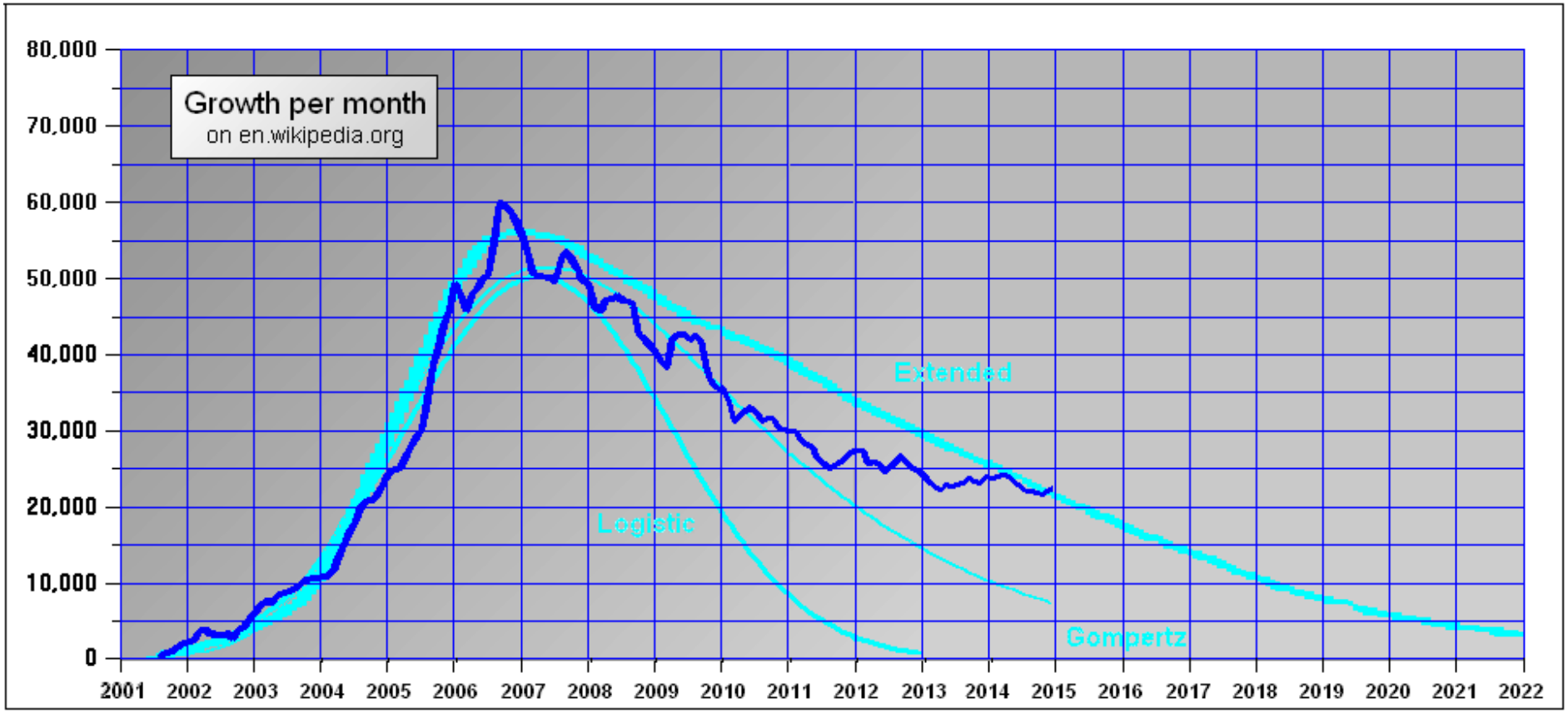
Impacts of Project Membership on Contribution to Wikipedia

Yan Chen, Iman Yeckehzaare,
Fangzhou Zhang

Introduction

- Flourishing rise of website supported by user-generated content in the past decade
- Wikipedia
  - over 5m articles
  - 28m registered editors
  - ranks among the most popular websites with Google and Facebook

Introduction



Introduction

- Sluggish growth in recent years
  - low rate of retention
  - power law distribution

- WP as public goods
  - Non-excludable and non-rivalrous
  - Underprovision

Introduction

- WikiProject: A collaboration platform
  - Platform where editors coordinate their efforts and connect to one another
  - A WikiProject typically focuses on a particular topics (WikiProject Military History, WikiProject Economics).
  - Most WikiProjects provide on its homepage a list of goals (articles of the month, articles to be promoted)
  - Free entry (joining by adding one's username to the list of members)

Introduction

- Research question
  - What is the impact of WikiProject membership on individual contribution?
  - What are the mechanism through which WikiProject membership encourages individual contribution?
- Main findings
  - By matching members to the non-members who have similar contribution behavior, we measure the impact of project membership on individual contribution.
  - Compared to those who do not join WikiProjects, project membership doubles the size of individual contribution.
  - The impact of joining WikiProject persists for at least six months.

Previous Literature

- The effect of WikiProject on retention rate
  - peripheral users' activity in the first year (Solomon and Wash 2014).
  - tenure diversity (Chen, Ren and Riedl, 2010)
- The effect of WikiProject on coordination
  - complex coordination structure (Kittur and Kraut, 2008; Ung and Dalle, 2010; Morgan et al., 2014)
  - group size and task performance (Kittur et al., 2007; Romero et al., 2010; Zhang and Zhu, 2011)

Econometric Methodology

- What we want to measure?
  - average treatment effect on the treated
  - treatment: joining at least one WikiProject
  - $\mathbb{E}\{Y_{i1t} - Y_{i0t} | D_{it} = 1\}$

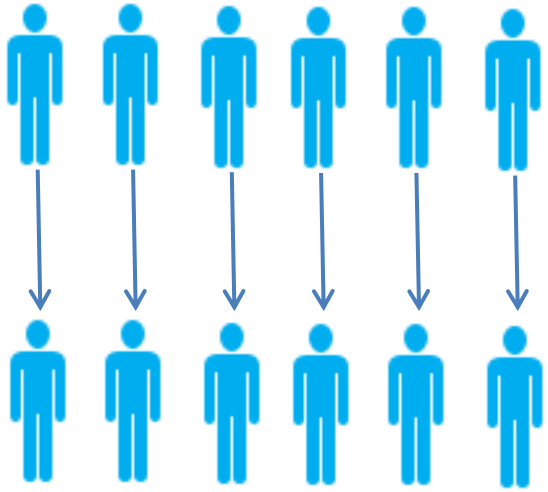
$$= \underbrace{\mathbb{E}\{Y_{i1t} | D_{it} = 1\}}_{\text{observed in the data}} - \underbrace{\mathbb{E}\{Y_{i0t} | D_{it} = 1\}}_{\text{not observed in the data}}$$

  - challenge: construct measure for $\mathbb{E}\{Y_{i0t} | D_{it} = 1\}$

Econometric Methodology

- How do we measure $\mathbb{E}\{Y_{i0t}|D_{it}=1\}$?
  - use $\mathbb{E}\{Y_{i0t}|D_{it}=0\}$ observed in the data
  - matching: match treated editors to untreated editors who are similar to them

Joining a WikiProject



$\mathbb{E}\{Y_{1it}|D_{it}=1\}$ observed in the data

Not Joining any WikiProject

$\mathbb{E}\{Y_{0it}|D_{it}=0\}$ replaced by $\mathbb{E}\{Y_{0it}|D_{it}=1\}$

Econometric Methodology

- Self-selection into WikiProject
  - $D_{it} = 1$ editors are different from $D_{it} = 0$ editors
  - Those who are more dedicated to WP are more likely to join WikiProject
- The validity of matching depends on
  - how editors self-select into WikiProject (identification assumption)
  - how we perform matching

Econometric Methodology

- Identification assumption
  - joining in WikiProject depends on the editors' tenure and activeness at Wikipedia.
  - long tenure and higher activeness increase exposure to WikiProject and likelihood of joining.
- Matching procedure
  - matching based on tenure and activeness in the past six months
  - matching performed month-by-month

Econometric Methodology

- Matching procedure
  - Matching performed for each month separately
  - Calculate propensity score (probit)
    $$\mathrm{P}(D_{it} = 1) = F(Tenure, activity\ in\ past\ 6\ months)$$
  - Observations include 1) treated editors who join his first project at t and 2) control editors who haven't joined any project as of t.
  - Each treated editor is matched to the control editor with the nearest estimated propensity score.

Data

- The sample includes the most active 9,134 register WP editors.
- Complete editing history including a snapshot of each edit – timestamp, title, namespace, size, etc.
- Editing behavior aggregated on a monthly basis
- How we identify when one joins his first project?
  - whether one edits the list of members for a project
  - This is usually indicated in the comment attached to the edit.

Data

	All Editors	Treated Editors	Control Editors
Tenure as of Apr. 2015	84.5 (34.8)	86.5 (34.1)	76.9 (36.4)
Tenure when joining		16.9 (18.8)	
Monthly edit - size (1000)	139.6 (391.8)	147.5 (393.8)	110.6 (383.2)
Before joining		98.8 (543.6)	
After joining		169.1 (366.3)	
Monthly edit – num	253 (340)	248 (281)	268 (500)
Before joining		199 (313)	
After joining		286 (363)	
Monthly add - size (1000)	102.8 (290.9)	107.6 (274.5)	85.2 (343.5)
Before joining		73.9 (394.4)	
After joining		123.7 (263.9)	

Results and Discussion

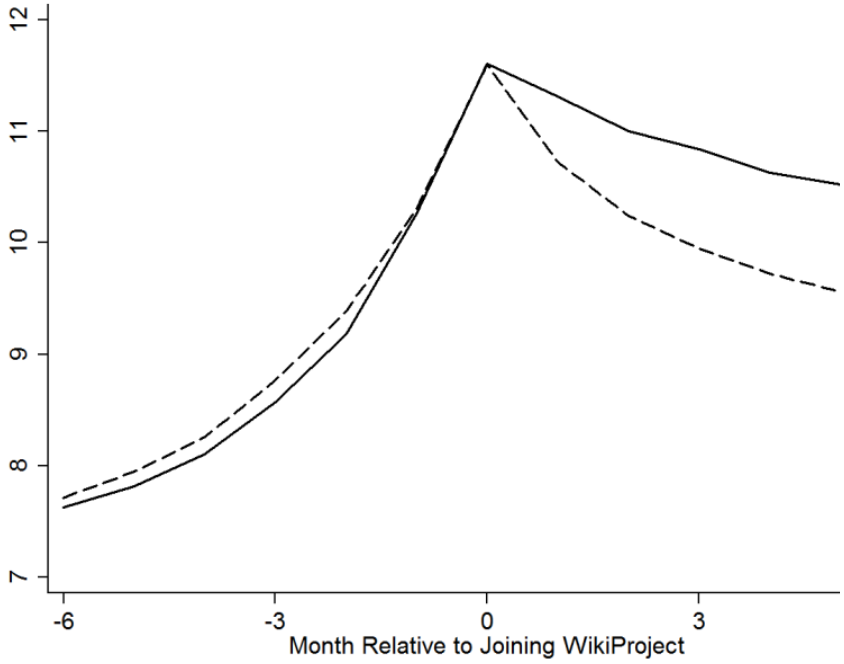
- Treatment effect on the first month after joining
- Estimates are based on log-arithmetic
- Joining project almost double the size of edits.

	Size of Edits		Number of Edits		Size of Additions	
Month relative to Joining	Treated	Matched Untreated	Treated	Matched Untreated	Treated	Matched Untreated
-6	7.611	7.824	3.417	3.620	7.337	7.532
-5	7.793	8.028	3.519	3.733	7.508	7.727
-4	8.060	8.330	3.683	3.898	7.771	8.033
-3	8.516	8.855	3.933	4.180	8.228	8.539
-2	9.144	9.441	4.282	4.496	8.848	9.112
-1	10.232	10.348	4.920	5.005	9.943	10.027
0	11.599	11.600	5.823	5.659	11.319	11.220
1	11.308	10.727	5.594	5.229	11.012	10.385

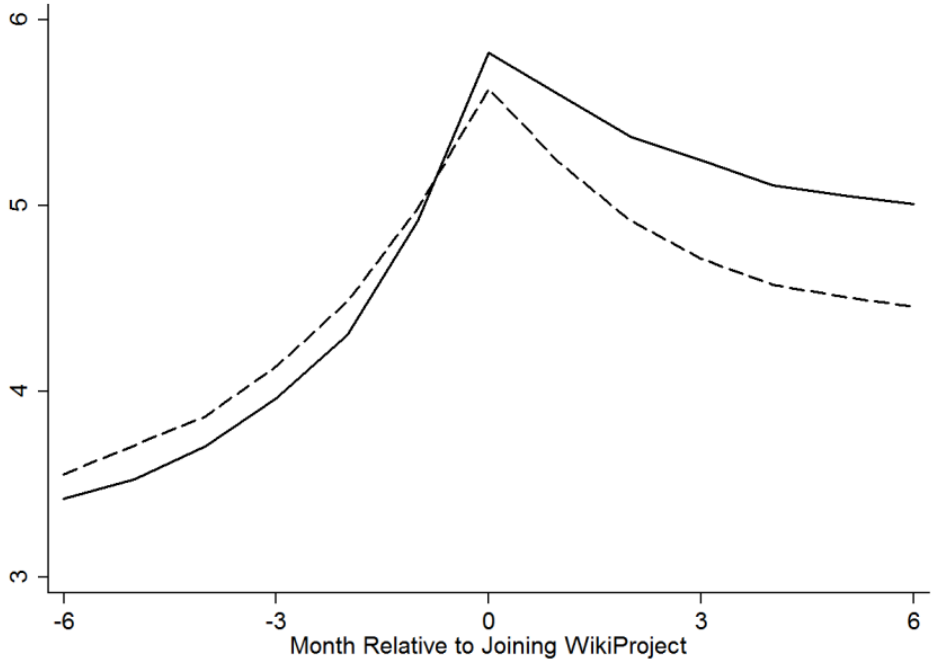
Results and Discussion

- Is the impact sustainable?

Cumulative Size of Monthly Edits



Number of Monthly Edits



Results and Discussion

- What's the mechanism that encourage contribution?
- WikiProjects usually provide a list of goals to be finished (articles to improve)
  - reduce searching cost
  - more concentrated contribution
- Use concentration of contribution rather than size of contribution as the outcome variable

Results and Discussion

- How to measure concentration of contribution
  - information entropy
  - For each entry, categorize the article under a project
  - calculate the proportion of entries categorized under each project
  - calculate the information entroy
- The higher the entropy is, the less concentrated one's contribution is.
- We expect that entropy is lower after joining WikiProject.