

# Signals Matter: Understanding Popularity and Impact of Users on Stack Overflow

Amul Agrawal

December 22, 2020

The paper investigates the role of the game elements in characterizing social qualities (specifically, popularity and impact) of its users. It operationalizes these attributes using common metrics and apply statistical modeling to empirically quantify and validate the strength of these signals. The results are based on a rich dataset of 3,831,147 users and their activities spanning nearly a decade since the site's inception in 2008. Stack Overflow has evolved from a simple QA site to a massive social community where knowledge seekers and knowledge providers of all levels of expertise interact with each other to solve programming difficulties. The paper tries to explain the role of virtual rewards in studying user qualities on crowd-learning platforms using the linear regression model. Signaling is a well-studied and popular solution to the problem of adverse selection. Signals are images, symbols and signs that allow users to communicate information and meaning with appropriate context. It considers game elements such as reputation scores and badges to be digital signals and investigate whether they are indicative of the performance and qualities of users. The data has been divided in two parts:

- Reputation: Earned (or lost) when a user's question or answer is upvoted (or downvoted), when an answer is marked accepted by the user who originally asked the question, when bounties are received (or spent), or when suggested edits are accepted.
- Badges: Awarded in addition to reputation scores when the corresponding pre-defined set of actions and/or reactions are performed.

After making the dataset, a survey was conducted to gauge the user views about the game elements and the platform in general. Some important survey insights:

- Popularity: Respondents with low reputation consider reputation scores to be better indicators of user popularity than badges; yet majority of respondents with high reputation only somewhat agree with this statement.
- Impact: Fifty one percent of respondents feel that badges are a good measurement of how helpful and knowledgeable a user is.

From the survey further regression model was made in order to deeply analyse the hypothesis made. Three models were made: Control Model, Reputation Model, Badges Model.

The analysis gives us the further insight that RM is a good predictor of popularity. That is, users with high reputation points tend to attract other users to their profile pages.

Some threats to validity: The focus specifically on reward-based features and do not incorporate content-based features. Past evidence has shown that women have faced significant barriers to participating on the site. This suggests that game elements may be biased against some users. In this paper, they present evidence that some of these game elements also act as reliable digital signals of social qualities such as popularity and impact. They reveal that certain non-trivial answer badges, high reputation scores and age of the user on the site indicate significant correlations. Also, they find differentiating characteristics that distinguish communities of popular and impactful users.

Future work could examine linguistic attributes of posts that affect performance. They models can be made by observing more diverse data and the survey results can also be more accurate by including more people for surveys.