

The Role of Emotions in Contributors Activity: A Case Study of the GENTOO Community

David Garcia, Marcelo Serrano Zanetti and Frank Schweitzer
Chair of Systems Design – www.sg.ethz.ch – ETH Zurich
dgarcia@ethz.ch

Abstract—We analyze the relation between the emotions and the activity of contributors in the Open Source Software project GENTOO. Our case study builds on extensive data sets from the project’s bug tracking platform BUGZILLA, to quantify the activity of contributors, and its mail archives, to quantify the emotions of contributors by means of sentiment analysis. The GENTOO project is known for a period of centralization within its bug triaging community. This was followed by considerable changes in community organization and performance after the sudden retirement of the central contributor. We analyze how this event correlates with the negative emotions, both in bilateral email discussions with the central contributor, and at the level of the whole community of contributors. We then extend our study to consider the activity patterns of GENTOO contributors in general. We find that contributors are more likely to become inactive when they express strong positive or negative emotions in the bug tracker, or when they deviate from the expected value of emotions in the mailing list. We use these insights to develop a Bayesian classifier that detects the risk of contributors leaving the project. Our analysis opens new perspectives for measuring online contributor motivation by means of sentiment analysis and for real-time predictions of contributor turnover in Open Source Software projects.

I. INTRODUCTION

Collaboration within an online environment is an everyday challenge for contributors of Open Source Software (OSS) projects. They need to interact with other contributors to decide about the direction of their project and, equally important, need to interact with users to learn about their demands. Communication within the contributor’s community and towards the user’s community both impact project reputation and the availability of resources, which are crucial to further develop the project. Thus, understanding how people interact, collaborate and communicate online is an important field of research that has the potential to improve the performance of OSS projects.

The relevance of OSS projects goes beyond research, and reaches wide industrial applications. The current technological landscape is constantly influenced by large OSS projects that generate important software products. For example, the APACHE server is used in more than 60% of the websites¹, and FIREFOX and CHROME have a combined market share of more than 50%². These are possible thanks to the efforts of OSS projects, in which potentially large amounts of contributors can participate by coding, proposing functionalities, or reporting and triaging bugs. All contributors benefit equally from the project, receiving the software product and its code as a result. These contributors are free to stop collaborating at any time; a decision that does not prevent them from profiting from

the project and using its products. In this sense, an OSS community is an example of a public goods game [1], in which participants have no punishment for free-riding, and they equally benefit from the common good. This poses a paradox, as the game theoretical result of the “tragedy of the commons” [2] implies that, when collaborators are purely rational, the expected outcome of the project is a complete failure.

In proprietary software projects, developers, analysts, and testers are bound by legal contracts that provide a mechanism to cope with conflicts and guarantee a certain level of collaboration. On the other hand, OSS projects are mostly composed of volunteer contributors, whose collaboration scheme can be fragile and suffer in the presence of disagreements or loss of motivation. For example, the PIDGIN project developed a program for instant messaging commonly used in LINUX distributions³. After the release of a new version, users, developers and other contributors disagreed on a change related to its user interface, leading to a heated discussion and the expression of negative emotions⁴. As a result, the project was divided (i.e. forked) into two different projects, which is equivalent to a large exodus of contributors. This example illustrates the impact that the emotional climate of an OSS community has on its success. Certain level of positive emotions seems necessary to sustain the intrinsic motivation of the collaborators, and strong instances of negative emotions pose a threat that trigger the turnover of important contributors.

The human factor of OSS projects composes the mechanism that make them possible, but also poses a threat that endangers their success. Often, the social component of the projects is analyzed through social network analysis [3], [4], [5], [6], but the psychological component of OSS interaction has not been explored so far. Thanks to the development of tools for sentiment analysis [7], we can quantify the emotions of OSS contributors, looking for relations between their activity and emotional expression. Furthermore, given the availability of large datasets of OSS development forums, this sentiment analysis can be extended to higher levels of aggregation in which collective emotions emerge from the interaction of individual contributors. This poses the opportunity to empirically analyze the conditions that lead to the turnover of OSS contributors, and to apply such findings in the creation of tools to monitor and predict the evolution of OSS projects [8], [9], [10].

In this article, we explore preconditions for contributor turnover, and their impact on the performance and cohesion of the community. We focus on the large GENTOO project,

¹<http://w3techs.com/technologies/details/ws-apache/all/all>

²<http://www.w3counter.com/globalstats.php?year=2013&month=04>

³<http://www.pidgin.im/>

⁴<https://developer.pidgin.im/ticket/4986>

by analyzing two disjoint datasets spanning about 10 years of activity and more than 35,000 contributors. The first dataset contains the records of bug triaging and processing (i.e. bug tracker) while the second contains messages exchanged within the developers' mailing list. To these data, we apply sentiment analysis, looking for indicators to predict the turnover of contributors from an OSS community. Our quantitative results show that the emotional expression of a contributor is an indicator for the likelihood of a contributor to remain active in the project. Finally, we apply these results to formulate a real-time prediction of contributors leaving the project, providing a tool that enables timely reactions against undesirable turnover events.

II. RELATED WORK

A. Social Dynamics of Open Source Software

The social organization in open source communities has been addressed in a number of relevant works. In [11], the focus was in division of labor. By analyzing a dataset composed of the APACHE and MOZILLA projects, the authors show that while coding efforts are concentrated on a few *core-developers*, maintenance activities, such as bug report triaging, are performed by a much larger community. In [3], this core periphery structure was also studied within a social network analysis framework. The social network framework was also recently applied by [5] for the study of the behavior of individuals within communities. Contributor motivation and its relationship to project performance is also an important topic, and was considered in a number of works reviewed in [12], [13], [14]. Finally, [15] proposes a framework to analyze the congruence between technical and social organization within a software project. In this way, the authors wish to answer the question of which social organization structure is the best performing given a particular technological scenario. Or the analogous, how to structure a technical architecture in order to fit an established social organization.

B. Emotions in Social Media

The most common mechanisms for communication in OSS projects are forums and bug trackers, which are special cases of social media. This allows the application of sentiment analysis tools [7], providing insights to the psychological experience [16] of OSS contributors, rather than just their social interaction. This approach has been proved useful for the analysis of collective emotions in forum discussions [17], emotional interaction in chatting communities [18], and to test previous hypotheses from psychology in online data [19], [20]. The attention to sentiment analysis is increasing due to its multiple applications in finance and marketing. For example, mood measures from social media have been used to predict the stock market [21], [22]. Sentiment analysis has also been applied to customer emotions in Amazon product reviews [23], and to the viral spread of information in TWITTER [24].

Different sentiment analysis tools are available, depending on the type of analysis and data to process. Supervised methods use training data to mine emotions and opinions from text [25], and word category frequencies can be used to measure collective mood [26], [21]. Regarding short and informal text, lexicon-based classification provides unsupervised methods to extract sentiment. The state-of-the-art tool in such situation is SENTISTRENGTH [7], [27], which we use in this article. The accuracy of its last version has been validated with human annotations of a wide variety of online communities [7].

Among its previous applications, SENTISTRENGTH has been used to analyze emotions about political topics in YOUTUBE and TWITTER [28], [29], product reviews [23], and YAHOO! ANSWERS [30].

C. Social Resilience and Contributor Motivation

The question of how groups are formed and disappear has been addressed for online social networks and scientific communities [31], [32]. This highlights the relevance of trust networks in social recommender systems [33], and how social movements in TWITTER grow and decay through spreading patterns and complex contagion [34]. The departure of individual users, commonly denoted as churn, has also been analyzed for the online communities like YAHOO! ANSWERS [35], and other social networks [36]. Furthermore, previous analysis provide insights on the decision of users to leave P2P networks [37], discussion boards [38], and online videogames [39]. These previous works focused on the relation between social indicators, like amount of contacts, with the likelihood of users to leave an online community. While useful as a first approximation, these analyzes did not take into account emotional expression and interaction, which are related in the psychology literature to motivation and social interaction [40], [12], [14].

The microscopic dynamics that drives the decisions of users to leave a community create the macroscopic effect of social resilience [41], or how strong is the community when facing disrupting periods. Such disrupting events have been characterized by text analysis on FACEBOOK [42], but their influence on the survival of an online community at large cannot be simply mapped to its social network [41]. The intrinsic motivation of the users and their individual decisions are key factors for the collective dynamics of the community. As an example, external incentives do not guarantee more efficient viral marketing campaigns [43]. On the other hand, information spread can be motivated by emotional content [24], leading to higher levels of user activity and interaction when emotions are involved.

III. GENTOO DATASETS

A. Bug reports

The GENTOO project adopts the BUGZILLA as its bug tracking system [44]. It is composed of an online database⁵ where each entry is organized around the notion of a *bug report*. A bug report status will change as its processing progresses towards a solution (e.g. *pending*, *reproduced*, *closed*, etc). In general, the modification of a bug report field (e.g. status) allows its author to leave simple text comments. Using the BUGZILLA API, we collect the time series of comments, along with the unique *username* of its author and the unique *id* of the respective bug report. In TABLE I, we summarize the main statistics of this dataset. To each of those comments, we apply the SENTISTRENGTH tool in order to quantify its positive and negative valence.

TABLE I. BASIC STATISTICS OF THE DATASETS USED FOR THIS STUDY. WE COVER ACTIVITY WITHIN GENTOO'S BUG TRACKER AND WITHIN THE GENTOO-DEV MAILING LIST.

Statistics	GENTOO BUGZILLA	GENTOO-DEV
Observation period	01/04/2002 to 04/26/2012	04/01/2001 to 29/06/2012
Messages	661,783	81,328
Discussions	140,216	14,070
Contributors	36,555	4,664

⁵<https://bugs.gentoo.org/>

B. Developer mailing list

While triaging and processing bug reports, contributors may rely on information exchange through mailing lists. In the case of GENTOO, this is mainly done via the *gentoo-dev* list⁶, which is the list subscribed by core-developers and code maintainers. Thus if contributors processing bug reports want to call the attention of a serious maintainer, that is the best place to start. Messages sent to this mailing list are stored in a database and can be retrieved at any time from their archive version, which is accessible via a HTML interface. Using this channel, we extract the time series of email messages sent to the *gentoo-dev* list, along with the unique *userid* of its author and message subject, which repeats for all messages sent to the same thread. Again, the textual content of each message is analyzed with SENTISTRENGTH yielding positive and negative valence scores (i.e. excluding content commented out with character “>” at the start of each new line).

C. Sentiment analysis

We process all comments and messages in the bug reports and the developer’s mailing list using SENTISTRENGTH [7]. SENTISTRENGTH is the state-of-the-art tool for lexicon-based analysis of social media messages, in particular for informal communication. It has been validated on test datasets including DIGG and other fora on specialized topics, which are similar communication media as the GENTOO bug tracker and mailing list. When classifying the polarity of forum messages, SENTISTRENGTH has an accuracy above 88% for DIGG, and above 90% in other fora [7]. It has high correlation values with human raters on these communities, providing sentiment scores that would be indistinguishable from a human rater, and providing not only an accurate, but also a *valid* estimation of the sentiment. For these reasons, previous works have applied it to YAHOO! ANSWERS [30], TWITTER messages [24], and chatroom communication [18].

SENTISTRENGTH uses a lexicon of emotional-bearing terms combined with the detection of negations, amplifiers and diminishers. Its output is composed of two values, a measure of positive sentiment $p \in [+1, +5]$, and a measure of negative sentiment $n \in [-1, -5]$. Following the rationale of [7], we can aggregate these two values to a measure of polarity. A message m is classified as positive ($s = +1$) if $p + n > 0$, negative ($s = -1$) if $p + n < 0$, or neutral ($s = 0$) if $p = n$ and both have an absolute value lower than 4. Comments with high and equal positive and negative sentiment do not map to this unidimensional simplification. Nevertheless, this approximation is valid in our data analysis, as only 265 messages were detected as in the cases of $[+4, -4]$ or $[+5, -5]$. These messages compose 0.032% of the total, and we discard them from our analysis.

TABLE II. MESSAGE RATIO PER POLARITY WITHIN GENTOO’S BUG TRACKER AND WITHIN THE GENTOO-DEV MAILING LIST.

Polarity	GENTOO BUGZILLA	GENTOO-DEV
positive	0.28	0.28
neutral	0.56	0.49
negative	0.16	0.23

IV. THE DEPARTURE OF A CENTRAL CONTRIBUTOR

In this section we quantify and discuss the role of emotions in a case study focused on the GENTOO-LINUX project. The GENTOO project is of particular interest due to a well documented centralization event followed by significant changes in

community performance [45]. In that work, we focus on the evolution of social organization within GENTOO’s bug triaging community. Using a quantitative methodology based on social network analysis [46], [6], we show that we can monitor drastic changes in social organization which are usually associated with increased risks. More specifically, the bug triaging community of GENTOO came to rely on a single person (i.e. named *Alice*) to help them in processing bug reports. This is in accordance with previous findings relating centrality to preference in collaboration [5]. Based on *Alice*’s activity, we divide the timespan of our dataset into three observation periods P_1 , P_2 , P_3 . In period P_1 , between January 2002 and October 27, 2004, *Alice* was not yet active and the community was growing. During the second period P_2 starting on October 28 2004, *Alice* gradually became the most central contributor. She unexpectedly left the community after her last contribution on March 29 2008, which marks the start of the third period P_3 in which *Alice* was not active anymore. In the next we discuss *Alice*’s impact on community performance and the possible effects of emotions on her motivation to leave the project.

A. Effect in performance

During P_2 , *Alice* concentrated most of the work related to bug triaging on herself, and as a result, the time to first reply and finally solve open bug reports were minimized. These are important metrics that correlate to the likelihood of a bug reporter in becoming a *long time contributor* to the project [47]. The main issue about *Alice*’s impact was that – due to personal conflicts, and dissatisfaction with the social environment of the project as a whole – she left the community suddenly. As we show in [45], after *Alice*’s retirement (i.e. period P_3) the community never managed to achieve the same levels of performance. Thus, besides monitoring changes in community social organization and its implied risks, we wish for quantitative measures that could give an early indication to individual loss of motivation or activity.

B. Changes in collective emotions

To measure the collective emotions in the discussions associated with comments to a bug report or e-mails to a thread in the mailing list, we aggregate the emotional values of the messages in the discussion. In this way, for the set of messages in a discussion M_d , we calculate the ratios of positive $P_d = \frac{\sum_{m \in M_d} s_m = 1}{|M_d|}$, negative $N_d = \frac{\sum_{m \in M_d} s_m = -1}{|M_d|}$, and neutral $U_d = \frac{\sum_{m \in M_d} s_m = 0}{|M_d|}$ messages. These measurements map the discussions to a simplex on the plane [28], where each discussion has a distance to the vertices of a triangle proportional to P_d , N_d , and U_d . Fig. 2 shows this representation separately for bug tracker and the mailing list, where each discussion d is a point of size proportional to $|M_d|$.

The ratio of the overall emotional expression in each medium (i.e. bug tracker or mailing list), $\bar{P}, \bar{N}, \bar{U}$, allow us to compare the emotions of a discussion with this ground state of the GENTOO community. We perform a set of nonparametric statistical tests to classify each discussion, consisting on three χ^2 tests at the 95% confidence interval:

- 1) Test of $U_d \simeq \bar{U}$: if this hypothesis cannot be rejected, the discussion is not considered to include collective emotions, and it is classified as *neutral*. If the $U_d > \bar{U}$ hypothesis is supported, we classify the discussion as *underemotional*. Examples of this kind of discussions are exchanges of computer code or error logs, which

⁶<http://archives.gentoo.org/gentoo-dev/>

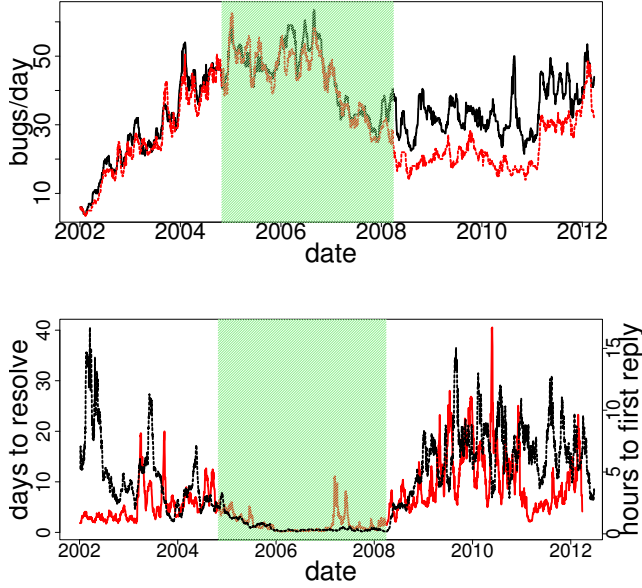


Fig. 1. Time series of performance metrics. On top, the median values of bugs reported (black curve) and solved (red curve) per day. On bottom, the median time to solve (red curve) and give the first reply (black curve) to a bug report. The green interval highlights $P2$, the period when *Alice* was active.

serve a technical purpose but do not compose emotional interaction. If the hypothesis $U_d < \bar{U}$, is supported, the discussion contained collective emotions, and the next two tests are evaluated to classify the emotions in this discussion.

- 2) Test of $P_d \simeq \bar{P}$: if the null hypothesis can be rejected and the data supports $P_d > \bar{P}$, we classify the discussion as *positive*.
- 3) Test of $N_d \simeq \bar{N}$: in the same way as the previous point, if the data supports $N_d > \bar{N}$, we classify the discussion as *negative*.

The above set of tests allows us to detect discussions that simultaneously contain positive and negative emotions, which will pass the second and third test. We classify these discussions as *bipolar*, representing collective emotions in which the authors of messages are polarized in different emotional states [48]. Additionally, a discussion might pass the first test, but not the second nor the third. These discussions contain more emotional content than the average of the community, but there is not enough data to classify the polarity of the emotions expressed in it. We classify these as *undetermined*.

Our statistical analysis highlights the presence of strong positive discussions in the bug reports, represented by points close to the lower left corner of the triangle. In these discussions, positive collective emotions are usually created as the result of fixing a software issue. The bug report system also shows some instances of underemotional discussions, represented by gray points close to the upper corner of the triangle. These threads are large exchanges of error logs and program outputs, and do not constitute a significant source of emotional interaction.

The lower row of Fig. 2 shows the collective emotions in the discussions of the developer’s mailing list (i.e. GENTOO-DEV). The structure of the emotions in these discussions is significantly different when compared against their bug reports

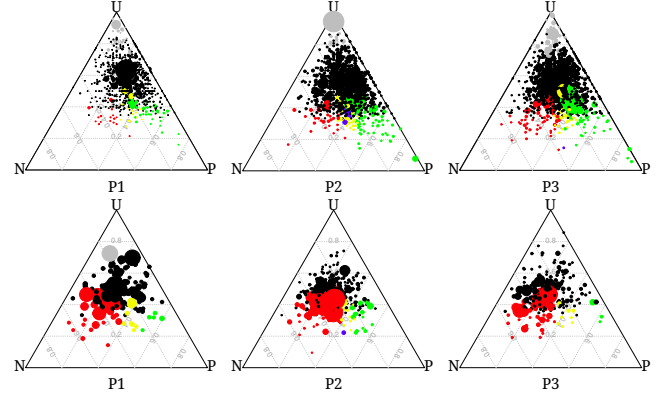


Fig. 2. Triangular representation of the emotions in GENTOO discussions in bug tracker (top) and the developer mailing list (bottom). Points represent discussions with at least 20 messages, with a size proportional to the amount of messages in the discussion, at a distance to the triangle vertices proportional to the ratios of positive, negative and neutral messages. Points are colored according to the classification of the discussion (i.e. black for *neutral*, green for *positive*, red for *negative*, gray for *underemotional*, blue for *bipolar* and yellow for *undetermined*).

counterpart: there are very few instances of positive discussions, and there are large discussions that elicited negative collective emotions.

These differences are possible due to the fact that these two communication channels (i.g. bug tracker and mailing list) shows very different styles of emotional interaction. In the bug tracker, positive emotions prevail. Users, developers and other contributors need to interact focusing on solving existing software issues. Thus, bug reports must be written as clear as possible. Moreover, contributors might need to write back to bug reporters in order to gather further information. This needs to be done in a smooth way that will lead to the identification of the locus of software issue as fast as possible. On the mailing list the situation can be quite different. Specially in the case of the developers’ private list, large instances of negative emotions can be observed. Likely, this is due to disagreements in collaboration processes and on competing agendas specifying how work and software should be organized.

The representation of collective emotions in Fig. 2 is useful to detect discussions that could trigger the decision of contributors to stop contributing to the OSS community. When comparing the three intervals depending on *Alice*’s presence, it is difficult to find differences in this representation. Periods $P2$ and $P3$ seem slightly more emotional, with some instances of bipolar discussions. For the case of the mailing list, negative emotions appear to be more salient in $P2$ and $P3$, but these observations require a quantitative validation. For that reason, we compute the time series of emotions in messages, using a moving average with $T = 30$ days range. Thus, M_T represents all messages found within such a time window. This allows us to calculate the respective mean positivity $p(t) = \sum_{m \in M_T} p_m / |M_T|$, mean negativity $n(t) = -\sum_{m \in M_T} n_m / |M_T|$, and mean polarity $s(t) = \sum_{\min |M_T|}^{\max |M_T|} s_m / |M_T|$ of messages.

Fig. 3 shows these time series for the bug reports and for the developer’s mailing list, divided into the three periods mentioned above. It can be noticed that there is no clear effect of the presence or absence of *Alice* in the bug tracker, but the developer’s mailing list seems to change. Before *Alice*’s presence, the mean polarity used to have positive values, and

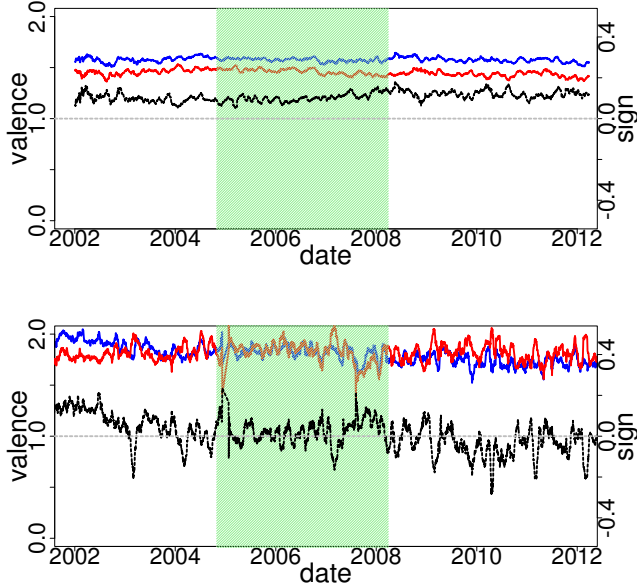


Fig. 3. Moving average applied to the time series of emotional expression within GENTOO BUGZILLA (top) and GENTOO-DEV (bottom). The red curve shows the daily mean negativity in the messages, the blue curve shows the mean positivity, and the black curve represents the mean polarity of all the messages. The green interval highlights $P2$, the period when *Alice* was active. during her activity this was close to 0. After her departure, there seems to be a period of stronger negativity. We statistically tested this observations, performing χ^2 test on the values of $s(t)$ across periods.

The results of these tests are reported in TABLE III, supporting our observation that – in the mailing list – $P3$ had more negative emotions than the periods $P1$ and $P2$ together. During that period, the community went through a complete reorganization, catalyzed by the creation of the *bug wranglers* project⁷. This was an initiative specially meant to cope with *Alice*'s sudden retirement. Thus, the negativity observed during $P3$ is likely to be due to the community struggle in restructuring its procedures. What about *Alice*'s presence during $P2$? Did *Alice* experience different sentiment expression within the bug tracker and mailing list? We separate the discussions in which *Alice* took part, from the remaining taking place within that period, and again calculated the different proportions of polarities. We show in TABLE III that the discussions in the mailing list that contained *Alice*'s messages were indeed more negative and less positive than the discussions not containing her messages, while the proportions of neutral polarity were roughly the same. Now focusing on the bug tracker, we observed that the proportion of negative polarity were roughly the same in *Alice*'s discussions when compared to the remaining discussions. Moreover, *Alice*'s discussions were more neutral and less positive.

V. EMOTIONS AND INACTIVITY

The above analysis of the departure of *Alice* serves as an example of the interplay between emotions and activity in the GENTOO community. In this section, we extend that analysis to contributors in general, exploring the role of emotions in their activity patterns. We continue by developing a method to predict long periods where an individual contributor is inactive.

⁷<http://www.gentoo.org/proj/en/qa/bug-wrangers/>

TABLE III. TEST FOR STATISTICAL SIGNIFICANCE OF DIFFERENCES IN PROPORTION OF POLARITIES. N REPRESENTS THE PROPORTION OF NEGATIVE MESSAGES, P FOR POSITIVE ONES AND U FOR NEUTRAL ONES. THE null hypothesis IS ALWAYS $Prop_1 = Prop_2$. THE SUBSCRIPTS $P1-P2$ AND $P3$ CORRESPONDS TO THE ANALYSIS PER PERIOD, WHILE *without Alice* AND *with Alice* TO THE ANALYSIS PER THREAD.

GENTOO BUGZILLA		
p-value of null hypothesis	alternative hypothesis	estimate
$1.04e - 033$	$N_{P1-P2} > N_{P3}$	0.011
$2.12e - 003$	$U_{P1-P2} > U_{P3}$	0.003
$1.29e - 040$	$P_{P1-P2} < P_{P3}$	0.014
$2.00e - 002$	$N_{without Alice} <> N_{with Alice}$	0.003
$2.06e - 130$	$U_{without Alice} < U_{with Alice}$	0.045
$6.62e - 188$	$P_{without Alice} > P_{with Alice}$	0.049
GENTOO-DEV		
p-value of null hypothesis	alternative hypothesis	estimate
$1.49e - 021$	$N_{P1-P2} < N_{P3}$	0.033
$5.08e - 006$	$U_{P1-P2} < U_{P3}$	0.017
$7.61e - 046$	$P_{P1-P2} > P_{P3}$	0.050
$1.61e - 026$	$N_{without Alice} < N_{with Alice}$	0.066
$5.50e - 001$	$U_{without Alice} <> U_{with Alice}$	0.004
$8.56e - 023$	$P_{without Alice} > P_{with Alice}$	0.001

A. Activity modes

For the case of *Alice*, determining when she became inactive is a trivial task, as she had no activity after a certain date. This is not necessarily the case for contributors in general, who might be inactive for a long period and then become active again. In general, contributors do not have a standard mechanism to inform the rest of the community if they are active or not, and the only way to detect their inactivity is when they do not produce messages for a period of time. To detect if a contributor became inactive, we use the theory of interevent time distributions [49], [50], which divides human communication in two modes: A bursty, *correlated* mode in which the time between the actions of a human is very short; and an *uncorrelated* mode that corresponds to the long times between bursts of activity. The *correlated* mode can be detected when the interevent times of a human follow a power-law distribution [49], [18], which emerges when humans reply to each other. The *uncorrelated* mode can be detected as an exponentially decaying regime, which can be explained as the result of a Poisson process of decoupled actions that start activity bursts [50].

For both datasets, we measured the interevent times τ between the messages of each contributor, and characterized the maximum inactivity period of each one through the maximum interevent time τ_{max} . Fig. 4 shows the distribution of τ_{max} for the mailing list and the bug tracker, with power-law fits to the head of the distributions. It can be noticed that these power-law regimes are not valid after a certain value, where both distributions show a tail that decreases much faster than a power-law. This shows the division between the two modes of activity mentioned above: the head of the distributions correspond to the *correlated* mode of active contributors, while their tails represent the *uncorrelated* time intervals when contributors are inactive. We found that the point between both modes is approximately $\tau = 30$ days. This indicates that when a contributor does not create any new message for a month, its behavior is uncoupled from the rest and it can be considered as inactive.

B. Contributor emotions

To produce the dataset that classifies contributor activity, we do the following: We collect all messages written by each contributor u , sorting messages by date. Then we iterate over the messages, starting from the earliest. If the contributor only posted a single message, we discard this *one time contributor* from our analysis. If the contributor posted more than one

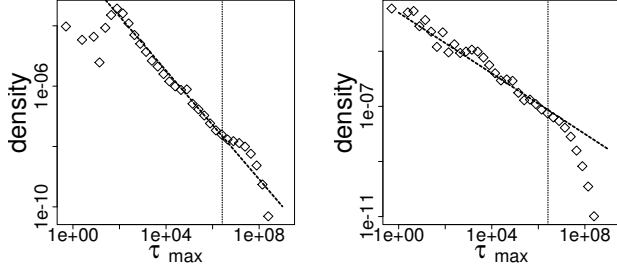


Fig. 4. Distributions of the maximum interevent times per contributor in the bug tracker (left) and the mailing list (right). The distributions are plotted in a log-log scale with exponentially increasing bins, and power-law fits to the head of the distributions. Dashed vertical lines show the mark at $\tau_{max} = 30$ days, where there is a regime change.

message, for each message m_t^u posted at time t , we measure the time interval τ between it and his next message $m_{t+\tau}^u$. If this time interval is shorter than 30 days, we label the interval I_t^u as *ACT*, meaning the contributor is active. Otherwise, we label the interval as *INA*, meaning that the contributor started a period of inactivity according to the theory explained above. For this analysis, we discard the last message posted by each contributor.

For each interval of contributor u , we compute the mean positivity score P_u and mean negativity score N_u of the messages of the contributor in the 5 days preceding the interval. This way, each data point is an interval between messages of the same contributor, with real-time measurements of the emotions expressed by that contributor in the days before the interval takes place. Our aim is to provide a predictor that identifies when a contributor is going to become inactive, as a tool that can warn community managers about the risk of losing contributors. This is not a simple task, as the ratios of each type of interval are very unevenly distributed. The prior probability of an interval being labeled as *INA* is 0.088 in the bug tracker, and 0.075 in the mailing list.

We calculate the conditional distributions of emotions given the label of an interval, $P(N_u|I)$ and $P(P_u|I)$, which we show in Fig. 5 for both datasets. An initial inspection shows the differences between the expression of emotions when a contributor is going to become inactive and when not. For both datasets, the distribution of emotional expression followed by an interval labeled as *INA* has larger variance than when followed by intervals labeled as *ACT*, showing signs of bimodality. Wilcoxon tests reveal that the conditional distributions of both emotions in the bug tracker are significantly different ($p < 1e-15$). In the mailing list, this is the case only for N_u ($p < 1e-15$), while the null hypothesis could not be rejected ($p = 0.21$) for P_u . This highlights the role of negative expression among developers, which differs more when one is going to become inactive, in comparison with active periods. Nevertheless, for the case of the mailing list, the failure to reject the null hypothesis for P_u does not imply that it is not informative, as we show below.

C. Activity tendencies

A notable difference in the distributions of Fig. 5 is the range where $P(N_u|I = \text{INA}) > P(N_u|I = \text{ACT})$ and $P(P_u|I = \text{INA}) > P(P_u|I = \text{ACT})$. For the case of the bug tracker, this condition is present only when N_u and P_u are above a certain value (i.e. in terms of absolute valence), while for the case of the mailing list, this is also

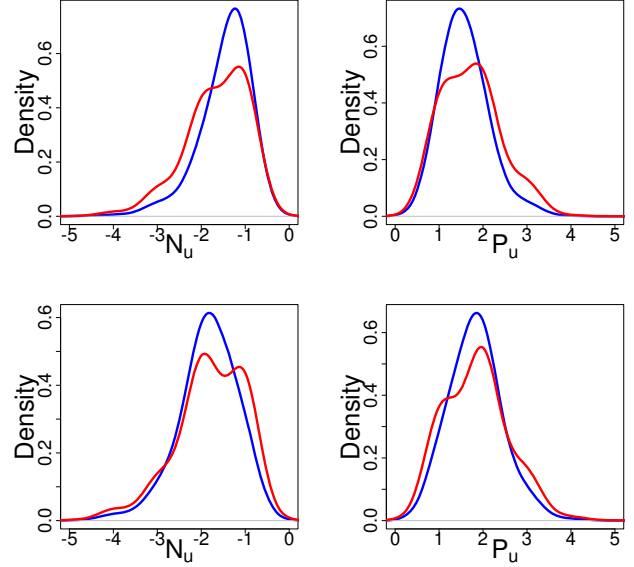


Fig. 5. Conditional distributions of contributor emotions $P(N_u|I)$ and $P(P_u|I)$ in the bug tracker (top) and the mailing list (bottom) for $I = \text{ACT}$ (blue) and $I = \text{INA}$ (red). Distributions were smoothed through a Gaussian kernel of width 0.35.

true when N_u and P_u have a sufficiently low absolute value. This indicates that strong emotions in the bug tracker, and deviations from the mean emotions in the mailing list inform about the likelihood of a contributor becoming inactive. To measure these effects, we compute the posterior distribution of a contributor becoming inactive at given time, considering his emotional expression in the last five days as

$$P(I = \text{INA}|N_u) = \frac{P(N_u|I = \text{INA}) \cdot P(I = \text{INA})}{P(N_u)} \quad (1)$$

and its equivalent for P_u . We bin P_u and N_u in five bins, using the ranges $[1, 5]$ and $[-1, -5]$ respectively, computing confidence intervals for the posterior distribution. Fig. 6 shows the posterior likelihood of becoming inactive for the first four bins, as the fifth one was not giving significant values due to the low probability of having $|P_u| > 4$ and $|N_u| > 4$. Our observation of the difference of the influence of emotions in both communication channels becomes clear: the likelihood of becoming inactive increases with P_u and N_u in the case of the bug tracker, while it grows with the distance to the mean for the case of the mailing list.

The distinctive v-shape of the likelihood for the emotions in the mailing list (Fig. 6 right) implies that lack of emotions can also serve as an indicator for contributors becoming inactive, but only when shared with others through the mailing list. It is remarkable that both N_u and P_u are informative to discriminate periods of inactivity in both datasets, suggesting that the decision to become inactive is more related to emotional intensity in general, rather than to positive or negative emotions alone. This is in line with the psychological theory which states that certain levels of arousal, or emotion intensity, are motivators for activity [48].

D. Real-time prediction

We apply the Bayesian analysis explained above to predict when contributors are going to start periods of inactivity, solely

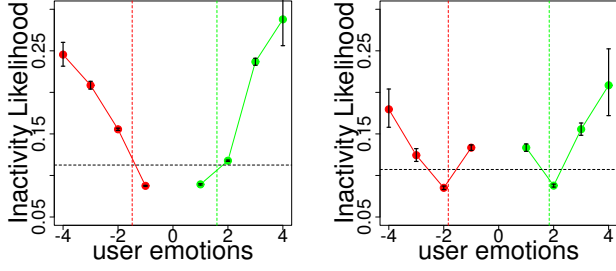


Fig. 6. Likelihood of an interval to be labeled as *INA* given contributor emotions, $P(I = \text{INA}|N_u)$ (red) and $P(I = \text{INA}|P_u)$ (green), for the bug tracker (left) and the mailing list (right). Error bars show confidence intervals, the horizontal dashed lines indicate $P(I = \text{INA})$ and vertical bars the means of N_u and P_u in each dataset.

based on the emotional content of their messages. Given the results shown in Fig. 6, we apply two different models:

- 1) bug tracker: if $|N_u| > \Theta_1$ or $|P_u| > \Theta_1$, then the next interval is predicted to be *INA*, and *ACT* otherwise.
- 2) mailing list: if $|N_u - \bar{N}_u| > \Theta_2$ or $|P_u - \bar{P}_u| > \Theta_2$, then the next interval is predicted to be *INA*, and *ACT* otherwise, where \bar{N}_u and \bar{P}_u are the average values of emotions expressed by this contributor.

We apply the above predictors with $\Theta_1 = 1.9$ and $\Theta_2 = 0.8$ to each point in our datasets, and compute values of *Precision* and *Recall* [51] over 20 bootstrapped samples, to ensure the robustness of our predictor. TABLE IV reports the means and standard deviations of *Precision* and *Recall* for each class and dataset. Both predictors have *Precision* significantly higher than the prior probabilities in the respective classes. In particular, the *Precision* of the minority class, *INA*, is sufficiently above the prior probability $P(I = \text{INA})$, showing that our method produce meaningful results when using contributor emotions to predict when these are at risk of becoming inactive. In addition, the values of *Recall* for both classes are well above 0.6, correctly classifying most of the existing instances.

TABLE IV. RESULTS OF THE PREDICTION OF CONTRIBUTORS BECOMING INACTIVE OR REMAINING ACTIVE IN BOTH DATASETS. STANDARD DEVIATIONS OF PRECISION AND RECALL VALUES ARE CALCULATED OVER 20 BOOTSTRAPPED SAMPLES OF THE DATASETS.

Dataset	measure	<i>Active</i>	<i>Inactive</i>
GENTOO BUGZILLA	Prior probability	0.886 ± 0.030	0.112 ± 0.030
	Precision	0.930 ± 0.024	0.196 ± 0.019
	Recall	0.673 ± 0.087	0.624 ± 0.012
GENTOO-DEV	Prior probability	0.892 ± 0.032	0.107 ± 0.032
	Precision	0.931 ± 0.026	0.175 ± 0.030
	Recall	0.654 ± 0.072	0.623 ± 0.021

VI. CONCLUSIONS

Based on a case study of the GENTOO project, we analyzed the relation between emotions and activity of its contributors. We gathered two disjoint datasets of communication within the community: (i) the bug reports stored in its bug tracking system BUGZILLA, and (ii) the messages posted in the developer’s mailing list. We provided a sentiment analysis of the messages written by all the contributors and related the emotional expressions to the activity patterns.

The first part of our case study investigated the emotional components related to the leave of a central contributor, named *Alice*. We show that her email discussions with other contributors were more negative than the rest, and that her departure was followed by higher stages of negativity in the community during its reorganization.

We extended this analysis to contributors in general, both in the bug tracker and the mailing list. To detect inactivity, we applied current state-of-the art theories on human correlated behavior, finding a mode of interevent times that indicates stages of contributor inactivity. This allowed us to statistically analyze the relation between a contributor’s emotional expressions and its individual intervals of inactivity. We reveal preconditions of emotional expressions that indicate when contributors feel demotivated to further contribute to the project. Based on this, we are able to estimate when a contributor becomes inactive, based on emotions expressed on his last messages. With this, we provide a tractable approach that can be applied by community managers to monitor emotional interaction within the community, and to foster timely reaction against undesirable turnover events of contributors.

Our contributions do not only focus on predictive results, but provide additional insights into the phenomenon at hand, in particular into fundamental relations between emotions and activity (and implicitly into motivation). We find that it is the emotional *intensity* which defines activity, rather than its *polarity* in terms of positive or negative emotions. Thus, in this work, we took a step forward by providing a methodology based on *sentiment* analysis, which sheds new light on GENTOO’s case study. This unveils a wide horizon of new quantitative approaches to the analysis of social dynamics within online communities, extending previous approaches to online emotional interaction [18], and social resilience [41].

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7-ICT-2008-3 under grant agreement no 231323 (CYBEREMOTIONS).

REFERENCES

- [1] J. Andreoni, “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving,” *The Economic Journal*, vol. 100, no. 401, p. 464, 1990.
- [2] G. Hardin, “The Tragedy of the Commons,” *Science*, vol. 162, no. 5364, pp. 1243–8, 1968.
- [3] K. Crowston and K. Wei, “Core and periphery in Free/Libre and Open Source software team communications,” *System Sciences, 2006.*, vol. 00, no. C, pp. 1–7, 2006.
- [4] K. Ehrlich, G. Valetto, and M. Helander, “Seeing inside: Using social network analysis to understand patterns of collaboration and coordination in global software teams,” *ICGSE 2007*, no. Icgse, pp. 297–298, Aug. 2007.
- [5] P. He, B. Li, and Y. Huang, “Applying Centrality Measures to the Behavior Analysis of Developers in Open Source Software Community,” *Second IEEE International Conference on Cloud and Green Computing*, pp. 418–423, 2012.
- [6] M. S. Zanetti, I. Scholtes, C. J. Tessone, and F. Schweitzer, “Categorizing bugs with social networks: A case study on four open source software communities,” in *Proceedings of the ICSE ’13*, 2013, pp. 1032–1041.
- [7] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment strength detection for the social web,” *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 163–173, 2012.
- [8] K. Crowston, K. Wei, J. Howison, and A. Wiggins, “Free/libre open-source software development: What we know and what we do not know,” *ACM Computing Surveys*, vol. 44, no. 2, pp. 1–35, 2012.
- [9] M. S. Zanetti, “The co-evolution of socio-technical structures in sustainable software development: Lessons from the open source software communities,” in *Proceedings of the 34th ICSE*. IEEE Press, 2012, pp. 1587–1590.

- [10] G. Valetto, M. Helander, K. Ehrlich, S. Chulani, M. Wegman, and C. Williams, "Using software repositories to investigate socio-technical congruence in development projects," in *MSR '07*. IEEE, 2007, pp. 25–25.
- [11] A. Mockus, R. T. Fielding, and J. D. Herbsleb, "Two case studies of open source software development: Apache and Mozilla," *ACM Transactions on Software Engineering and Methodology*, vol. 11, no. 3, pp. 309–346, 2002.
- [12] J. Lerner and J. Tirole, "Some simple economics of open source," *Journal of Industrial Economics*, pp. 197–234, 2002.
- [13] C. Gacek and B. Arief, "The many meanings of open source," *IEEE Software*, vol. 21, pp. 34–40, 2004.
- [14] G. von Krogh and E. von Hippel, "The promise of research on open source software," *Management Science*, vol. 52, no. 7, p. 975, 2006.
- [15] M. Cataldo, J. D. Herbsleb, and K. M. Carley, "Socio-technical congruence: a framework for assessing the impact of technical and work dependencies on software development productivity," in *ESEM*, 2008.
- [16] A. Kappas, "Social regulation of emotion: messy layers." *Frontiers in psychology*, vol. 4, no. February, p. 51, 2013.
- [17] A. Chmiel, J. Sienkiewicz, M. Thelwall, G. Paltoglou, K. Buckley, A. Kappas, and J. A. Hoyst, "Collective Emotions Online and Their Influence on Community Life," *PLoS ONE*, vol. 6, no. 7, p. e22207, 2011.
- [18] A. Garas, D. Garcia, M. Skowron, and F. Schweitzer, "Emotional persistence in online chatting communities," *Scientific Reports*, vol. 2, p. 402, 2012.
- [19] D. Garcia, A. Garas, and F. Schweitzer, "Positive words carry less information than negative words," *EPJ Data Science*, vol. 1, no. 1, p. 3, 2012.
- [20] D. Garcia and F. Schweitzer, "Modeling online collective emotions," in *Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media - DUBMMSM '12*, 2012, p. 37.
- [21] J. Bollen, H. Mao, and X.-j. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, pp. 1–8, 2011.
- [22] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction," *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 800–807, 2011.
- [23] D. Garcia and F. Schweitzer, "Emotions in Product Reviews Empirics and Models," *Proceedings of 2011 IEEE International Conference on Social Computing, SocialCom*, pp. 483–488, 2011.
- [24] R. Pfitzner and A. Garas, "Emotional divergence influences information spreading in Twitter," in *AAAI ICWSM 2012*, 2012, pp. 2–5.
- [25] G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, and D. Thalmann, "Sentiment analysis of informal textual communication in cyberspace," in *In Proc. Engage 2010*, 2010, pp. 13–25.
- [26] L. Zhang, Y. Jia, B. Zhou, and Y. Han, "Microblogging Sentiment Analysis Using Emotional Vector," *2012 IEEE Second International Conference on Cloud and Green Computing*, pp. 430–433, 2012.
- [27] M. Thelwall, K. Buckley, G. Paltoglou, M. Skowron, D. Garcia, S. Gobron, J. Ahn, A. Kappas, D. Kuster, and A. Janusz, "Damping Sentiment Analysis in Online Communication : Discussions , Monologs and Dialogs," in *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, 2013, pp. 1–12.
- [28] D. Garcia, F. Mendez, U. Serdült, and F. Schweitzer, "Political polarization and popularity in online participatory media," in *Proceedings of the workshop on Politics, elections and data*, 2012, pp. 3–10.
- [29] Y. Wu, J. Wong, Y. Deng, and K. Chang, "An Exploration of Social Media in Public Opinion Convergence: Elaboration Likelihood and Semantic Networks on Political Events," *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 903–910, 2011.
- [30] O. Kucuktunc, B. B. Cambazoglu, I. Weber, and H. Ferhatosmanoglu, "A large-scale sentiment analysis for Yahoo! answers," in *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*, New York, New York, USA, 2012, p. 633.
- [31] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks," in *Proceedings of the 12th international conference on Knowledge discovery and data mining - KDD '06*, 2006, p. 44.
- [32] E. Zheleva, H. Sharara, and L. Getoor, "Co-evolution of social and affiliation networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*. ACM Press, 2009, p. 1007.
- [33] F. E. Walter, S. Battiston, and F. Schweitzer, "Personalised and dynamic trust in social networks," in *Proceedings of the third ACM conference on Recommender systems - RecSys '09*, 2009, pp. 197–204.
- [34] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno, "The dynamics of protest recruitment through an online network." *Scientific reports*, vol. 1, p. 197, 2011.
- [35] G. Dror, D. Pelleg, O. Rokhlenko, and I. Szpektor, "Churn prediction in new users of Yahoo! answers," in *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*, New York, New York, USA, 2012, p. 829.
- [36] S. Wu, A. Das Sarma, A. Fabrikant, S. Lattanzi, and A. Tomkins, "Arrival and departure dynamics in social networks," in *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, New York, New York, USA, 2013, p. 233.
- [37] O. Herrera and T. Znati, "Modeling Churn in P2P Networks," in *40th IEEE Annual Simulation Symposium (ANSS'07)*, 2007, pp. 33–40.
- [38] M. Karnstedt, T. Hennessy, J. Chan, and C. Hayes, "Churn in Social Networks: A Discussion Boards Case Study," in *2010 IEEE Second International Conference on Social Computing*, 2010, pp. 233–240.
- [39] J. Kawale, A. Pal, and J. Srivastava, "Churn Prediction in MMORPGs: A Social Influence Based Approach," in *IEEE International Conference on Computational Science and Engineering*, 2009, pp. 423–428.
- [40] J. F. Hall, *Psychology of motivation*. Lippincott, 1961.
- [41] D. Garcia, P. Mavrodiev, and F. Schweitzer, "Social Resilience in Online Communities: The Autopsy of Friendster," 2013, under review. [Online]. Available: <http://arxiv.org/abs/1302.6109>
- [42] I. P. Cvijikj and F. Michahelles, "Monitoring Trends on Facebook," *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 895–902, 2011.
- [43] R. Michalski, J. Jankowski, and P. Kazienko, "Negative Effects of Incentivised Viral Campaigns for Activity in Social Networks," *2012 IEEE Second International Conference on Cloud and Green Computing*, pp. 391–398, 2012.
- [44] N. Serrano and I. Ciordia, "Bugzilla, itracker, and other bug trackers," *Software, IEEE*, vol. 22, no. 2, pp. 11–13, 2005.
- [45] M. S. Zanetti, I. Scholtes, C. J. Tessone, and F. Schweitzer, "The rise and fall of a central contributor: Dynamics of social organization and performance in the gentoo community," in *Proceedings of the CHASE '13 ICSE Workshop*, 2013, pp. 49–56.
- [46] M. S. Zanetti, E. Sarigol, I. Scholtes, C. J. Tessone, and F. Schweitzer, "A quantitative study of social organisation in open source software communities," in *Proceedings of the ICCSW '12 - Imperial College Computing Student Workshop*, vol. 28. Schloss Dagstuhl, 2012, pp. 116–122.
- [47] M. Zhou and A. Mockus, "What make long term contributors: Willingness and opportunity in oss community," in *ICSE*, 2012, pp. 518–528.
- [48] F. Schweitzer and D. Garcia, "An agent-based model of collective emotions in online communities," *The European Physical Journal B*, vol. 77, no. 4, pp. 533–545, 2010.
- [49] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics." *Nature*, vol. 435, no. 7039, pp. 207–11, 2005.
- [50] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber, "Evidence for a bimodal distribution in human communication." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 44, pp. 18 803–8, 2010.
- [51] P. Hooimeijer and W. Weimer, "Modeling bug report quality," in *Proceedings of the 22nd IEEE/ACM international conference on Automated software engineering*, 2007, pp. 34–43.