2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

How Popularity Shapes User Interactions in Tech-Related Online Communities

Abduljaleel Al Rubaye
Department of Computer Science
University of Central Florida
Orlando, FL, US
aalrubaye@knights.ucf.edu

Gita Sukthankar
Department of Computer Science
University of Central Florida
Orlando, FL, US
gitars@eecs.ucf.edu

***Abstract*—Tech-related online communities on GitHub, Reddit, and Stack Overflow are an invaluable resource for software engineers, allowing them to find solutions to problems and connect with other professionals. Much of the discourse on these platforms is conducted using commenting mechanisms in which one user responds to content posted by another user. Even though these communities lack formal organizational structures, these technologists are often followed by other software developers who monitor their posts; users who regularly post useful solutions are recognized using platform-specific mechanisms such as stars or karma points. This paper investigates the relationship between popularity and discourse in tech-related online communities. To do this, we create comment timelines from sequences of user interactions and extract commenting networks from comment response patterns. Although there are some commonalities, there are distinct differences between the commenting behavior of GitHub users vs. Reddit and Stack Overflow. By understanding how popularity affects user interactions, we can design communities that are more effective at supporting learning, collaboration, and knowledge sharing.**

***Index Terms*—social coding platforms, tech-related forums, popularity, knowledge sharing**

I. Introduction

Modern software developers rely on knowledge sharing communities, including social coding platforms, question and answer sites, and news forums, to keep up with ever-changing technologies. Although many users passively search for solutions without contributing, others actively participate in these knowledge sharing forums by posting code and articles, commenting, and upvoting good solutions. Most platforms have mechanisms in place to allow users to publicly endorse the contributions of others. Expert software developers who regularly provide valuable content may be viewed as "gurus" or "wizards" by the community. Identifying the experts in a online forum can make it easier to find high quality posts; however high quality code and accurate technical responses are often generated by users who possess few followers. Although many high status users earn their popularity through the regular production of high quality content, some users more actively "game" platform popularity measures [1].

Popular users in online communities can influence others' actions and attitudes. Their views and actions can shape the conversation, setting the agenda, tone, and behavior of others. When a popular user expresses a strong viewpoint or shares specific content, it can trigger a chain reaction of similar opinions and content among other users. This paper tackles the research question: how does popularity affect the commenting behavior of other users within the same community? We hypothesize that popular users affect both *comment timelines* and *commenting networks*. This paper analyzes: 1) the correlation between author popularity and the commenting behavior of other users and 2) the influence of popular commenters on comment timelines.

II. Related Work

Previous studies have been conducted on user commenting behavior on GitHub [2], Reddit [3], [4], and Stack Overflow [5], [6]. Destefanis et al. [2] studied how users contribute to GitHub projects through issues. They found that there are different patterns of user comments, and that these patterns can be grouped based on how the comments are expressed. Choi et al. [3] note that a small number of users drive the most critical conversations on Reddit. They also discovered that users who are active in multiple areas are more likely to be active in conversations. Zhang et al. [5] investigated how commenting on Stack Overflow supports knowledge sharing. They found that commenting behavior is related to user characteristics, such as experience level and social activity. Sengupta et al. [6] further studied the impact of commenting on community interactions by analyzing the content of comments. They

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

ASONAM '23, November 6-9, 2023, Kusadasi, Turkey
© 2023 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0409-3/23/11…$15.00
https://doi.org/10.1145/3625007.3627296

categorized comments into groups based on how they support knowledge sharing and learning. Buntain et al. [4] studied Reddit communities to understand how users' behaviors can be used to evaluate community stability and provide recommendations. They focused on identifying the social role of question answerers, and showed that this role could be classified based on users' activities without the need for content-based analysis. Our work extends on these previous studies by presenting an in depth analysis of comment timelines.

III. Method

A. Dataset:

To perform a cross-platform analysis of user commenting behavior, we gathered datasets from platforms that host tech-related discussions: GitHub, Reddit, and Stack Overflow. GitHub is a social coding platform for collaboration between software developers that offers code version control. Collaborative communication on GitHub is facilitated through discussion forums and issue reporting. Stack Overflow is a community-based Q & A website dedicated to answering technical problems. Reddit allows for asynchronous discussions between users; it comprises communities called subreddits dedicated to specific topics, including software development.

We developed a Python crawler utilizing platform-specific REST APIs. This crawler was then used to retrieve *posts*: GitHub issues, Reddit submissions, and Stack Overflow questions. The current popularity of AI has led to a boom in posting activity in this area so we selected keywords related to artificial intelligence, machine learning, and robotics (Figure 1). The crawler used the same set of keywords to retrieve posts across all three platforms.

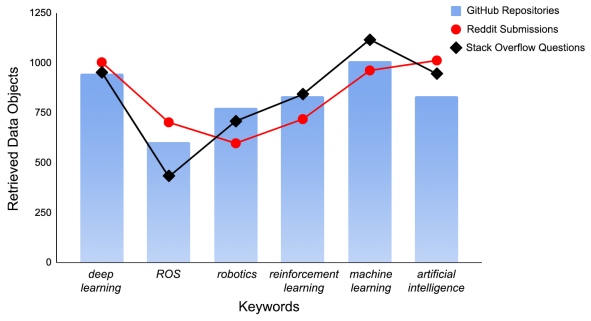

Fig. 1. Retrieved posts grouped by search term.

B. General Statistics:

The commenting mechanisms on online social networks let users leave comments in response to an initial post. Users can also comment on other users' comments. Table I provides some comment-specific statistics about our dataset. It can be observed that our GitHub dataset has a higher number of comments per post, but that those comments are generated from a smaller number of commenters. On average, 20-23% of the comments in Reddit and Stack Overflow are made by the posts' authors, while authors in GitHub showed a slightly higher rate of commenting, submitting 32% of all the collected comments. This is unsurprising given that GitHub comment chains are often focused on resolving code issues and involve in-depth discussion between a small number of developers.

TABLE I
Comment-related statistics

	GitHub	Reddit	Stack OF
Avg. Comment Count per Post	17.24	6.38	7.29
Avg. Authors' Comments	5.53	1.46	1.50
Avg. Commenters Count	3.87	5.34	4.82
% One-Time Commenters	38%	89%	74%

C. Comment Timeline:

To explore time-related commenting behaviors, we constructed comment timelines. In our dataset, GitHub commenters spent around four months on average between creating and closing issues. This time interval is substantially larger than the time required to close questions on Stack Overflow. On Reddit, the majority of comments are submitted over a few days. Comment density indicates how closely or widely dispersed the comments in a specific post's comment timeline are. The density of comments in open source software platforms may be affected by project characteristics such as code size and project age, as well as other factors such as code functionality and code quality [7].

D. User Popularity

The mechanisms for quantifying popularity on social media are platform-specific. For instance, GitHub allows users to follow each other, much like most online networks. We utilize the follower count to measure users' popularity on GitHub as was done in other studies [8], [9]. On Reddit, users' popularity can be measured by *karma*. Users can vote on a submission or comment, and Reddit uses an algorithm to calculate *karma* from user votes. *Karma* encourages and enables community engagement, and users seek to boost their scores by sharing their opinions, knowledge, and expertise [1]. On Stack Overflow, users earn rewards and gain *reputation* through their activities [10]. Users gain (or lose) reputation based on how many people vote on their postings. Posting good questions and meaningful replies is the most effective strategy to acquire reputation. Various studies on the Stack Overflow platform (e.g., [11]) suggest that *reputation* score is highly correlated to users' popularity. We gather popularity statistics for all authors and commenters in our dataset.

E. Role Categorization

To analyze the effects of user features on community discourse, users were clustered into four social roles using popularity and seniority. These can be colloquially described as follows:

- *Newbies* (SR1): new users who also have low popularity scores.

- *Rising Stars* (SR2): users who rapidly achieve a high level of popularity. This is unusual since there is often a correlation between popularity and account age.
- *Longstanding Members* (SR3): longstanding members of tech communities who do not enjoy a high level of popularity. Many of them post infrequently.
- *Tech Gurus* (SR4): experienced users who have gained the confidence of others through their interactions.

Table II shows the distribution of commenter social roles on the three platforms. On GitHub and Stack Overflow, most commenters fall in the category of members (SR3); whereas on Reddit, commenters are split between newbies (SR1) and members (SR3). Across all three platforms, rising stars (SR2) are the rarest category. Gurus (SR4) are most commonly found commenting on Stack Overflow. Figure 2 depicts the distribution of commenters' count ratios based on their social role classification across all three platforms.

TABLE II
ROLE DISTRIBUTION

	GitHub	Reddit	Stack OF
SR1 Newbies	19.4%	38.2%	9.8%
SR2 Stars	5.0%	7.8%	2.2%
SR3 Members	61.8%	36.5%	65.4%
SR4 Gurus	13.8%	17.5%	22.6%

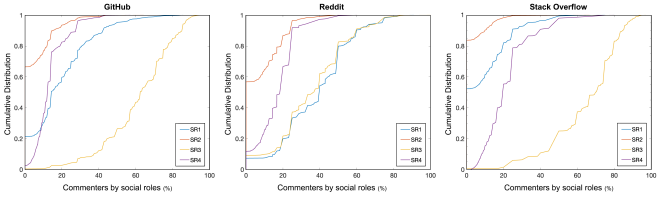

Fig. 2. Cumulative commenter distribution broken down by role.

IV. RESULTS

Our study examined two aspects of social commenting behavior. First, we looked at the general patterns of commenting behavior among users. Second, we compared the impact of popularity on the timelines of posts.

A. Commenting Behaviours

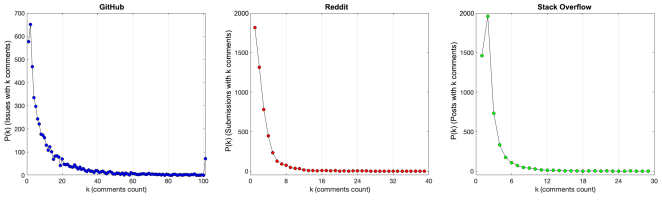

Fig. 3. Distribution of posts based on extracted comments.

Figure 3 shows the distribution of the posts in our dataset based on the comment counts. According to the illustrated data, on GitHub, over 82% of the issues had 20 or fewer comments, whereas, on Reddit, nearly 95% of the submissions had eight or fewer comments. Similarly, 96% of the questions on Stack Overflow received only five or fewer comments. Moreover, we observed that there are a considerable number of posts that received zero or one comments. Like many social media datasets [12], [13], our data follows a power law distribution in which a small number of posts have very long comment chains but most do not.

Commenters can be divided into two groups based on the extent of their commenting participation:

- One-time commenters: Users that contribute by leaving only one comment on a post.
- Multi-time commenters: Active users who comment more than once during a post's timeline.

The investigation reveals that individuals comment on GitHub issues an average of 4.45 times (the ratio of the average number of commenters over the comment count), indicating a significant number of repeat commenters. Approximately 38% of GitHub users are one-time commenters. In contrast, this ratio is substantially higher on the other two platforms, where 89% and 74% of Reddit and Stack Overflow users, respectively, leave only one comment.

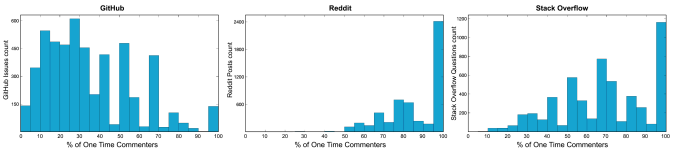

Fig. 4. Distribution of posts based on the ratio of single commenters.

Figure 4 depicts the distribution of captured posts based on the commenting ratio of their contributors. Reddit and Stack Overflow have right-skewed distributions that contain a larger number of one-time commenters, whereas commenters on GitHub are more likely to make repeated comments.

This reveals clear platform specific differences in user commenting behavior. On social coding platforms that involve groups of developers and reviewers working as a team (e.g., GitHub), the results suggest that users exhibit conversational behavior when posting. Posts are utilized as part of an ongoing conversation to review work and exchange information. On Reddit and Stack Overflow, commenters are more likely to express their viewpoint once and refrain from further comments.

B. User Popularity Effect

We investigated the popularity of two user categories, authors and most popular commenters (MPCs), to understand the impact of popularity on other communication-related features.

*1) **Authors' Popularity Effect**:*

Table III details the correlation of the author's popularity to the following measurements: 1) number of comments left on the same post, 2) the total number of commenters that engaged with the post, 3) the popularity of the commenters, and 4) the author's comments count. Across all three platforms, we observe a statistically significant positive correlation between

TABLE III
AUTHOR'S POPULARITY CORRELATIONS TO OTHER COMMENT-RELATED MEASUREMENTS.

	GitHub		Reddit		Stack OF	
	r-val	*p-val*	*r-val*	*p-val*	*r-val*	*p-val*
Comments Count	0.30	0.01	0.12	0.00	0.39	0.02
Commenters Count	0.21	0.00	0.10	0.00	0.31	0.03
Author's Comment Count	-0.05	0.29	-0.03	0.49	-0.11	0.09
Commenters' Popularity	0.38	0.01	0.08	0.57	0.40	0.01

the authors' popularity and the number of comments to posts they initiate, as well as to the total number of commenters engaging with their posts. The correlation between authors popularity and their own comment count is not statistically significant; it isn't the case that simply commenting more guarantees an increase in popularity. On GitHub and Stack Overflow, we found that popular authors are more likely to receive comments from popular users. However, this trend is not observed on Reddit.

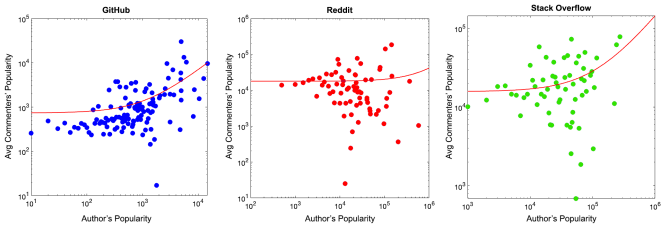

Fig. 5. Log-log plot of authors' popularity correlation with the average commenters' popularity

*2) **MPCs' Popularity Effect**:*

MPCs are the commenters with the highest popularity all the users who contribute to a post. Like other commenters, MPCs may contribute to a post by commenting more than once. Our data shows that 85% of GitHub posts include MPCs with multiple comments per post, resulting in 27% of the total number of comments. On the other two platforms, MPCs are less likely to make multiple comments. Most of our collected posts in Reddit and Stack Overflow (79% and 62%, respectively) have only one MPC contribution. This indicates that MPCs drive more of the conversation in GitHub.

MPC's Comment Tail: From the comment timelines, we extract a comment tail in order to quantify changes in the number and the popularity of the commenters before and after the MPC's comment.

We define three types of comment tails based on their length:

- *Short comment tails*: MPCs comment closer to the end of the comment timeline. In these conversations, a smaller number of individuals leave comments after the MPCs compared to the ones who comment before the MPC:

$$|CT_P| < |\{c_0, ..., c_{k-1}\}|$$

- *Medium comment tails*: MPCs contribute closer to the middle of the conversation, and a similar number of comments are observed before and after the MPC's comment:

$$|CT_P| \simeq |\{c_0, ..., c_{k-1}\}|$$

- *Long comment tails*: MPCs tend to comment earlier in the conversation, leading to a longer comment tail. Hence most commenters engage with the post after the MPC comments:

$$|CT_P| > |\{c_0, ..., c_{k-1}\}|$$

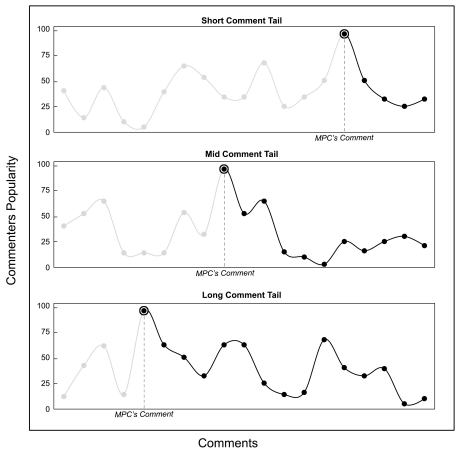

Fig. 6. Three sample comment timelines, each showing one of the comment tail length categories: long, medium, and short. The comment tail includes all comments that are posted after the MPC comments.

TABLE IV
MPC COMMENT TAIL STATISTICS

	GitHub	Reddit	Stack OF
Short comment tail ratio	69.14%	40.39%	43.19%
Medium comment tail ratio	11.79%	20.56%	18.37%
Long comment tail ratio	19.07%	39.05%	38.44%
Comments per user before MPC	3.093	1.026	2.022
Comments per user after MPC	2.181	1.040	1.416
Authors' comments count before MPC	5.938	0.905	1.040
Authors' comments count after MPC	1.795	1.001	0.997

From the comment timelines, we extracted each post's MPC-related comment tail and categorize them based on length. Table IV details the ratio of each category and the commenting rate per user, including the authors, before and after the MPCs comment.

We observed a similar pattern of comment tail lengths on Reddit and Stack Overflow, with similar proportions of short, medium and long tails. There is no noticeable MPC-related change in users' comments count in Reddit and Stack Overflow. In GitHub, however, more than two third of the posts fall in the category of short comment tails. Short comment tails indicate that the MPCs have commented mainly toward the end of the conversation. In other words, GitHub users' commenting behavior changes after an MPC comments; MPCs seem to have the "final word" in the conversation, bringing commenting to a close. We see a similar pattern in the authors' comment rates. In GitHub, there is a noticeable reduction in

TABLE V
MPC POPULARITY CORRELATIONS TO OTHER USERS' COMMUNICATIONS FEATURES.

	GitHub		Reddit		Stack OF	
	r-val	*p-val*	*r-val*	*p-val*	*r-val*	*p-val*
Comment Count	0.22	0.00	0.16	0.01	0.22	0.01
Authors' Popularity	0.07	0.02	0.00	0.83	0.10	0.03
Commenters' Popularity pre MPC	0.41	0.00	0.22	0.01	0.48	0.01
Commenters' Popularity post MPC	0.35	0.00	0.11	0.00	0.33	0.04

authors comments after the MPC comment.

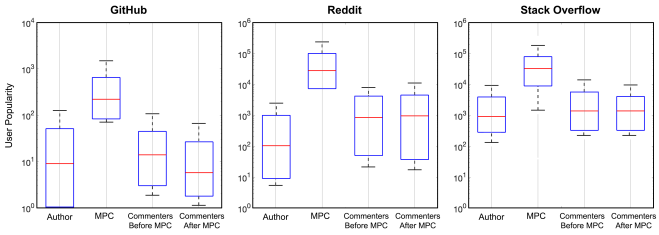

Fig. 7. Popularity of authors, MPCs (the Most Popular Commenters), and commenters (partitioned by those who comment before and after the MPC)

Figure 7 depicts the popularity distribution of different types of users. The data show that the posts on all platforms exhibit a similar pattern, where the popularity of the authors is slightly lower than most commenters. The users who comment before or after MPCs have similar levels of popularity on Reddit and Stack Overflow. However, on GitHub, the average popularity of the commenters decreases after the contribution of MPCs.

Table V details the correlation between MPCs' popularity to other communication features. There is a positive relationship between the popularity of the post's MPC and the number of comments a post receives on all three platforms. It is likely that a post will receive a higher number of comments if its MPC's popularity is relatively high. Also the other commenters (both before and after the MPC's comment) are likely to have a high popularity themselves. The more popular MPCs are, the higher the likelihood of popular users' participation in the same post. However, MPC popularity is not strongly correlated with the post author popularity, particularly on Reddit. Popular commentators often comment on posts that interest them, regardless of the popularity of the original poster.

V. CONCLUSION AND FUTURE WORK

The objective of this research study was to assess the engagement of individuals with technology-related posts on online social platforms. Our findings revealed clear differences in commenting behavior on different social platforms, particularly for GitHub users. GitHub users seem to use the platform in a more conversational manner when communicating through posts. For example, GitHub users post more comments and are more likely to comment multiple times. Additionally, GitHub discussion timelines are longer, and feature conversations spanning months rather than days.

We aim to understand how the popularity of users can affect the trajectory of discourse. To achieve this goal, we evaluated the impact of popularity on two types of users: the authors of posts and the most popular commenters (MPCs). Our data suggests that posts initiated by popular users tend to elicit more comments and involve a larger group of people in the discussion. The participation of high-popularity MPCs is positively correlated with the number of comments and the participation of other high-popularity commenters. By studying comment timelines, we observed that users' commenting rates and involvement change before and after MPCs contribute to discussions on GitHub. However, this pattern was not evident on Reddit or Stack Overflow.

By understanding how popularity affects user interactions, we can design communities that are more effective at supporting collaboration. One of our primary concerns was that popular users had the implicit authority to prematurely terminate productive technical discussions. On GitHub, we found that commenting behavior changes after an MPC comments, typically ending the discussion. This pattern is not observed on Reddit or Stack Overflow. This may occur when GitHub MPCs also possess the formal technical responsibility for code review, which gives their comments more weight.

REFERENCES

[1] A. Richterich, "Karma, precious karma!: Karmawhoring on Reddit and the front page's econometrisation," *Journal of Peer Production*, vol. 4, no. 1, pp. 1–12, 2014.

[2] G. Destefanis, M. Ortu, D. Bowes, M. Marchesi, and R. Tonelli, "On measuring affects of GitHub issues' commenters," in *Proceedings of the International Workshop on Emotion Awareness in Software Engineering*, 2018, pp. 14–19.

[3] D. Choi, J. Han, T. Chung, Y.-Y. Ahn, B.-G. Chun, and T. T. Kwon, "Characterizing conversation patterns in Reddit: From the perspectives of content properties and user participation behaviors," in *Proceedings of the ACM Conference on Online Social Networks*, 2015, pp. 233–243.

[4] C. Buntain and J. Golbeck, "Identifying social roles in Reddit using network structure," in *Proceedings of the International Conference on World Wide Web*, 2014, pp. 615–620.

[5] H. Zhang, S. Wang, T.-H. Chen, and A. E. Hassan, "Reading answers on Stack Overflow: Not enough!" *IEEE Transactions on Software Engineering*, vol. 47, no. 11, pp. 2520–2533, 2019.

[6] S. Sengupta and C. Haythornthwaite, "Learning with comments: An analysis of comments and community on Stack Overflow," in *Proceedings of the Hawaii International Conference on System Sciences*, 2020.

[7] O. Arafat and D. Riehle, "The comment density of open source software code," in *International Conference on Software Engineering-Companion Volume*. IEEE, 2009, pp. 195–198.

[8] A. Al-Rubaye and G. Sukthankar, "Scoring popularity in GitHub," in *International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2020, pp. 217–223.

[9] K. Blincoe, J. Sheoran, S. Goggins, E. Petakovic, and D. Damian, "Understanding the popular users: Following, affiliation influence and leadership on GitHub," *Information and Software Technology*, vol. 70, pp. 30–39, 2016.

[10] D. Movshovitz-Attias, Y. Movshovitz-Attias, P. Steenkiste, and C. Faloutsos, "Analysis of the reputation system and user contributions on a question answering website: Stack Overflow," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2013, pp. 886–893.

[11] A. Merchant, D. Shah, G. S. Bhatia, A. Ghosh, and P. Kumaraguru, "Signals matter: understanding popularity and impact of users on Stack Overflow," in *The World Wide Web Conference*, 2019, pp. 3086–3092.

[12] A.-L. Barabási, "The new science of networks," *Cambridge: Perseus*, 2002.

[13] L. Takac and M. Zabovsky, "Data analysis in public social networks," in *International Scientific Conference and Workshop Present Day Trends of Innovations*, vol. 1, no. 6, 2012.