HAL open science

Why Do People Give Up FLOSSing? A Study of Contributor Disengagement in Open Source

Courtney Miller, David Gray Widder, Christian Kästner, Bogdan Vasilescu

To cite this version:

Courtney Miller, David Gray Widder, Christian Kästner, Bogdan Vasilescu. Why Do People Give Up FLOSSing? A Study of Contributor Disengagement in Open Source. 15th IFIP International Conference on Open Source Systems (OSS), May 2019, Montreal, QC, Canada. pp.116-129, 10.1007/978-3-030-20883-7_11 . hal-02305702

HAL Id: hal-02305702

https://inria.hal.science/hal-02305702v1

Submitted on 4 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Why do People Give Up FLOSSing? A Study of Contributor Disengagement in Open Source

Courtney Miller[1⋆], David Gray Widder[2], Christian Kästner[2], and Bogdan Vasilescu[2]
[1] New College of Florida, USA [2] Carnegie Mellon University, USA
courtney.miller17@ncf.edu, {dwidder, kaestner, vasilescu}@cmu.edu

Abstract. Established contributors are the backbone of many free/libre open source software (FLOSS) projects. Previous research has shown that it is critically important for projects to retain contributors and it has also revealed the motivations behind why contributors choose to participate in FLOSS in the first place. However, there has been limited research done on the reasons why established contributors disengage, and factors (on an individual and project level) that predict their disengagement. In this paper, we conduct a mixed-methods empirical study, combining surveys and survival modeling, to identify the reasons and predictive factors behind established contributor disengagement. We find that different groups of established contributors tend to disengage for different reasons; however, overall contributors most commonly cite some kind of transition (e.g., switching jobs or leaving academia). We also find that factors such as the popularity of the projects a contributor works on, whether they have experienced a transition, when they work, and how much they work are all factors that can be used to predict their disengagement from open source.

1 Introduction

Contributor disengagement in open source is widely known as a costly and critical issue [9, 19, 49], as it can directly affect the sustainability of projects. For example, in a recent study Coelho et al. reported that 41 % of failed open source projects cited a reason involving the developer team, such as lack of interest or time of the main contributor [9]. Such local (project-level) sustainability issues in open source can have cascading effects on the entire ecosystem because of project interdependencies [12, 53]. So-called "core", i.e., established, contributors are particularly critical for the sustainability of open source projects [19, 57].

There are many reasons why established contributors disengage. Some may be unavoidable, whereas others could perhaps be prevented through interventions or by providing better community support. Likely there are various dynamics in play, including the role of volunteers as compared to corporate employees [44], the role of external events such as family planning and job changes, and the role of perceived purpose, community support, and stress. Effects might include abruptly leaving the project, but also slow disengagement, or causing rippling frustrations through delays or cynicism.

The goal of our research is to better understand disengagement factors and which established contributors are at risk and when; this will enable us to build and validate a

⋆ Part of this work was carried our during the author's REU program at CMU.

conceptual framework and theory. Moreover, we pursue a data-driven approach, operationalizing uncovered factors based on publicly available trace data. This way, we can identify at-risk open source contributors and communities, and help guide resources (e.g., volunteers, sponsors) toward projects and contributors in need, enhancing the sustainability of the overall ecosystem.

We identify potential disengagement factors from literature on turnover and open source retention, cross-validate them with results from a survey among contributors who recently stopped all open source activities on GitHub, operationalize select factors with public trace data, and finally conduct survival modeling among a set of 206 GitHub users to triangulate the survey results.

Among others, we identify the degree to which contributors work outside of typical office hours and to what degree they engage in support activities as important moderating factors. According to Claes et al. [8], 33 % of open source contributors do not follow typical working hours, but instead work nights and weekends. Our survey shows that contributors who work nights and weekends proportionally tend to disengage for different reasons than those working regular hours. In addition, our survey reveals that the most common reasons for complete disengagement relate to transitions in employment, such as graduating from academia, changing employers, and changing roles.

To validate disengagement factors beyond our survey, we model to what degree hypothesized factors—such as working hours, engagement in support activities, and team size, which can be measured in public trace data of contributor activities—can predict the later disengagement of those contributors. To that end, we use the quantitative statistical method of survival modeling. As a key factor in our model, derived from our survey results, we incorporate transitions identified from public CVs of developers. Specifically, we analyze which contributor populations are more resilient to transitions such as job changes.

We find that working predominantly during office hours and experiencing a transition both increase a contributors risk of disengagement. Conversely, we find that increased levels of activity and working on more popular projects both decrease a contributors risk of disengagement.

In summary, we contribute (1) a survey revealing the reasons behind contributor disengagement; (2) a comparison between different groups of contributors; (3) measures to differentiate between groups, which could be used to help identify at-risk groups and better target support interventions; (4) a novel operationalization of transition data; and (5) a survival model demonstrating which factors are able to predict contributor disengagement.

2 Related work

Turnover. Prior work has shown that the turnover rate of a project profoundly affects its survival probability [33, 46] and code quality [21]. Approximately 80 % of open source projects fail due to contributor turnover related issues [46]. Even within projects that do not outright fail, contributor turnover has a significant adverse effect on software quality [21]. On a project level, contributor disengagement results in knowledge loss, which is a particularly expensive issue [33].

Employee turnover and *retention* have been broadly studied across many fields [31, 35]. In professional settings, early turnover research has focused often on personal characteristics (e.g., ability, age) and employee satisfaction, measured with hiring tests and

surveys, whereas later research has explored many more nuanced factors, such as labor market (e.g., job opportunities), non-work values, and organizational commitment [31]. Research has shown that, while far from all turnover can be explained by dissatisfaction and similar factors [38], there are positive and negative factors that can buffer against shocks such as external job offers [6, 20]. Turnover among volunteers is less explored: Although some research suggests that similar personal and environmental factors influence their decisions to quit [41], other researchers point out that satisfaction and achievement, compatible working hours, training, challenging work, and role identity may play particularly strong roles [25, 34, 50].

Whereas reasons for joining open source [5, 24, 37, 44, 48, 54, 55] and interventions to improve the onboarding experience for new developers [7, 18, 30, 52] have been studied in depth, studies of contributor retention are rarer. Prior research has focused primarily on testing basic attributes [11, 39, 40, 46, 49, 53, 58] . For example, they have shown that retention is higher for contributors that have participated longer [39, 49], contributed more code changes [11, 39], and communicated more [11]. However, there has been a limited amount of prior research has also explored more nuanced factors, like whether a developers gender and social network effect their risk of disengagement [43]. Using surveys, researchers further associated ratings of general dissatisfaction and lack of community identification with higher perceived turnover and turnover intentions [32, 56]. Zhou et al.'s case study of three projects further suggests that commercial participation can crowd out volunteers [58].

Long working hours, lack of sleep, and lack of recovery on weekends are often discussed as stressors. Many studies confirm the importance of "mentally switching off" [1, 4, 51]. In software engineering, several studies have shown the influence of time-related factors, such as late-night commits and long working sessions being more likely to contain bugs [17, 45], sleep deprivation reducing code quality [22], Monday commit comments using more negative language [29], and time pressure is often seen as an important stressor [36].

Open source practitioners reporting stress. In addition to the academic literature, open source practitioners also spoke out about frustrations, funding concerns, stress, and even burnout. Often, there are high expectations and copious amounts of pressure placed on established open source contributors.

Many stories via blog posts from maintainers who disengaged have a similar narrative that describes the growing pressures and responsibilities they experienced that lead to their disengagement. One such blog post describes how *"as [my project's] popularity rose and rose, my drive to continue to create new projects, fell. All while the burden of supporting the needs of the massive user bases of my successful projects and the pressure of maintaining those projects grew."*[1]

In addition to blog posts, there were also participants from the survey we ran who explicitly cited a lack of support as a reason for their disengagement. For example *"[The open source project] is increasingly depended upon by other projects, but very few external developers are interested/willing enough to [understand the company] let alone contribute improvements/fixes. The support burden is a good problem to have (people are finding [the project] useful), but it does impose a productivity (and sometimes a motivation) burden."* (P35)

Contributors are broadly expected to maintain their projects. Having a seemingly never-ending list of tasks is another commonly cited reason for disengagement among the aforementioned blog posts and survey respondents. As described in a blog post by

[1] `https://www.kennethreitz.org/essays/the-reality-of-developer-burnout`

a now-retired developer, *"working long hours for endless months"* was a critical reason for their disengagement.[2]

3 Overview: Mixed-Method Research

Our mixed-method empirical study follows a sequential exploratory design [14], combining qualitative and quantitative analysis of survey and GitHub trace data.

Step 1: Survey (Sec. 4). Although the turnover literature (Sec. 2) provides several starting points for potential disengagement factors, there has been only limited research on the actual reasons *why* open source contributors disengage. Therefore, we decided to ground our research by conducting an open-ended survey among developer who recently disengaged from all public GitHub activities. We furthermore analyze the frequency of self-reported reasons for disengagement regarding whether different populations disengage for different reasons.

Step 2: Survival analysis (Sec. 5). We test to what degree the potential disengagement factors identified statistically explain disengagement. To that end, we operationalize several disengagement factors, including when and what contributors worked on as well as job transitions in historic trace data and public CVs, and use survival modeling [42] to test their significance.

4 Self-Reported Reasons for Disengagement (Survey)

4.1 Survey Methodology

To ground our analyses, we surveyed a sample of open source contributors who recently disengaged from all public GitHub activities, asking about their reasons.

Recently Disengaged Established Contributors. We invited open source contributors who stopped all public activity on GitHub after being active for at least 18 month. We identified such contributors from GHTorrent [26] trace data (version 2018-08). We then constructed six-month panels aggregating contributions (commits and issue/pull request events) per person, and selected those contributors who contributed at least 100 commits per six-month period for three consecutive periods, but at most 5 commits in the following period (the five commit threshold allows for some residual activity). This way, we identified a total of 702 contributors who disengaged (i.e., stopped contributing publicly) within the last year and had public email addresses listed on their GitHub profile pages.

We specifically sampled only previously active contributors with at least 100 commits per period across all of GitHub. Previous research has shown that within a single project, there are many different kinds of contributors, with one of the most popular models being the onion model [15]. With our threshold we target contributors who are likely very active in at least one project, rather than more peripheral or episodic contributors, which may have different motivations [2].

Survey Design. We designed a simple, single-question, open-ended survey, asking *"Could you help us understand your reasons for reducing your contributions to GitHub projects?"* We chose the open-ended format to avoid priming the participants to ensure organic but relevant responses. We use the single-question format without external survey software,

[2] `https://hackernoon.com/what-is-programmer-burnout-651aa48984ef`

because it reduces the barrier to participation. We invited all 702 identified candidates and received 151 valid answers (21.5 % response rate). Our response rate is in line with other GitHub surveys, e.g., [27].

Card Sorting Analysis. We used card sorting, a qualitative content analysis [47] method, to analyze the survey answers. Two researchers reviewed the cards and organized them into mutually agreed upon categories using a ground-up process resulting in 17 subgroups. These subgroups were then further grouped into three overarching themes: Technical, Social, and Occupational. Note that many participants cited multiple reasons, resulting in 239 reasons from 151 responses.

Quantitative Analysis. In addition to identifying common self-reported reasons for disengagement from the survey responses, we additionally explore whether different populations report different reasons. Based on the literature and reports from open source practitioners (cf. Sec 2), we specifically investigate whether contributors (a) working mostly "regular" office hours or (b) performing more support activities report disengaging for different reasons.

Working hours: Analyzing GitHub data, we measure what percentage of contributions are made between 7am and 7pm local time, Monday through Friday, captured as *indexWorkHours* (the slightly wider interval than the traditional 9am to 5pm increases robustness to daylight savings [8]). To detect the contributor's local time, we adjusted the UTC times in GHTorrent with the average time zone offset for each developer, collected from a small random sample of their commits after cloning repositories locally. We then separate our survey participants into two groups, *Office Hours* (more likely paid contributors) and *Nights and Weekends* (more likely volunteers), based on whether they perform more or less relative amount in the office hour window described above than average (average $indexWorkHours = 0.6$; design following prior research [39]).

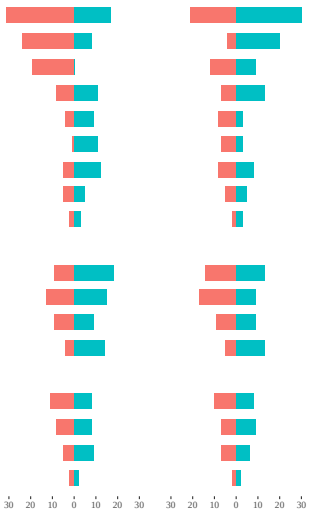
Support activity: We also measured *indexSupport* as the percentage of support activities among all activities, i.e., all non-commit GHTorrent events related to managing issues and pull requests. We distinguish between *High Support Work* and *Low Support Work* relative to the mean ($indexSupport = 0.2$).

Note that given the different ways in which we aggregate the survey responses and the relatively small sample size overall, we cannot draw sound statistical conclusions about differences between the (sub)groups. While we report exact numbers, readers should focus on qualitative differences.

Threats to Survey Validity. As usual for surveys, our results may be affected by a selection bias: contributors who did not answer may have had different reasons for disengaging. To identify contributors who had disengaged, we used public GitHub data, which covers much but not all open source activities, as also visible in 10 (of 151) survey responses that indicate changing platforms. Deriving the survival model data from survey participants enabled modeling only contributors confirmed to have disengaged. Note that we consider moving to private repositories (12 answers) still as disengagement from public open source activities. Furthermore, our approach to identify disengagement looks for sudden disengagement (within a six-month window) and results may not generalize to contributors who disengage more gradually. Contributors may also deliberately or unconsciously self-censor in their answers, providing socially acceptable reasons rather than real—a common concern in turnover research [31]. Note however, that our survival model (discussed later) is built entirely on historic trace data rather than self-reported answers, and thus reduces this threat.

Table 1. Self-Reported Reasons for Disengagement in Survey

Subgroup	Count	Office Hrs vs Nights&We	More Support vs Less
Occupational reasons			
Got new job that doesn't support FLOSS	37		
Changed role/project	25		
Left job where they contributed to FLOSS	16		
No time: new job	15		
No time: existing job	10		
Left school where they contributed to FLOSS	12		
No time: in school	12		
FLOSS in school, now job doesn't support FLOSS	7		
Too much coding at work	4		
Social reasons			
Lost interest in FLOSS	24		
No time: personal	23		
Lack of peer support	16		
No time: nondescript	15		
Technical reasons			
Issues w GitHub or industry	14		
Individually moved to private repos	12		
Changed platform	10		
Feature complete project	3		



4.2 Results from Survey

In Table 1, we show the survey results. The most common self-reported reason for disengagement was changing jobs to a job that does not support open source work and occupational reasons were generally the most frequent.

Furthermore, we observe differences across populations: *Contributors who work nights and weekends tend to disengage for different reasons than those who work during office hours:* contributors who worked nights and weekends most commonly cited social reasons, whereas those who worked during office hours most commonly cited occupational reasons; the largest difference is between those who cited *Left job where they contributed to OSS*, with 19 % and 0 % citing it respectively.

Next, we turn to the aggregation by type of work, noting *Contributors who do less support work tend to disengage for different reasons than those who do more:* In particular, only 67 % of the *More Support Work* group cited at least once Occupational reason, compared to 72 % of the *Less Support Work* group. The difference between these two groups may be because since they are less stressed when major life changes occur (i.e., getting a new job or leaving school), they are better able to cope with transitions.

Finally, we emphasize a surprising result. For all contributors, occupational reasons such as major life changes (e.g., getting a new job or leaving school) were the most cited (with 106 citations), significantly more than lacking peer support or losing interest that are more commonly discussed in the literature. This motivated us to consider transitions explicitly in our survival analysis below.

5 Modeling Disengagement Factors (Survival Analysis)

5.1 Survival Model Methodology

We use survival analysis to triangulate the survey results and model the relative strengths of the effects of the three main factors emerging from the survey analysis on the risk of disengagement from public GitHub activity (*Work Hours* vs *Nights and Weekends*; *High Support Work* vs *Low Support Work*; and *Job Transitions*). Survival analysis is a statistical modeling technique that specializes in time-to-event data [42], particularly suited for modeling right censored data. In our study, the event is public GitHub *disengagement*; right censorship can occur for contributors whose last recorded event may be very close to the end of the observation period, for which it is not clear whether they will return to contribute more. In particular, we use a Cox Proportional Hazards regression model [13]. The estimated regression coefficients describe each variable's *hazard ratio* (HR), which is analogous to an odds ratio in for multiple logistic regression analysis. Briefly, an HR > 1 indicates an increased risk of observing the event, and an HR < 1 indicates a decreased risk, relative to a one unit change in a predictor variable (or flipping the value, in case of binary variables), while holding all other predictors constant.

Data. We collect GitHub data on several variables for the open source contributors who disengaged and responded to our survey (the 'treatment' group), as well as for an equal sized 'control' group of contributors who did not disengage. With this design, a survival model estimates which factors are statistically useful for distinguishing groups.

For job transition data, we collect publicly available CV data from contributors by following links on their GitHub profiles. Since our data collection is not yet fully automated, we can currently only assemble a dataset of moderate size, therefore we only collected data for our our survey participants (plus the control group), because their survey answers validate that they actually disengaged. For non-CV data, we use GHTorrent (Sec. 4). We discard 34 participants for which we cannot find CVs or similar information from which we can deduce past transitions, leaving us with a dataset of 206 contributors of which 103 disengaged. By construction, both groups contributed actively for 18 months (at least 100 commits per six-month period for three consecutive periods; Sec. 4); the 'control' group contributors then remained active for at least another six months at similar levels or higher, while the 'treatment' group contributors made at most five commits in the following period, i.e., they *disengaged*.

Model factors and operationalization. We compute:

- *Activity level:* Prior work has shown that more active contributors are less likely to disengage [11], hence we control for the average quarterly activity level by counting all activities (commits and support) per person.
- *Working hours and support:* We use the two factors *indexWorkHours* and *indexSupport* as introduced in Section 4.1 to characterize the degree of work outside regular working hours (more likely volunteers) and the degree of support activities, both identified as stressors by practitioners (cf. Sec. 2). We compute dummy variables indicating being above or below the mean.
- *Organizational affiliation:* - Previous research has shown that on a project scale, having an organizational affiliation can help increase developer retention rates [58]. We test whether organizational affiliation has the same affect on engagement on an individual scale as it does on a project scale. Using GHTorrent, we record whether contributors had an Organizational Affiliation listed on their GitHub public profile.

- *Team size:* Turnover research regularly reveals social embedding in a team as an antidote to turnover [19]. We operationalize this as the number of contributors per project. Since a contributor may be part of multiple projects, we consider only their main projects (for a contributor, taking all projects with the highest number of contributions that together constitute at least 50 % of all contributions) and record the average team size among those projects. 'Teams' comprise everyone who authored at least one commit.
- *Project popularity:* To control for whether contributors are more likely to disengage from small or very popular projects, we use the number of stars a project has on GitHub as a proxy for its popularity (standard measure in GitHub research [16]). We model popularity in addition to activity level because previous research has shown that the popularity of a project influences its survival probability [53], and we are interested in whether the popularity of a project also affects the survival probability of its contributors on an individual level. For contributors working on multiple projects, we consider the max popularity of the contributor's active projects (see team size).
- *Transition found:* Finally, to operationalize a contributor's transition data, identified as very important in our survey, we went to their linked publicly available CV and created a binary variable that recorded whether there was a transition present in the last year or not. We considered a transition to be either the stopping or starting of a job or educational program.

Model diagnostics. We performed the standard model diagnostics: We log transformed variables with highly skewed distributions, as necessary, to reduce heteroscedasticity [23]. We tested for multicollinearity using the variance inflation factor (VIF $<$ 3) [10]. We also inspected Schoenfeld residual plots to graphically diagnose Cox regression modeling assumptions [28].

Threats to Model Validity. Regarding the survival model, statistical power is limited by the small sample size, which is limited by our design of modeling only survey participants with public CV data (due to confirming disengagement with the survey and manual effort required, as discussed). Since our treatment group was limited to the survey respondents, our survival model also has the risk of suffering from selection bias. As usual, our operationalization of factors in our survival model can only capture part of the concept to be measured. While we experimented with different operationalizations of our factors to ensure construct validity and robustness, one needs to be careful in generalizing our results beyond our specific operationalizations.

5.2 Results from Survival Modeling

Table 2 presents the results from the two survival models created; a base model without the novel transition found variable, and a full model with.

The *base model* had a goodness of fit of $R^2 = 0.21$. The controls behave as expected. *Total activity* had a hazard rate of 0.36, meaning it decreases a contributor's risk of disengaging by a factor of 0.38. Similarly, contributors who work on more popular projects are less likely to disengage (*Max number of stars* has a hazard ratio of 0.85).

As predicted based on previous research, the *workHours* dummy affects a contributor's risk of disengaging, having a hazard ratio of 1.56. This suggests that working during business hours more than the average contributor increases the risk of disengaging by a factor of 1.56. Surprisingly, we do not observe any statistically significant

Table 2. Survival models for contributor disengagement.

	Base model			Full model		
activity	0.36	(0.21)***	29.00***	0.36	(0.21)***	27.92***
orgAffiliation	0.90	(0.21)	0.27	0.92	(0.21)	0.17
maxTeamSize	1.17	(0.08)	3.59	1.17	(0.08)	3.41
maxNumStars	0.85	(0.05)**	10.01**	0.86	(0.05)**	9.08**
highSupportWork_TRUE	1.29	(0.26)	0.96	1.43	(0.27)	1.74
workHours_TRUE	1.56	(0.21)*	4.52*	2.20	(0.30)**	5.59*
jobTransition				2.48	(0.31)**	8.15**
workHours:jobTransition				0.55	(0.42)	
R^2	0.21			0.25		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

effects of doing more support work than average (the *highSupportWork* dummy), perhaps due to our operationalization or relatively small sample size.

The *full model* fits the data better ($R^2 = 0.25$), meaning that adding in the *jobTransition* variable helped increase the explanatory power of the model. The *jobTransition* variable has a hazard ratio of 2.48, meaning, as suggested by the survey results, that experiencing a transition significantly increases a contributor's risk of disengagement by a factor of 2.48.

6 Discussion and Conclusions

In this research, we have looked at the reasons why established open source contributors disengage, using a survey with 151 responses and a survival model to quantify factors which predict disengagement. From the grouped analysis of survey results, we learned that the *Nights and Weekends* and *Office Hours* groups tend to cite different reasons for their disengagement, and so do more the *Less Support Work* and *More Support Work* groups.

Importantly, our study shows that operationalizations of different disengagement risk factors using publicly observable trace data are plausible. For example, since occupational reasons were the most commonly cited, we used online public CVs to operationalize the *jobTransition* variable; however, other commonly cited reasons from the survey may also be operationalizable. Another commonly cited reason was 'no time, personal circumstance', more specifically people often cited having children or getting married. Such circumstances may be observable on social networking platforms. This suggest that a data-driven systems could be developed to help identify at-risk groups on a significantly larger scale, instead of having to rely on relatively expensive survey data. This information could be useful to different stakeholders, such as open source foundations and other funding agencies, looking to target support interventions. Overall, support interventions targeted more appropriately could significantly increase the sustainability of open source ecosystems.

We aim to work on these extensions of the research and more, to better understand the reasons why different kinds of established contributors disengage, since defining the problem is the first step to solving it [3].

Acknowledgements. This work was supported through CMU's REU in SE, NSF (1318808, 1552944, 1717022, and 1717415), and AFRL and DARPA (FA8750-16-2-0042). We thank our survey participants, and colleagues at CMU, especially Jim Herbsleb, Chris Bogart, Marat Valiev, and Sophie Rosas-Smith.

References

[1] Akira Bannai and Akiko Tamakoshi. The association between long working hours and health: a systematic review of epidemiological evidence. *Scandinavian journal of work, environment & health*, pages 5–18, 2014.

[2] Ann Barcomb, Andreas Kaufmann, Dirk Riehle, Klaas-Jan Stol, and Brian Fitzgerald. Uncovering the periphery: A qualitative survey of episodic volunteering in free/libre and open source software communities. *IEEE Transactions on Software Engineering*, 2018.

[3] Eugene Bardach and Eric M Patashnik. *A practical guide for policy analysis: The eightfold path to more effective problem solving*. CQ press, 2015.

[4] Carmen Binnewies, Sabine Sonnentag, and Eva J Mojza. Recovery during the weekend and fluctuations in weekly job performance: A week-level study examining intra-individual relationships. *Journal of Occupational and Organizational Psychology*, 83(2):419–441, 2010.

[5] Andrea Bonaccorsi and Cristina Rossi. Comparing motivations of individual programmers and firms to take part in the open source movement: From community to business. *Knowledge, Technology & Policy*, 18(4):40–64, 2006.

[6] James P Burton, Brooks C Holtom, Chris J Sablynski, Terence R Mitchell, and Thomas W Lee. The buffering effects of job embeddedness on negative shocks. *Journal of Vocational Behavior*, 76(1):42–51, 2010.

[7] Gerardo Canfora, Massimiliano Di Penta, Rocco Oliveto, and Sebastiano Panichella. Who is going to mentor newcomers in open source projects? In *Proc. Int'l Symposium Foundations of Software Engineering (FSE)*, page 44, New York, 2012. ACM Press.

[8] Maëlick Claes, Mika V Mäntylä, Miikka Kuutila, and Bram Adams. Do programmers work at night or during the weekend? In *ICSE*, pages 705–715. ACM, 2018.

[9] Jailton Coelho and Marco Tulio Valente. Why modern open source projects fail. In *ESEC/FSE*, pages 186–196. ACM, 2017.

[10] Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology Press, 2014.

[11] Eleni Constantinou and Tom Mens. An empirical comparison of developer retention in the rubygems and npm software ecosystems. *Innovations in Systems and Software Engineering*, 13(2-3):101–115, 2017.

[12] Eleni Constantinou and Tom Mens. Socio-technical evolution of the Ruby ecosystem in GitHub. In *SANER*, pages 34–44. IEEE, 2017.

[13] David Roxbee Cox. *Analysis of survival data*. Routledge, 2018.

[14] John W Creswell and Vicki L Plano Clark. *Designing and conducting mixed methods research*. Wiley Online Library, 2007.

[15] Kevin Crowston, Hala Annabi, James Howison, and Chengetai Masango. Effective work practices for software engineering: free/libre open source software development. In *Proceedings of the 2004 ACM workshop on Interdisciplinary software engineering research*, pages 18–26. ACM, 2004.

[16] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. Social coding in github: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1277–1286. ACM, 2012.

[17] Jon Eyolfson, Lin Tan, and Patrick Lam. Do time of day and developer experience affect commit bugginess? In *Proceedings of the 8th Working Conference on Mining Software Repositories*, pages 153–162. ACM, 2011.

[18] Fabian Fagerholm, Alejandro Sanchez Guinea, Jay Borenstein, and Jürgen Münch. Onboarding in open source projects. *IEEE Software*, 31(6):54–61, 2014.

[19] Yulin Fang and Derrick Neufeld. Understanding sustained participation in open source software projects. *Journal of Management Information Systems*, 25(4):9–50, 2009.

[20] Daniel C Feldman and Thomas WH Ng. Careers: Mobility, embeddedness, and success. *Journal of management*, 33(3):350–377, 2007.

[21] Matthieu Foucault, Marc Palyart, Xavier Blanc, Gail C Murphy, and Jean-Rémy Falleri. Impact of developer turnover on quality in open-source software. In *ESEC/FSE*, pages 829–841. ACM, 2015.

[22] Davide Fucci, Giuseppe Scanniello, Simone Romano, and Natalia Juristo. Need for sleep: the impact of a night of sleep deprivation on novice developers' performance. *IEEE Transactions on Software Engineering*, 2018.

[23] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

[24] Rishab A Ghosh, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. Free/libre and open source software: Survey and study – part 4: Survey of developers. Technical report, International Institute of Informatics, University of Maastricht, Maastricht, 2002.

[25] Benjamin Gidron. Predictors of retention and turnover among service volunteer workers. *Journal of Social Service Research*, 8(1):1–16, 1985.

[26] Georgios Gousios. The GHTorent dataset and tool suite. In *MSR*, pages 233–236. IEEE, 2013.

[27] Georgios Gousios, Andy Zaidman, Margaret-Anne Storey, and Arie Van Deursen. Work practices and challenges in pull-based development: the integrator's perspective. In *ICSE*, pages 358–368. IEEE, 2015.

[28] Patricia M Grambsch and Terry M Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

[29] Emitza Guzman, David Azócar, and Yang Li. Sentiment analysis of commit comments in github: an empirical study. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 352–355. ACM, 2014.

[30] Christoph Hannebauer and Volker Gruhn. On the relationship between newcomer motivations and contribution barriers in open source projects. In *Proc. Int'l Symposium on Open Collaboration*, OpenSym '17, pages 2:1–2:10, New York, NY, USA, 2017. ACM.

[31] Peter W Hom, Thomas W Lee, Jason D Shaw, and John P Hausknecht. One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3):530, 2017.

[32] Dirk Homscheid and Mario Schaarschmidt. Between organization and community: Investigating turnover intention factors of firm-sponsored open source software developers. In *Proc. Conf. Web Science (WebSci)*, pages 336–337, New York, NY, USA, 2016. ACM.

[33] Daniel Izquierdo-Cortazar et al. Using software archaeology to measure knowledge loss in software projects due to developer turnover. In *HICSS*, pages 1–10. IEEE, 2009.

[34] Irma Browne Jamison. Turnover and retention among volunteers in human service agencies. *Review of Public Personnel Administration*, 23(2):114–132, 2003.

[35] Hyosu Kim and Dennis Kao. A meta-analysis of turnover intention predictors among us child welfare workers. *Children and Youth Services Review*, 47:214–223, 2014.

[36] Miikka Kuutila, Mika V Mäntylä, Maëlick Claes, and Marko Elovainio. Reviewing literature on time pressure in software engineering and related professions: Computer assisted interdisciplinary literature review. In *Emotion Awareness in Software Engineering (SEmotion), 2017 IEEE/ACM 2nd International Workshop on*, pages 54–59. IEEE, 2017.

[37] Karim Lakhani and Robert G Wolf. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. Technical report, MIT Sloan working paper, 2003.

[38] Thomas W Lee and Terence R Mitchell. An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review*, 19(1):51–89, 1994.

[39] Bin Lin, Gregorio Robles, and Alexander Serebrenik. Developer turnover in global, industrial open source projects: Insights from applying survival analysis. In *Global Software Engineering (ICGSE), 2017 IEEE 12th International Conference on*, pages 66–75. IEEE, 2017.

[40] Vishal Midha and Prashant Palvia. Retention and quality in open source software projects. *AMCIS 2007 Proceedings*, page 25, 2007.

[41] Lynn E Miller, Gary N Powell, and Joseph Seltzer. Determinants of turnover among volunteers. *Human Relations*, 43(9):901–917, 1990.

[42] Rupert G Miller Jr. *Survival analysis*, volume 66. Wiley, 2011.

[43] Huilian Sophie Qiu, Alexander Nolte, Anita Brown, Alexander Serebrenik, and Bogdan Vasilescu. Going farther together: The impact of social capital on sustained participation in open source. In *International Conference on Software Engineering*, ICSE. IEEE, 2019.

[44] Jeffrey A Roberts, Il-Horn Hann, and Sandra A Slaughter. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science*, 52(7):984–999, 2006.

[45] Ariel Rodriguez, Fumiya Tanaka, and Yasutaka Kamei. Empirical study on the relationship between developers working habits and efficiency. In *Proc. Conf. Mining Software Repositories (MSR)*, 2018.

[46] Andreas Schilling, Sven Laumer, and Tim Weitzel. Who will remain? an evaluation of actual person-job and person-team fit to predict developer retention in FLOSS projects. In *HICCS*, pages 3446–3455. IEEE, 2012.

[47] Margrit Schreier. *Qualitative content analysis in practice*. Sage Publications, 2012.

[48] Sonali K Shah. Motivation, governance, and the viability of hybrid forms in open source software development. *Management Science*, 52(7):1000–1014, 2006.

[49] Pratyush N Sharma, John Hulland, and Sherae Daniel. Examining turnover in open source software projects using logistic hierarchical linear modeling approach. In *IFIP International Conference on Open Source Systems*, pages 331–337. Springer, 2012.

[50] Andrea Galiette Skoglund. Do not forget about your volunteers: A qualitative analysis of factors influencing volunteer turnover. *Health and Social Work*, 31(3):217, 2006.

[51] Sabine Sonnentag, Carmen Binnewies, and Eva J Mojza. Staying well and engaged when demands are high: The role of psychological detachment. *Journal of Applied Psychology*, 95(5):965, 2010.

[52] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. Social barriers faced by newcomers placing their first contribution in open source software projects. In *Proc. Conf. Computer Supported Cooperative Work (CSCW)*, pages 1379–1392, New York, NY, USA, 2015. ACM.

[53] Marat Valiev, Bogdan Vasilescu, and James Herbsleb. Ecosystem-level determinants of sustained activity in open-source projects: A case study of the PyPI ecosystem. In *ESEC/FSE*. ACM, 2018.

[54] Georg Von Krogh, Stefan Haefliger, Sebastian Spaeth, and Martin W Wallin. Carrots and rainbows: Motivation and social practice in open source software development. *Mis Quarterly*, 36(2):649–676, 2012.

[55] Joel West and Scott Gallagher. Challenges of open innovation: the paradox of firm investment in open-source software. *R&d Management*, 36(3):319–331, 2006.

[56] Yiqing Yu, Alexander Benlian, and Thomas Hess. An empirical study of volunteer members' perceived turnover in open source software projects. In *Proc. Hawaii Int'l Conf. System Sciences (HICSS)*, pages 3396–3405. IEEE, 2012.

[57] Minghui Zhou and Audris Mockus. Developer fluency: Achieving true mastery in software projects. In *ESEC/FSE*, pages 137–146. ACM, 2010.

[58] Minghui Zhou, Audris Mockus, Xiujuan Ma, Lu Zhang, and Hong Mei. Inflow and retention in OSS communities with commercial involvement: A case study of three hybrid projects. *ACM Trans. Softw. Eng. Methodol. (TOSEM)*, 25(2):13, 2016.