Signals Matter: Understanding Social Status on Stack Overflow

Arpit Merchant, Daksh Shah, Gurpreet Singh Bhatia, Anurag Ghosh, Ponnurangam Kumaraguru

arpit.merchant@helsinki.fi, daksh17336@iiitd.ac.in, gurpreetb@iiitd.ac.in, anurag.ghosh@research.iiit.ac.in, pk@iiitd.ac.in Code & Data — https://github.com/precog-iiitd/Signals Matter TheWebConf 2019

Adverse Selection.

What is *adverse selection*?

- Users may not always know how audiences will react to their posts or actions.
- Audiences may not always know about users' abilities.
- Due to information asymmetry, users tend to have a bias towards specific kinds of actions.
- In particular, actions (e.g. low-hanging fruit) that help them to stand out in a large crowd.

Signals that are costly to generate for the signaller and cognitively easy to process for the observer tend to be reliable solutions to the problem of adverse selection.

On Stack Overflow (SO), Reputation Points and Badges are important digital signals.

Dataset Description.

Total number of active users	3,831,147
Total number of questions	15,711,957
Total number of answers	24,492,236
Mean reputation of active users	111
Mean number of badges earned by active users	22
Mean year when users joined the site	2015

In addition to reputation points, badges are awarded to users based on objective, pre-defined metrics.



- 91 different badges available on SO.
- Multiple badges of each type may be earned by a user.

Insights from a user survey reveal that, in general, some users find reputation points to be better indicators of popularity and impact compared to badges.

Characterizing the Effects of Signals.

Positive correlation with Popularity and Impact Scores.

Features	Control Model	Reputation Model	Badges Model	Popularity S views on a us
Age on the site	0.319	0.225	0.191	
Number of questions	0.055	0.074	0.008	
Number of answers	0.250	0.047	0.075	
Number of upvotes	0.122	0.123	0.021	(A) Users wit
Number of downvotes	0.115	0.092	0.048	attract others
Reputation score		0.313		
Nice Answer Badges			0.062	
Populist Badges			0.052	(B) Reputatio
Enlightened Badges			0.029	badges add si
Necromancer Badges			0.039	
Good Answer Badges			0.031	power compa
	$R^2 = 0.911$	$R^2 = 0.939$	$R^2 = 0.957$	

Regression models for predicting Popularity of users.

Badges are better indicators than Reputation Points.

Features	Control Model	Reputation Model	Badges Model	
Age on the site	0.321	0.225	0.065	
Number of questions	0.129	0.129	0.015	
Number of answers	0.250	0.094	0.119	
Number of upvotes	0.085	0.067	0.013	
Number of downvotes	0.033	0.049	0.006	
Reputation score		0.394		
Great Answer Badges			0.069	
Revival Badges			0.055	
Enlightened Badges			0.071	
Necromancer Badges			0.188	
Good Answer Badges			0.043	
	$R^2 = 0.685$	$R^2 = 0.767$	$R^2 = 0.858$	

Regression models for predicting Impact of users.

Users with Answer Badges tend to be more popular and impactful.

Badge











Badge



core = Number of distinct ser's profile page.

th high reputation tend to to their profile pages.

on points and number of ignificant explanatory ared to control features.

Impact Score = Number of distinct views on a user's questions and 'useful' answers (cf. Meta Stack Exchange).

(A) Combining all positive actions into a single reputation score dilutes it's effect.

(**B**) Badges capture a more nuanced summary of a user's contributions.

Popular and Impactful Users.



Question, Good Answer, Necromancer, Populist and Enlightened.





high scoring question

Answers drive popularity, but questions offer more influence.

Feature	HPLI	LPHI	t-statistic	Sig
Questions	54.65	42.02	-6.23	***
Answers	452.63	137.89	-39.97	***
Question Scores	233.15	286.61	4.38	**
Answer Scores	1190.07	679.83	-24.09	***
Reputation	16304.64	8672.31	-30.20	***
Necromancer Badges	2.47	6.32	29.6	***
Populist Badges	0.174	0.218	4.041	**
Great Answer Badges	0.682	0.887	7.84	***

Table 3: Differentiating between HPLI and LPHI users. ** = p < 0.01, *** = p < 0.001 represents statistical significance of Welch's t-statistic after Bonferroni correction (p/14).

Conclusion: Our results offer insights on combating inefficiencies arising due to adverse selection and encourage further exploration of the role of game elements as symbols of social status on online platforms.



NTERNATIONAL INSTITUTE OF **INFORMATION TECHNOLOGY** HYDERABAD

Presentation on Thursday (16.05.2019): 2:00pm - 3:30pm

Early adopters of SO enjoy a strong standing in the community.

LPLI, LPHI, HPLI, and HPHI users can be identified by the presence of badges indicating well-received posts with long-term value, such as Good



The presence of costly to earn and hard to observe signals qualitatively and quantitatively differentiate highly impactful users from highly popular users.