Identifying reputation collectors in community question answering (CQA) sites: Exploring the dark side of social media

Pradeep K. Roy[a], Jyoti P. Singh[a], Abdullah M. Baabdullah[b], Hatice Kizgin[c], Nripendra P. Rana[c,*]

[a] Computer Science and Engineering Department, National Institute of Technology Patna, Bihar, 800005, India
[b] Department of Management Information Systems, King Abdulaziz University, Jeddah, Saudi Arabia
[c] School of Management, Swansea University, Fabian Way, Swansea, SA1 8EN, UK

ARTICLE INFO

ABSTRACT

This research aims to identify users who are posting as well as encouraging others to post low quality and duplicate contents on community question answering sites. The good guys called *Caretakers* and the bad guys called *Reputation Collectors* are characterised by their behaviour, answering pattern and reputation points. The proposed system is developed and analysed over publicly available Stack Exchange data dump. A graph-based methodology is employed to derive the characteristic of *Reputation Collectors* and *Caretakers*. Results reveal that *Reputation Collectors* are primary sources of low-quality answers as well as answers to duplicate questions posted on the site. The Caretakers answer limited questions of challenging nature and fetches maximum reputation against those questions whereas *Reputation Collectors* answers have so many low quality and duplicate questions to gain the reputation point. We have developed algorithms to identify the *Caretakers* and *Reputation Collectors* of the site. Our analysis finds that 1.05% of *Reputation Collectors* post 18.88% of low-quality answers. This study extends previous research by identifying the *Reputation Collectors* and how they collect their reputation points.

1. Introduction

Community Questions Answering (CQA) sites such as Yahoo! Answers (YA),[1] Stack Overflow (SO),[2] Stack Exchange,[3] Quora[4] etc. are Web 2.0 based services, which allow people to seek information by asking questions and share knowledge by providing answers to questions asked by rest of the community members (Luo, Zhang, Hu, & Wang, 2016; Roy et al., 2018). Some CQA sites allow users to ask questions without any topic restriction such as YA, Quora, while other CQA systems are devoted to a specific area such as SO. SO was primarily developed for software developer to make it a useful resource of conceptual or code review questions for them. The content of this site sometimes supplements the official software documentation as well (Serna, Bachiller, & Serna, 2017; Treude, Barzilay, & Storey, 2011). Therefore, the quality of content on this site is the most important thing (Aladwani, 2017; Hashim and Tan, 2015; Jin, Zhou, Lee, & Cheung, 2013). Any compromise with quality of the content on this site will make it useless and people will be afraid using it. As digital and social

media platforms and applications continue to disseminate, both positive and negative aspects associated with them are becoming increasingly apparent (AlAlwan, Rana, Dwivedi, & Algharabat, 2017; Aswani, Kar, Ilavarasan, & Dwivedi, 2018; Dwivedi, Kapoor, & Chen, 2015; Dwivedi et al., 2016; Dwivedi, Rana, Jeyaraj, Clement, & Williams, 2017; Dwivedi, Rana, Janssen et al., 2017; Dwivedi et al., 2018; Ismagilova, Dwivedi, Slade, & Williams, 2017; Kamboj, Sarmah, Gupta, & Dwivedi, 2018; Kapoor, Tamilmani, Rana, Patil, & Dwivedi Nerur, 2017; Kapoor and Dwivedi 2015; Plume, Dwivedi, & Slade, 2016; Rathore, Ilavarasan, & Dwivedi, 2016; Shareef, Mukerji, Dwivedi, Rana, & Islam, 2017, Shareef, Mukerji, Alryalat, Wright, & Dwivedi et al. 2018 ;Tamilmani, Rana, Alryalat, Alkuwaiter, & Dwivedi, 2018). For example, every social media platform such as Facebook, Twitter and YouTube is now facing the problem of some naughty users who are trying to dilute these forums. A number of researchers have started finding the notorious activities on these forums (Garcia and Sikström, 2014; Fox and Moreland, 2015; Kaplan and Haenlein, 2010; Krasnova, Widjaja, Buxmann, Wenninger, & Benbasat, 2015; Stieglitz, Mirbabaie, Ross, &
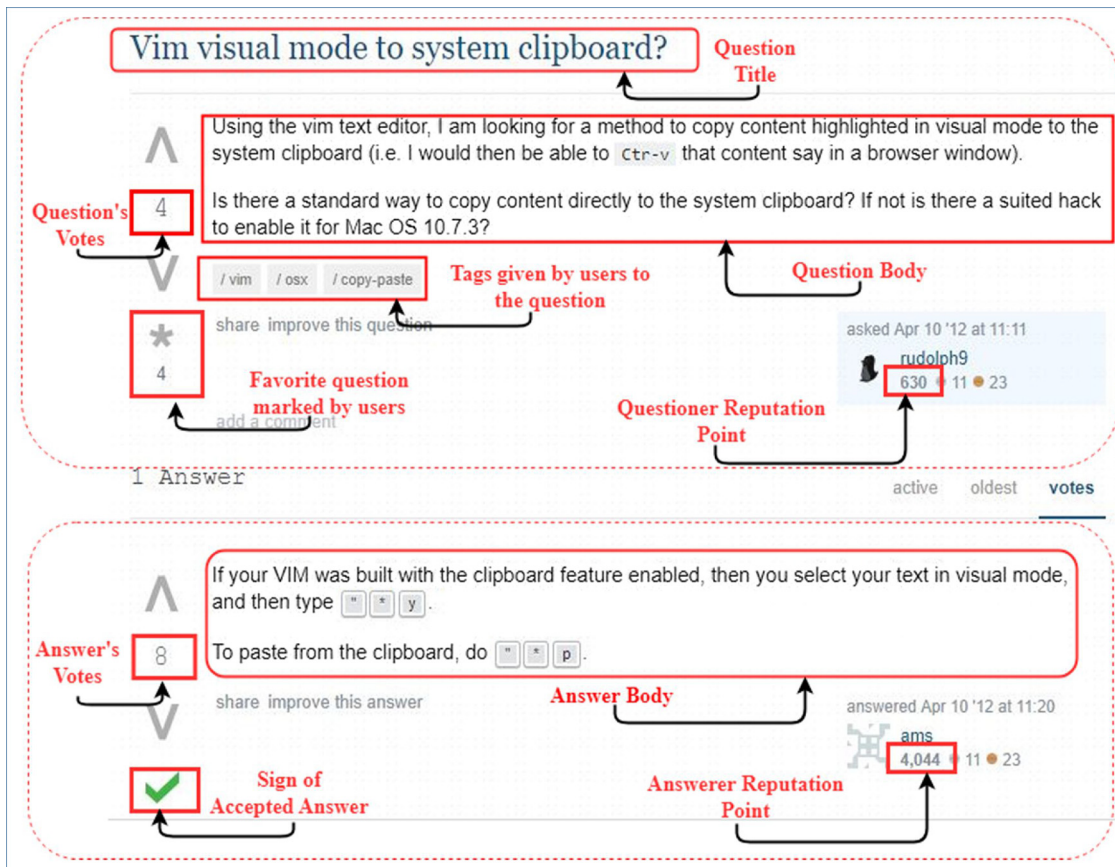
Fig. 1. A sample of Stack Exchange post with different attributes.

Neuberger, 2018). CQA site such as Stack Exchange is not an exception to this list and some people have started posting abusing content, duplicate questions and below-quality answers to this forum. This forum is highly dependent on its quality content, hence posting of any below-quality questions, or duplicate questions and their low-quality answers may be seen an act of negative activity on this site. These activities will kill the very purpose of the forum for which it was being developed and used. Traditionally the quality of the posts (questions as well answers) is evaluated and maintained by the community users only. Users can vote up or vote down a post (questions or answers) to express their views on posts. Fig. 1 shows a sample post of CQA site with different attributes.

To encourage user participation in the site, reputation points and badges system are in use. For example, the questioner gets $+2$ reputation points by accepting an answer while the answerer gets $+15$ reputation points. Other activity that a user can do is, he/she can comment on any posted answers they are not satisfied. User can vote positive or negative with their satisfactory level. Every activity that a user performs with a posted answer, the reputation point is updated in respective questioner/answerer reputation point. Table 1 presents the summary of activities along with their reputation point change. If a questioner has not received any answer for his/her question, and he/she needs the answer immediately then he/she may assign a bounty on that particular question to attract other user to answer the question. If the answer of a bounty questions is accepted, the bounty is added to the answerer's reputation and the same reputation point is subtracted from the questioner's reputation. There is no limit on the bounty, any number of reputation point can be added to the question as bounty. However, a user can earn a maximum of 200 reputation (except bounty) points per day, according to Stack Overflow community policy rules.

The user's privilege keeps on increasing as they earn more

Table 1
Stack exchange reputation scheme for community users.

Action	Reputation Change
Answer is voted up	$+10$
Question is voted up	$+5$
Answer is accepted	$+15$ ($+2$ to acceptor)
Question is voted down	-2
Answer is voted down	-2 (-1 to voter)
Experienced Stack Exchange user	Onetime $+100$
Accepted answer to bounty	$+$bounty
Offer bounty on question	-bounty

reputation points. There are five classes of privilege[5] such as (i) *documentation privilege*, (ii) *creation privilege*, (iii) *communication privilege*, (iv) *moderation privilege*, and (v) *milestone privilege*. Among these privileges, the milestone is the highest level privilege (maximum control over the site content) whereas documentation is the lowest level privilege: i) with *documentation privilege*, a user has authority to approve or reject the changes made on the posts, can comment on the proposed changes and so on ii) the main authority *creation privilege* is to create a new *tag* for the site, iii) with communication privilege, a user can create gallery chat rooms where only specific users may chat, iv) the main authority of *moderation privilege* is *marked questions as protected*. A protected question prevents answers being added by anonymous and new users, v) finally, a *milestone privilege* makes a user as a trusted user of the site. Hence, they can delete the questions having negative votes, also delete the low-quality answers if there is no hope to improve it further and so on. A user with this privilege may get a special access to the data collected from the community users. A group of users has

---

[5] https://stackoverflow.com/help/privileges.