All who wander:
On the Prevalence and Characteristics of Multi-community Engagement
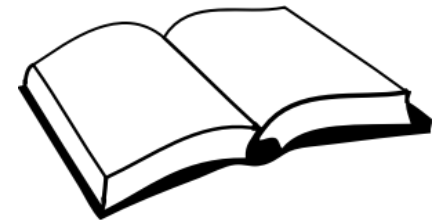
Chenhao Tan

Cornell University

Joint work with Lillian Lee

We have many chances to engage with many communities

A variety of organizations, and social circles exist on a university campus

We have many chances to engage with many communities

A variety of organizations, and social circles exist on a university campus

Computer science conferences (DBLP)

We have many chances to engage with many communities

A variety of organizations, and social circles exist on a university campus

Computer science conferences (DBLP)

Online communities, e.g., reddit.com

r/Politics

Babybumps

Existing work

Single community setting: e.g., predicting user survival (churn prediction) [Danescu-Niculescu-Mizil et al. 2013, Dasgupta et al. 2008, Dror et al. 2012, Rowe 2013]
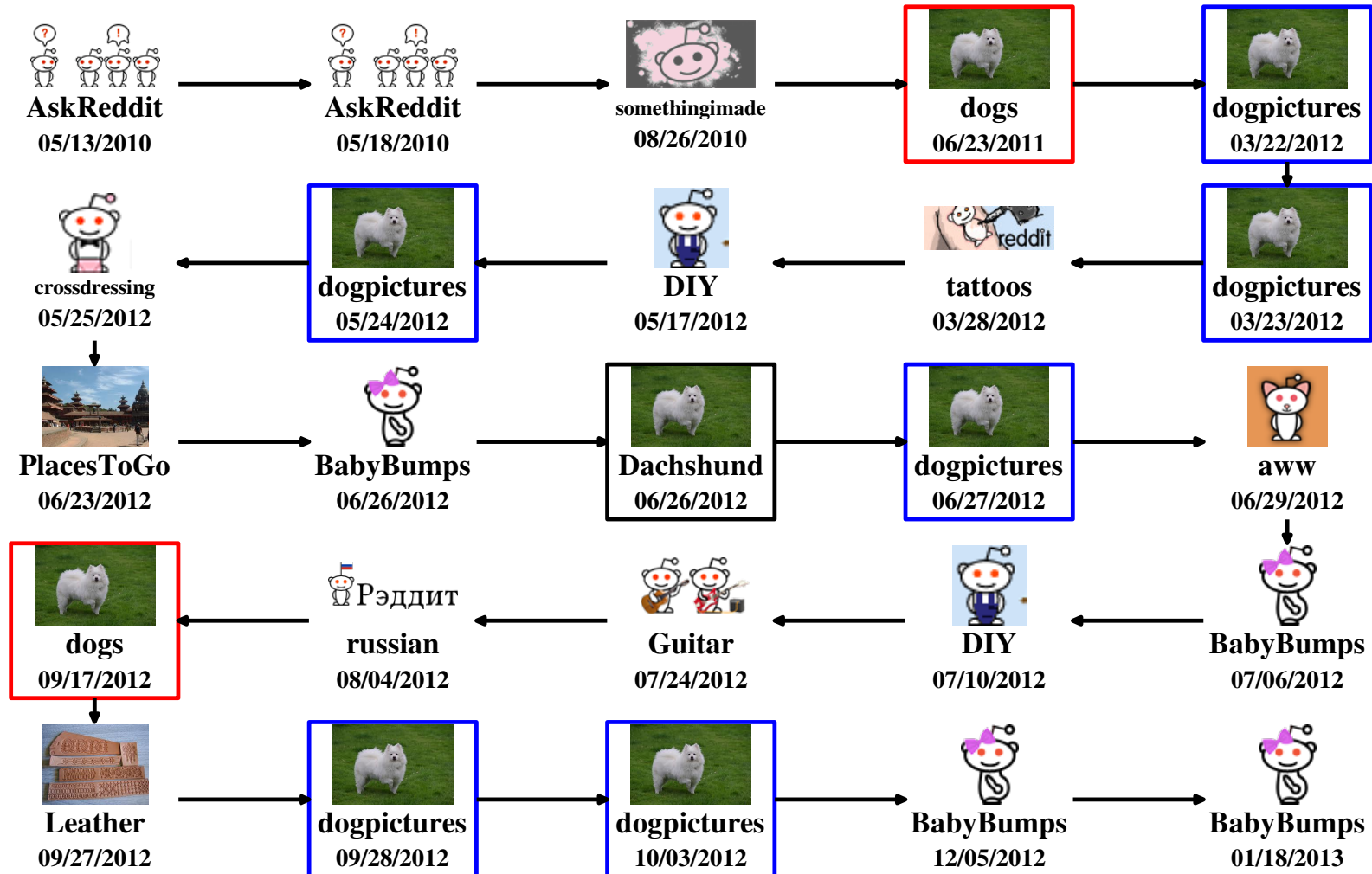
Single-community success prediction [Iriberri and Leroy 2009, Kairam, Wang and Leskovec 2012, Ludford et al. 2004, Zhu et al 2014]

Babybumps

Objects of interest:
user trajectories across communities
An example from a user on Reddit

AskReddit 05/13/2010 → **AskReddit** 05/18/2010 → **somethingimade** 08/26/2010 → **dogs** 06/23/2011 → **dogpictures** 03/22/2012

crossdressing 05/25/2012 ← **dogpictures** 05/24/2012 ← **DIY** 05/17/2012 ← **tattoos** 03/28/2012 ← **dogpictures** 03/23/2012

PlacesToGo 06/23/2012 → **BabyBumps** 06/26/2012 → **Dachshund** 06/26/2012 → **dogpictures** 06/27/2012 → **aww** 06/29/2012

dogs 09/17/2012 ← **russian** 08/04/2012 ← **Guitar** 07/24/2012 ← **DIY** 07/10/2012 ← **BabyBumps** 07/06/2012

Leather 09/27/2012 → **dogpictures** 09/28/2012 → **dogpictures** 10/03/2012 → **BabyBumps** 12/05/2012 → **BabyBumps** 01/18/2013

How much do users explore new communities?
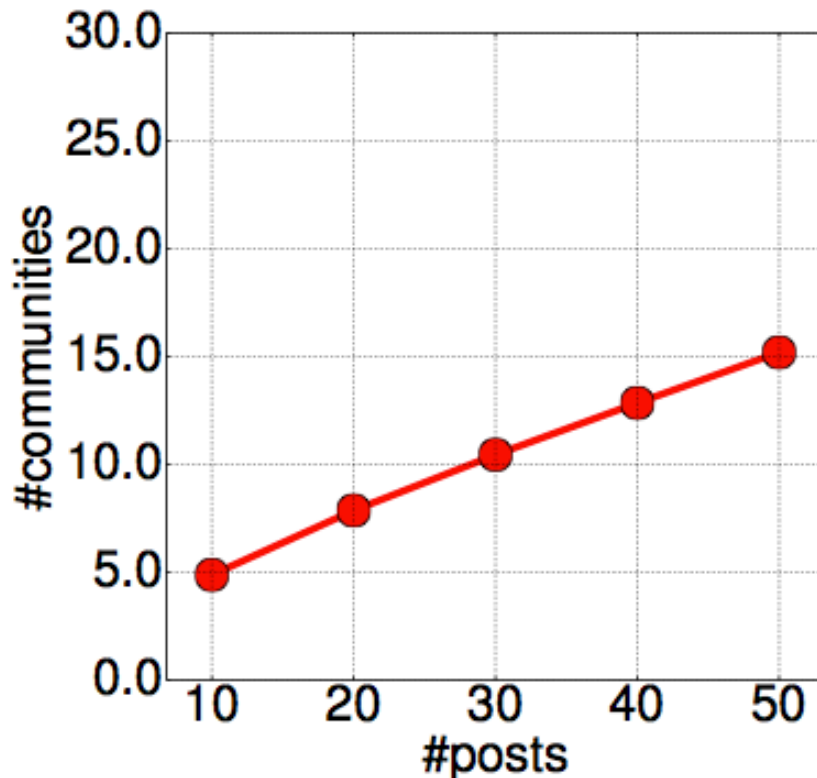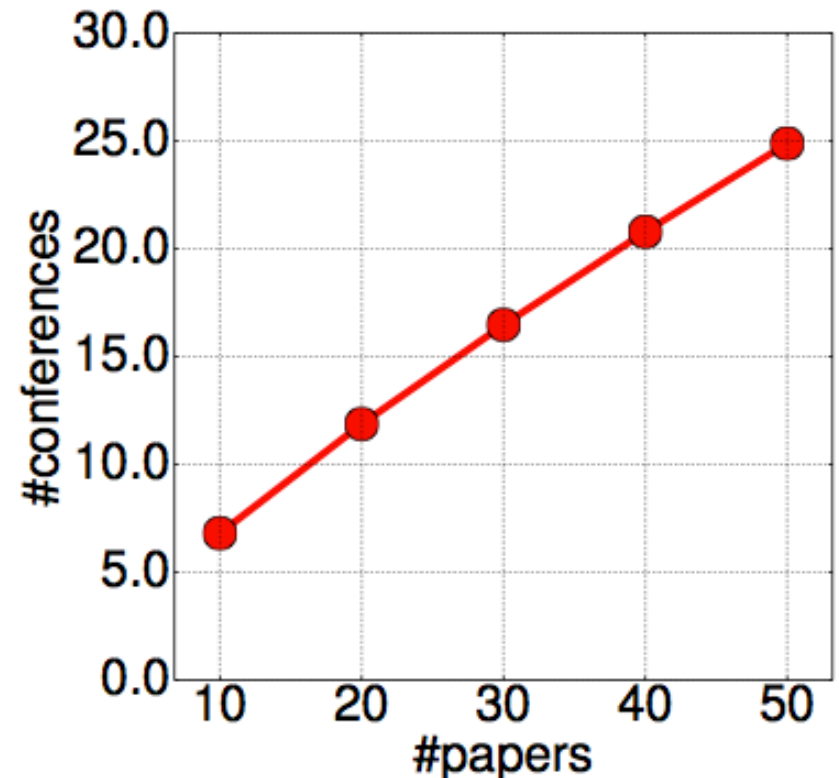
#communities

Age

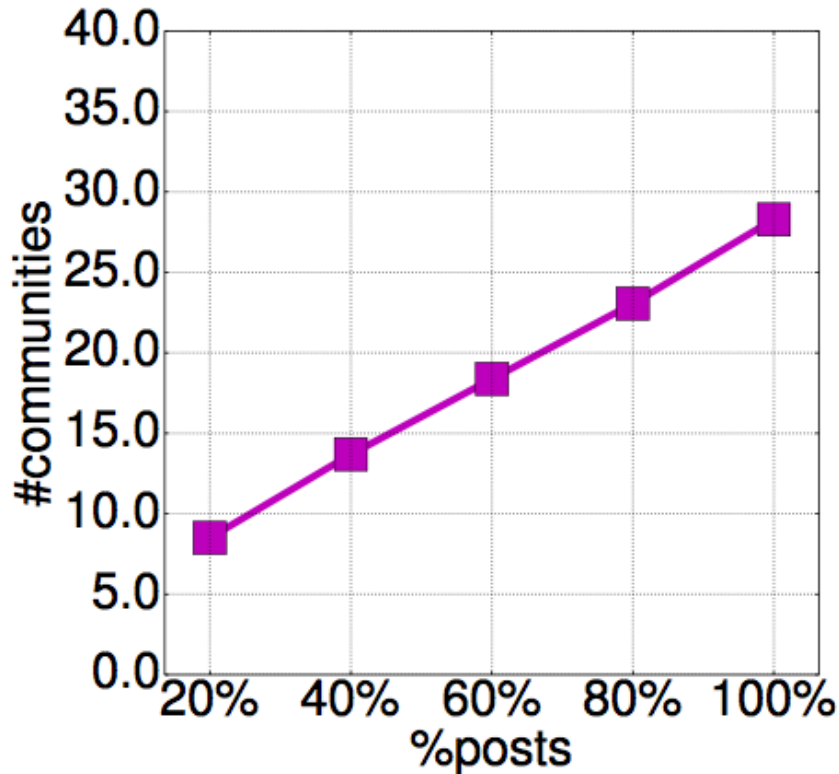# First 50 posts on Reddit and DBLP

Reddit

DBLP

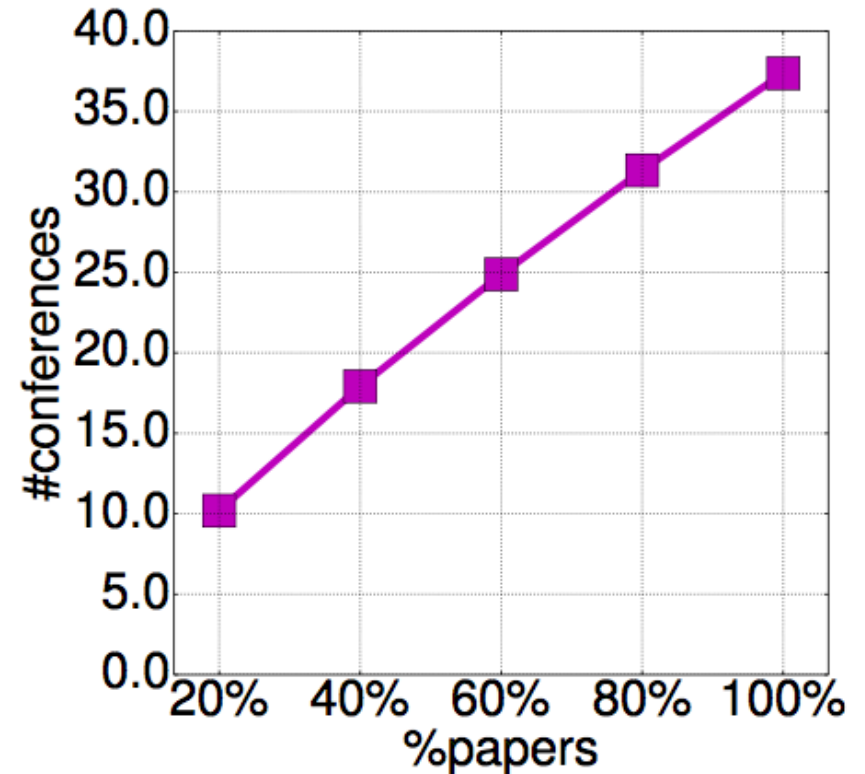Error bars (tiny) show standard error.

The average time to accumulate 50 contributions is 456.0 days on Reddit, 15.6 years on DBLP.

Lifetime on Reddit and DBLP

Reddit

DBLP

Error bars (tiny) show standard error.

Main dataset: reddit

- Many sub-communities (subreddits)
- An active platform where users submit posts, make comments and rate posts

- All 76.6M posts ever submitted to Reddit from its inception until Jan 2014
- 157K "50+" posters who first posted between Jan 2008 and Jan 2012 [Danescu-Niculescu-Mizil et al. 2013]

Link: *https://chenhaot.com/pages/multi-community.html*

Three aspects of user trajectories

- (How) does the wandering pattern change over time?

- Do people adapt their language in each community over time?

- Do people receive better evaluations over time?

OR?

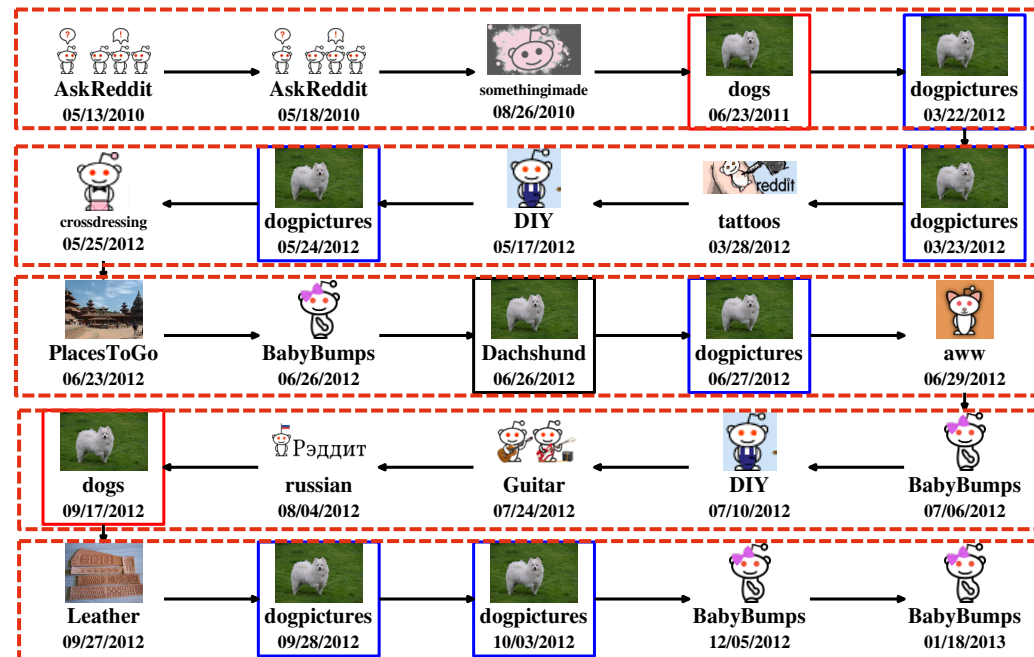evaluations= ups - downs

2876
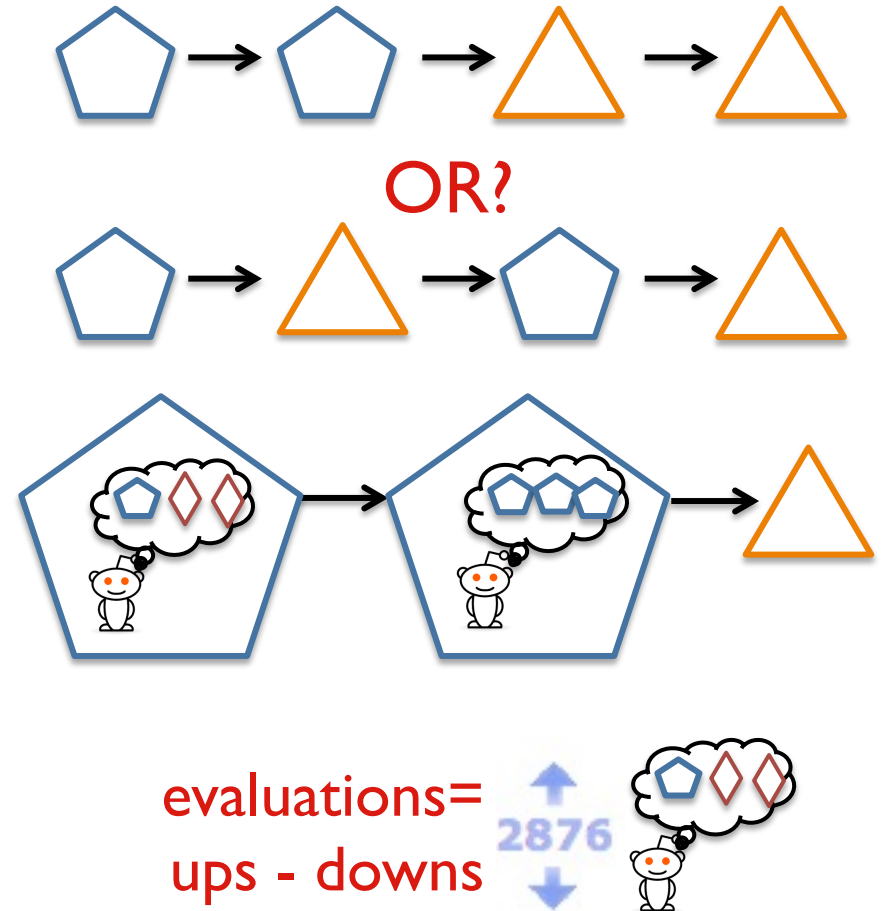
Can these be used to predict future user activity?

A framework for understanding properties of the trajectory

Split the trajectory into windows of the same size
(w=10 in main experiments)

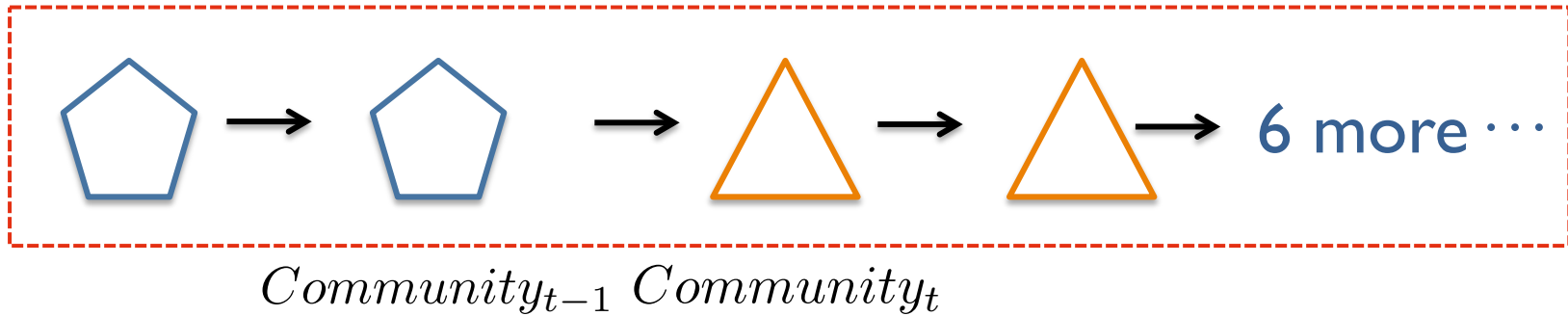Define a function (*F*) on a window to capture different properties and obtain a time series

e.g., window size = 5

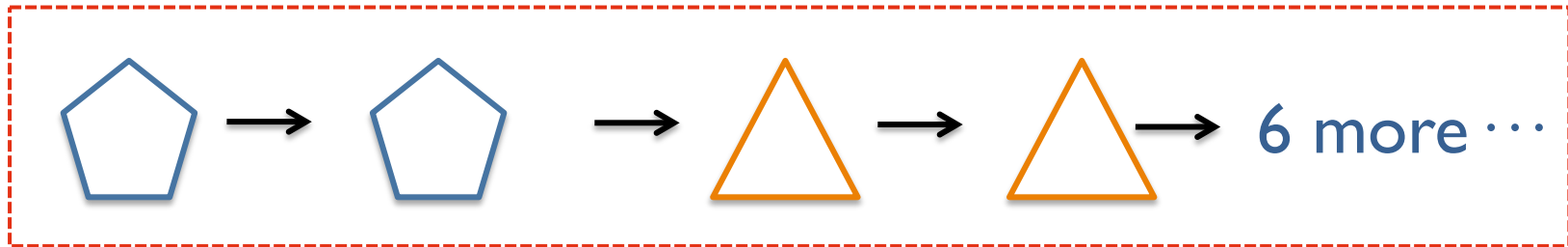$$Community_t, words_t, evaluations_t$$

Do users "jump" more over time?

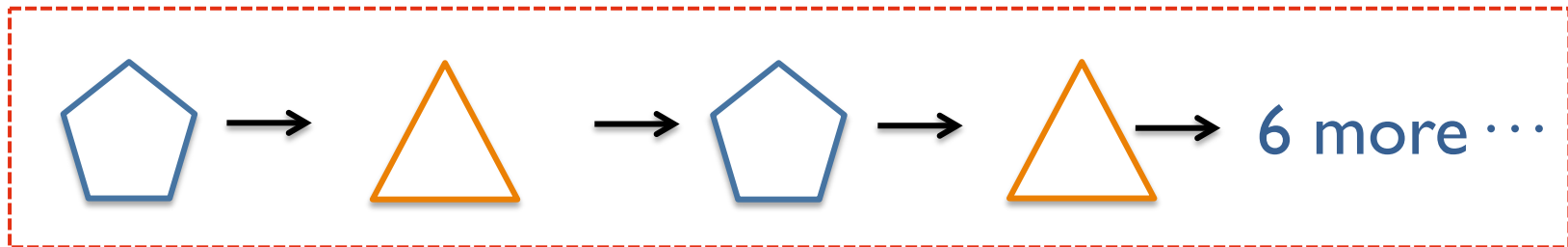F: Count of $Community_t \neq Community_{t-1}$ in a window (w=10)

$Community_{t-1}$ $Community_t$

Do users "jump" more over time?

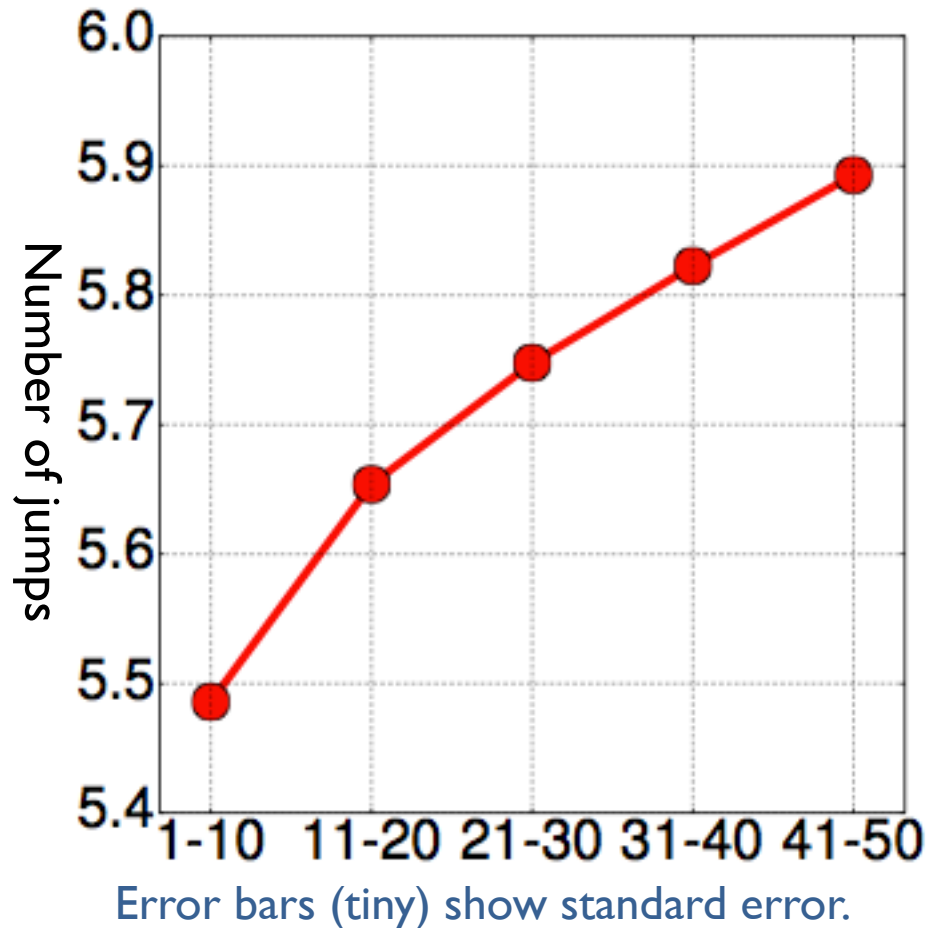*F*: Count of $Community_t \neq Community_{t-1}$ in a window (w=10)

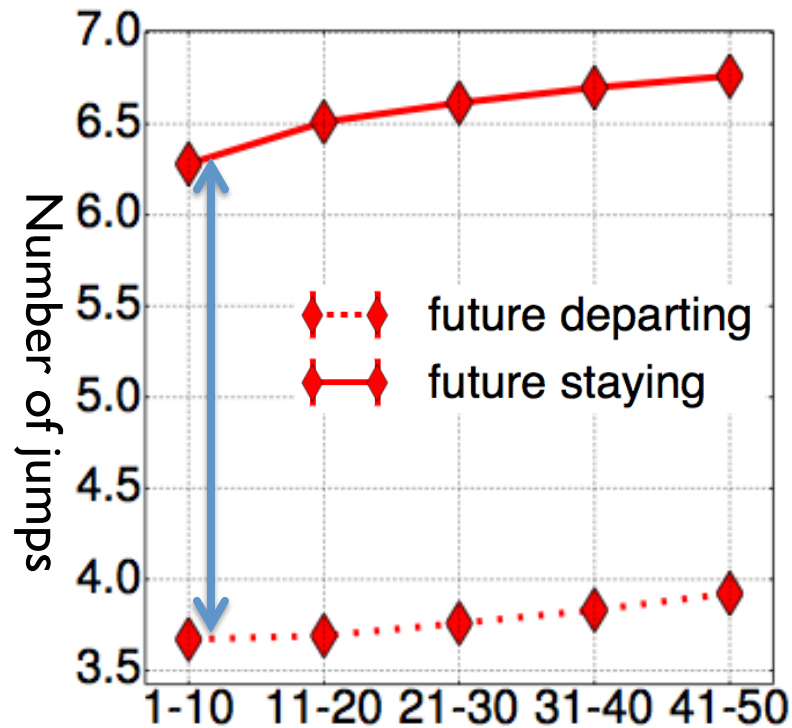$$Jumps = 1$$

$$Jumps = 3$$

Users "jump" more over time

F: Count of $Community_t \neq Community_{t-1}$ in a window (w=10)

Error bars (tiny) show standard error.

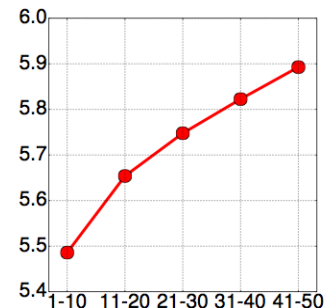# Users "jump" more over time; future departing users less so

Future departing users: stopped posting in the entire reddit in the last 6 months (44K)

Future staying users: stay *active* in the last 6 months (76K)

Error bars (tiny) show standard errors.

Users get more adventurous over time; future departing users less so

Many more different perspectives on the wandering pattern:

number of unique communities, ⬆

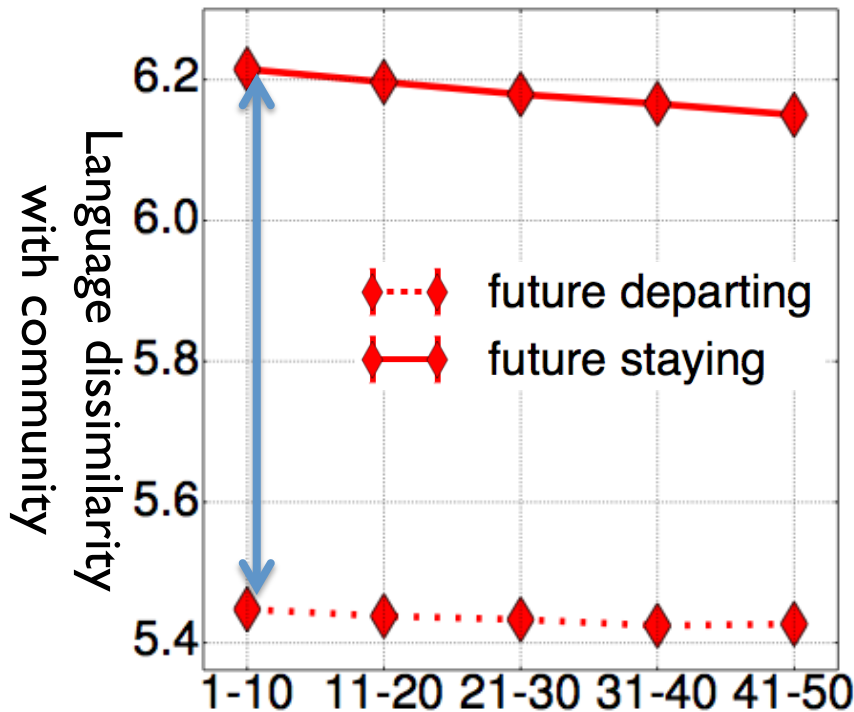level of concentration, ⬇

visible community size, ⬇

community similarity ⬇

In our data, people do not settle down at all!

Users keep adopting each community's language; future departing users more so

F: average cross entropy of $words_t$ vs language in $Community_t$

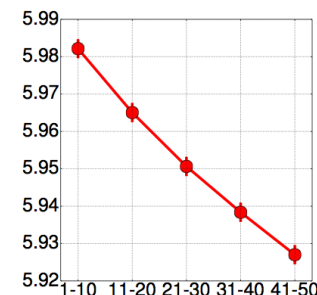A larger value indicates larger dissimilarity

Error bars (tiny) show standard error.

Users stay young:
Different from "users get old" in single community setting
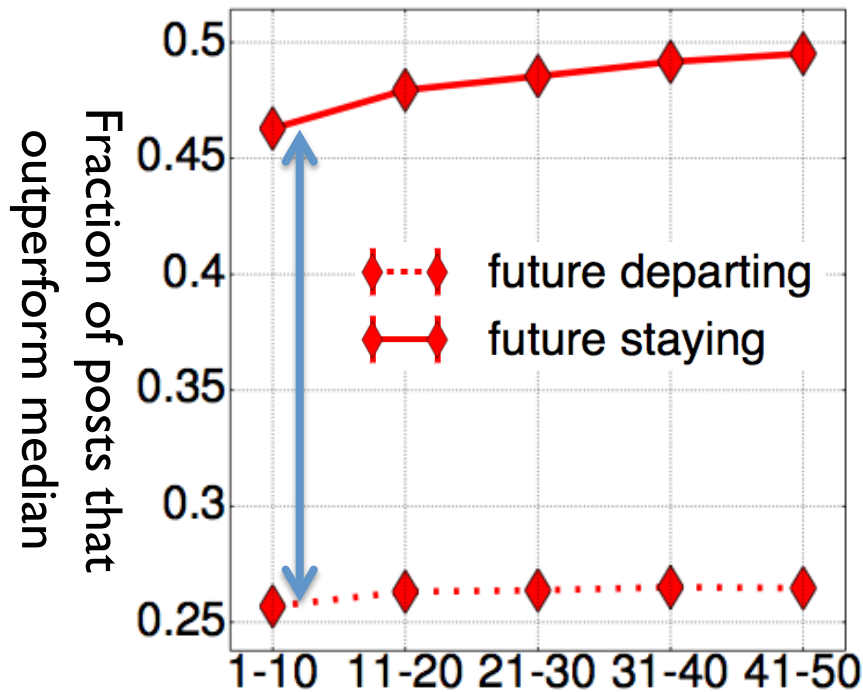[Danescu-Niculescu-Mizil et al. 2013]

Overall trend

Users get more positive evaluations over time; future departing users less so
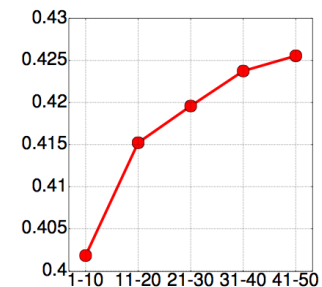
F: fraction of $evaluations_t$ that outperform the median in $Community_t$

A larger value indicates better evaluations

Error bars (tiny) show standard error.

Overall trend

Using the first 50 posts to predict future departing status

- Feature sets
  - Baseline: average time gap [Danescu-Niculescu-Mizil et al. 2013, Dror et al. 2012, Yang et al. 2010]
  - Wandering pattern
  - Language
  - Evaluations
  - Combination of the above features
- 30 randomized train-test samples, logistic regression, F1 on departing users for evaluation
- Measure performance using the first x posts

Features from trajectories outperform time-gap baseline

Users are destined to leave from the beginning!

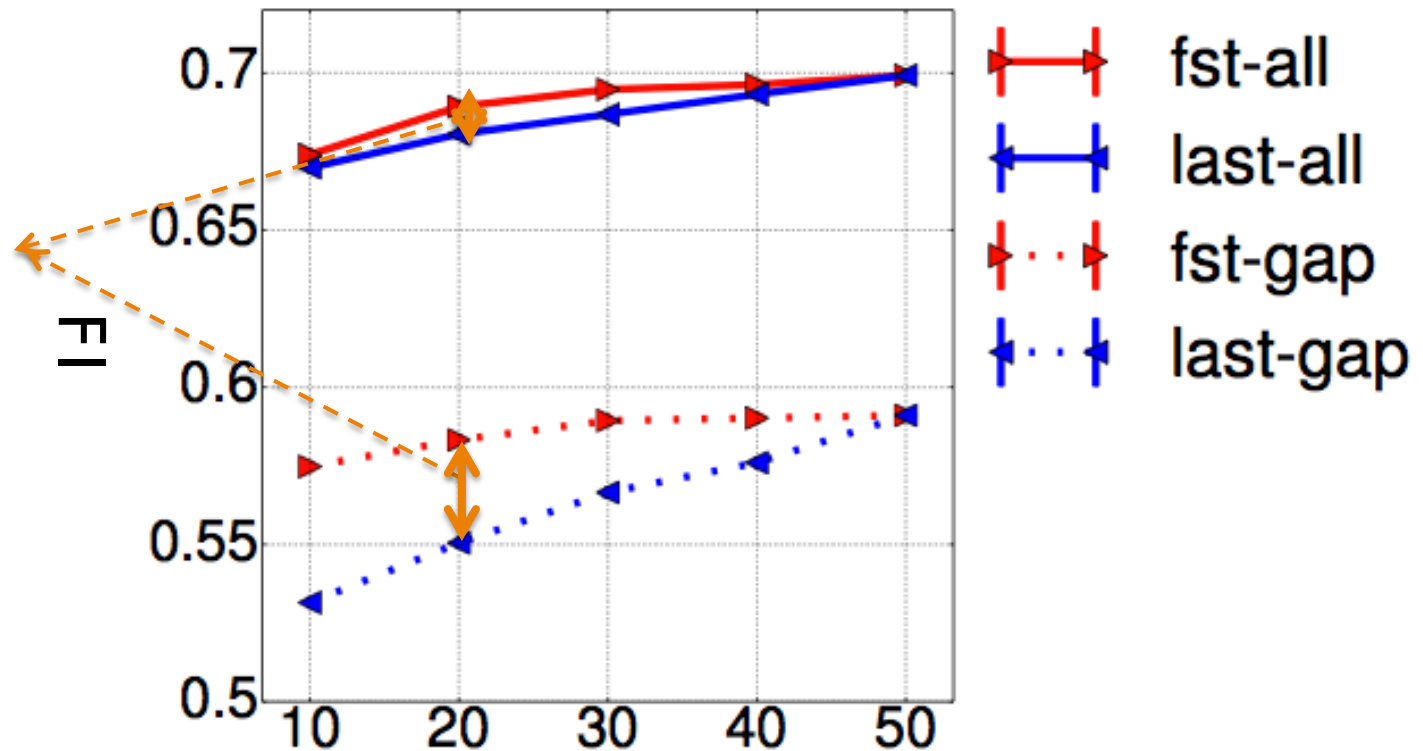Features from first 10 posts outperform baseline with all 50 posts.
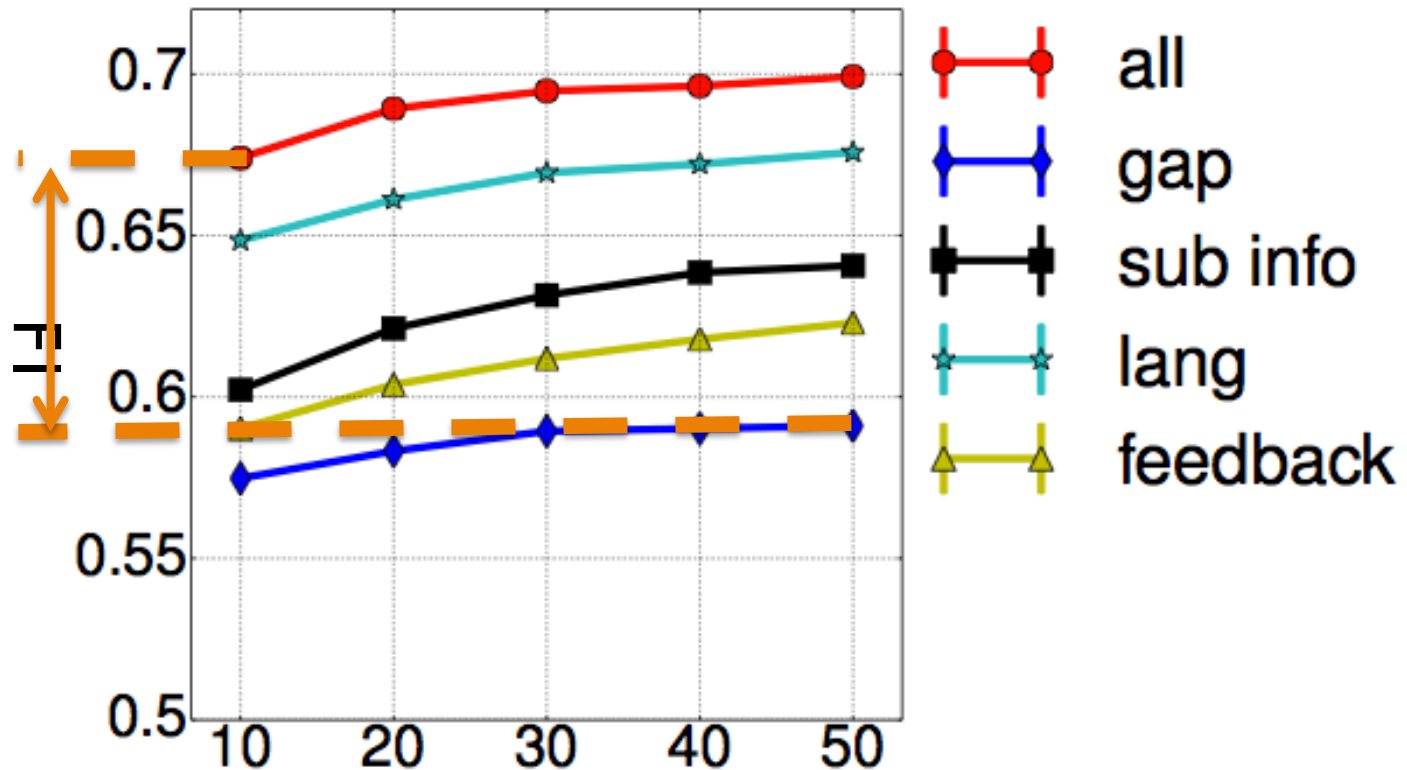
All differences along x-axis are significant (p<0.001) according to Wilcoxon signed rank test.

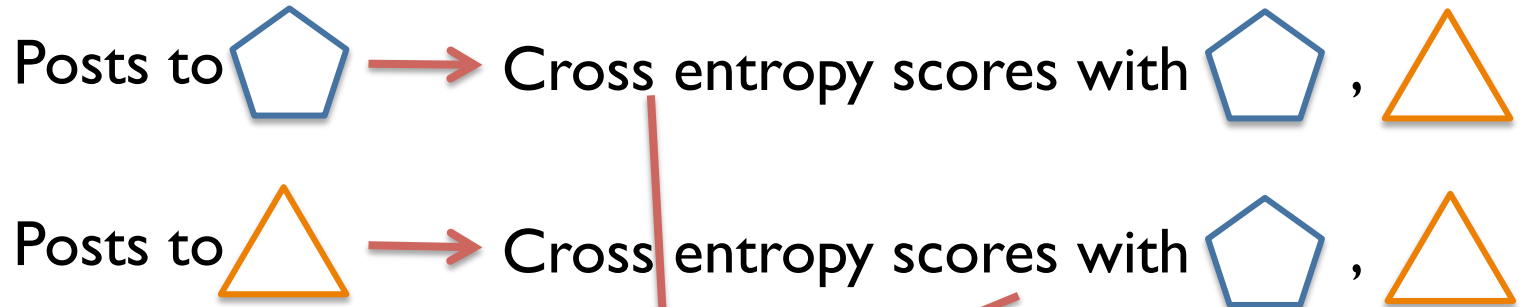# Is recent information more important or how you start more important?

First 10, 20, 30, 40, 50 vs Last 10, 20, 30, 40, 50

How you start is more important!

All differences along x-axis are significant (p<0.001) according to Wilcoxon signed rank test.

Do people speak differently in different communities?

Posts to ⬠ → Cross entropy scores with ⬠ , △

Posts to △ → Cross entropy scores with ⬠ , △

Which community were posts made to?

Focus on non-content words for language style

[Chung and Pennebaker 2007, Argamon and Levitan 2005, Danescu-Niculescu-Mizil, Gamon and Dumais 2011]

V	accuracy
parts of speech	62.5%
most frequent 100 words	56.0%
most frequent 500 words	61.4%

Summary

- Users' multi-community engagement is an interesting problem; lots of room for future work
- Design implications:
  - First impressions matter
  - Give people choices to move to
- Life lessons: People, unlike trees, thrive by relocation 人挪活，树挪死

Thank you!

Contact: *chenhao@cs.cornell.edu*, *@ChenhaoTan*
Data: *https://chenhaot.com/pages/multi-community.html*