Encouraging User Behaviour with Achievements: An Empirical Study

Scott Grant
School of Computing, Queen's University
Kingston, Ontario, Canada
scott@cs.queensu.ca

Buddy Betts
OUYA, Inc.
Santa Monica, California, USA
buddy@ouya.tv

***Abstract*—Stack Overflow, a question and answer website, uses a reward system called badges to publicly reward users for their contributions to the community. Badges are used alongside a reputation score to reward positive behaviour by relating a user's site identity with their perceived expertise and respect in the community. A greater number of badges associated with a user profile in some way indicates a higher level of authority, leading to a natural incentive for users to attempt to achieve as many badges as possible. In this study, we examine the publicly available logs for Stack Overflow to examine three of these badges in detail. We look at the effect of one badge in context on an individual user level and at the global scope of three related badges across all users by mining user behaviour around the time that the badge is awarded. This analysis supports the claim that badges can be used to influence user behaviour by demonstrating one instance of an increase in user activity related to a badge immediately before it is awarded when compared to the period afterwards.**

I. Introduction

Stack Overflow is a question and answer website created in 2008 that is primarily used by computer programmers. As stated in the FAQ, if it is related to coding, it should be on Stack Overflow [1]. Users are actively encouraged to participate in the community by creating public user profiles, engaging in discussion by asking good questions, and providing helpful and relevant answers. This desirable user behaviour is rewarded with the combination of a numerical score called reputation and a goals framework called badges. Reputation is awarded when individual activities are performed, such as voting on the quality (or lack of quality) of questions and answers, and providing content that is also voted favourably itself. Badges are awarded when larger-scale goals are completed, including the *Civic Duty* badge for voting 300 times or the *Fanatic* badge for visiting the site each day for 100 consecutive days. There is a hierarchy of badges, with bronze badges being relatively common and easy to achieve, silver badges being more difficult, and golden badges awarded for long term dedication and recognition from the community. Reputation and badges are treated as an estimate of how much the community trusts each user. This leads to a natural incentive for users to attempt to achieve as much reputation and as many badges as possible to demonstrate their expertise and respect in the programming community.

Awarding badges for user behaviour is a form of gamification, defined as the process of game-thinking and game mechanics to engage users and solve problems [8]. For example, one large study of the reputation mechanism used by eBay showed a clear incentive for semi-anonymous users to obtain meaningful incentive through reputation-garnering [7]. In the gaming community, this type of reward is often referred to as an achievement [6]. The Xbox Live[1] and Steam[2] communities highlight achievements prominently on public user profiles, resulting in a strong desire by users to perform in-game feats that they may not otherwise have attempted [5]. Game designers understand that achievements have a predictable positive impact on play time, and online games like World of Warcraft appear to understand how to manipulate reward structures, including the player achievement system, to motivate users who are driven to achieve [3], [4]. In fact, one such achievement[3] involves repeatedly performing difficult and time-consuming tasks to raise a user's reputation with in-game factions, ultimately receiving an achievement and the ability to add "the Insane" as a suffix to the character's name.

At the time of writing, there are 78 unique named badges available on Stack Overflow. Badges have appeared and disappeared over the site's history, and it is likely that new badges will appear in the future. Some of these badges can be awarded multiple times, which helps to explain how users like #22656 (the prolific Jon Skeet, who is well known on the site[4]) have accrued hundreds or thousands of these awards. Figure 1 shows the number of badges awarded to users who have been awarded at least one badge.

Stack Overflow releases a set of the user-generated contributed content as a *cc-by-sa* licensed data dump[5] [2]. This data contains information about the users, their comments, posts, and related activities, a subset of voting history, and which badges were awarded. We use this data to compile user activity with the goal of identifying patterns that suggest significant shifts in behaviour specifically designed to obtain a badge. The goal is to use the data mined from the logs to demonstrate the shift in behaviour motivated by the badge reward system. In this study, we focus specifically on a subset

[1] http://www.xbox.com/
[2] http://store.steampowered.com/
[3] http://www.wowpedia.org/Insane_in_the_Membrane
[4] http://meta.stackoverflow.com/questions/9134/jon-skeet-facts
[5] http://blog.stackoverflow.com/category/cc-wiki-dump/

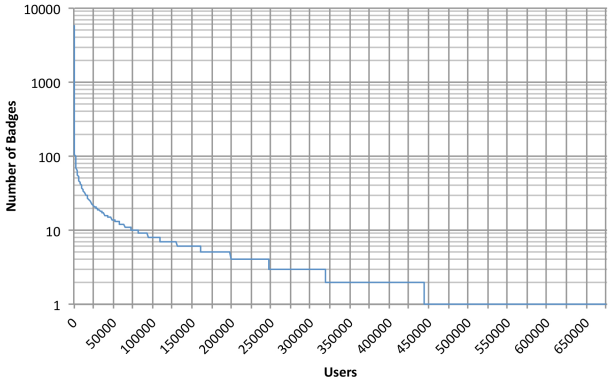

Fig. 1. Number of Stack Overflow users who have been awarded at least one badge. Of the 1,295,620 total user accounts represented in the logs, there are 676,770 users (52.2%) with at least one badge.

of three badges related to editing posts (*Strunk & White*, *Copy Editor*, and *Archaeologist*.

II. Methodology

The data dump for the main Stack Overflow site consists of nearly 40GB of logs. We parse this data line by line and extract individual activities associated with a user id that are used to form user activity profiles. This data is loaded into a single MySQL[6] database indexed by user id, and the history or behaviour for each user is generated by selecting all rows matching the user id and ordering by timestamp.

To examine how often badges are pursued by users, whether actively or not, we look at the number of users who have achieved at least one instance of a particular badge, and how their activity is affected around the time that the badge is awarded. To identify situations where user behaviours appear to have been influenced by a badge, we plot the history of that user around the time that they received the badge. For our analysis, we focus on four pieces of data for each user account: badge awards, new posts, edits and other modifications to posts, and user comments.

In Figure 2, we show an example of the visualization used to highlight these interesting behaviours extracted from the log files. This visualization abstracts away details about the content and leaves a representation of the type of behaviour that quickly shows how an individual uses the site in regular use and around the time that a badge is achieved. Figure 2 represents a linear timeline progressing from left to right. To focus on activity around a badge, we use a four-month window centered on the moment that the badge is achieved. In this diagram, the left-most side corresponds to a point in time two months before the badge was achieved, while the right side corresponds to two months after the badge was achieved. Red bars in the first row indicate that one of the badges was awarded to the user, orange bars in the second row indicate that the user left a comment on a post, blue bars in the third

[6]http://www.mysql.com/

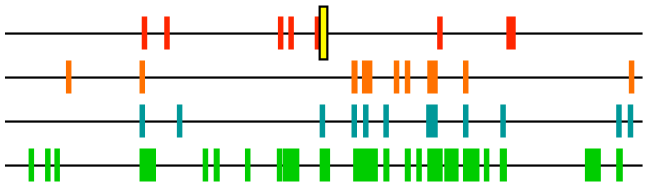

Fig. 2. Timeline of user activity for user #17343. A description of the visualization can be found in Section II.

row indicate new posts by this user, and green bars in the bottom row indicate modifications to a post, such as edits. The yellow highlighted bar in the center of the top row indicates a particular noteworthy event. In this case, it is the moment that the *Copy Editor* badge is awarded to the user after editing 500 posts.

From the existing badges, we identified three that may not necessarily be awarded as a result of typical user behaviour, and may demonstrate unusual behaviour that can be explained if the user made a conscious effort to complete the badge pre-conditions. The silver *Strunk & White* badge which references the authors of the classic writing guide *The Elements of Style* is awarded to users who have edited 80 posts. The golden *Copy Editor* badge is awarded to users who have made 500 edits to posts. The silver *Archaeologist* badge is related to the other two, and is awarded to a user who has edited 100 posts that were inactive for six months. In the dataset used for this study with 1,295,620 user accounts, 2,726 of those had achieved the *Strunk & White* badge, 470 had received a *Copy Editor* badge, and 221 had been awarded an *Archaeologist* badge.

III. Analysis

A. Individual User Behaviour

Stack Overflow promotes the idea that user-submitted content belongs to the community instead of the user, and as a result, encourages community editing. This shared responsibility leads to questions that are rephrased to be accessible to a wider number of users encountering a similar problem, and answers that are well written, clear, and helpful. To encourage this behaviour which might otherwise be a selfless task, users gain reputation for edits that are received favourably by other users, and over time become eligible for badges.

Figure 3 provides examples of the four-month window of activity for four user accounts that have been rewarded with the *Copy Editor* badge for making 500 edits to existing posts. Edits to posts are shown as vertical green bars on the fourth row of the visualization of each user's history. Unlike the visualization in Figure 2, which shows a relatively uniform pattern of behaviour across the events in the user's history, these four accounts show behaviour that suggests an influence from the badge. In each case, users have actively edited posts, and in many cases, show activity across all of the logged behaviours. Finally, when the users achieve the badge as indicated by the highlighted vertical yellow bar on the top row, a significant change in behaviour is observed. In these

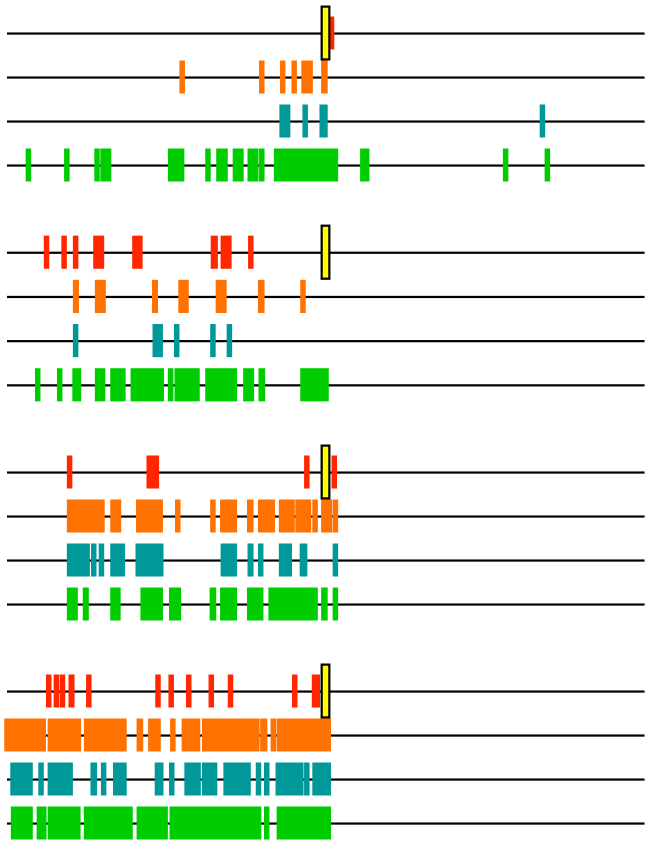

Fig. 3. Four Stack Overflow user accounts centered on a four-month window around the time they were awarded the *Copy Editor* badge for making 500 edits. Edits to posts are displayed using a green bar in the fourth row. In each case, users maintain a very active behaviour pattern involving a large amount of editing which is directly related to the badge in question, until finally completing the tasks required to achieve the badge. At that point in time, a change in their behaviour is observed. The four users, listed from top to bottom, are #163809, #366898, #647772, and #1219006.

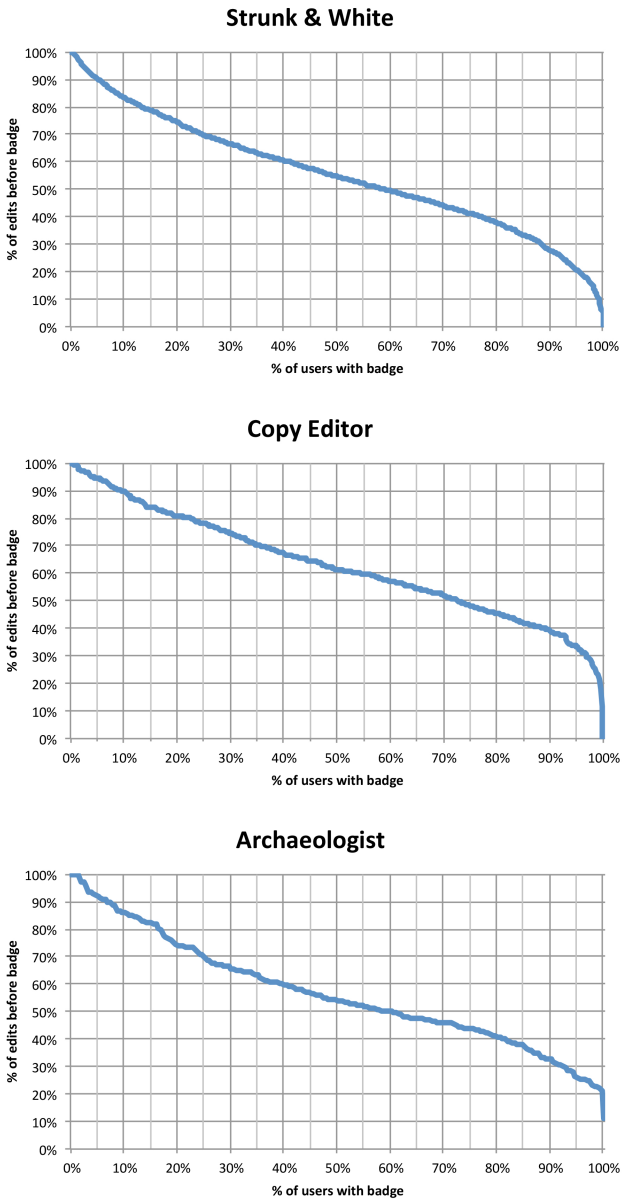

Fig. 4. The percentage of edits made in the two-month period before a badge is awarded compared to the two-month window after a badge is awarded. In each of the three edit-related badges, users are on average more likely to make edits before the badge than after the badge, suggesting that the badge is successfully motivating users to perform desirable behaviour.

cases, it appears as though the users have actively shifted their focus toward completion of the badge in the time before it is awarded, and upon reaching their achievement, see no immediate need to continue the labour-intensive task.

Each of the three badges that we investigated in detail had users who fit this pattern to some degree. Although there are other badges awarded in the diagrams, these are typically badges like *Great Question* or *Nice Answer* that can be achieved multiple times and are more difficult to game.

B. Global Influence

The three editor badges referenced in Section III-A that reward otherwise selfless edits should result in a higher number of edits, on average, in the time before a badge is awarded when compared to the time after the badge is awarded. For example, if user behaviour is influenced by the possibility of receiving the *Copy Editor* badge, and the incentive to perform edits is no longer there after the badge is awarded, we would expect to see less edits on average after the badge was awarded for all users who had received it when compared to the time period before the badge.

In Figure 4, graphs for each of the three badges are provided to show the distribution of edits made around the four months that a badge is awarded. A higher value in the y-axis indicates a higher percentage of edits made in the two-month period before the badge was awarded, and values below 50% indicate that more edits were made in the two-month period after the badge. The x-axis is a list of user accounts ordered by the percentage of edits made in the period prior to receiving the

badge. For example, for each of the 2,726 users who received the *Strunk & White* badge, a corresponding data point is plotted in its graph in Figure 4. These users are sorted along the x-axis based on how many edits they made before receiving the badge compared to afterwards. If one of the users made 90 edits in the two months before the badge and 10 edits in the two months after the badge, their corresponding y-axis value would be 90%, and their x-axis position would be the position in the sorted list of users by y-axis value.

As shown in Figure 4, each of the three badges results in a greater number of users spending more time editing posts before the badge. In the two months before the *Strunk & White* badge was awarded to users who make 80 edits, 58.4% of user accounts who received this badge made fewer edits after the badge had been received. The motivation to edit was more pronounced with the golden *Copy Editor* badge, at 72.9% of edits made in the two months before the badge when compared to the two months after the badge. It seems plausible that the increased effect is due in part to the greater amount of work required to achieve this badge, with 500 edits required instead of 80, and due to the fact that the badge is golden (and therefore more prestigious) instead of silver. Additionally, the *Archaeologist* badge which requires edits to posts older than six months corresponded to 60.6% of edits made in the two-month period before the badge when compared to the two months after the badge.

Each of the three badges rewarding edits made to existing posts can be observed to correspond with an increase in edits before the badge when compared to the time period after the badge. We believe that this leads to the conclusion that the badges are working as intended, and motivating users to actively improve the site.

IV. Threats to Validity

One issue in this study is that the researcher must demonstrate that badges influence user behaviour instead of badges being a side effect of typical user activity. For example, very active users who enjoy the site will receive badges as a natural part of their use. In this study, we have attempted to identify cases where user behaviour deviates significantly enough for a brief time and is concluded upon achievement of a particular badge that explains the behaviour. While this is not a proof of intent, we believe that it is sufficient in most cases to show the influence of badges. As seen in Figure 5, aggressive users of the site, such as #22656 referenced earlier, achieve

Fig. 5. Timeline of user activity for user #22656 who holds the highest reputation and greatest number of badges.

badges purely as a side effect of their regular activity. In this initial study, we only look at three of the nearly eighty unique named badges available, and we do not have enough evidence to suggest that these results are generalizable. In addition, many badges, including the *Great Question* and *Nice Answer* badges referenced earlier, are awarded simply on the basis of providing good content, and it is difficult to imagine a way to convincingly extract data showing that a user is driven by rewards.

Many users who use the site regularly may shift their behaviour slightly to complete a badge that they are close to fulfilling instead of exclusively performing an activity such as editing posts. However, in the majority of cases, user behaviour is not as polarized as Figure 3 suggests. This makes it difficult to look at users on an individual basis to draw conclusions about the effectiveness of the badge. We believe that the global analysis in Section III-B suggests the badges are effective on a large scale without relying on single users to show the trend.

V. Conclusions and Future Work

In this paper we demonstrate how one way in which badges, a collection of visible awards associated with public user profiles, are used to influence user behaviour. For Stack Overflow, the goal of providing a high-quality question and answer service requires a community that subscribes to the goal of producing good content through submission, editing, and review. Three badges specifically designed to encourage users to edit posts for quality are used to examine user behaviour using the publicly available data dump released by the site, and are shown to promote the desired ability on both an individual and global level. We plan to expand this initial exploratory study to examine how a greater subset of the badges affects other aspects of user behaviour, and how badges can be compared and evaluated to measure their effectiveness.

Acknowledgements

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Stack Overflow - Frequently Asked Questions. http://stackoverflow.com/faq.

[2] Alberto Bacchelli. Mining Challenge 2013: Stack Overflow. In *The 10th Working Conference on Mining Software Repositories (MSR '13)*, page to appear, 2013.

[3] Jonathan Schaffer Bell, Swapneel Kalpesh Sheth, and Gail E Kaiser. A Large-Scale, Longitudinal Study of Player Achievements in World of Warcraft. 2011.

[4] Thomas Debeauvais, Bonnie Nardi, Diane J Schiano, Nicolas Ducheneaut, and Nicholas Yee. If you build it they might stay: Retention mechanisms in World of Warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games (FDG '11)*, pages 180–187. ACM, 2011.

[5] Kai Huotari and Juho Hamari. "Gamification from the perspective of service marketing. In *Proceedings of the Gamification: Using Game Design Elements in Non-Gaming Contexts Workshop*, 2011.

[6] Mikael Jakobsson. The achievement machine: understanding Xbox 360 achievements in gaming practices. *Game Studies*, 11(1), 2011.

[7] Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9(2):79–101, 2006.

[8] Gabe Zichermann and Christopher Cunningham. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. 2011.