Crowdsourcing, Attention and Productivity

Bernardo Huberman[†], Daniel M. Romero[*], and Fang Wu[†]

† Social Computing Lab, HP Laboratories, Palo Alto, CA 94304
∗ Cornell University, Ithaca, NY, 14850

June 7, 2009

Introduction

- Content generation and consumption dynamics have changed.
- Free Riding vs. Contributing.
- Small ratio of contributors to consumers yet the amount of content increases dramatically.
- Conjecture: Contributing is considered a private good rewarded in the form of attention.


YouTube
Broadcast Yourself ™
Worldwide | English
Sign Up | QuickList (0) | Help | Sign In
Home
Videos
Shows
Channels
Community
social networks
Search
Upload
Privacy and Social Networks
0:05 / 2:40
HQ
46 ratings
28,668 views
Favorite
Share
Playlists
Flag
MySpace
Facebook
Twitter
(more share options)
PrivacyComm
November 07, 2007
(less info)
Subscribe
What does a friend of a friend of a friend know about you?
Video generated by the Office of the Privacy Commissioner of Canada, Jennifer Stoddart.
For more information, please contact Colin McKay, Director of Communications, at cmckay@privcom.gc.ca
Category: News & Politics
Tags: privacy commissioner privacy identity social network facebook bebo orkut myspace Jennifer Stoddart
URL
http://www.youtube.com/watch?v=X7gWEgH
Embed
<object width="425" height="344"><param na
More From: PrivacyComm
Related Videos
Social Networking in Plain English
428,283 views
leelefever
1:47
Featured Video
Youthprivacy.ca Contest

The Data

- 1 million users and 10 million videos uploads including: user ID, time and date of upload, number of views, ranking, number of comments.

- For each users we partition time into weeks and define:
  $n(t) =$ number of uploads during active period t
  $v(t) =$ average number of views during active period t
  period = 2 weeks

Detrending $v(t)$

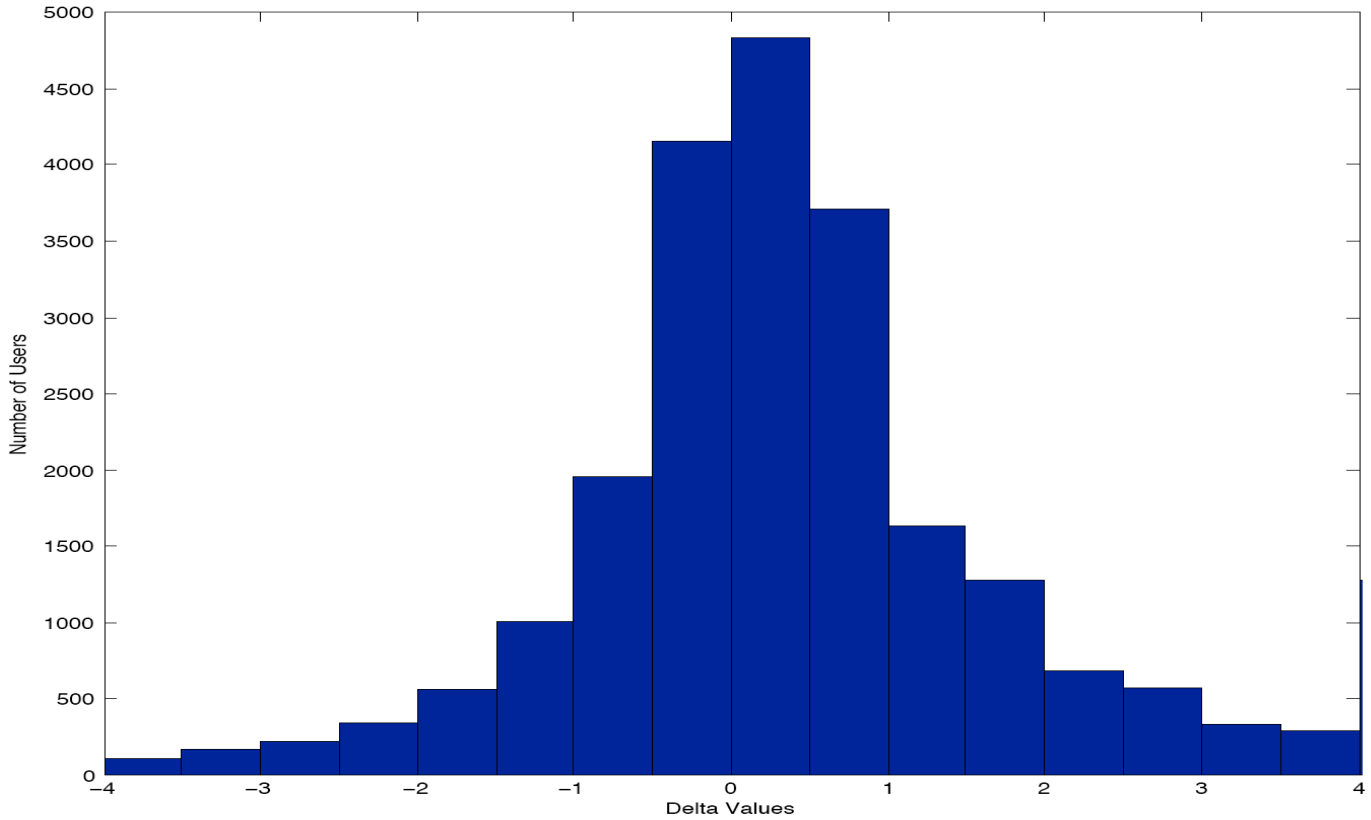
- Older videos have been on the website for a longer time, they naturally will have more views
- We need to detrend the final view count data
- Linear regression of $v(t) \sim at + b \Longrightarrow a = -28.80$ and $b = 404,650$
- Define $v_d(t) = v(t) - (at + b)$
- Results hold when using $v_d(t)$ or $v(t)$

Linear Fit

- $\{n_{t+1}\}_{t=1}^{T} \sim \alpha\{\log_{10} v_t\}_{t=1}^{T} + \beta$ for $T > 10$ periods.
- 76,462 α values
- Null Hypothesis: The α values come from a normal distribution with non-positive mean.
- t-test $\Longrightarrow$ p-value $<$ .001
- Null hypothesis can be rejected

A More Direct Approach

- $v^m = 10,000$ (Average view count of all videos in the data set)
- $G = \{s : v_s > v^m\}$, $B = \{s : v_s < v^m\}$
- $n^G =$ average $\{n_{s+1}\}_{s \in G}$
  $n^B =$ average $\{n_{s+1}\}_{s \in B}$
- $\Delta = n^G - n^B$



2-9 active periods. p-value $< .001$

Constant Threshold

Number of Active Periods	Number of Users	Δ-Mean	p-value
2-9	20061	.65	$< .001$
10-19	24517	.53	$< .001$
20-29	7789	.38	$< .001$
30-39	2153	.20	.18
40-70	515	.09	.50

Table 1: $v_m = 10000$.

What if v_m is not constant?

- $v^m = \text{median}\{v_t\}_{t=1}^{T-1}$, where T is the user's number of active periods.
- $G = \{s : v_s > v^m\}$, $B = \{s : v_s < v^m\}$
- $n^G = \text{average } \{n_{s+1}\}_{s \in G}$

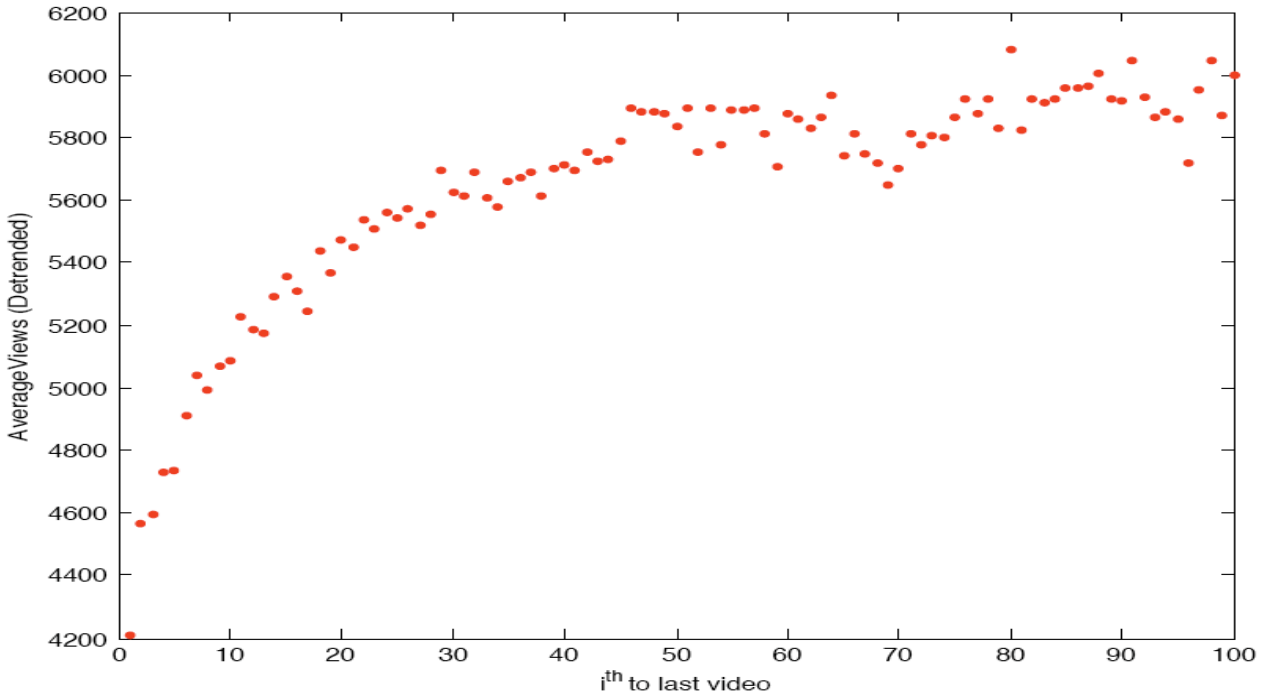
  $n^B = \text{average } \{n_{s+1}\}_{s \in B}$
- $\Delta = n^G - n^B$

Varying Threshold

Number of Active Periods	Number of Users	Δ-Mean	p-value
2-9	85949	.05	$.15$
10-19	68317	.20	$< .001$
20-29	14757	.23	$< .001$
30-39	3303	.30	$< .001$
40-70	673	.43	$< .01$

Table 2: v_m = median of $\{v(t)\}_{t=1}^{i}$.

Lack of Attention Yields Quitting



Granger Causality

- The Granger causality test determines causality in terms of prediction accuracy.
- X_1 G-causes X_2 if past values of X_1 contain information that helps predict future values of X_2.
- It is important to note that Granger causality is only meaningful if only found in one direction.

Granger Causality

- Define $v_a(i)$ and $n_a(i)$ as the average of all user's views and uploads, respectively, during their i^{th} active week
- The Granger causality test of the hypothesis that $v_a(i)$ G-causes $n_a(i)$ resulted in a p-value of 0.01, and of the hypothesis that $n_a(i)$ G-causes $v_a(i)$ which gave a p-value of .61.
- The Granger causality test supports the hypothesis that views "cause" more uploads.

Summary

- Productivity in crowdsourcing strongly depends on attention.
- Lack of attention yields to a decrease in the amount of contributions.
- Long term users tend to compare their current performance with their previous performance, while short term users tend to compare their performance with other user's performance.