Stack Overflow and Gamification - Something from Nothing?

Andrew Marder

February 10, 2015

"Stack Exchange is a network of 130+ Q&A communities including Stack Overflow, the preeminent site for programmers to find, ask, and answer questions about software development... Stack Overflow Careers connects Stack Overflow users with employers." On January 20, Spoelsky (2015) announced that Stack Exchange had raised $40 million in venture capital funding. Stack Exchange gives users who ask questions access to expert technical help for free. Users who answer questions build their reputation for technical expertise, and use that reputation to find better jobs.

Although Stack Overflow's creation was strongly rooted in computer science, the social sciences will play a large role as the platform matures. There are a number of big picture questions to consider. How should Stack Overflow design its platform to maximize profits? Is the firm's profit-maximizing strategy efficient, does it maximize total surplus? How equitable is the profit-maximizing strategy? Stack Overflow creates value by improving job matches, how much of that value is captured by the platform, employers, and employees? What motivates users to contribute to Stack Overflow, what do user preferences look like? How important are self-interested motives versus pro-social motives? What combination of reputation rules and badge awards maximize the quantity and quality of contributions on the site?

This paper takes a first step along the path of applying econometric analysis to publicly available Stack Overflow data. Specifically, I study how users behave around the time they are awarded badges. By taking a closer look at user actions we start to get an idea of how badges motivate user contributions.

What is Stack Overflow?

Deterding et al. (2011) define "*gamification* as the use of game design elements in non-game contexts." Stack Overflow is a question and answer website designed for programmers that employs a few game elements:

1. A user earns reputation points when another user votes on her posts (5 points when a question is voted up, 10 points when an answer is voted up, 15 points when an answer is accepted, and 2 points when an edit is approved).
2. As a user earns reputation points she unlocks privileges on the site. For instance, a user must have at least 15 reputation points to vote up a question or answer. A list of all privileges is available here.
3. Users are awarded badges for special achievements. One receives the "Informed" badge by reading the entire tour page. A list of badges is available here.

Grant and Betts (2013) present empirical evidence that three of the badges awarded for various editing accomplishments are effective in encouraging users to make more edits in the two months preceding receipt of the badge compared to the two months after receiving the badge. This paper builds on their findings by:

1. Looking at the impact of badges on all types of user activity (posting questions, posting answers, and editing posts).
2. Comparing the impact of different badges. Looking at three different badges, each aimed at promoting a different type of user activity.

How do users behave when earning badges?

Grant and Betts (2013) find that users who receive a badge for editing make more edits in the two-month window before receiving the badge compared to the two-month window after receiving the badge. I extend their work by exploring, on average, how many questions, answers, and edits a user posts around the time of receiving a badge. Let y_{it} be the number of edits user i makes on day t. Following the approach of Jacobson, LaLonde, and Sullivan (1993) define the dummy variable

$$D_{it}^{k} = \begin{cases} 1 & \text{if user } i \text{ earns the badge on day } t+k, \\ 0 & \text{otherwise.} \end{cases}$$

I regress the number of edits user i makes on day t on a user fixed effect α_i and a set of dummy variables indicating whether the user received the badge of interest on day $t+k$

$$\log(1+y_{it}) = \alpha_i + \sum_{k=-29}^{30} D_{it}^{k}\delta_k + \epsilon_{it}.$$

The model parameters are estimated using an ordinary least squares regression, and standard errors are clustered at the user level.

Let t_i^* denote the day user i recieves the badge. Figure 1 plots the expected number of actions taken on the k'th day since receiving the badge

$$f(k) = E\left[\log(1+y_{it}) \mid t = t_i^* + k\right].$$

Strunk & White Edited 80 posts Copy Editor Edited 500 posts (excluding own or deleted posts and tag edits) Archaelogist Edited 100 posts that were inactive for 6 months Curious Asked a well-received question on 5 separate days, and maintained a positive question record Inquisitive Asked a well-received question on 30 separate days, and maintained a positive question record

The 95% confidence interval is tight around the line, standard errors were calculated using the delta method (Williams 2012). Figure 1 confirms the findings of Grant and Betts (2013), editing increases gradually before receiving the badge with a large jump in activity in the day immediately before earning the badge. We also see that editing drops quickly immediately after receiving the badge and gradually declines over time. It's interesting to see how few questions were asked by the recipients of the editing badges in the two months around receiving the badge, and to see that the rate of answering questions tends to be constant throught the two month window.

Figure 2 plots user activity around receiving badges for asking questions. User activity around question badges differs in interesting ways to badges for edits:

1. All types of actions tend to increase in the thirty days leading up to earning the badge. In Figure 1, only edits increased. In Figure 2, the number of questions, answers, and edits posted all increased.
2. User activity stops almost immediately after receiving the badge. This makes me think something might be wrong with the data. It's too stark.

Conclusion

As a fan of Holland and Rubin's motto "no causation without manipulation", it is important to note that this paper does not identify the causal effect of badges (Holland 1986). To reliably estimate the causal impact of

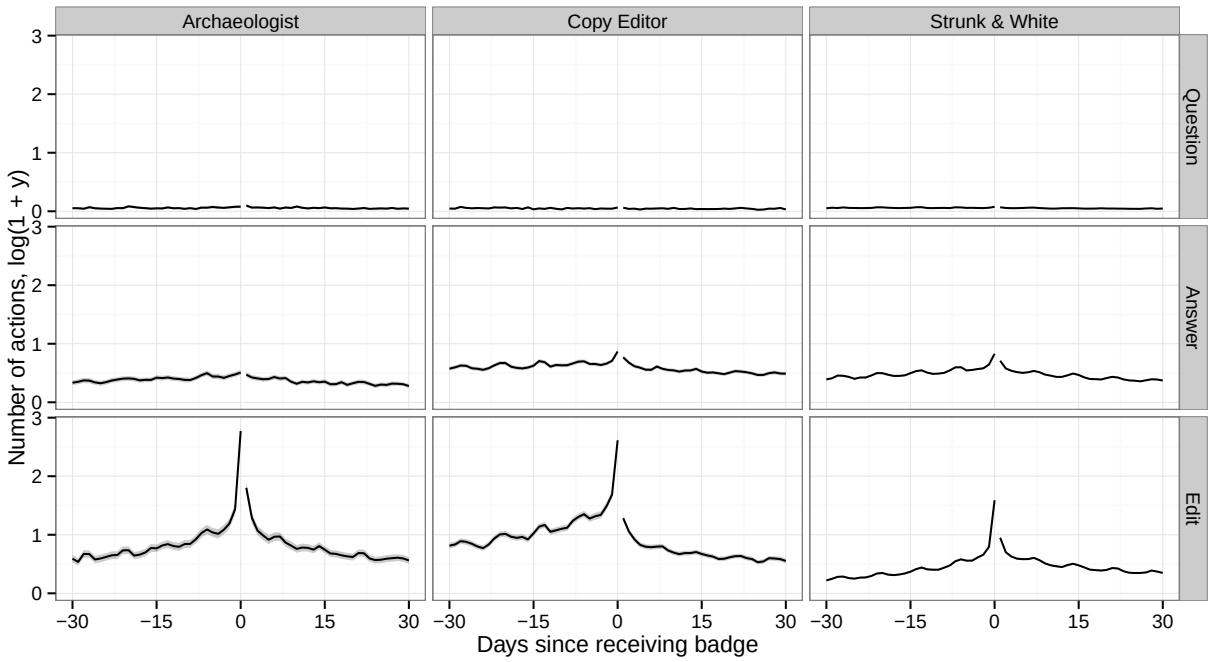

Figure 1: User activity over time - badges for edits

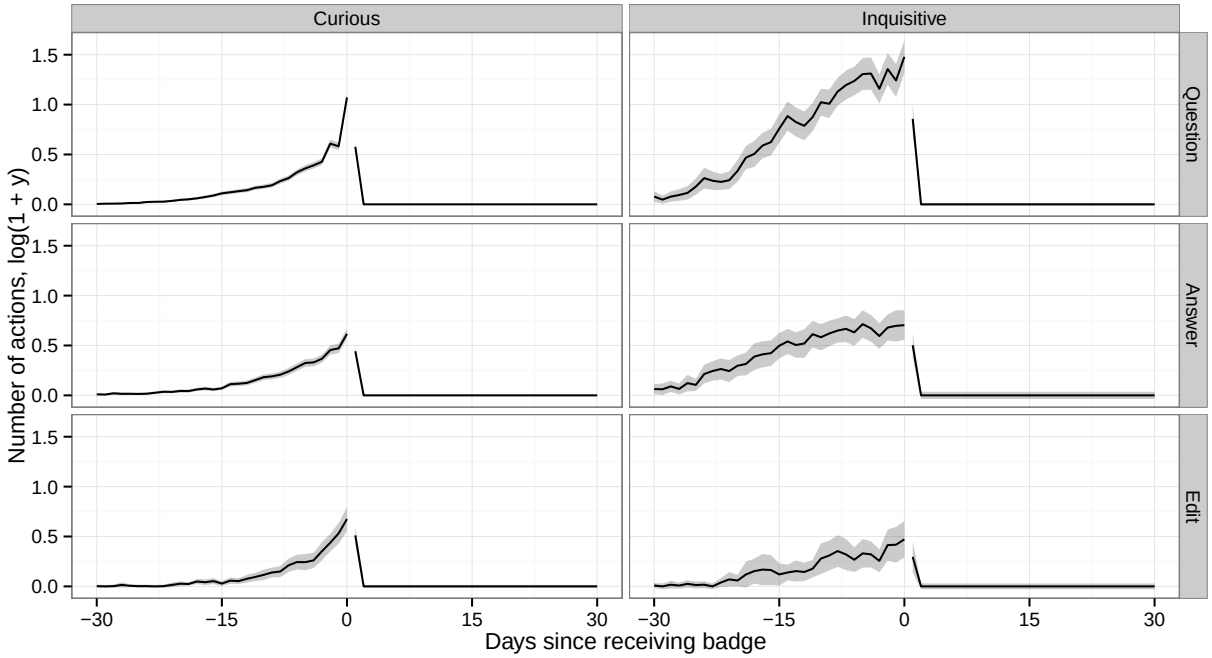

Figure 2: User activity over time - badges for questions

badges of user activity we need a source of exogenous variation (Miller 2013). This paper contributes to the literature by describing more clearly how users behave around receiving badges.

Antin and Churchill (2011)

References

Antin, Judd, and EF Churchill. 2011. "Badges in social media: A social psychological perspective." In *CHI 2011 Gamification Workshop Proceedings*, 1–4. http://uxscientist.com/public/docs/uxsci_2.pdf.

Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. "From game design elements to gamefulness." In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, 9–11. doi:10.1145/2181037.2181040.

Grant, Scott, and Buddy Betts. 2013. "Encouraging user behaviour with achievements: An empirical study." In, 65–68. doi:10.1109/MSR.2013.6624007.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. doi:10.1080/01621459.1986.10478354.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 1993. "Earnings losses of displaced workers." *The American Economic Review*, 685–709. http://www.jstor.org/stable/2117574.

Miller, Ashley. 2013. "Principal turnover and student achievement." *Economics of Education Review* 36 (October): 60–72. doi:10.1016/j.econedurev.2013.05.004.

Spoelsky, Joel. 2015. "Andreessen Horowitz Invests in Stack Exchange." http://blog.stackoverflow.com/2015/01/andreessen-horowitz-invests-in-stack-exchange/.

Williams, Richard. 2012. "Using the margins command to estimate and interpret adjusted predictions and marginal effects." *Stata Journal* 12 (2). http://econpapers.repec.org/article/tsjstataj/v_3a12_3ay_3a2012_3ai_3a2_3ap_3a308-331.htm.