2018 ACM/IEEE 3rd International Workshop on Emotion Awareness in Software Engineering

On Measuring Affects of GitHub Issues' Commenters

Giuseppe Destefanis[1], Marco Ortu[2], David Bowes[1], Michele Marchesi[2], Roberto Tonelli[2]
University of Hertfordshire, Hatfield, UK[1]
{g.destefanis,d.h.bowes}@herts.ac.uk
University of Cagliari, Italy[2]
{marco.ortu,michele,roberto.tonelli}@diee.unica.it

ABSTRACT

In this study, we analyzed issues and comments on GitHub projects and built collaboration networks dividing contributors into two categories: users and *commenters*. We identified as commenters those users who only post comments without posting any issues nor committing changes in the source code. Since previous studies showed that there is a link between a positive environment (regarding affectiveness) and productivity, our goal was to investigate commenters' contribution to the project concerning affectiveness.

We analyzed more than 370K comments from 100K issues of 25K contributors from 3 open source projects. We then calculated and compared the affectiveness of the issues' comments written by users and commenters in terms of sentiment, politeness, and emotions. We provide empirical evidence that commenters are less polite, less positive and in general they express a lower level of emotions in their comments than users. Our results also confirm that GitHub's contributors consist of different groups which behave differently, and this provides useful information for future studies in the field.

KEYWORDS

mining software repositories, human factors, software engineering

ACM Reference Format:
Giuseppe Destefanis[1], Marco Ortu[2], David Bowes[1], Michele Marchesi[2], Roberto Tonelli[2]. 2018. On Measuring Affects of GitHub Issues' Commenters. In *Proceedings of (SEmotion'18).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Open source software development is a collaborative activity and tools such GitHub allowed researchers to understand better how such activity is affected by various aspects. Recently, research has focused on understanding how human emotions and mood affect the working environment, the processes, and therefore the final results of an activity. The effects have been studied concerning the application of software engineering practices and software development phases. It is well recognized that psychological aspects play a crucial role in understanding the phenomena. The foundation of emotions mining resides in psychology studies which map individual words to emotions.

SEmotion'18, May 2018, Gothenburg, Sweden

ACM ISBN ACM ISBN 978-1-4503-5751-7/18/06…$15.00
https://doi.org/10.475/123_4

In this paper, we empirically analyze more than 370K comments from 100K issues posted by 25K contributors on GitHub.

With the generic term "contributor" we refer to all the users related to a specific project under development on the GitHub platform. However, it is clear that contributors are a container of more specialized groups such us developers, users and, as we explained in this paper, *commenters*. But are these different roles characterized by distinct expression of emotions? Can we understand to which group a contributor belongs to by looking at how emotions are expressed? In previous studies [7, 8, 25, 26] it has been shown how positive emotions and high level of politeness have a positive impact on developers' productivity concerning lower issue fixing time and how developers tend to avoid impolite or "negative" states (from the emotions point of view) [24]. Since commenters only post comments and do not submit any code, it is interesting to study how they affect the general mood. Understanding how different contributors belonging to different groups express emotions when commenting issues, can provide new suggestions and shed new light for managers and/or developers who need to deal with team composition.

In this paper, we analyze communication patterns differentiating between users and commenters. We extracted affective metrics for emotion, sentiment, and politeness to study how these two different groups express emotions and aiming at answering the following questions:

RQ1: Do users and commenters communicate differently with respect to sentiment and politeness?
The answer is yes; commenters were less polite and more negative than users in the comments posted on issues.

RQ2: Do users and commenters communicate differently with respect to emotions?
Commenters expressed fewer emotions than users while they communicated with higher levels of Arousal, Valence, and Dominance.

The rest of the paper is organized as follows: we first discuss related work (Section 2). In Section 3, we describe how we measure affectiveness by measuring emotions, sentiment, and politeness in contributors' comments. Section 4 introduces how we built the issue collaboration graph and the methodology we followed for performing our experiments. In Section 5 we present and discuss our finding and analysis of threats to validity in Section 7. We finally draw our conclusions in Section 7.

2 RELATED WORK

Recent results on emotion analysis in software development, the presence of publicly available data repositories, and the availability of well-developed analysis tools are the basis of this study.

2.1 Emotion in Software Engineering

An important point in approaching emotional studies is represented by the work of Murgia et al. [20]. The authors studied whether issue reports carried emotional information about software development and found, by manually analyzing the Apache Software Foundation issue tracking system, that developers did express emotions like sadness, joy, and gratitude.

Feldt et al. [9] focused on personality as a relevant psychometric factor, presenting results from an empirical study about correlations between personality and attitudes to software engineering processes and tools. Authors defined the personality dimensions and found that higher levels of "conscientiousness" correlated with attitudes towards work style, openness to changes and task preference.

Acuna et al. [1], performed empirical research examining the work climate within software development teams. The authors attempted to understand if team climate (defined as the shared perceptions of teamwork procedures and practices) bore any relation to software product quality. They found that high team vision preferences and high participative safety perceptions of the team were significantly related to better software.

At the same time, researchers studied the relationship between affect and work-related achievements, including performance [18] and problem-solving processes, focusing on particular aspects such as creativity [2].

Since the work of Russell, to better monitor emotions, the amount of words that have been mapped to the Valence-Arousal space has increased considerably. In 2013, Warriner et al. [30] used 1,865 participants to rate 13,915 English words to create the most extensive collection of individual words mapped to a Valence-Arousal-Dominance space.

In 2013, Bazelli et al. [3] analyzed questions and answers on stackoverflow.com to determine the developer personality traits, using the Linguistic Inquiry and Word Count [27]. The authors found that the top reputed authors were more extroverted and expressed less negative emotions than authors of downvoted posts. Questions and answers on stackoverflow.com are the core of another ongoing study, proposed in 2014 by Novielli et al. [21], which aims at clarifying if emotions in question influence the quality of the answer, which may decrease the success of the question. To do this, the authors individuated elements which describe the independent variable (question) and described elements to define the dependent variable. In their opinion, recognizing the role of emotion lexicon could enable a better user experience and the development of new intelligent tools to better support on-line communities.

In 2014, Graziotin et al. [11] reported the results of an investigation with 42 participants about the relationship between the affective states, creativity, and analytical problem-solving skills of software developers. The results offered support for the claim that happy developers were better problem solvers regarding their analytical abilities. The authors provided a better understanding of the impact of affective states on the creativity and analytical problem-solving capacities of developers, introduced and validated psychological measurements, theories, and concepts of affective states, creativity, and analytical-problem-solving skills in empirical software engineering, and raised the need for studying the human factors of software engineering by employing a multi-disciplinary viewpoint. In another work, Graziotin et al. [13] conducted a qualitative interpretive study based on face-to-face open-ended interviews, in-field observations and e-mail exchanges. This enabled the authors to construct a novel explanatory theory of the impact of affects on development performance. The theory was explicated using an established taxonomy framework. The proposed theory built upon the concepts of events, affects, attractors, focus, goals, and performance.

2.2 Mining emotion from software repositories

Guzman et al. [14] investigated the relationship between affect and programming languages. Utilizing a sentiment analysis on Github's commit comments, they showed Java as the language with most associated negative affect, and no correlation between the number of Github stars and the affect of the commit messages.

Ortu et al. [23] analyzed the relation between sentiment, emotions, and politeness of developers for Jira comments and the issue resolution time. The results showed that positive emotions and politeness were related to shorter issue fixing time. On the other hand, negative emotions were linked with longer issue fixing time.

Murgia et al. [19] demonstrated the feasibility of a machine learning classifier for identifying issue comments containing gratitude, joy, and sadness. Such a classifier, using emotion-driving words and technical terms, obtained a good precision and recall for identifying the emotion love, while for joy and sadness a lower recall was obtained.

Islam et al. [15] studied how developers' emotions affect the software development process. They reconstructed the emotional variations using a sentiment analysis on commit messages acquired using the Boa public API and examining it with SentiStrength. They found that emotions influenced differently different typologies of activity. Results showed that in bug-fixing and refactoring activities, positive emotions are dominant. Energy-aware tasks are characterized by neutral emotions.

According to Jongeling et al. [16], every typology of sentiment analysis must take in account that currently available tools do not provide the human perception of sentiment. Also, the selection of sentiment analysis tools impact on the study results.

3 BACKGROUND

GitHub is a popular platform for social development. It has rapidly become one of the largest and most used software repositories in the world [5]. It offers a large variety of features including, versioning control system, issue tracking system, wikis and pull request mechanisms. It is mostly focused to ease contributor collaboration offering a simple code and functionality management. It is a flexible tool enabling users to decide the development workflow that best suits their needs. Its issue tracking system is designed for allowing projects contributor to share their discussion during the development process in a social fashion, e.g., using emoji [1] in contributors' comments.

3.1 Measuring Affects

3.1.1 Emotions. To measure emotion we used a machine learning classifier proposed by Ortu et al. [22] and extended by Murgia et al. [19]. As reported in [19], they parsed Apache's Jira-based repository in July 2013, fetching all the issue reports since the 19th of October 2000 building a classifier model able to detect *love* (with 0.82 F1 measure), *joy measure* (with 0.7 F1 measure), *anger* (with 0.82 F1 measure) and *sadness* (with 0.84 F1 measure).

We adopted their best classifier for detecting emotions such as *love, joy, sadness* and *anger* in comments related to issues posted on GitHub.

3.1.2 Politeness. To evaluate the politeness of the contributors present in our dataset, we used the tool developed by Danescu et al. [6]. Given an input text, the tool calculates its overall politeness and provides a binary output, i.e. *polite* or *impolite*.

The tool has been trained and validated through machine learning using a set of over 10,000 manually labeled requests from StackOverflow[2] and Wikipedia[3]. The set included comments written by authors from all over the world while the annotators were selected among U.S. residents, based on a linguistic background questionnaire. It was built to include comments written by authors from all over the world while the annotators were selected among U.S. residents, based on a linguistic background questionnaire. We analyzed GitHub commit messages written by contributors, therefore the content of the messages and the language used is very similar to the one used on StackOverflow.

3.1.3 Sentiment. We used Sentistrength [29] for computing the sentiment of comments posted by contributors in issue reports. Given a text, Sentistrength associate a score to each word based on the emotional content conveyed by the word. It is able to report two sentiment strengths, or polarity: negative that ranges from *-1 (not negative)* to *-5 (extremely negative)* and positive from *1 (not positive)* to *5 (extremely positive)*. Given the strength of each word, we then compute the overall sentiment of an issue's comment as the average of these two sentiment strengths.

3.1.4 Valence, Arousal, and Dominance. Valence, Arousal, and Dominance (VAD) are emotional dimensions able to derive a person's interest (attraction), level of activation and perceived level of control for a particular situation from textual communication. To evaluate VAD metrics we adopted the same approach proposed by Mäntylä et. al. [17]. All measures of VAD are based on a list of words that have manually been analyzed and assigned a VAD score. Warriner et al.'s [30] leading lexicon contains 13,915 English words with VAD scores for Valence, Arousal, and Dominance. To calculate the corresponding VAD scores for a piece of text (i.e., a list of words $\bar{w} = [w_1, w_2, ..., w_n]$), the Range of the words' individual VAD scores is computed by taking the two words with the Max and Min Valence, Arousal or Dominance. For the special cases when Max has lower than average value or when Min has higher than the average value we set the Max or Min to the average of all words of the lexicon ($\bar{W} = [W_1, W_2, ..., W_N]$, where N is 13,915).

[1] https://en.wikipedia.org/wiki/Emoji
[2] http://stackoverflow.com
[3] https://en.wikipedia.org/wiki/Main_Page

4 EXPERIMENTAL SETUP

This research aims to understand whether users and commenters express different emotions than other contributors while posting comments on GitHub. We define as a *commenter* a project contributor without neither commits nor issues posted on a project, and define a *user* a contributor that does not match the definition we provided of a commenter. We built a collaboration network based on issues and comments posted by contributors, distinguishing between commenters and users, then we computed the affects metrics.

4.1 The dataset

We used the GHTorrent dataset [10] identifying about ~25K GitHub's contributors (both users and commenters), ~100K issues and ~380K comments from three open source systems.

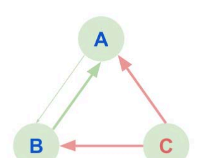
Project	Age	Language	# Users	# Commenters	# Issues	# Comments
Rails	13 years	Ruby	6010	4499	31731	138039
Homebrew	8 years	Ruby	6192	3443	25235	95342
Salt	6 years	JavaScript	3922	1651	43846	142345
tot	-	-	16124	9593	100812	377450

Table 1: Dataset statistics

Table 1 shows the language, number of issues, numbers of contributors and number of comments of the selected projects. We verified that there are no cross-projects contributors, namely each contributor worked for only one of the considered projects.

4.2 Issue Collaboration Graph

GitHub is a social development platform, and as such, its issue tracking system manly focuses on collaboration to help contributors to have a easy and direct way of keeping track of tasks, enhancements, bugs and issues in general.



	In Degree	Out Degree
A	4	1
B	3	2
C	0	4

Figure 1: Issue based Collaboration Network.

Figure 1 presents a simplified scenario which shows how we built the collaboration network based on issues and comments. We generate the issue collaboration network considering each contributor as a node. When a contributor *A* comments on an issue created by contributor *B* then an edge goes from node *A* to node *B*. The more contributor *B* comments on issues created by contributor *A* the thicker the edge from *B* to *A* becomes.

The right part of Figure 1 shows a table which summarizes the in-degree and out-degree of the network. In-degree represents the number of arrows (edges) incidents to a node, while out-degree represents the number of edges going out from a node. In this example we identify *C* as a *Commenters*, and *A* and *B* as *Users*. Table 1 shows in details the number of *users* and *commenters* we analyzed.

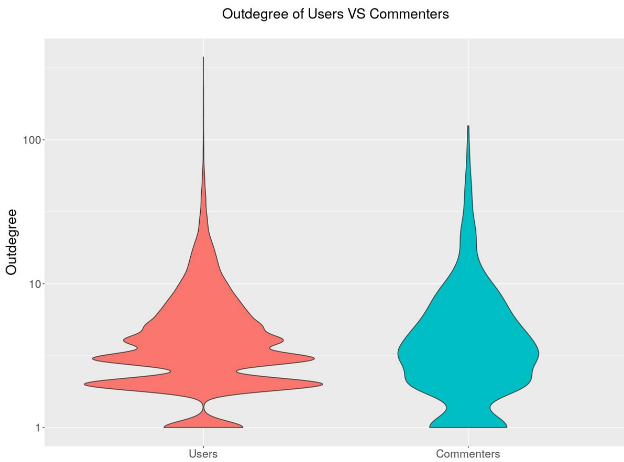


Figure 2: Contributors and commenters outdegree.

Figure 2 shows the out-degree for users and commenters in a log-scale. As it is possible to observe, users and commenters are two different groups in term of out-degree. Users are characterized by a higher distribution of number of comments for higher level of out-degree and they are *clustered* in roughly three groups. Commenters are mostly distributed in lower values of out-degree than users.

5 RESULTS

Project	Politeness			Sentiment		
	User	Commenter	p-value	User	Commenter	p-value
Rails	0.453	0.365	***	0.053	0.011	***
Homebrew	0.453	0.396	***	0.091	0.031	***
Salt	0.481	0.422	***	0.068	-0.029	***
All projects	0.447	0.384	***	0.068	0.024	***

Table 2: Developers and users Politeness and Sentiment means.
p-values codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

RQ1: Do users and commenters communicate differently with respect to sentiment and politeness?
Motivation. Research in the last years has shown that the sentiment and politeness expressed in text comments by contributors of open source software are linked with productivity (regarding issue fixing time), performance [12], and also with the attractiveness of a project for new developers [20, 22, 26]. Previous studies considered the general concept of contributors to an issue tracking system such as Jira and GitHub, without distinguishing between roles (e.g., developer, commenter). In Section 4.2 we suggested that there is a difference on how users and commenters communicate, here we quantitatively analyzed such differences regarding sentiment and politeness affective metrics.

Approach. In Section 4 we describe the process to construct the issue collaboration graph exploiting the GitHubTorrent dataset. We retrieved all the comments related to about 100K issues and computed the sentiment and politeness of all contributors' comments. Then, we distinguished between *users* and *commenters*.

Finally, we analyzed the distributions of sentiment and politeness, as explained in Section 4.

To compare sentiment and politeness metrics between the two groups, we used the Wilcoxon test Siegel [28], which is non-parametric and thus can be used with no restrictions nor hypotheses on the statistical distribution of the sample populations. The test is suitable for comparing differences in the averages or the medians of two populations when their distributions are not Gaussian. For the analysis, we use the one-sided Wilcoxon rank sum test using the 5% significance level (i.e., p-value < 0.05).

Findings. Commenters were less polite and more negative than Users in the comments posted on issues.

Figure 3 shows that commenters (represented in blue) are more negative than users (represented in red). It can be observed that commenters are largely distributed from neutral to extremely negative values of sentiment, on the contrary users are more distributed toward positive values of sentiment (from positive to extreme positive values). This means that commenters tend to comment on issues with more negative sentiment than users.

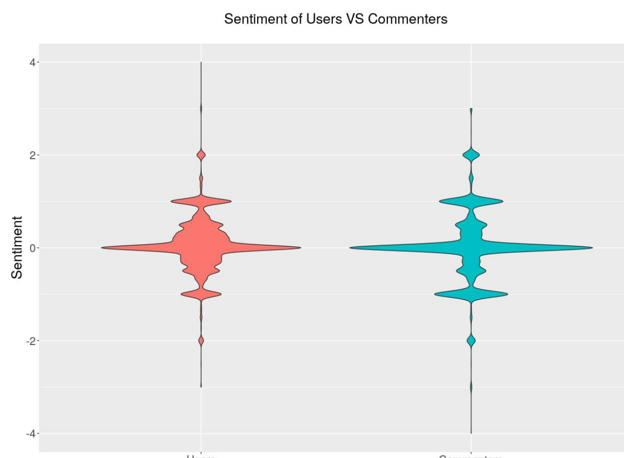


Figure 3: Contributors and commenters sentiment.

Considering politeness, Figure 4 shows that commenters (represented in blue) are more impolite than users (represented in red). A value of 1 means polite comments, a value of 0.5 same number of polite/impolite comments and 0 means impolite comments. We can observe that commenters are primarily distributed with impolite comments, on the contrary users are more distributed toward neutral and polite comments. This means that commenters tend to comment on issues with more impolite comments than users.

Project	Love			joy			anger			sadness		
	User	Commenter	p-value	User	Commenter	p-value	User	Commenter	p-value	User	Commenter	P-value
Rails	0.209	0.106	***	0.086	0.078	***	0.392	0.292	***	0.030	0.030	***
Homebrew	0.246	0.161	***	0.095	0.078	***	0.487	0.390	***	0.031	0.0177	***
Salt	0.23	0.112	***	0.096	0.057	***	0.388	0.335	***	0.025	0.019	***
All projects	0.218	0.134	***	0.091	0.080	***	0.030	0.023	***	0.434	0.336	***

Table 3: Developers and users out-emotion metrics means. p-values codes: 0 '**' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1**

Project	Arousal			Valence			Dominance		
	User	Commenter	p-value	User	Commenter	p-value	User	Commenter	p-value
Rails	1.277	1.569	***	1.853	2.259	0.1	1.847	2.212	0.06
Homebrew	1.199	1.299	0.1	1.766	1.878	0.1	1.765	1.87548	0.5
Salt	1.29	1.39	0.06	1.835	1.955	***	1.826	1.944	**
All projects	1.239	1.43	***	1.811	2.077	*	1.807	2.05	**

Table 4: Contributors and users VAD metrics means. p-values codes: 0 '**' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1**

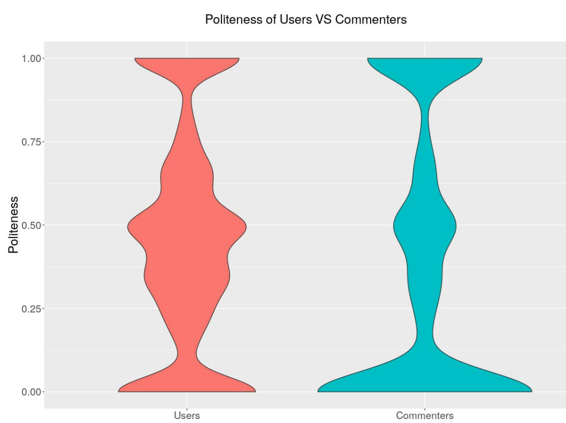


Figure 4: Contributors and commenters politeness.

Results show that commenters and users communicate differently and that they are characterized by different levels of sentiment and politeness.

Table 2 presents in more details the difference in sentiment and politeness of commenters and users, showing the two affects metrics means (cells in red higher values, cells in grey lower values) along with the p-value of the Wilcoxon test.

Results are confirmed for both single projects and when considering all the projects together. For all projects, sentiment and politeness means are lower for commenters. For Salt project it can be observed that, differently from other projects, sentiment mean is negative, meaning that for that particular project commenters are remarkably more negative than users.

RQ2: Do users and commenters communicate differently with respect to emotions?

Motivation. Emotions, e.g., love, joy, anger, and sadness, expressed in text comments by contributors of open source software are linked with productivity (regarding issue fixing time) performance [22]. Emotion dimensions, e.g., arousal, valence, and dominance, have a different influence on issues characteristics [17] such as issue resolution and priority. Commenters and users communicate differently, here we are interested in quantifying such differences in terms of emotions measured as explained in Sections 3.1.1 and 3.1.4.

Approach. As for answering the previous research question, we divided developers into two groups: commenters and users. We then compared the communication metrics between the two groups characterizing four emotions, *love*, *joy*, *anger* and *sadness* and the VAD metrics. We than compared means of the two groups using the Wilcoxon Test Siegel [28].

Findings. Commenters expressed fewer emotions than users while they communicated with higher levels of Arousal, Valence, and Dominance.

Considering emotions, Table 3 shows that users express more emotions in their comments than commenters. This result is confirmed when analyzing both single projects and all projects together.

Considering VAD metrics, Table 4 shows that commenters present higher level of VAD metrics in their comments than users. In this case, we can observe clear results only when considering all projects together while, when considering single projects, we found significant differences just for *arousal* in Rails and *valence* and *dominance* in Salt projects.

6 THREATS TO VALIDITY

Threats to external validity correspond to the generalization of experimental results. In this study, we used several empirical approaches to evaluate the collaboration network of three projects from GitHub repositories and computed the affects of about 370K comments related to the GitHub projects. We considered the datasets as a representative sample of the open source world. Hence, replications on commercial and other open source projects are needed to confirm our findings.

Threats to internal validity concern confounding factors that can influence the obtained results. Based on empirical evidence, we supposed a causal relationship between the emotional state of developers and what they write in issue reports [30].

Another threat is related to the reliability of emotions, sentiment and politeness tools applied in the software engineering domain. For the emotion detection we used a tool specifically trained using Jira comments [19] while for detecting politeness we used the tool developed by Danescu et al. [6] which was trained using StackOverflow questions and answers following the same approach of Celefato et al. [4] for sentiment detection in software engineering.

Threats to reliability validity correspond to the degree to which the same data would lead to the same results when repeated. This research manually investigate the affect of developers and users during software development by means of collaboration networks, no previous studies in this field exists to compare our findings.

7 CONCLUSIONS

In this study, we examined affects of GitHGub contributors based on collaboration networks. We analyzed more than 370K comments from 100K issues of 25K users and commenters from 3 open source projects and discovered that users and commenters express emotions differently. We then mined contributors sentiment, politeness and emotions in issues' comments. We built upon previous results in the field of human factors in software engineering performing finer grain analysis and differentiating between users and commenters. The results showed that the considered groups play different roles in a project and communicate differently with different affects.

REFERENCES

[1] Silvia T Acuña, Marta Gómez, and Natalia Juristo. 2008. Towards understanding the relationship between team climate and software quality - a quasi-experimental study. *Empirical software engineering* 13, 4 (2008), 401–434.

[2] Teresa Amabile, Sigal G. Barsade, Jennifer S. Mueller, and Barry M. Staw. 2005. Affect and Creativity at Work. *Administrative Science Quarterly* 50, 3 (Sept. 2005), 367–403. DOI:http://dx.doi.org/10.2189/asqu.2005.50.3.367

[3] Blerina Bazelli, Abram Hindle, and Eleni Stroulia. 2013. On the personality traits of Stack Overflow users. In *Software Maintenance (ICSM), 2013 29th IEEE International Conference on.* IEEE, 460–463.

[4] Fabio Calefato, Filippo Lanubile, Federico Maiorano, and Nicole Novielli. 2017. Sentiment Polarity Detection for Software Development. *Empirical Software Engineering* (2017), 1–31.

[5] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work.* ACM, 1277–1286.

[6] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of ACL.*

[7] Giuseppe Destefanis, Marco Ortu, Steve Counsell, Stephen Swift, Michele Marchesi, and Roberto Tonelli. 2016. Software development: do good manners matter? *PeerJ Computer Science* 2 (2016), e73.

[8] Giuseppe Destefanis, Marco Ortu, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. 2017. On the randomness and seasonality of affective metrics for software development. In *Proceedings of the Symposium on Applied Computing.* ACM, 1266–1271.

[9] Robert Feldt, Richard Torkar, Lefteris Angelis, and Maria Samuelsson. 2008. Towards individualized software engineering: empirical studies should collect psychometrics. In *Proceedings of the 2008 international workshop on Cooperative and human aspects of software engineering.* ACM, 49–52.

[10] Georgios Gousios. 2013. The GHTorrent dataset and tool suite. In *Proceedings of the 10th Working Conference on Mining Software Repositories (MSR '13).* IEEE Press, Piscataway, NJ, USA, 233–236. http://dl.acm.org/citation.cfm?id=2487085.2487132

[11] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2014. Happy software developers solve problems better: psychological measurements in empirical software engineering. *PeerJ* 2 (2014), e289.

[12] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2014. Software Developers, Moods, Emotions, and Performance. *IEEE Software* 31, 4 (2014).

[13] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. 2015. How do you feel, developer? An explanatory theory of the impact of affects on programming performance. *PeerJ Computer Science* 1 (2015), e18.

[14] Emitza Guzman, David Azócar, and Yang Li. 2014. Sentiment analysis of commit comments in GitHub: an empirical study. In *Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014.* ACM Press, New York, New York, USA, 352–355. DOI:http://dx.doi.org/10.1145/2597073.2597118

[15] M. R. Islam and M. F. Zibran. 2016. Towards understanding and exploiting developers' emotional variations in software engineering. In *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA).* 185–192. DOI:http://dx.doi.org/10.1109/SERA.2016.7516145

[16] R. Jongeling, S. Datta, and A. Serebrenik. 2015. Choosing your weapons: On sentiment analysis tools for software engineering research. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME).* 531–535. DOI:http://dx.doi.org/10.1109/ICSM.2015.7332508

[17] Mika Mäntylä, Bram Adams, Giuseppe Destefanis, Daniel Graziotin, and Marco Ortu. 2016. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?. In *Proceedings of the 13th International Conference on Mining Software Repositories.* ACM, 247–258.

[18] A. G. Miner and T. M. Glomb. 2010. State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes* 112, 1 (May 2010), 43–57. DOI:http://dx.doi.org/10.1016/j.obhdp.2009.11.009

[19] Alessandro Murgia, Marco Ortu, Parastou Tourani, Bram Adams, and Serge Demeyer. 2017. An exploratory qualitative and quantitative analysis of emotions in issue report comments of open source systems. *Empirical Software Engineering* (2017), 1–44.

[20] Alessandro Murgia, Parastou Tourani, Bram Adams, and Marco Ortu. 2014. Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In *Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014.* ACM Press, New York, New York, USA, 262–271. DOI:http://dx.doi.org/10.1145/2597073.2597086

[21] Nicole Novielli, Fabio Calefato, and Filippo Lanubile. 2014. Towards discovering the role of emotions in Stack Overflow. In *Proceedings of the 6th International Workshop on Social Software Engineering.* ACM, 33–36.

[22] Marco Ortu, Bram Adams, Giuseppe Destefanis, Parastou Tourani, Michele Marchesi, and Roberto Tonelli. 2015. Are Bullies More Productive? Empirical Study of Affectiveness vs. Issue Fixing Time. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories.* IEEE, 303–313. DOI:http://dx.doi.org/10.1109/MSR.2015.35

[23] Marco Ortu, Giuseppe Destefanis, Bram Adams, Alessandro Murgia, Michele Marchesi, and Roberto Tonelli. 2015. The JIRA Repository Dataset. In *Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering - PROMISE '15.* ACM Press, New York, New York, USA, 1–4. DOI:http://dx.doi.org/10.1145/2810146.2810147

[24] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. 2016. Arsonists or firefighters? Affectiveness in agile software development. In *International Conference on Agile Software Development.* Springer, 144–155.

[25] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. 2017. How diverse is your team? Investigating gender and nationality diversity in GitHub teams. *Journal of Software Engineering Research and Development* 5, 1 (2017), 9.

[26] Marco Ortu, Giuseppe Destefanis, Mohamad Kassab, Steve Counsell, Michele Marchesi, and Roberto Tonelli. 2015. Would you mind fixing this issue? An Empirical Analysis of Politeness and Attractiveness in Software Developed Using Agile Boards. In *Agile Processes, in Software Engineering, and Extreme Programming.* Springer, 129–140.

[27] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71 (2001), 2001.

[28] Sidney Siegel. 1956. Nonparametric statistics for the behavioral sciences, New York, 1956. *Sisdpalvelusohjesaant& (The manual of Interior Duty), Helsinki* (1956).

[29] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment in Short Strength Detection Informal Text. *J. Am. Soc. Inf. Sci. Technol.* 61, 12 (Dec. 2010), 2544–2558.

[30] Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45, 4 (dec 2013), 1191–1207. DOI:http://dx.doi.org/10.3758/s13428-012-0314-x