# On Validation and Predictability of Digital Badges' Influence on Individual Users

Tomasz Kuśmierczyk, Kjetil Nørvåg
{tomaszku, noervaag} @idi.ntnu.no

NTNU – Trondheim
Norwegian University of Science and Technology

## Threshold Badges

A *badge* is a formal indicator of some accomplishment or skill that when shown to the others confirms the status of its owner.

*Threshold badges* are awarded after a user performs a certain number of desired actions.

## Notation

User $u \in \mathcal{U}$ in context of the badge $b$ can be represented by a tuple:

$$(\underset{\text{start/eligibility}}{s_u},\ \overset{\text{end/censoring}}{e_u},\ \underset{\text{user features}}{\vec{x}_u},\ \overset{\text{action times}}{\{t_u\}},\ \underset{\text{badge awarding}}{b_u},\ \overset{\text{badge influence}}{i_u})$$

According to the *latent* variable $i_u$, two user types can be distinguished:

- → $i_u = 0$: *user not attracted by the badge*
- → $i_u = 1$: *user attracted by the badge*

## Dataset & Badges

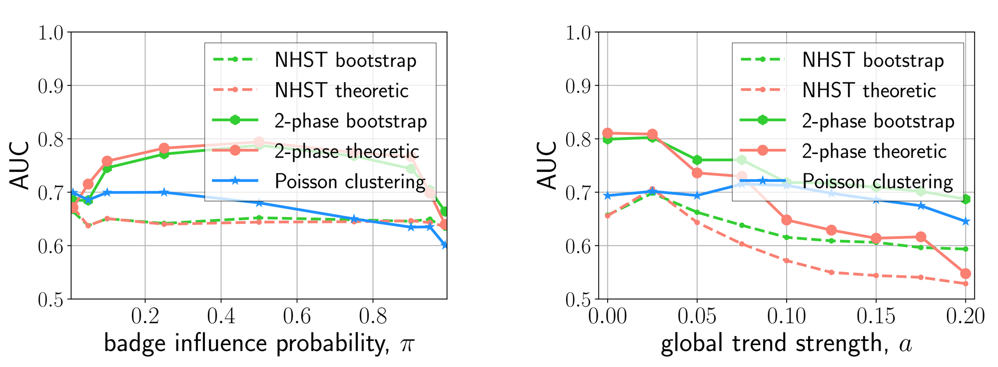
We used a *StackOverflow.com* dataset, that contains timestamped events from between 07/2008 and 09/2014 and some basic information about users:

- user age and location
- total number of user page views, posted comments, and votes

We demonstrate the effectiveness of the proposed approaches for two *threshold badges*:

- *Research Assistant*: awarded to users who edited at least 50 wiki sites describing tags (wiki tag edits). Users with reputation 1500 or higher can perform these actions.
- *Copy Editor*: awarded to users who performed a total of 500 post (e.g., question or answer) edits. Users with reputation 100 or higher can perform these actions.

## Synthetic Data Results



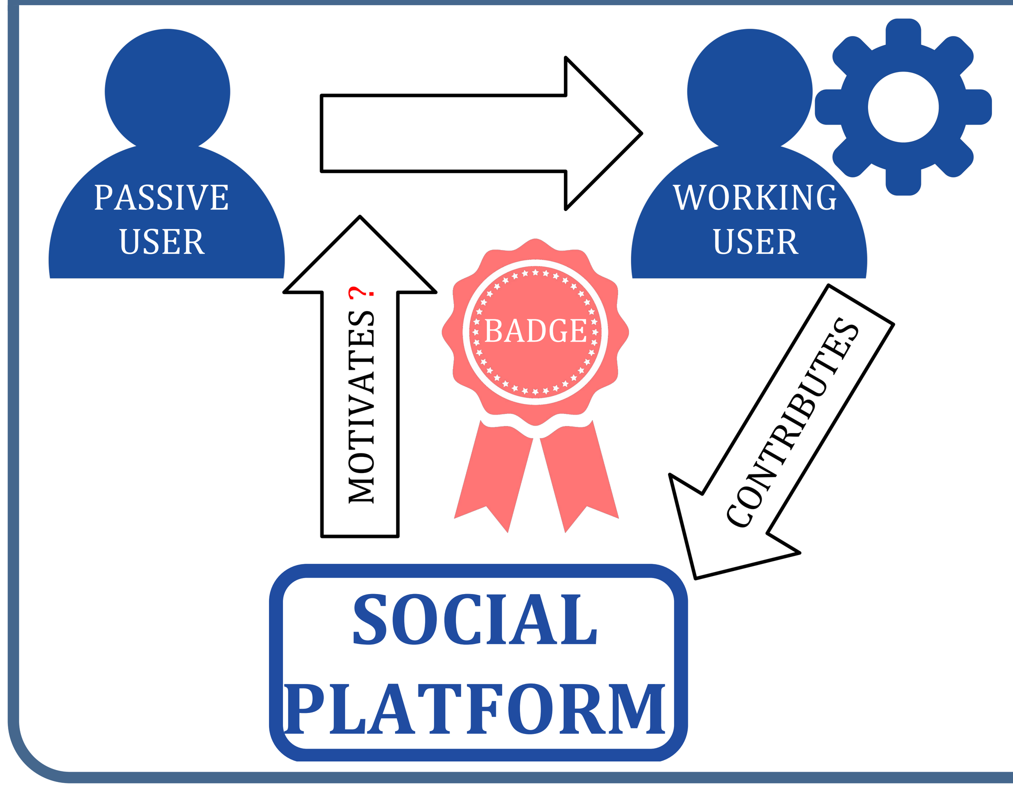

We studied the performance (AUC) of the methods for varying badge impacts, user clusterizations, and in the presence of the global trend in users' activeness levels, finding that:

– *2-phase bootstrap* is the best method in most of the cases,

– *Poisson processes clustering* degrades the least when the global trend is imposed,

– Class imbalance has a low impact on the performance of the methods.

## Problem



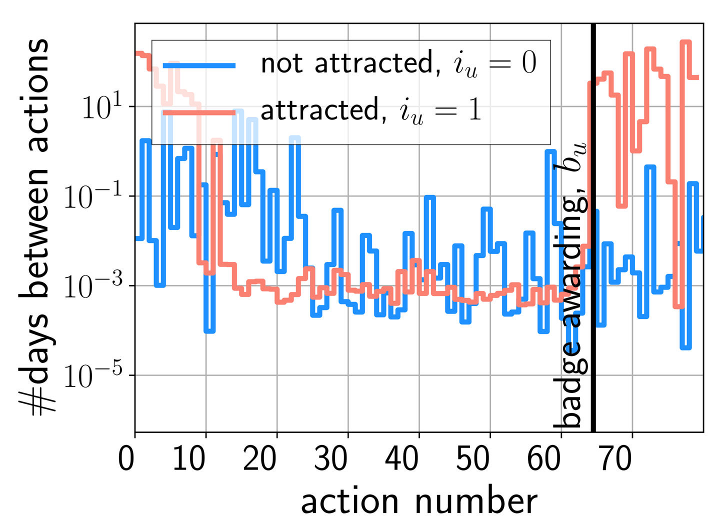

**If the promise of a digital badge award motivated / will motivate a social platform user to perform desired actions?**

**Challenges:**

– Random temporal fluctuations
– Users heterogeneity
– No ground truth

## Observations & Models

**Observation 1:** *Attracted users change their mean behavior around the badge awarding time $b_u$.*




**Model 1:** User temporal trace is a *Poisson process*:

$$\{t_u\} \sim PP(\lambda_u^{i_u}(t))$$

with *intensity* depending on the latent variable $i_u$:

→ $i_u = 0$: *user not attracted by the badge* $\Longrightarrow$ user does not change the behavior over time and intensity is a constant:

$$\lambda_u^0(t) = \lambda^0(u)$$

→ $i_u = 1$: *user attracted by the badge* $\Longrightarrow$ actions mean intensity changes when user is awarded the badge at $b_u$:

$$\lambda_u^1(t) = \begin{cases} \lambda_0^1(u) & \text{if } s_u < t \le b_u \\ \lambda_1^1(u) & \text{if } b_u < t \le e_u \end{cases}$$

**Observation 2:** *Influenceable users have similar characteristics, e.g., similar values of $\vec{x}_u$.*

**Model 2:** Users form clusters in covariates space, e.g., users with $i_u = 1$ attracted to the badge can be partially separated from not attracted ones having $i_u = 0$.

## Joint Models of Temporal Traces and Covariates

**Idea:** Map between $i_u$ and clusters of users formed in either point processes (I.) or covariates (II.) space.

### I. Clustering Poisson Processes

**Generative process:**

**1.** Assign user $u$ to one of the clusters:

$$i_u \sim \text{Bernoulli}(\pi_u)$$

with priors being a function of $\vec{x}_u$:

$$\pi_u = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}_u}}$$

**2.** Draw intensities and point processes using cluster-dependent parameters:

→ for cluster assigned $i_u = 0$:

$$\lambda^0(u) \sim \text{Gamma}(\alpha^0, \beta^0)$$

→ for cluster assigned $i_u = 1$:

$$\lambda_0^1(u) \sim \text{Gamma}(\alpha_0^1, \beta_0^1)$$

$$\lambda_1^1(u) \sim \text{Gamma}(\alpha_1^1, \beta_1^1)$$

To fit $\alpha$-s, $\beta$-s and $i_u$, we integrate out intensities $\lambda$ and perform EM. Additionally, in every M step we optimize weights $\vec{w}$ to fit $\pi_u$-s.

### II. 2-phase Bootstrap

**Phase 1**: Classify users into positives $\mathcal{P}$ and negatives $\mathcal{N}$ using *significance testing* (NHST):

$$H_0 : \lambda_u^{i_u} = \lambda_u^0(t) \quad \text{vs.} \quad H_1 : \lambda_u^{i_u} = \lambda_u^1(t)$$

with log-likelihood ratio (LLR) test statistic and *virtual badges bootstrapping* to estimate the test statistic distribution under $H_0$ (=simulate badges at $b'_u \neq b_u$ to get $F_{LLR'}$).

**Phase 2**: To refine classification results, perform clustering with Gaussian mixtures in covariates space using Dirichlet hyperpriors, different for each user group $G \in \{\mathcal{N}, \mathcal{P}\}$:

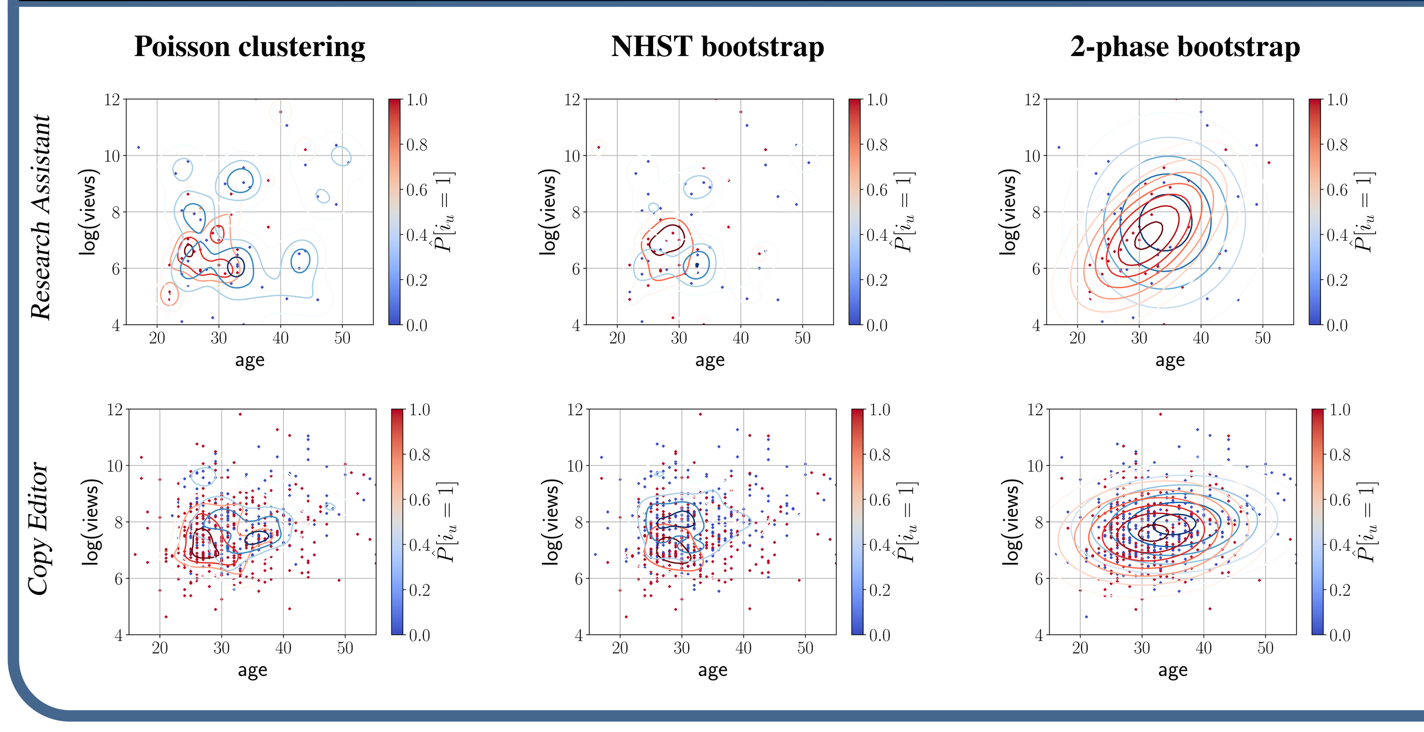
$$\vec{\pi}_G \sim \text{Dirichlet}(\alpha_G^c, ..., \alpha_G^K)$$

and employ results from Phase 1:

$$\alpha_G^c = \begin{cases} \sigma \frac{|\mathcal{P}| \cdot FPR}{|C^0|} & \text{if } c \in C^0 \wedge G = \mathcal{P} \\ \sigma \frac{|\mathcal{P}| \cdot (1-FPR)}{|C^0|} & \text{if } c \in C^1 \wedge G = \mathcal{P} \\ \sigma \frac{|\mathcal{N}| \cdot (1-FNR)}{|C^1|} & \text{if } c \in C^0 \wedge G = \mathcal{N} \\ \sigma \frac{|\mathcal{N}| \cdot FNR}{|C^1|} & \text{if } c \in C^1 \wedge G = \mathcal{N} \end{cases}$$

– $FPR/FNR$=false positives/negatives rate,
– $C^0/C^1$=clusters assigned $i_u = 0 / i_u = 1$,
– $\sigma$=a parameter weighting priors strength.

## StackOverflow.com Results




– The classification results from different methods agree to a high degree.

– Results suggest that only about half of the users intentionally performed actions needed to receive the badge.

– Features derived from location best discriminate between classes, for example, users from USA receive badges more often by chance.