Refactoring from 9 to 5? What and When Employees and Volunteers Contribute to OSS

Luiz Felipe Dias[1], Caio Barbosa[2], Gustavo Pinto[3], Igor Steinmacher[4], Baldoino Fonseca[5], Márcio Ribeiro[5], Christoph Treude[6], Daniel Alencar da Costa[7]
[1]VCU, USA [2]PUC-RJ, BR [3]UFPA, BR [4]NAU, USA [5]UFAL, BR [6]U Adelaide, AUS [7]U Otago, NZ

***Abstract*—In this paper we characterize the contributions made by employees (developers that work for GitHub, the company) and volunteers (developers that use GitHub, the platform) to OSS projects maintained by GitHub (the company) on GitHub (the platform). By mining activities performed in five well-known company-owned OSS projects, we investigate *what* they do and *when* they do it. We found that the majority of the volunteers' contributions are related to reengineering (e.g., refactoring), while employees focus more on management (e.g., documentation). When it comes to the working hours, we found that contributions are made mostly from 9am–5pm, even for the volunteers.**

***Index Terms*—Open Source Software, Employees, Volunteers**

I. Introduction

OSS development was often regarded as a voluntary activity [1], in which developers spend their own free time to construct, design, test, and refactor projects. However, the last few years introduced fundamental changes to how OSS is developed. These changes are mostly due to the wave of software companies supporting, maintaining, and releasing OSS projects on a regular basis [2]. For instance, Apple released Swift as an open source project, while Microsoft released the ASP.net framework. OSS contributors are now a mix of both volunteers (i.e., developers that contribute to a project in their free time) and employees (i.e., developers hired by the company to maintain the OSS projects).

While the individual motivations of employee and volunteer contributors have been examined before [3], [4], [5], researchers have not explored their interplay. In this paper we compared these two contributors in terms of *what* and *when* they contribute to OSS projects. We explored these differences in five GitHub-owned projects: `atom`, `electron`, `hubot`, `git-lfs`, and `linguist`. We chose these projects because they were initially developed by GitHubers (GitHub employees) and are maintained on GitHub (the infrastructure). Using unique features available through the GitHub API, we could differentiate whether a contribution comes from a GitHub employee or from a volunteer. Through an extensive analysis of ∼12k pull requests submitted and ∼25k issues reported, we derived the following findings:

- volunteers and employees work on different tasks, but volunteers focus mostly on reengineering, while employees invest more effort on management tasks;
- volunteers contribute more on weekends than employees (15% of volunteers' pull requests versus 8% from employees) in four of the five studied projects. Both groups follow a similar 9am to 5pm work shift.

II. Related Work

Zhou *et al.* [6] studied how industry involvement influences the onboarding of developers in hybrid settings. They found that while high intensity of commercial involvement was associated with a decrease of external inflow and with improved retention, a shared control mechanism was associated with increased external inflow. Homscheid and Schaarschmidt [7] investigated the drivers that explain organizational and community turnover intentions of volunteer developers who are paid by third-party companies. Atiq and Tripathi [8] explored how the developers perceive the differences of rewards in OSS projects, and found that OSS projects where only some people get directly paid may fail if they are mismanaged. Riehle *et al.* [9] analyzed more than 5,000 active OSS projects, from 2000 to 2007, and found that around 50% of all contributions have been paid work. Dias et al. [10] found that both employees and volunteer developers are rather active in company-owned OSS projects, when it comes to the number of pull requests submitted and the participation in the code review process. Pinto et al. [11] analyzed pull request rejections, and provided initial evidence on the use of coding best practices, and the time taken to process a pull request, comparing employees and volunteers. Our study complements the literature, by comparing what kinds of activities they perform and when they perform them.

III. Research Questions

RQ1: *What is the nature of the contributions that employees and volunteers perform?* Here we studied the description of the commit messages to understand the intention behind them using the approach of Hattori and Lanza [12].

RQ2: *When do employees and volunteers perform their contributions?* Here we used two different approaches: we studied the weekday on which contributions were performed (i.e., from Sunday to Saturday) and the working hours in which the contributions were placed (i.e., from 12:00am to 11:59pm).

IV. Research Approach

Employee or Volunteer? The first step was to classify the contributors as employees or volunteers. To this end, we used an approach employed in a previous study [10], which is relying on GitHub's `site_admin` flag. This flag is set to true for GitHub users that are GitHub employees, and to false for those who do not work for GitHub. Consequently, we used this flag to categorize employees and volunteers in the analyzed OSS projects, which are owned by GitHub (the company).

978-1-5386-5541-2/18/$31.00

Company-owned Projects In this study, we leveraged a sample made available in a previous study [11], which is composed of projects developed by GitHub (the company) on GitHub (the platform): `atom`, a cross-platform text editor (∼43,000 stars), `electron`, a tool to build cross platform desktop apps (∼56,000 stars), `hubot`, a customizable life embetterment robot (∼13,700 stars), `git-lfs`, a git extension for versioning large files (∼5,300 stars), and `linguist`, a library to detect blob languages (∼5,400 stars). For each company-owned project, we collected pull requests and issues data using the GitHub REST API. For each pull request, we considered if it is: *open*: waiting for code reviews and/or a final decision; *closed*: the code reviews are done, but the pull request was not accepted (the status in GitHub is closed/unmerged); and *merged*: the code reviews are done, and the pull request was accepted (the status in GitHub is closed/merged).

Data Overview The data reported in this paper is based on pull requests and commits created from the beginning of the studied projects up to June of 2018 (when we collected data).[1] Our data comprises 2,962 volunteers and 94 employees. These employees and volunteers submitted 11,885 pull requests (5,143 submitted by employees, and 6,742 by volunteers).

V. Results

A. RQ1: What is the nature of the contributions?

Here we present an analysis of commit messages based on the approach of Hattori and Lanza [12], in which commits are divided into four major categories of activity: (1) **Forward engineering**, e.g., adding new features; (2) **Reengineering**, e.g., refactoring activities; (3) **Corrective**, e.g., fixing bugs; and (4) **Management**, e.g., updating documentation.

To automatically classify commits into these categories, we compared the content of commit messages against predefined word banks for each commit type based on the earliest match found. For instance, consider the following commit message: "*Add: custom pageSize for printToPDF.*"[2] Unpacking the semantics of this commit message requires extensive domain knowledge and that is difficult to automate. However, the word "add" is a match in the word bank for forward engineering; it is reasonable to assume (in this case) that the commit is adding a new feature to the software.

As we can see in Table I, the intention of the contributions varied greatly. Interestingly, employees' contributions to Management are the most common in three out of the five studied project. Project `linguist` is the only exception to this rule; 28% of the employees' contributions are forward engineering. This pattern can also be seen when considering the contributions made by volunteers. For volunteers, 48% of the contributions made to `linguist` were intended to add new features. However, for three out of the four other studied projects (`atom`, `electron`, and `git-lfs`), reengineering was the most common intent. This finding corroborates with the literature that suggests that contributions made by volunteers to open source projects are far from being trivial [13], [14]. Overall, the presence of uncategorized contributions was not negligible, accounting for 23% of the overall contributions (29% when considering only employees' contributions, and 17% for volunteers)

B. RQ2: When do they perform the contributions?

To provide a general overview, Figure 1 shows the percentage of pull requests placed by day of the week (from Monday to Sunday). As this figure shows, volunteers have a more active presence in contributing during the weekends. In particular, about 15% of the contributions made by volunteers were performed during the weekends (8% on Saturday and 7% on Sundays). For employees, the percentage of contributions on weekends is lower: 8.4% (5.1% on Saturdays and 3.3% on Sundays). Project `electron` is the one that received most contributions by volunteers on weekends (17.2%), whereas project `git-lfs` is the one with the greatest presence of employees working on weekends (15.4%). On the other end of the spectrum, project `hubot` is the one with the least presence of contributions made during weekends (3.8% of the employees' contributions and 5.9% of the volunteers' contributions). Interestingly, none of the studied projects received more contributions from employees than volunteers during the weekends. These findings substantiate the belief that there are volunteers that work in their spare time.

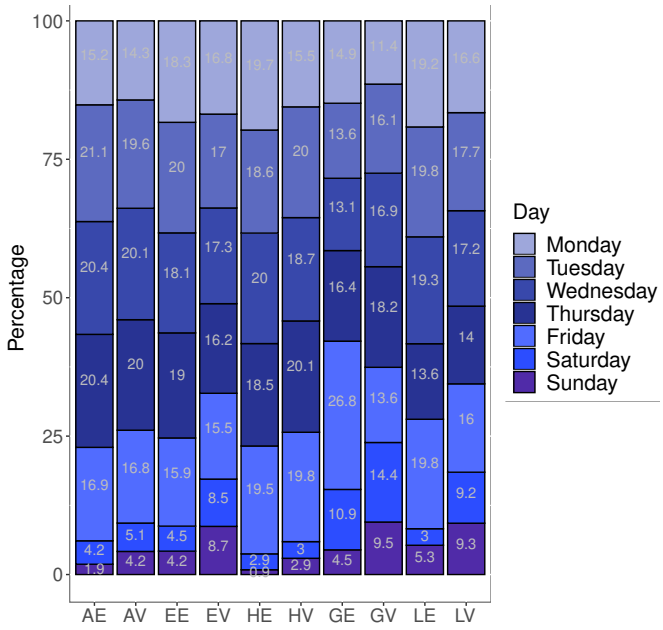

Fig. 1. Percentage of pull requests placed by weekday. There are two bars for each studied project. The first bar (that ends with "E") is for employees, whereas the second one (that ends with "V") is for volunteers. The letter "A" in the beginning stands for `atom`, whereas "E" stands for `electron`, "H" stands for `hubot`, "G" stands for `git-lfs`, and "L" stands for `lingust`.

We now analyze the contributions by paying particular attention to the working hours in which they were placed. To do that, we used the field `date` that is available in each pull request accessed through the GitHub API (normalized to the timezone GMT-7). To conduct our analysis, we adjusted time to match the timezone where most of GitHub's operation

[1] Available online at: https://github.com/fronchetti/VLHCC-2020.

[2] https://github.com/electron/electron/commit/fc6628d

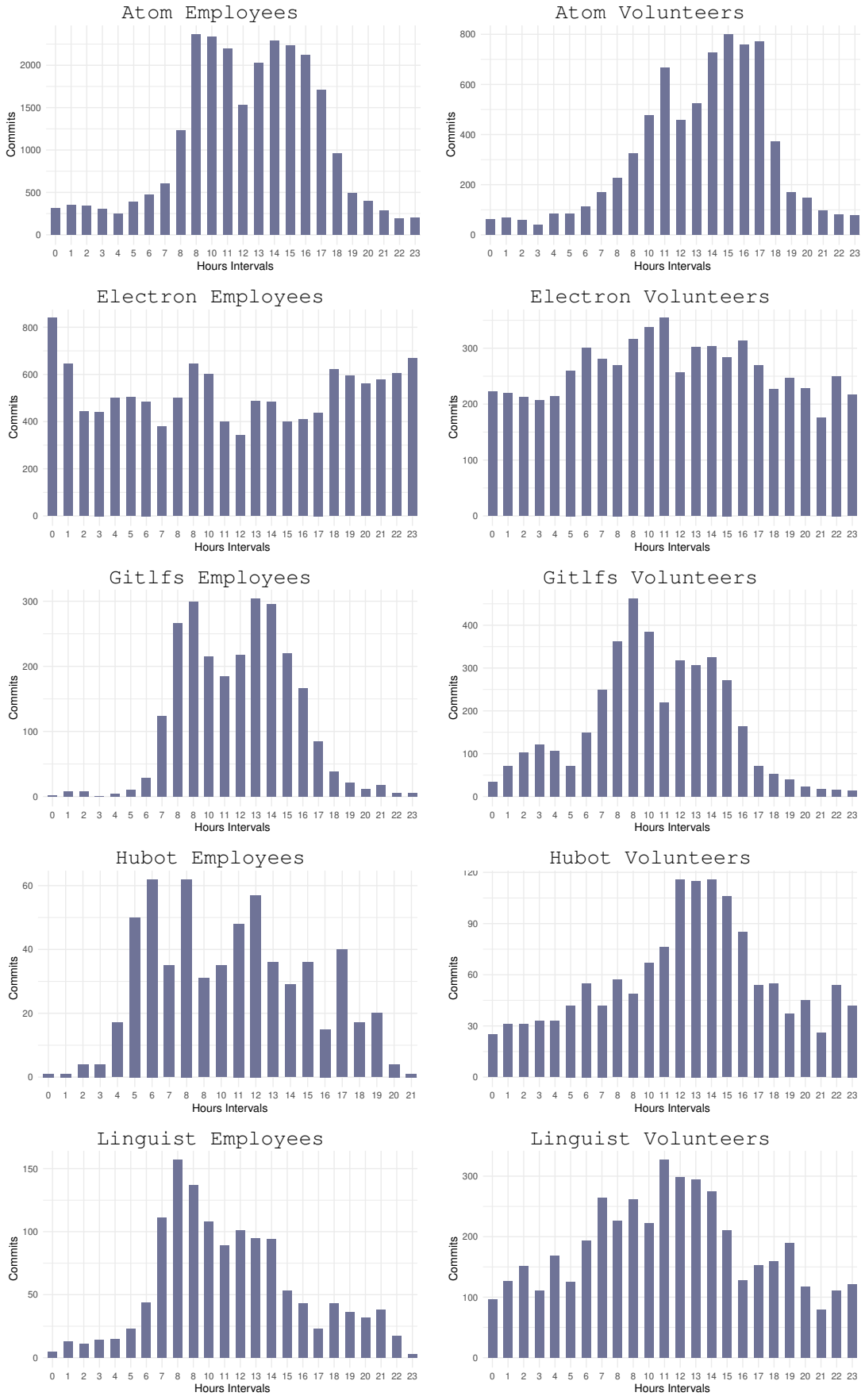

Fig. 2. The working hours of the employees and volunteers.

TABLE I
THE HATTORI-LANZA CLASSIFICATION SCHEME [12].

	Employees					Volunteers				
	Atom	Electron	Git-Lfs	Hubot	Linguist	Atom	Electron	Git-Lfs	Hubot	Linguist
Forward Eng	18%	10%	19%	5%	28%	15%	14%	20%	22%	48%
Corrective Eng	9%	21%	9%	24%	6%	24%	28%	22%	18%	9%
Reengineering	13%	18%	18%	20%	15%	29%	30%	20%	32%	17%
Management	25%	25%	25%	23%	21%	13%	12%	20%	11%	10%
Uncategorized	35%	26%	29%	28%	30%	19%	16%	18%	17%	16%

happens, in San Francisco, California. Figure 2 shows the number of pull requests organized by their timestamp.

We note that the number of contributions made by employees is similar to the number of contributions by volunteers. Overall, employees are responsible for 47.7% of the pull requests. In terms of their working hours, we found an interesting pattern: projects `atom`, `git-lfs`, `hubot`, and `linguist` are more likely to follow a traditional 9am to 5pm work practice (66.8%, 74.9%, 55% and 55% of their pull requests were created, respectively, between 9am and 5pm); 30% of the contributions to `electron`, however, were performed between 9am and 5pm. This pattern occurs regardless of whether the contribution was made by volunteers or employees. Interestingly, even the developers classified as volunteers have a significant proportion of their contributions placed during the 9am to 5pm time window. For instance, 64% of the contributions that volunteers performed to `atom` were placed between 9am and 5pm (the results for `git-lfs` and `linguist` are, respectively, 61% and 45%).

We also manually inspected the profiles of the top-10 volunteers and employees in terms of merged pull requests. For the employees, we found that more than 80% of the pull requests from the top-10 are from US-based developers, for all projects except `electron`. For `electron`, 65% of the pull requests come from a developer based in Asia, which should explain the different pattern. For the volunteers, when we analyze `atom`, `git-lfs`, and `linguist`, we also found that most of the pull requests are concentrated on US-based developers. However, for `electron` and `hubot`, we noted different behaviors. For the first, the contributions are spread almost equally among developers from Europe, US, and Asia-Pacific; for the latter, the contributions are split between developers in US and Europe (and 2% from Australia). Thus, we see that the many contributions that are apparently spread throughout the day, may be actually made in working hours, given the geographical distribution of the analyzed teams.

VI. IMPLICATIONS

Research. The results of **RQ2** suggest that most of the contributions performed by volunteers are done in the traditional working hours (e.g., from 9am to 5pm). Therefore, chances are that external developers are being paid by software companies other than GitHub to contribute to the studied OSS projects. Still, since there can be a shift from intrinsic to extrinsic motivations, it is necessary to revisit the common motivations to join and remain in OSS projects. Moreover, analyzing how the companies get involved with projects, and the way they interact with each other may be interesting research topics.

Practice. The results support maintainers (owners) to understand how the external community gets involved in a project, and takes different responsibilities. We showed that the contributions made by volunteers include new features and corrective engineering, but, for three out of five projects we found that the main focus is on Reengineering tasks; employees work on management tasks (being the most common task in 3 projects, **RQ1**). Companies could then understand that, when there is a critical mass of users interested in keeping the software evolving, a community will grow around the project. Therefore, it is important for the companies that own a product to create a welcoming environment, enabling the community to take active part in the process.

VII. LIMITATIONS

First, we rely on a very specific approach to verify whether a contributor is an employee or a volunteer. To minimize this threat, we manually investigated the affiliation of the contributors, by searching their profiles on their personal websites or other social networks. We found that two members classified as volunteer presented GitHub as their organizations. We further analyzed their profile, and found that they left GitHub and are now working in other companies. For those classified as employee members, all listed themselves as GitHub staff in their profile. Second, as we analyzed only five projects from the same company, our results cannot be generalized to other OSS projects driven by other software companies. Still, our approach limits us to scale up the number of studied projects. However, this small sample enabled us to better evaluate and understand the limitations of our approach. Nevertheless, our sample of studied projects is diverse in terms of (i) popularity (from 5k stars to 56k stars), (ii) domain (from a text editor, to a versioning tool, to libraries) and (iii) programming language (from JavaScript to Ruby and Go).

VIII. CONCLUSIONS

In this paper we investigate how the contributions from employees and volunteers differ in terms of what they do and when they do it. We found that 25% of the volunteers' contributions might be related to reengineering tasks (e.g., refactorings) whereas employees might concentrate a substantial part of their effort on management activities. Moreover, although most of the contributions made to the studied projects follow a 9am to 5pm work habit, we observed that volunteers are more likely to provide contributions during weekends (15% of the volunteers' pull requests were placed on weekends) than employees (8% of their pull requests were on weekends).

Acknowledgements. This work is partially supported by FAPESPA, UFPA, CNPq (309032/2019-9, 430642/2016-4, and 406308/2016-0), NSF (1900903) and ARC (DE180100153).

REFERENCES

[1] G. Pinto and F. Kamei, "The census of the brazilian open-source community," in *Open Source Software: Mobile Open Source Technologies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 202–211.

[2] G. Pinto, I. Steinmacher, L. F. Dias, and M. Gerosa, "On the challenges of open-sourcing proprietary software projects," *Empirical Software Engineering*, Mar 2018.

[3] G. von Krogh, S. Haefliger, S. Spaeth, and M. W. Wallin, "Carrots and rainbows: Motivation and social practice in open source software development," *MIS Quarterly*, vol. 36, no. 2, pp. 649–676, Jun. 2012.

[4] A. Hars and S. Ou, "Working for free? motivations of participating in open source projects," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE, 2001, pp. 1–9.

[5] K. Lakhani and R. Wolf, *Perspectives on Free and Open Source Software*. Cambridge, Mass.: The MIT Press, 2005, ch. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects, pp. 1–22.

[6] M. Zhou, A. Mockus, X. Ma, L. Zhang, and H. Mei, "Inflow and retention in oss communities with commercial involvement: A case study of three hybrid projects," *ACM Trans. Softw. Eng. Methodol.*, vol. 25, no. 2, pp. 13:1–13:29, Apr. 2016.

[7] D. Homscheid and M. Schaarschmidt, "Between organization and community: investigating turnover intention factors of firm-sponsored open source software developers," in *WebSci '16*. ACM, 2016, pp. 336–337.

[8] A. Atiq and A. Tripathi, "Impact of financial benefits on open source software sustainability," in *Proceedings of the International Conference on Information Systems - Digital Innovation at the Crossroads, ICIS 2016, Dublin, Ireland, December 11-14, 2016*, 2016, pp. 1–10.

[9] D. Riehle, P. Riemer, C. Kolassa, and M. Schmidt, "Paid vs. volunteer work in open source," in *Proceedings of the 2014 47th Hawaii International Conference on System Sciences*, ser. HICSS '14, 2014, pp. 3286–3295.

[10] L. F. Dias, I. Steinmacher, and G. Pinto, "Who drives company-owned oss projects: Internals or externals members?" *J. Braz. Comp. Soc.*, vol. 24, no. 1, pp. 2:1–2:11, 2018.

[11] G. Pinto, L. F. Dias, and I. Steinmacher, "Who gets a patch accepted first?: comparing the contributions of employees and volunteers," in *Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering, CHASE@ICSE 2018, Gothenburg, Sweden, May 27 - June 03, 2018*, 2018, pp. 110–113.

[12] L. P. Hattori and M. Lanza, "On the nature of commits," in *Proceedings of the 23rd IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE'08, 2008, pp. III–63–III–71.

[13] G. Pinto, I. Steinmacher, and M. Gerosa, "More common than you think: An in-depth study of casual contributors," in *IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering, SANER 2016, Suita, Osaka, Japan, March 14-18*, 2016, pp. 112–123.

[14] M. Ahokas and P. Laurila, "Benefiting from communities in open source business," *New Business in Computer-Mediated Communities*, pp. 51–64, 2004.