we don't use timestamped data when calculating this relationship, so it is possible that a user becomes a contributor after they post the tweet. Data inconsistencies and missing entries may also exist in GHTORRENT, potentially causing misidentified relationships. To estimate the severity of this threat, we manually examined tweets in the *Contributor* and *Other* categories, as we suspect they are most likely to be misidentified. We randomly sampled 40 tweets from each group and manually checked the contributor status and issue and pull request information in the linked repository webpage. Among the tweets in *Contributor* group, 31 tweets match our definition (78 %), but 4 of the 31 are the top 1 contributor of the repository (based on the number of commits). There are also 5 tweets that the contributor posts the link on the same day as their first contribution to the project, which mostly link to the contribution they made (*e.g.*, the issue they sent); behaviors of those contributors will also be different from other contributors. Besides the 31 'valid tweets', there are 4 tweets posted before the user's first contribution to the project; by definition they should be in the *Other* or *Follower* categories. The 5 tweets remaining are misidentified as *Contributor* tweets due to inconsistent GHTORRENT data, a wrong Twitter-GitHub user profile link, and unavailable GitHub page (we cannot manually check validity). Among the 40 tweets in the *Others* group, we cannot verify 2 of them due to inaccessible data in GitHub or Twitter. Among the 38 tweets left, 23 tweets match our definition of *Others* (61 %), 6 of 38 (16 %) are misidentified due to incorrectly linked Twitter-GitHub account pairs; misclassification of the rest of the tweets is due to missing data in GHTORRENT (21 %). Finally, we also acknowledge that our analysis was conducted based on recent tweets in the period January 1, 2018 to July 1, 2019, which doesn't represent the complete historical tweeting behaviors.

7 CONCLUSION

Cross-linking data across online platforms where open source developers participate (in this paper GitHub and Twitter) is feasible and it can help provide deeper insights into how the activities of developers on one platform are moderated by the roles they play on the other. The social media activities of developers may also have direct impact on software development outcomes, including open-source project success. We provide a large dataset mapping 70,427 open-source developer GitHub and Twitter accounts, as basis for future (quantitative) empirical research in these directions.

ACKNOWLEDGEMENTS

The authors kindly acknowledge support from NSF awards 1546393, 1633083, 1717415, and 1901311, and an award from the Alfred P. Sloan Foundation.

REFERENCES

[1] Sadika Amreen, Audris Mockus, Russell Zaretzki, Christopher Bogart, and Yuxia Zhang. 2020. ALFAA: Active Learning Fingerprint based Anti-Aliasing for correcting developer identity errors in version control systems. *Empirical Software Engineering* 25, 2 (2020), 1136–1167.

[2] Ali Sajedi Badashian, Afsaneh Esteki, Ameneh Gholipour, Abram Hindle, and Eleni Stroulia. 2014. Involvement, Contribution and Influence in GitHub and Stack Overflow. In *International Conference on Computer Science and Software Engineering (CASCON)*. IBM Corp., 19–33.

[3] Andrew Begel, Robert DeLine, and Thomas Zimmermann. 2010. Social Media for Software Engineering. In *FSE/SDP Workshop on Future of Software Engineering Research (FoSER)*. ACM, 33–38.

[4] Sue Black, Rachel Harrison, and Mark Baldwin. 2010. A Survey of Social Media Use in Software Systems Development. In *International Workshop on Web 2.0 for Software Engineering*. ACM, 1–5.

[5] Sue Black and Joanne Jacobs. 2010. Using Web 2.0 to Improve Software Quality. In *International Workshop on Web 2.0 for Software Engineering*. ACM, 6–11.

[6] Hudson Silva Borges and Marco Tulio Valente. 2019. How Do Developers Promote Open Source Projects? *Computer* 52, 8 (2019), 27–33.

[7] Gargi Bougie, Jamie Starke, Margaret-Anne Storey, and Daniel M German. 2011. Towards Understanding Twitter Use in Software Engineering: Preliminary Findings, Ongoing Challenges and Future Questions. In *International Workshop on Web 2.0 for Software Engineering*. ACM, 31–36.

[8] Georgios Gousios. 2013. The GHTorrent Dataset and Tool Suite. In *International Conference on Mining Software Repositories (MSR)*. IEEE, 233–236.

[9] Erik Kouters, Bogdan Vasilescu, Alexander Serebrenik, and Mark G.J van den Brand. 2012. Who's who in GNOME: using LSA to merge software repository identities. In *International Conference on Software Maintenance and Evolution (ICSM)*. IEEE, 592–595.

[10] Dorothy Leonard-Barton and Deepak K Sinha. 1993. Developer-user interaction and user satisfaction in internal technology transfer. *Academy of Management Journal* 36, 5 (1993), 1125–1139.

[11] David Nichols and Michael Twidale. 2003. The Usability of Open Source Software. *First Monday* 8, 1 (2003).

[12] Abhishek Sharma, Yuan Tian, and David Lo. 2015. What's Hot in Software Engineering Twitter Space?. In *International Conference on Software Maintenance and Evolution (ICSM)*. IEEE, 541–545.

[13] Vishal Sharma and Curtis Dyreson. 2018. LINKSOCIAL: Linking User Profiles Across Multiple Social Media Platforms. In *International Conference on Big Knowledge (ICBK)*. IEEE, 260–267.

[14] Giuseppe Silvestri, Jie Yang, Alessandro Bozzon, and Andrea Tagarelli. 2015. Linking Accounts across Social Networks: the Case of Stack Overflow, GitHub and Twitter. In *International Workshop on Knowledge Discovery on the Web (KDWeb)*. 41–52.

[15] Leif Singer, Fernando Figueira Filho, and Margaret-Anne Storey. 2014. Software Engineering at the Speed of Light: How Developers Stay Current using Twitter. In *International Conference on Software Engineering (ICSE)*. ACM, 211–221.

[16] Margaret-Anne Storey, Leif Singer, Brendan Cleary, Fernando Figueira Filho, and Alexey Zagalsky. 2014. The (R)Evolution of Social Media in Software Engineering. In *ICSE 2014 Track on the Future of Software Engineering (FOSE)*. ACM, 100–116.

[17] Margaret-Anne Storey, Christoph Treude, Arie van Deursen, and Li-Te Cheng. 2010. The Impact of Social Media on Software Engineering Practices and Tools. In *FSE/SDP Workshop on Future of Software Engineering Research (FoSER)*. ACM, 359–364.

[18] Yuan TIAN, Palakorn Achananuparp, Ibrahim Nelman Lubis, David LO, and Ee Peng LIM. 2012. What does software engineering community microblog about?. In *International Conference on Mining Software Repositories (MSR)*. IEEE, 247–250.

[19] Yuan Tian and David Lo. 2014. An Exploratory Study on Software Microblogger Behaviors. In *International Workshop on Mining Unstructured Data (MUD)*. IEEE, 1–5.

[20] Bogdan Vasilescu, Vladimir Filkov, and Alexander Serebrenik. 2013. Stack Overflow and GitHub: Associations Between Software Development and Crowd-sourced Knowledge. In *International Conference on Social Computing (SocialCom)*. IEEE, 188–195.

[21] Bogdan Vasilescu, Alexander Serebrenik, Mathieu Goeminne, and Tom Mens. 2014. On the variation and specialisation of workload—A case study of the GNOME ecosystem community. *Empirical Software Engineering* 19, 4 (2014), 955–1008.

[22] Yu Wu, Jessica Kropczynski, Patrick C Shih, and John M Carroll. 2014. Exploring the Ecosystem of Software Developers on GitHub and Other Platforms. In *ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. ACM, 265–268.

[23] Xuan Yang, Daning Hu, and Davison M Robert. 2013. How Microblogging Networks Affect Project Success of Open Source Software Development. In *Hawaii International Conference on System Sciences (HICSS)*. IEEE, 3178–3186.

[24] Muhammad Yasir, Kevin Michael, Bastin Tony Roy Savarimuthu, and Sherlock A Licorish. 2018. Formal in the Informal: A Multi-Level Analysis of Core Python Developers' Tweets. In *Australasian Software Engineering Conference (ASWEC)*. IEEE, 151–160.