Users' Behaviors of Submitting and Commenting in Education-Related Subreddits

Khadijah Alghamdi, Majmaah University, Saudi Arabia

Abstract

This study is an investigation of users' behaviors of posting and commenting in five education-related subreddits: r/teaching, r/Teachers, r/education, r/teachingresources, and r/CSEducation. The data were analyzed for the period from January 2018 until May 2019. Comparative analysis revealed some differences among the five results showing that the r/Teachers subreddit has the most engagement scores and active users' percentages while the r/education subreddit has the lowest score of engagement although it has the second-highest number of contributions (submissions and comments).

Keywords: education, CS, reddit, subreddits, user behavior, online communities

iafor

The International Academic Forum
www.iafor.org

Introduction

Despite the major focus in educational research on ways to improve students' learning and teachers' performance, teachers still face pedagogical challenges, curriculum challenges, and structural challenges (Margot & Kettler, 2019). One of the solutions to these challenges is collaboration among teachers in online communities (Margot & Kettler, 2019), which can be implemented on social media platforms.

Online communities can be valuable resources that offer teachers continuous professional development (Duncan-Howell, 2010). However, some computer science teachers have complained that online communities for sharing resources and learning from each other by posting questions and receiving answers often disappear (Yadav et al., 2016). They are looking for a sustainable and trustworthy source of information and support.

Social media platforms such as Twitter, Facebook, and Reddit are widely used by educators and learners (Carpenter et al., 2018; Willet & Carpenter, 2019). Srinivasan and Mekala (2014) defined social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchange of user-generated content" (p. 155). Drawing on these affordances, many teachers use social media platforms for formal and informal professional development (Margot & Kettler, 2019). One of the social media platforms most widely used by teachers in the United States is Reddit, which was created in 2005 by Alexis Ohanian and Steve Huffman (Vella, 2015). Reddit members submit such content as links, texts, photos, and videos for discussion. There is also a function for raising content to the top (upvoting) if a member likes it or dropping content down (downvoting) if not.

The Reddit platform has many subreddits, each of which constitutes a specific online community. This social media platform thus serves different perspectives and different people. Although there are many subreddits that are concerned with education and supporting teachers, researchers have paid little attention to the different kinds (Willet & Carpenter, 2019). Moreover, Chen et al. (2014) emphasized that social media conversations provide us with rich data about teachers' experiences with, opinions on, difficulties with, or feelings about learning. In order to bridge this gap in the literature, this study focuses on five of the education-related subreddits: r/teaching, r/Teachers, r/education, r/teachingresources, and r/CSEducation, the websites of which indicate that they are all focused on education but different aspects. The earliest subreddit among the five targeted for this research is r/education, which was initiated in May 2008 as a space for educational stakeholders to discuss topics related to educational policy, research, technology, and politics. In December 2008, r/teachers subreddit was begun as a community for all teachers in different disciplines to share their experiences, stories, insights, and inspirations with others. The r/Teaching subreddit was launched in May 2009, also for teachers to share experiences and concerns. The r/CSE subreddit was begun in 2012 to target computer science educators and researchers and provide them with best practices and pedagogies for teaching computer science.

The most recently created subreddit in this study is r/Teaching Resources, which was initiated in February 2013. Its goal is to provide classroom teachers with resources and technologies in order to improve their knowledge delivery to students. All of the five subreddits are public and target mostly K-12 teachers in all disciplines, except for the r/CSEducation subreddit, which specifically targets computer science educators. The aim of this paper is to explore the behaviors of users of the five subreddits and to compare them in terms of their activities and

engagement. This work expands the work of Willet and Carpenter (2019) by adding r/CSEducation, which was initiated in response to computer science educators' quest for online communities to support their professional development (Yadav et al., 2016).

Related Work

There are many studies about online communities for teachers in general and computer science teachers in particular. However, to the best of the author knowledge and search, few qualitative studies on subreddit communities for teachers were found, in which human-coded content analysis was used, such as, Carpenter et al. (2018) and Chang-Kredl and Colannino (2017). In these studies, however, only small numbers of comments or posts were analyzed because of humans' limited ability to analyze the large amount of data available from social media platforms.

Chang-Kredl and Colannino (2017) analyzed 600 posts on a Reddit platform initiated to address the public image of teachers. The posts were selected via on a search on the Reddit platform using the following key phrases: "best teacher," "worst teacher," "favorite teacher," and "least favorite teacher". Based on this search, eight threads were selected, four about best teachers and four about worst teachers. Similar to Carpenter et al. (2018), inductive content analysis was applied. This process resulted in the identification of 15 codes, which were classified into four broad themes: teachers' professional qualities, teachers' personal qualities, relational qualities between teachers and students, and implications for students' least favorite teachers. The researchers concluded that the best teachers are those who know how to present, care about their relations with their students, and have an impact on students, while the worst teachers are those who are not qualified in their teaching skills, do not care about their relations with their students, and do not affect their students' lives.

Carpenter et al. (2018) conducted a qualitative study to analyze the top 25 posts and the associated responses in the r/teachers subreddit for a 20-day period. This sub-Reddit had the second highest number of readers. Inductive thematic analysis was used in this study in which human-coding was applied to examine the content of the posts. They found that the most frequent codes in r/teachers subreddit posts were the following: advice, teachers, a community post, specific content area, venting, new teacher, and students. Also, they found that users who posted to r/teachers most frequently sought or gave technical and emotional advice. Regarding analysis of responses, the study revealed that the top-rated responses tended to provide advice in line with what was sought in the original advice-seeking posts. In addition, sometimes the advice provided emotional support.

In an exploratory study of conversations between teachers, Willet and Carpenter (2019) collected 351,301 posts, including 114,524 original posts and 236,777 responses, from four different education-related subreddits, r/education, r/Teaching, r/teaching, and r/Teaching Resources, between July 1, 2017 and June 30, 2018. In this study, the *active percentage* was defined as "the number of contributors to a subreddit in the past year divided by the number of subscribers to that subreddit, as of September 2018," and the *engagement score* "represented the number of responses to questions in a subreddit in the past year divided by the number of threads, or clustered conversations, in that subreddit during the same span of time" (p. 2789). Based on the traced data, they found that a large number of subscribers were not necessarily associated with the most active contributors. They also calculated a text-polarity sentiment score for every contribution in each of the subreddits for the voting score, which is the sum of all upvotes and downvotes. In addition, they created a word cloud for each subreddit to explore

the most popular words. They found that the subreddit r/education included broader conversations about education while the subreddit Teaching Resources was more concerned with resources than with what happens in schools.

Research Questions

The aim of this study was to investigate users' participation in five education-related subreddits: r/education, r/Teachers, r/Teaching, r/CSEducation, and r/Teaching Resources by answering the following questions:
1. What are the engagement scores for each subreddit?
2. What are the percentages of active users in each subreddit?
3. What are the users' behaviors of posing and commenting for each subreddit?
4. What are the most used words in submissions and comments for each subreddit?
5. Are the targeted subreddits growing or not?

Dataset

A dataset of 6,870 contributions was analyzed, the total number of original posts (submissions) and responses to them (comments) in five subreddits related to education: r/Teaching, r/Teachers, r/education, r/TeachingResources, and r/CSEducation during the period of January 2018 to May 2019. This dataset was only 1% of the original data and was downloaded from the server of Dr. Riddell, the course instructor, to my laptop.

In this study, a submission is a post first submitted by a user, referred to as the submitter, in the form of either a question or a statement communicated to other users and to which other users, referred to as commenters, can respond. Each submission with its comments is called a thread.

Method

In order to answer the research questions, the dataset drawn from five education-related subreddits was analyzed using Python. First, 34 parquet files were downloaded from the server to my laptop. The first 17 files had 1% of the comments of all the subreddits, and each file contained data for one month, starting from January 2018 to May 2019. The last 17 files included 1% of all subreddit submissions. The submissions and comments of the five targeted subreddits were extracted in order to find the number of comments, the number of submissions, the number of users who submitted, the number of users who commented, and the number of questions in the submissions data for each subreddit. The percentage of active users of each subreddit was calculated by dividing the total sum of commenters and submitters to that subreddit during the 17- month period by the total number of users as of December 2019. Engagement scores were also calculated for each subreddit by dividing the number of comments in a subreddit during the 17-month period by the number of submissions during the same span of time. These definitions of active percentages and engagement scores were adopted from Willet and Carpenter (2019).

In order to classify each submission post as questions or statements, if a question mark in each post was present, the submission was classified as a question; otherwise, the submission considered to be a statement. Then the submissions and comments were merged and then words counted separately in order to find the most common 50 words in each of the five subreddits, which I then used to create word clouds. The Word Cloud tool has the feature of filtering

out all punctuation and stop words, such as articles, pronouns and conjunctions, before forming clouds.

Evaluation

To assess the validity of the coding, two alternative samples were used from two subreddits. One was one of the largest subreddits r/AskReddit (Cole et al., 2019), a forum in which users can ask questions on different topics. It was started in January 2008 and has more than 25 million users. The other subreddit was r/FIFA, a sports channel started in 2009, which has more than 3,000 users. The results were robust in support of sampling data from other months in the same time period showing similar observations.

Results and Discussion

In this part, the findings of this study of users' behaviors in five education-related subreddits are presented. Descriptive statistics for the overall contributions to the five subreddits were calculated (Table 1). The results show that all subreddits had more comments than submissions except for the r/education subreddit, in which there were more submissions than comments, yielding a 0.40 engagement score, indicating that the users of this particular subreddit were not highly engaged with or not cognizant of the content discussed. Another possible explanation is that the submissions, which are the main posts in a forum, were more often in the form of declarative statements rather than questions and so did not elicit answers, and people are more likely to be engaged when submissions are questions.

This explanation is supported by the two highest engagement scores, 11.7 for the r/Teachers subreddit followed by 8.2 for r/teaching. Both of these subreddits had submissions with more questions than statements (Table 2), which might be a reason for the higher scores for engagement with the content. This explanation is also supported by the finding that the subreddits with the lowest engagement scores, CSEducation, and r/teachingresources, had submissions with fewer questions and more statements. These results are aligned with what Willet and Carpenter (2019) found in their study.

Table 1
Engagement Score of the Subreddits

Subreddit	Comments	Submissions	Engagement
r/education	425	1102	0.40
r/teachers	4060	349	11.71
r/Teaching	512	59	8.2
r/CSEducation	9	3	3
r/teachingresources	29	23	1.34

Regarding the type of submission, all five subreddits have fewer question type submissions and more of statement submissions and the r/CSEducation subreddit does not have any questions at all (Table 2).

Table 2
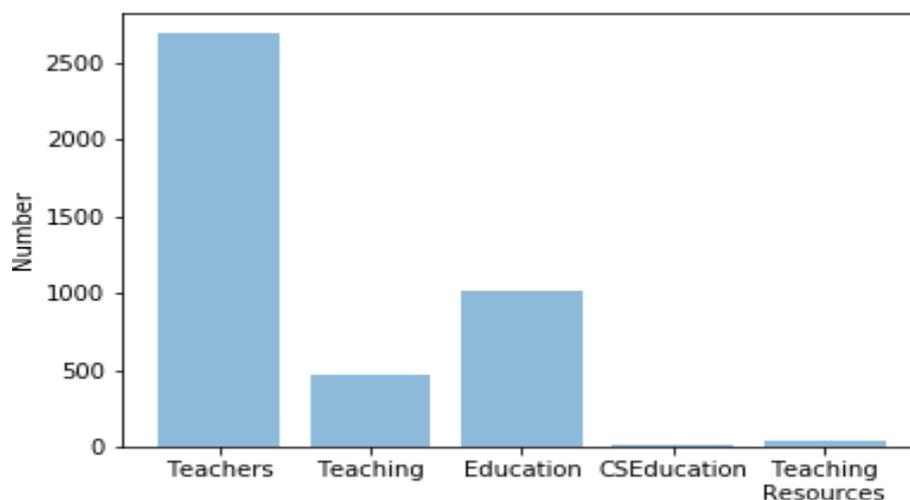
Questions and Statements in Submissions

Subreddit	Questions	Submissions
r/education	9	1093
r/teachers	155	194
r/Teaching	26	33
r/CSEducation	0	12
r/teachingresources	3	20

While all of the five subreddits had very low percentages of active users (Table 3), r/Teaching subreddit had the highest active users' percentage (2.1%), followed by r/education, r/teaching, r/teachingresources, and r/CSEducation consecutively (Figure 1).
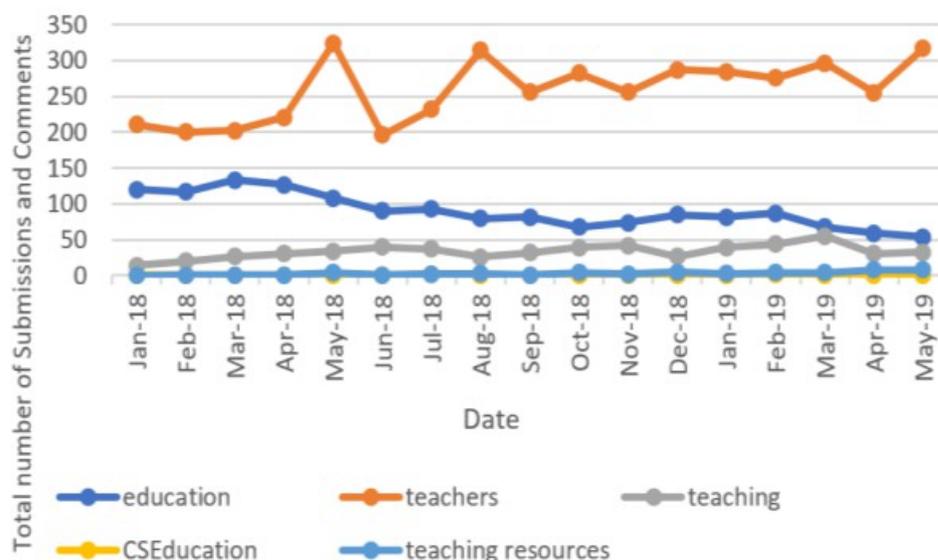
Table 3

Active Users Percentages for Subreddits

Subreddit	Commenters	Submitters	Subscribers	Active (%)
r/education	311	706	96,600	1.05%
r/teachers	2433	263	128,000	2.10%
r/Teaching	418	55	50,400	.93%
r/CSEducation	9	3	10,900	.11%
r/teachingresources	26	12	18,800	.23%

Figure 1
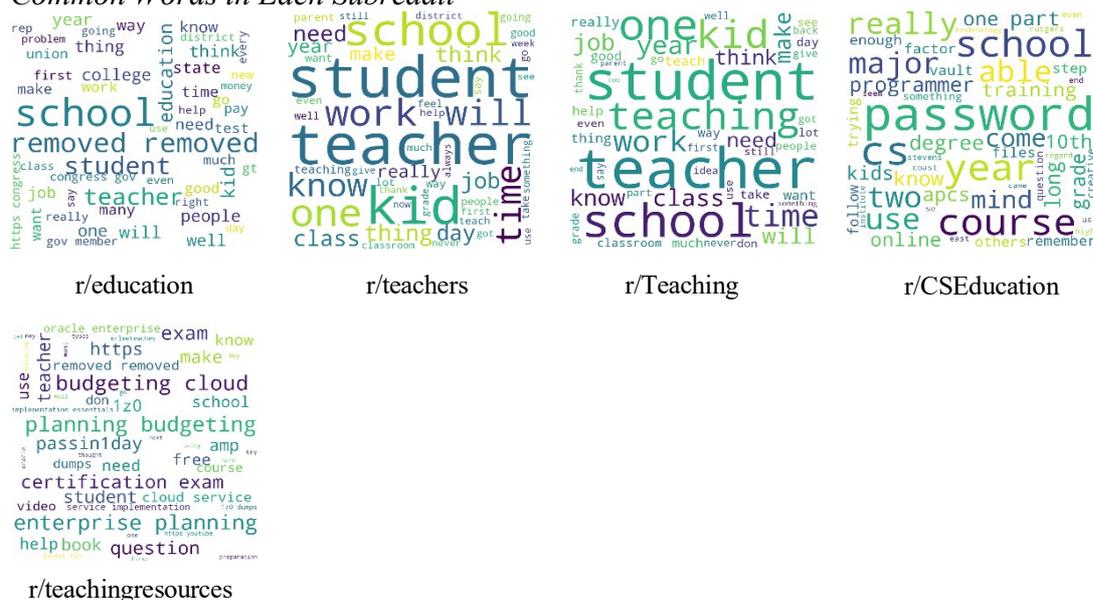
Number of Active Users of Each Subreddit

To explore the growth of subreddits during the 17 months starting from January 2018, a line graph of the total number of submissions and comments was created. The graph illustrates that the r/Teachers subreddit was more active at the beginning of this period and remained the most active subreddit until the end, while activity in the r/education subreddit decreased over the 17 months. On the other hand, the r/teachingresources and r/CSEducation subreddits, which began with low activity rates, remained relatively stable throughout the data collection period (Figure 2).

Figure 2
Are the Subreddits Growing?

In regard to the common words in each subreddit (Figure 3), the word clouds show that in general the words were related to educational discussions about K-12 contexts. The word clouds based on the most common words show a clear similarity in the topics discussed in the r/Teachers and r/teaching subreddits, for instance, "student," "teacher," "school," "kid," "work." In the r/education subreddit, the most common word is "school." Also, in the r/education subreddit the word "removed" is shown twice in large font, indicating that a comment or submission was removed either by Reddit moderators or by spam filters. This suggests that the moderators of this subreddit actively observe the published content, and that this subreddit's users posted inappropriate content. Moreover, "https" is one of the common words, showing that in general users like to share links. The word cloud for the r/teachingresources subreddit includes "budgeting," "cloud," "exam," "question," "planning," and "certification" as the most common words. This suggests that a main concern in the teaching resources subreddit is about resources that might support students' achievement in exams and ways to use school resources to promote learning. The word cloud for the r/CSEducation subreddit shows "password," "CS (computer science) programmer," and "training" as the most common words. These words emphasize that computer science education and a focus on students as programmers are main concerns of users of this subreddit. Also, the word "training" shows the desire for professional development by CS teachers.

Figure 3
Common Words in Each Subreddit

r/education   r/teachers   r/Teaching   r/CSEducation

r/teachingresources

Conclusion

Four of the five subreddits in this study can be considered as low-activity communities, the exception being the r/Teachers subreddit. Therefore, based on this as well as previous studies, it can be inferred that many CS teachers suffer the disadvantage of not having an online community on which they can rely for substantive interaction and support. Thus, the findings of this study suggest the need for more research attention to education-related subreddits, especially those with low engagement scores and percentages of active users, and ways to create and sustain more dynamic communities for CS educators.

Future Work and Limitations

Toward this end, this study can be expanded by including more education-related subreddits, for instance, subject-related subreddits in K-12 contexts to examine other users' behaviors and compare them with those of the r/CSEducation subreddit users. Moreover, similar studies might be conducted to explore other online communities for teachers, such as #ComputerScienceEduction on X platform, and compare these with Reddit communities in order to find the best practices for improving users' engagement with the content and interactions in subreddits. In addition, network analysis studies can increase understanding of the relations among users and how they respond to each other. Future research should also investigate the most active subreddits related to the field of education to explore the reasons for their high rates of activity.

There are some limitations to this study. One is the researcher's limited knowledge of computational linguistics and big data analysis. Another limitation is reliance on bots-generated content, which might have affected the data on which the results of the study were based.

Acknowledgements

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that no AI or AI-assisted technologies have been used to generate, refine, or correct the content in the manuscript. The ideas, design, procedures, findings, analyses, and discussion are originally written and derived from careful and systematic conduct of the research.

References

Carpenter, J., McDade, C., & Childers, S. (2018, March). Advice seeking and giving in the Reddit r/Teachers online space. In *Society for Information Technology & Teacher Education International Conference* (pp. 2207–2215). Association for the Advancement of Computing in Education (AACE).

Chang-Kredl, S., & Colannino, D. (2017). Constructing the image of the teacher on Reddit: Best and worst teachers. *Teaching and Teacher Education*, *64*, 43–51.

Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining social media data for understanding students' learning experiences. *IEEE Transactions on Learning Technologies, 7*(3), 246–259.

Cole, J., Ghafurian, M., & Reitter, D. (2019). Word adoption in online communities. *IEEE Transactions on Computational Social Systems*, *6*(1), 178–188.

Duncan-Howell, J. (2010). Teachers making connections: Online communities as a source of professional learning. *British journal of educational technology*, *41*(2), 324–340.

Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM Education, 6*(1).

Srinivasan, D. B., & Mekala, P. (2014). Mining social networking data for classification using REPTree. *International Journal of Advance Research in Computer Science and Management Studies, 2*(10).

Vella, M. (2015). Reddit reboots. *Time Magazine, 186*(3), 46–50.

Willet, K. B. S., & Carpenter, J. (2019, March). Educators on the Front Page of the Internet: Education-Related Subreddits as Learning Spaces. In *Society for Information Technology & Teacher Education International Conference* (pp. 2787–2795). Association for the Advancement of Computing in Education (AACE).

Yadav, A., Gretter, S., Hambrusch, S., & Sands, P. (2016). Expanding computer science education in schools: Understanding teacher experiences and challenges. *Computer Science Education*, *26*(4), 235–254.

Contact email: k.alghamdi@mu.edu.sa