Online Social Networks and Media 2 (2017) 45–59

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

The good, the bad and the deviant in community question answering

Imrul Kayes[a,*], Nicolas Kourtellis[b], Adriana Iamnitchi[c]

[a] Sonobi Inc., 444 W New England Ave #215, Winter Park, FL 32789, USA
[b] Telefonica Research, Plaça d'Ernest Lluch i Martin, 5, 08019 Barcelona, Spain
[c] Computer Science and Engineering, University of South Florida, Tampa, FL, USA

ARTICLE INFO

Article history:
Received 29 April 2017
Revised 9 August 2017
Accepted 10 August 2017

Keywords:
Community question answering
Content abusers
Flagging
Crowdsourcing

ABSTRACT

Community question answering (CQA) are collaborative online places where members ask questions for others to answer. Community members on these platforms share their expertise on various topics, from mechanical repairs to parenting. As a crowd-sourced service, such platforms not only depend on user-provided questions and answers, but also rely on their users for monitoring and flagging content that violates community rules.

This study focuses on user-reported flags to characterize the behavior of the good guys and bad guys in a popular community question answering, Yahoo Answers. Conventional wisdom is to eliminate the users who receive many flags. However, our analysis of a year of traces from Yahoo Answers shows that the number of flags does not tell the full story: on one hand, users with many flags may still contribute positively to the community. On the other hand, users who never get flagged are found to violate community rules and get their accounts suspended. This analysis, however, also shows that abusive users are betrayed by their network properties: we find strong evidence of homophilous behavior and use this finding to detect abusive users who go under the community radar. Based on our empirical observations, we build a classifier that is able to detect abusive users with an accuracy as high as 83%.

1. Introduction

Community-based Question-Answering platforms (CQA), such as Yahoo Answers, Quora and Stack Overflow, are collaborative online platforms where users ask and answer questions. Over the past decade, they have become rich and mature repositories of user-contributed questions and answers. For example, Yahoo Answers, launched in December 2005, has more than one billion posted answers [1]. Quora, one of the fastest growing CQA sites, has seen three times growth in 2013 [2]. A study [3] on Yahoo Answers revealed that about 2% of web searches performed by Yahoo Answers users lead to a question posted to the community.

In order to preserve the health and usefulness of online communities, CQAs define community rules and expect users to obey them. To enforce these rules, published as community guidelines and terms of services, these platforms provide users with tools to *flag* inappropriate content. In addition to community monitoring, some platforms employ human moderators to evaluate abuses and determine the appropriate responses, from removing content to suspending user accounts. These digital recordings of unethical behaviors enables the study of human behaviors at much larger scale than what is possible in lab experiments; helps to understand the effect that communication channels has on user behavior; has the potential of guiding the design of mechanisms that foster good behavior.

This article investigates the reporting of rule violations in Yahoo Answers (YA), one of the oldest, largest, and most popular CQAs. Our dataset contains about 10 million editorially curated abuse reports posted between 2012 and 2013. Out of the 1.5 million users who submitted content during the one-year observation period, about 9% of the users got their accounts suspended. We use suspended accounts as a ground truth of bad behavior in YA, and we refer to these users as *content abusers* or *bad guys* interchangeably.

We discover that, although used correctly, flags do not tell accurately which users should be suspended: while 32% of the users active in our observation period have at least one flag, only 16% of them are suspended during this time. Even considering the top 1% users with the largest number of flags, only about 50% of them deserve account suspension. We find that, unlike in other environments where abusers are clearly the bad guys (e.g., cheaters in online games [4]), the situation is not black and white. That is, users flagged many times for rule violations contribute positively to the community by increasing user engagement and providing best answers. Complicating an already complex problem,

* Corresponding author.
E-mail address: imrulkayes11@gmail.com (I. Kayes).

http://dx.doi.org/10.1016/j.osnem.2017.08.002
2468-6964/

we find that 40% of the suspended users have not received any flags.

To reduce this large gray area of questionable behavior, we employ social network analysis tools in an attempt to understand the position of content abusers in the *YA* community. We learned that the follower-followee social network tunnels user attention not only in terms of generating answers to posted questions, but also in monitoring user behavior. More importantly, it turns out that this social network divulges information about the users who go under the community radar and never get flagged even if they seriously violate community rules. This network-based information, combined with user activity, leads to accurate detection of the bad guys: our classifier is able to distinguish between suspended and fair users with an accuracy as high as 83%. Thus, the outcomes of this study can aid human moderators with automated tools in order to maintain the health of the community.

This work is a follow up on previous study of the social world of content abusers in the community question answering [5]. We extend it by performing new analyses of the users who report abuses and help to keep the environment clean, even without getting any benefits from the platform. We refer to these users as *good guys*. Our new investigation shows that good guys are also good contributors of questions and answers. Although very few, they report abuses consistently on diverse topics. Our spatial and linguistic analyses of abuse reports show that while community monitoring exists in diverse languages, cultural differences are evident in abuse reporting statistics. In our new analyses, we also find that on average, suspended users do not occupy central positions in the follower-followee social network.

The article is structured as follows. Section 2 discusses previous analysis of CQAs and the existing body of work on unethical behavior in online communities in general. Section 3 presents the *YA* functionalities relevant to this study and the dataset used. We introduce a deviance score in Section 4 that identifies the pool of bad users more accurately than the number of flags alone. Section 5 demonstrates that deviant users are not all bad: despite their high deviance score, in aggregate their presence in the community is beneficial. This section also presents an analysis of the good guys, who voluntarily flag the abusive content. Section 6 shows the effects of the social network on user contribution and behavior. Section 7 shows the classification of suspended and fair users. We discuss the impact of these results in Section 8.

2. Related work

CQA has attracted much research interest from diverse communities as information science, HCI and information retrieval. We collate past research on community-based question answering in five categories depending on whether it has dealt with content, users, applications, bad behavior in online settings, or CQA communication networks.

Content. Research in this area has investigated textual aspects of questions and answers. In so doing, it has proposed algorithmic solutions to automatically determine: the quality of questions [6,7] and answers [8,9], the extent to which certain questions are easy to answer [10,11], and the type of a given question (e.g., factual or conversational) [12].

Users. Research on CQA users has explored how users interface with the platform. Dearman and Truong [13] asked why users of *YA* do not answer questions and found that active answerers (who contribute most of the answers) do not want to get reported for abuse and potentially lose access to the community. Liu et al. [3] asked why users ask questions. They concluded that a vast majority of the askers are failed searchers; when web search fails they become askers. Kayes et al. [14] analyzed the influence of national culture on users' online questioning and answering behavior and found that national cultures differ in Yahoo Answers along a number of cultural dimensions. They also analyzed users' privacy settings and found that privacy-concerned users have higher qualitative and quantitative contributions [15].

Applications. Research on applications has developed techniques and tools to improve system performance and to provide better usability. Researchers have proposed effective ways of recommending questions to the most appropriate answerers [16,17]. Shtok et al. [18] used the repository of past answers to answer new open questions in order to reduce the number of unanswered questions.

Bad behavior in online settings. Qualitative and quantitative studies of bad behavior in online settings have been done before, including online chat communities [19], online multiplayer video games [20], and geosocial networks [21]. A body of work has also investigated the impact of the bad behavior. Researchers have found that bad behavior has negative effects on the community and its members: it decreases community's cohesion [22], performance [23] and participation [24]. In the worst case, users who are the targets of bad behavior may leave or avoid online social spaces [24].

Communication networks. Research on communication networks analyzed users' social networks on the CQAs and attempted to understand the interplay between users' social connections and their Q&A activities. Wang et al. [25] analyzed the social network of Quora and found that users who contribute more and better answers tend to have more followers. Panovich et al. [26] evaluated the impact of tie strength in question answers. They found that stronger ties (close friends) contribute a subtle increase in answer quality compared to weak ties. Kayes et al. [5] used user-provided rule-violation reports and exploited the social networks to detect content abusers in CQAs.

This article sheds light on abusive behavior in CQA communities by studying YA, one of the largest and oldest such communities. It quantifies how YA's networks channel user attention, and how that results in different behavioral patterns that can be used to limit abusive behavior.

3. Yahoo answers

After 10 years of activity, Yahoo Answers has 56 million monthly visitors from the U.S. [27], along with many others from the rest of the world.

3.1. The platform

YA is a CQA in which community members ask and answer questions on various topics. Users ask questions and assign them to categories selected from a predefined taxonomy, e.g., *Business & Finance, Health*, and *Politics & Government*. Users can find questions by searching or browsing through this hierarchy of categories. A question has a title (typically, a short summary of the question), and a body with additional details.

A user can answer any question but can post only one answer per question. Questions remain open for 4 days for others to answer. However, the asker can select a best answer before the end of this 4-day period, which automatically *resolves* the question and archives it as a *reference* question. The best answer can also be rated between one to five, known as *answer rating*. If the asker does not choose a best answer, the community selects one through voting. The asker can extend the answering duration for an extra 4 days. The questions left unanswered after the allowed duration are deleted from the site. In addition to questions and answers, users can contribute comments to questions already answered and archived.