Two Fun Topics + Quick Overview

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

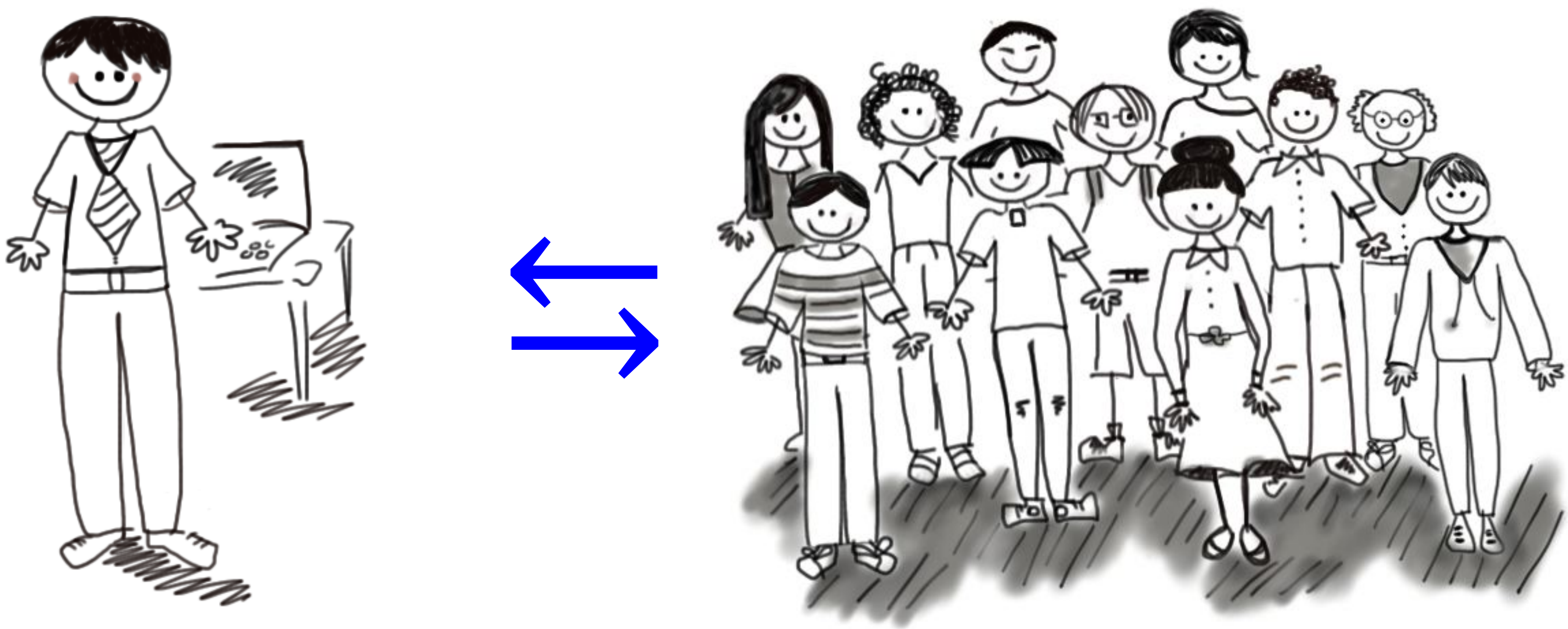
http://cs224w.stanford.edu

Users & Online Communities

Online Communities:

- Understanding the co-evolution of users and communities
- Steering user behavior

Online Communities



Modeling the relation between a user and a community

Users and Communities

Questions:

- How does a user become member of a community?
- How do user and community practices co-evolve?
- Can we predict when a user will leave the community?

Insight: Linguistic Change

- **Language practices (norms, etiquette, …)**
  - build collective identity
  - foster individual expression
- **Linguistic change captures the relation between users and communities**
  - Framework for tracking linguistic change
  - Measures of user reaction to linguistic change
  - Predict when user will leave the community

Online Communities

- **People discuss beer:**

Ba

Tballz420

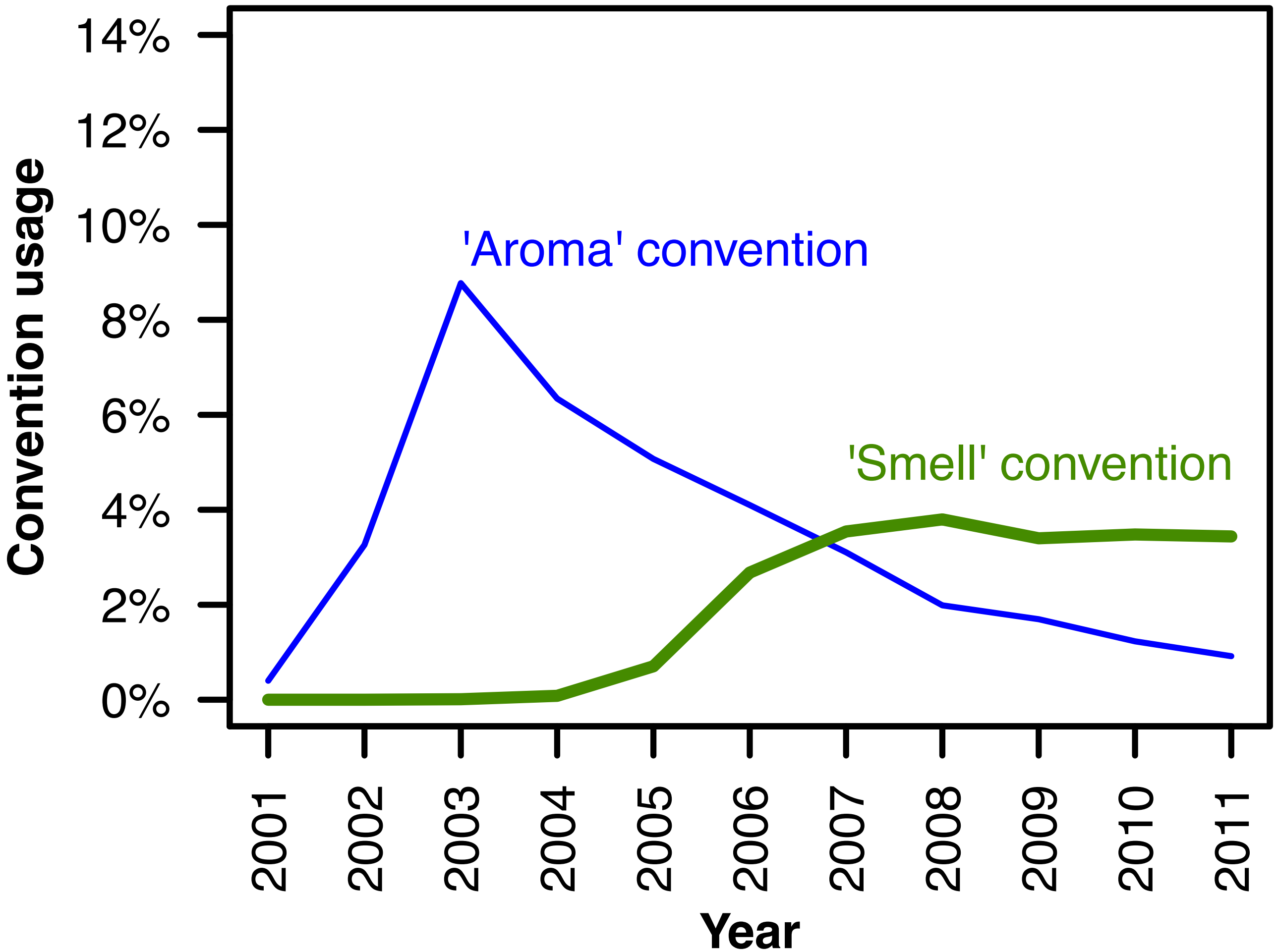
4/5 rDev +1.8%

look: 3 | smell: 4 | taste: 4 | feel: 3.5 | overall: 4.5

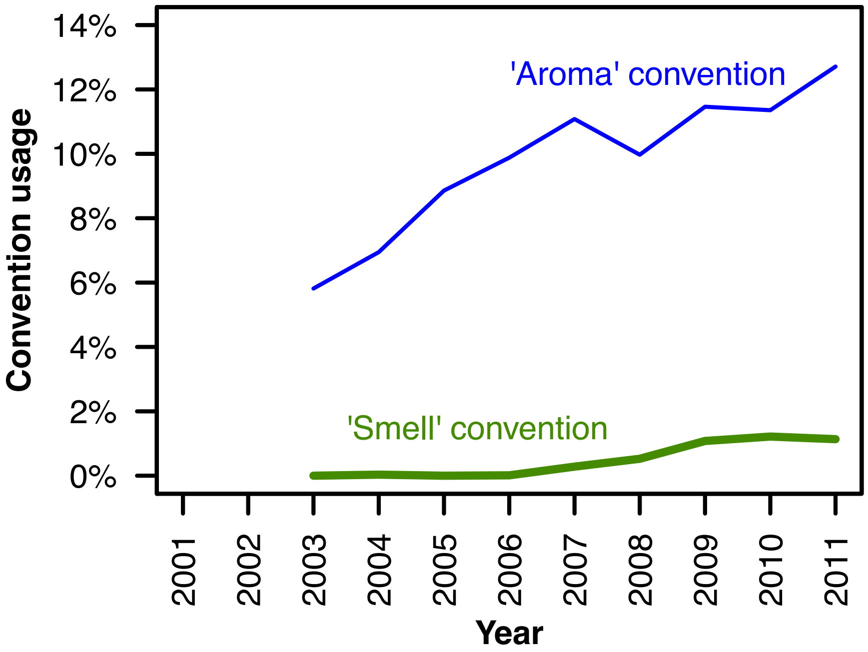
Clear copper colored brew, medium cream colored head. Floral hop nose, caramel malt. Caramel malt front dominated by a nice floral hop backround. Grapefruit tones. Very tasty hops run the show with this brew. Thin to medium mouth. Not a bad choice if you're looking for a nice hop treat.

- **10 years of complete linguistic data**
  - **RateBeer:** 3M posts, 30k users
  - **BeerAdvocate:** 1.6M posts, 33k users

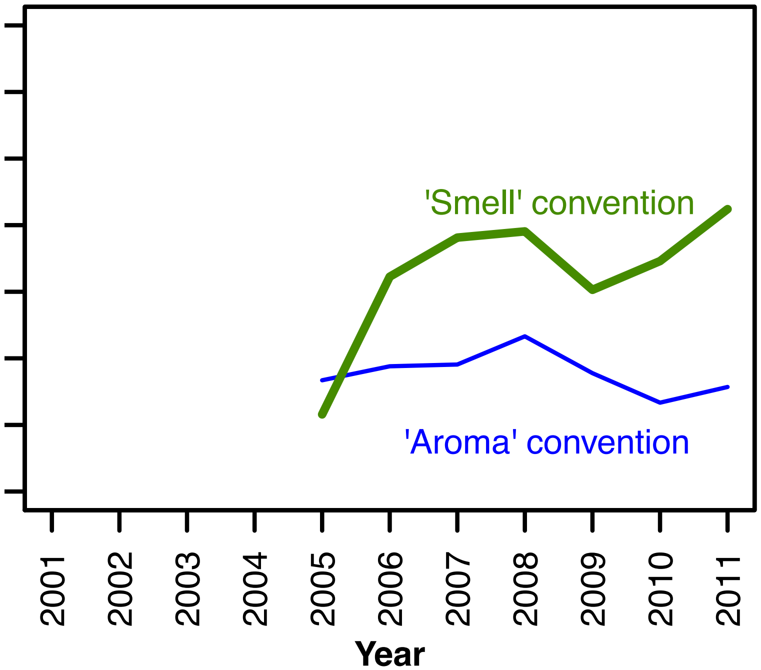
"Aroma" vs. "Smell"



Young Adopt Innovations



Users who joined in 2003



Users who joined in 2005

New users are more likely to use "smell" than users who have been part of the community for a long time.

Community / User Change

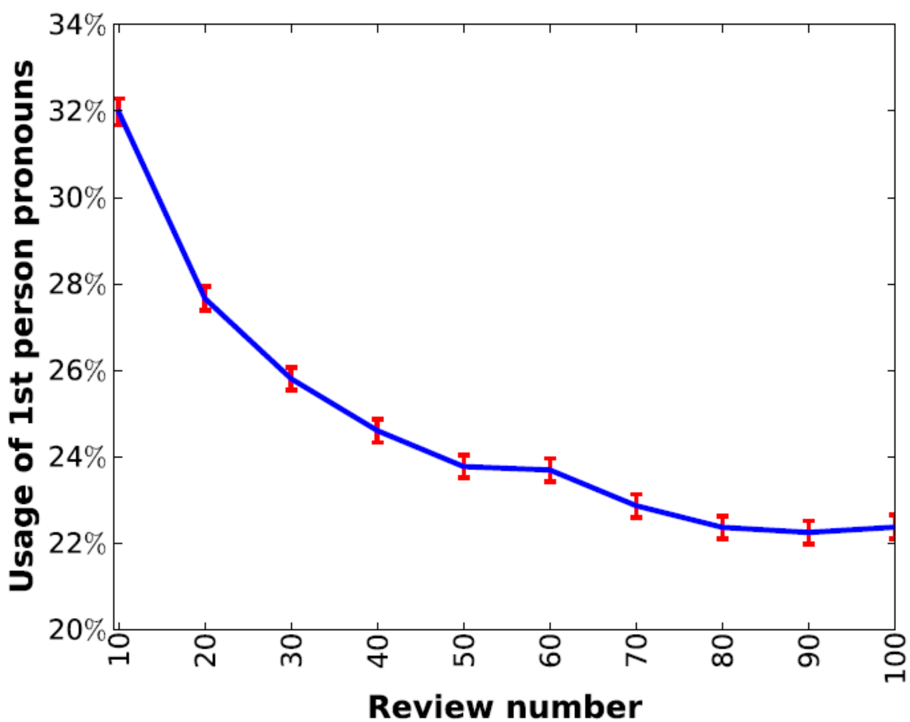
User:

"life stage"

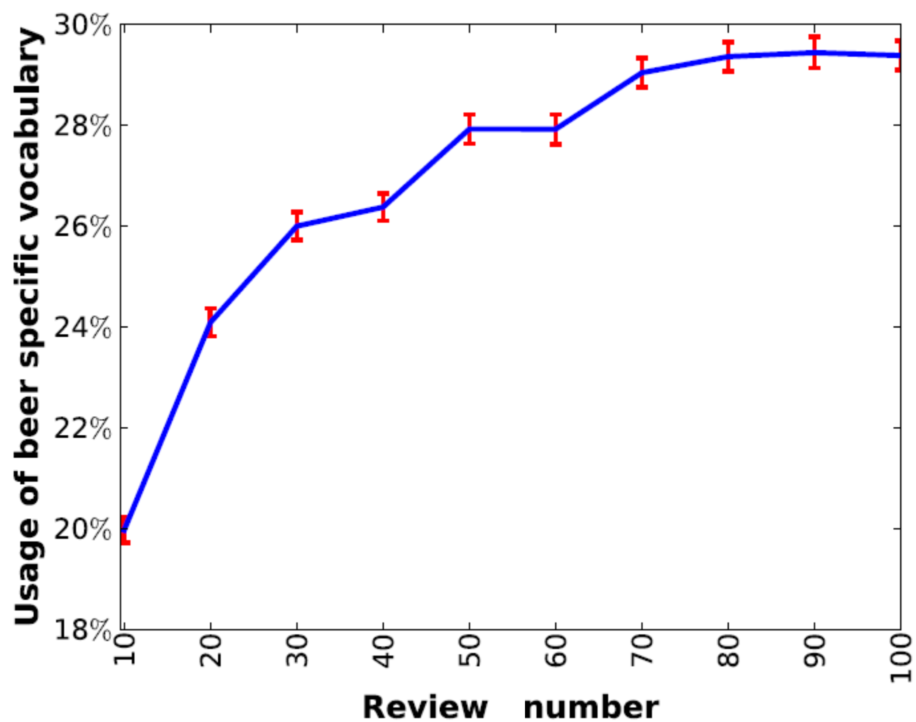
Community:

2001 2011

User Language Change



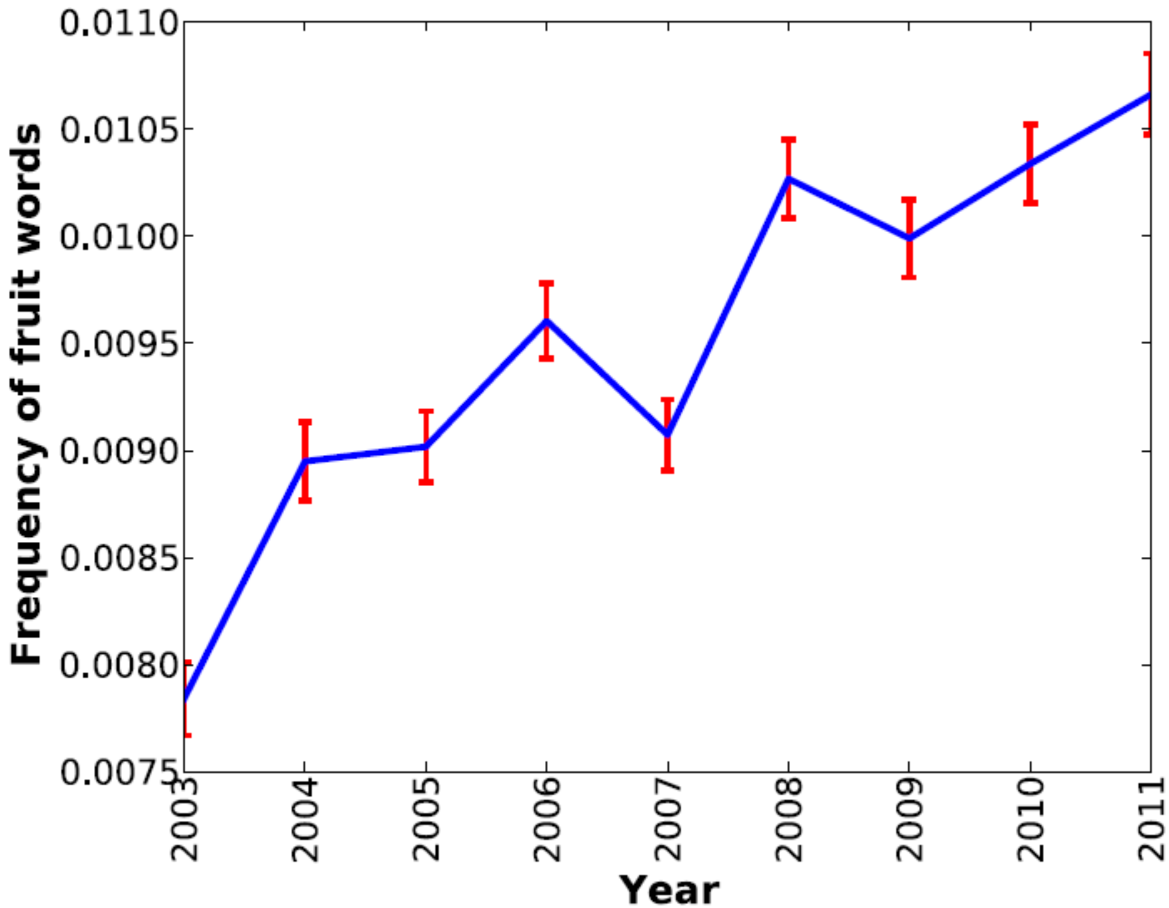
(a) First person sing. pronouns



(b) Beer specific vocabulary

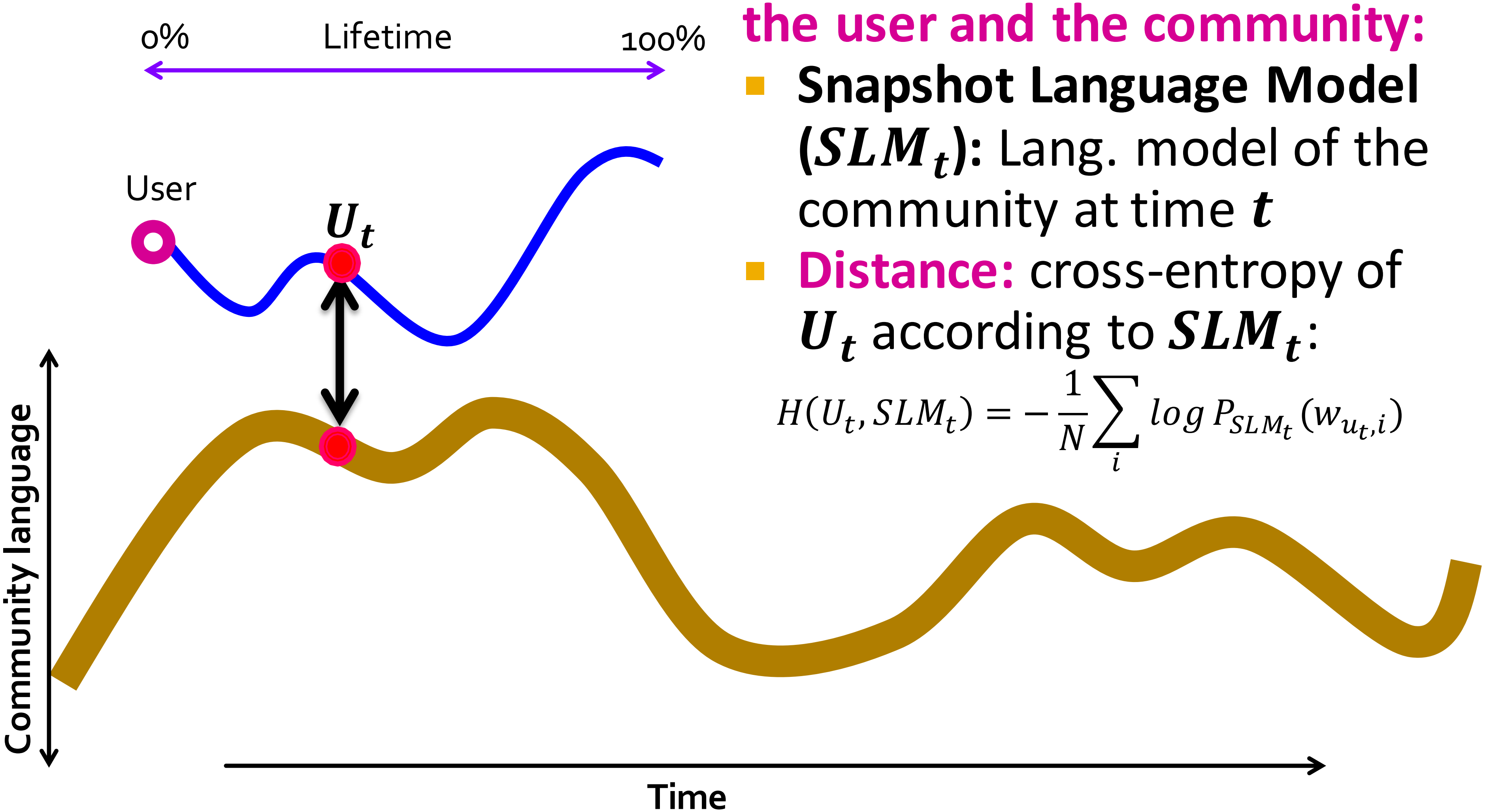
- **A sign of increasing identification with the community** [Pennebaker 2007; Sherblom 2009]

Community Language Change



- **Fruit words (*peach, pineapple, berry, …*) are getting ever more popular**

User-Community Change



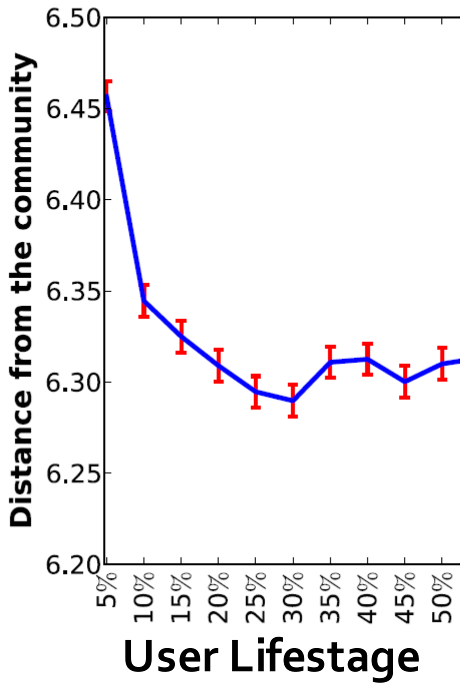
Language distance between the user and the community:

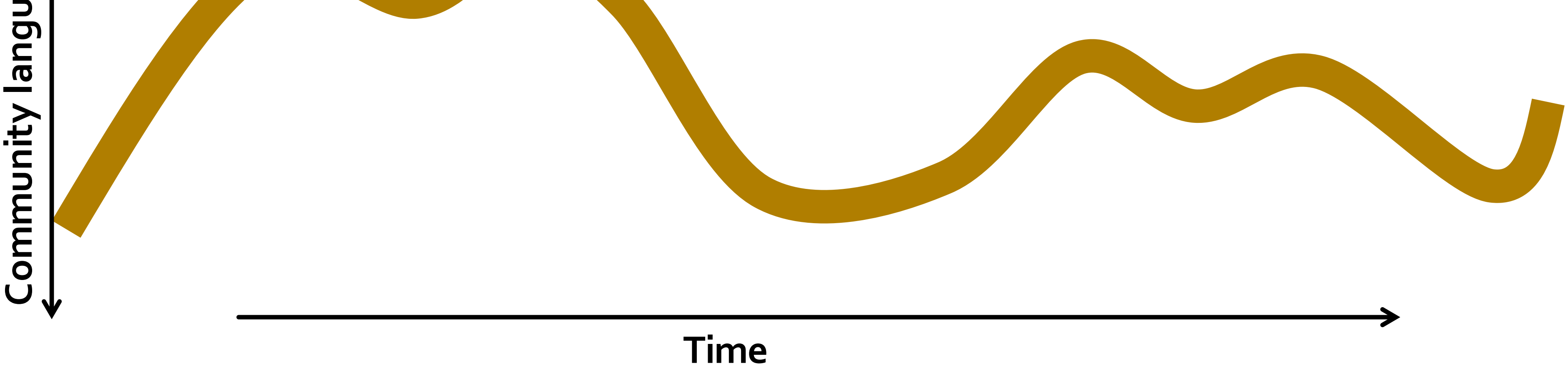
- **Snapshot Language Model (SLM_t):** Lang. model of the community at time t
- **Distance:** cross-entropy of U_t according to SLM_t:

$$H(U_t, SLM_t) = -\frac{1}{N}\sum_i log\, P_{SLM_t}(w_{u_t,i})$$

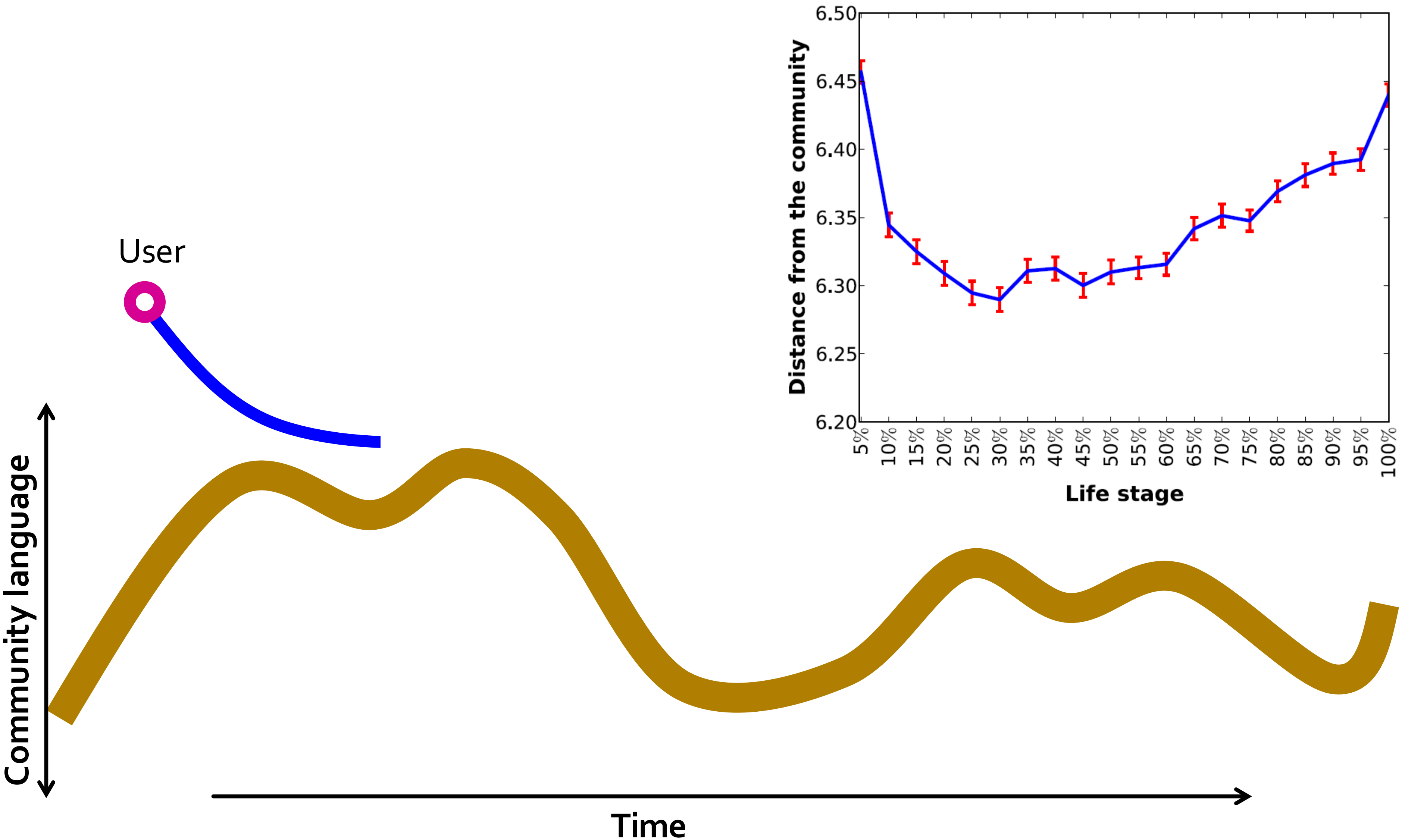
User-Community Change

Stage 1: User **assimilates** language of the community



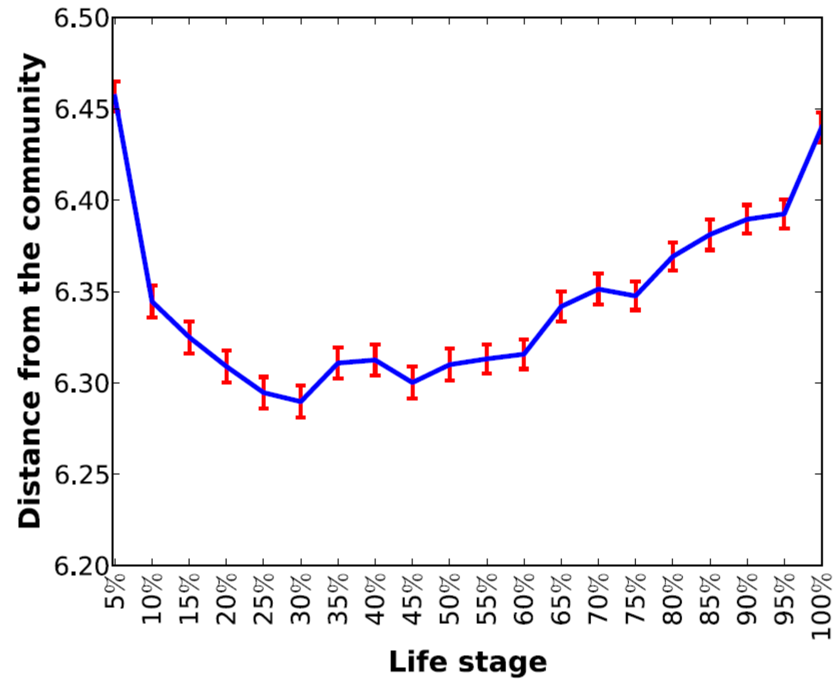


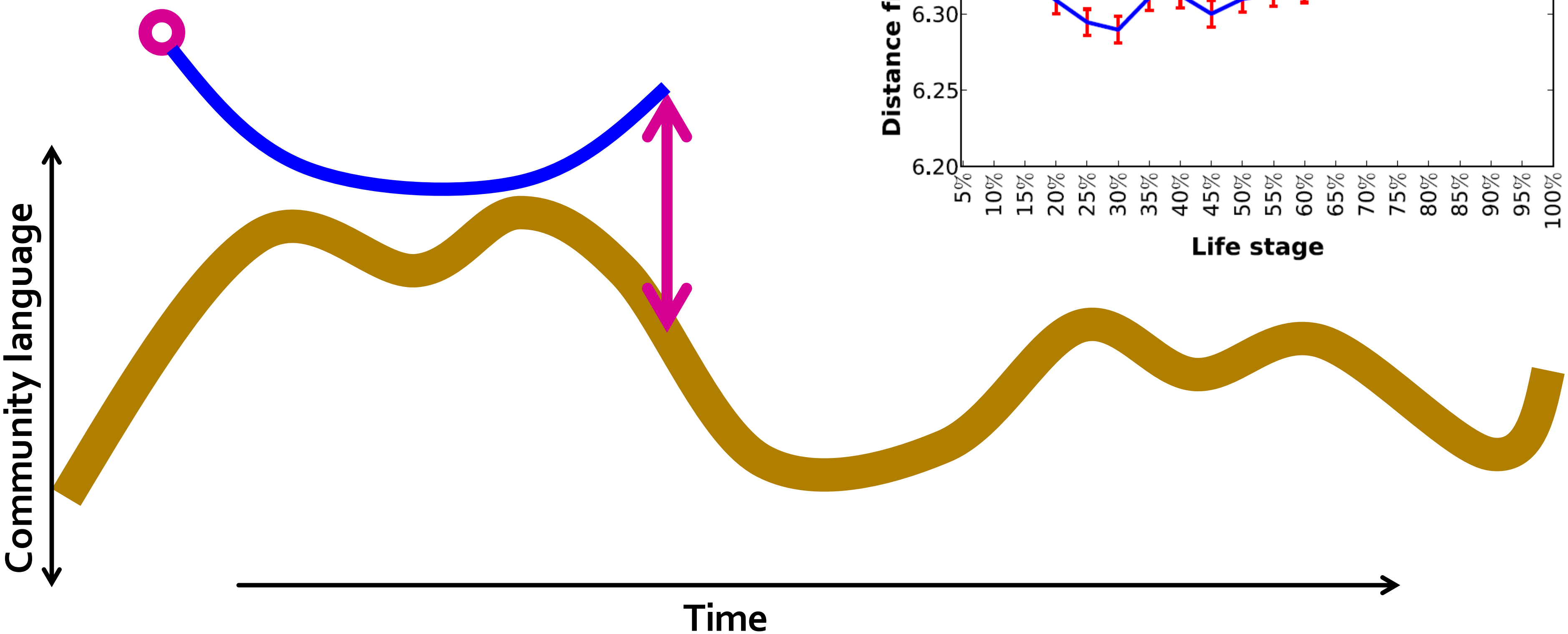
User-Community Change



User-Community Change

Stage 2: User's **language** distances itself from that of the community

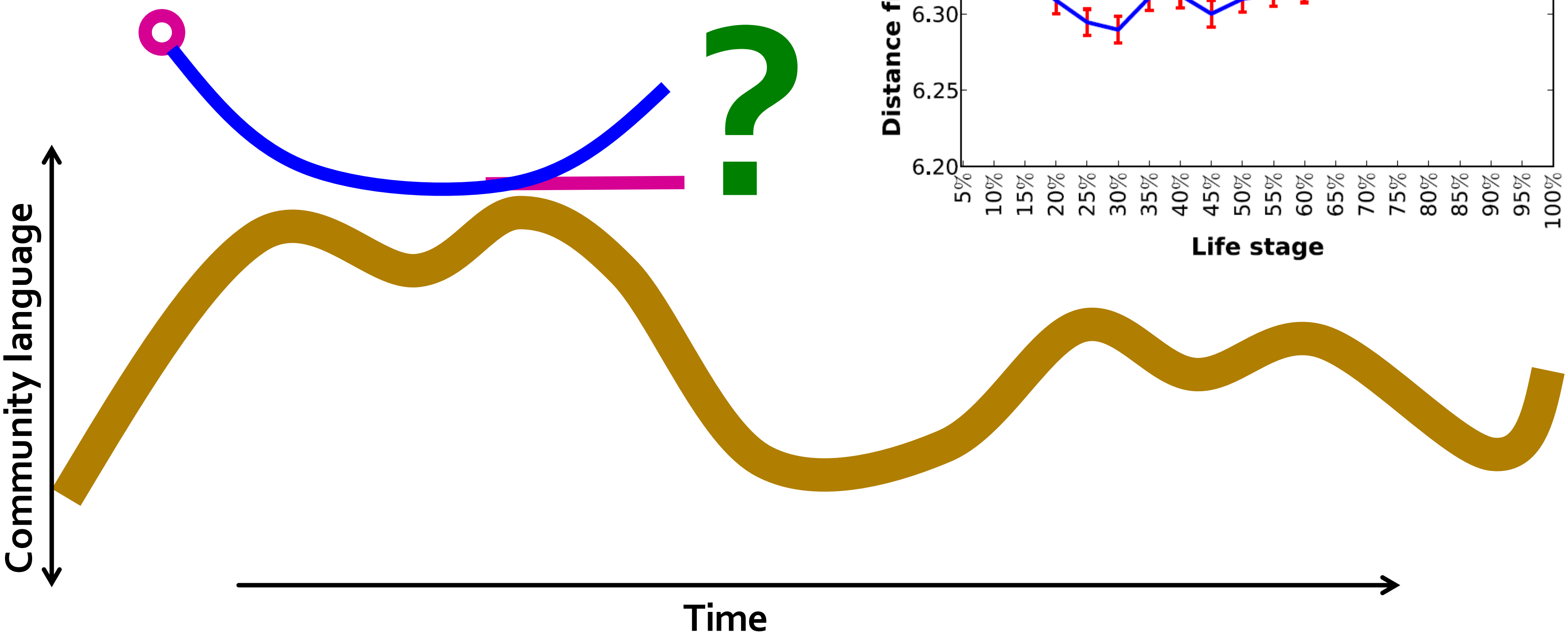


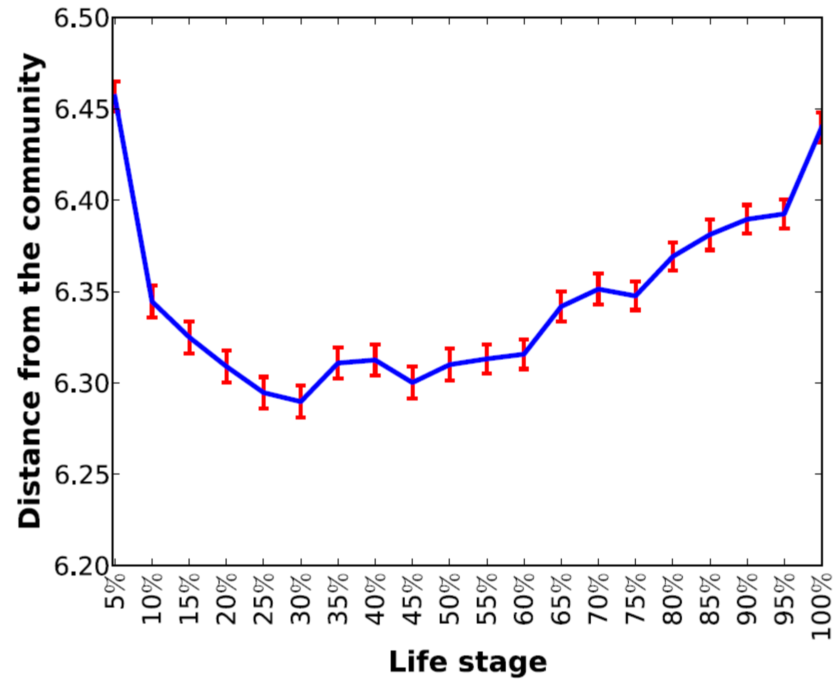


User-Community Change

H1: User moves away

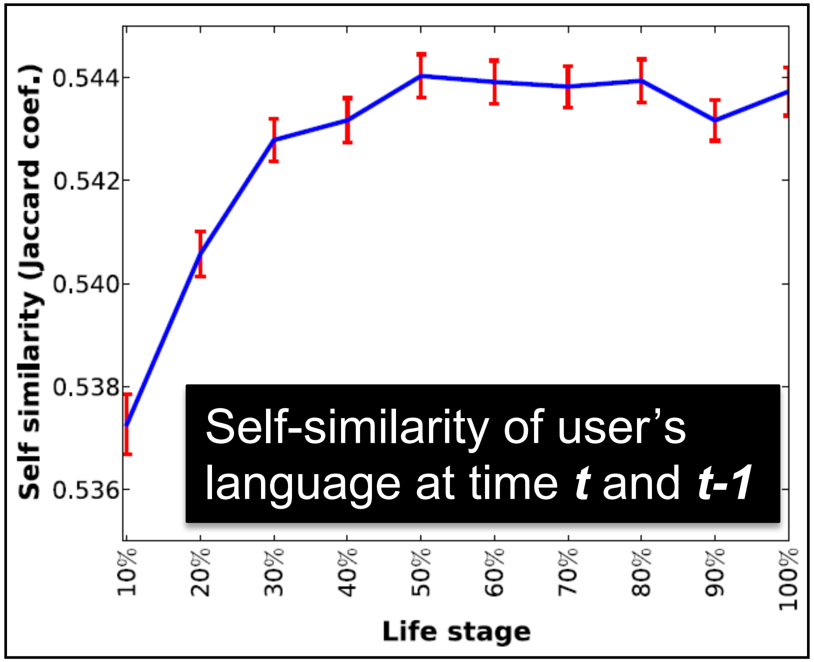
H2: User stops adapting

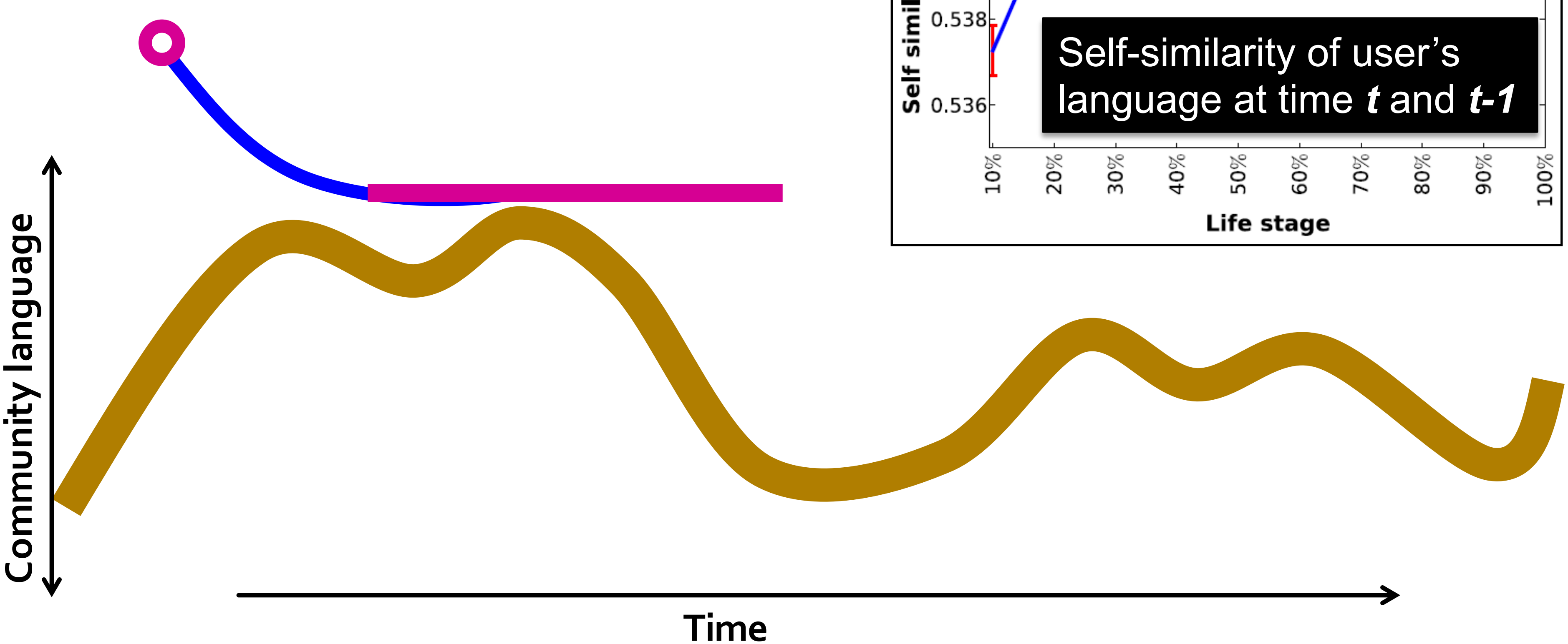




User-Community Change

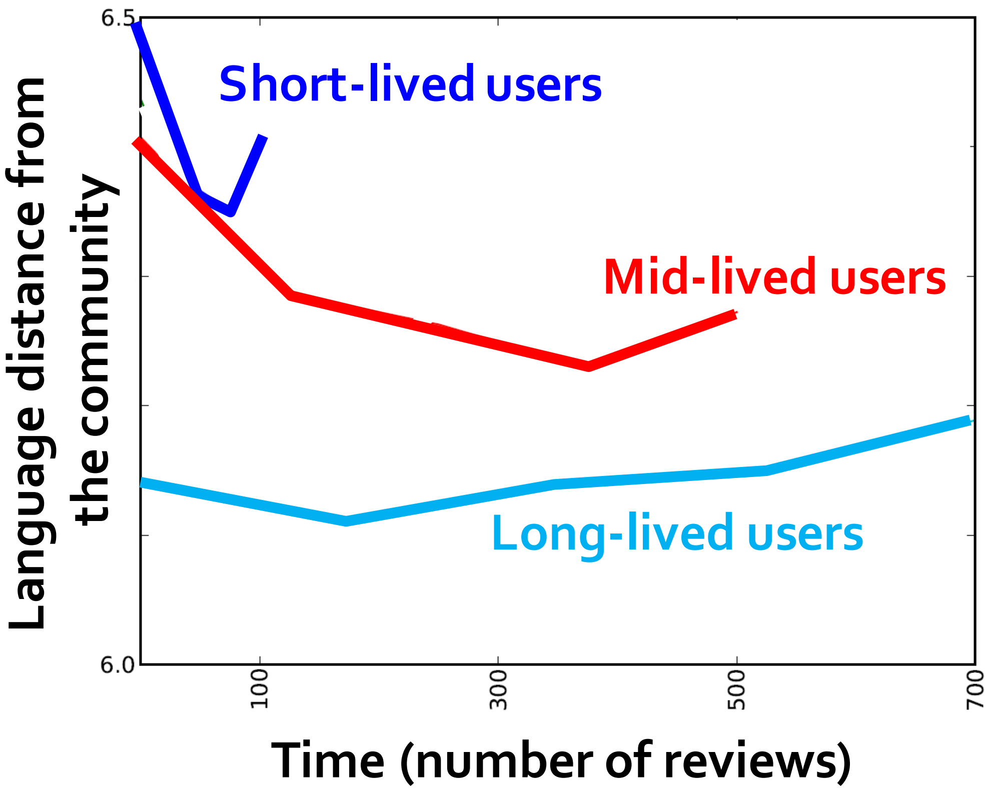
H2: Users get stuck in the past!





Elastic Lifecycle

- So far we stretched lifetimes to 0-100%
- **What about user's absolute lifetime?**



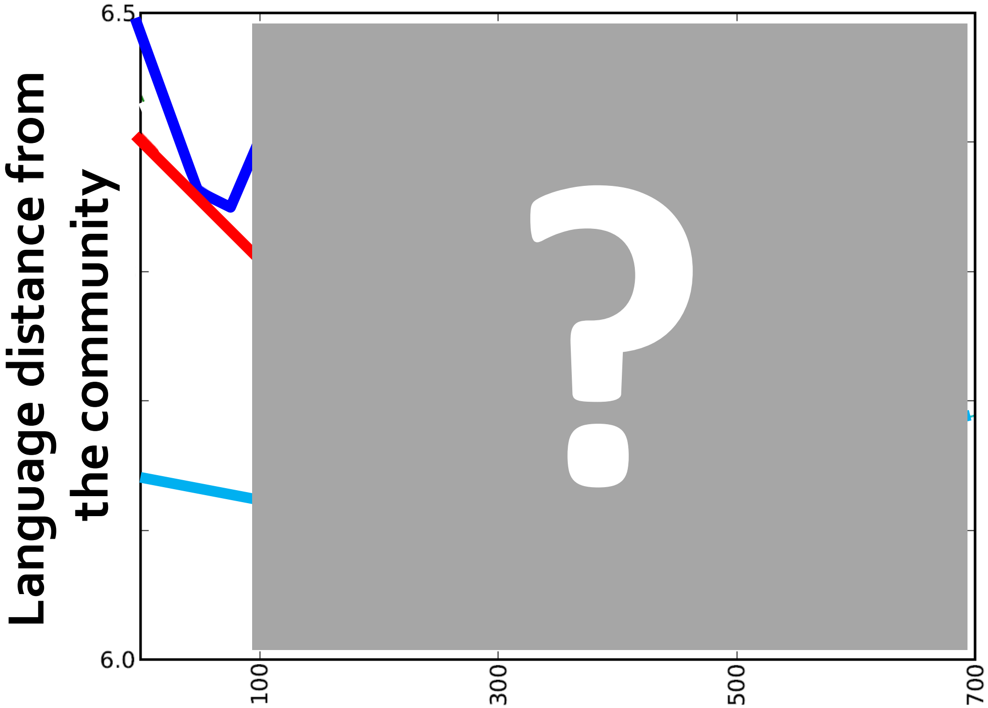
→ **Similar lifecycle in spite of different lifespans**
"All users die old"

→ **End of the adaptation phase is a function of the ultimate lifespan**

→ **Level of receptivity is related to the ultimate lifespan**

Elastic Lifecycle

- So far we stretched lifetimes to 0-100%
- **What about user's absolute lifetime?**



→ **Similar lifecycle in spite of <u>different lifespans</u>**
"All users die old"

→ **End of the adaptation phase is a function of the <u>ultimate lifespan</u>**

→ **Level of receptivity is related to the <u>ultimate lifespan</u>**

Predict user's ultimate lifespan

How do we influence social systems?

People work amazingly hard to earn badges

"Give me enough medals and I'll win you any war."

– Napoleon

KHAN ACADEMY

BADGE TYPES

Meteorite badges are common and easy to earn when just getting started.

Moon badges are uncommon and represent an investment in learning.

Earth badges are rare. They require a significant amount of learning.

Sun badges are epic. Earning them is a true challenge, and they require impressive dedication.

Black Hole badges are legendary and unknown. They are the most unique Khan Academy awards.

Challenge Patches a special awards for completing topic challe

WIKIPEDIA
Den fria encyklopedin

foursquare®

stackoverflow

Badges

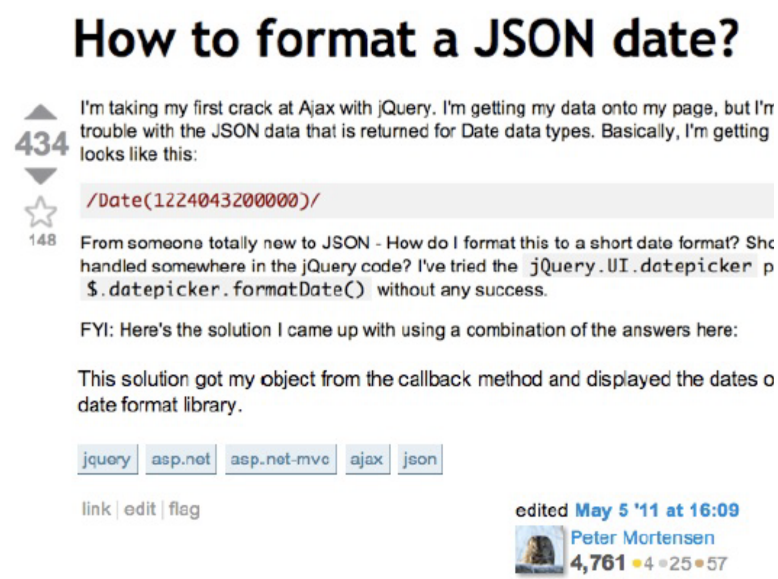
- **Multiple roles of badges:**
  - Can recognize a wide range of activities:
    - Total effort, Single high-impact contribution, …
  - Serve both as **credentials** and **incentives**
- **Incentive aspects of badges:**
  - Trend toward gamification [Deterding et al. '11]
  - Customer loyalty programs [Kopalle-Neslin '03]
  - Simple for users to understand, and less based on competition with others

Badges & Behavior Change

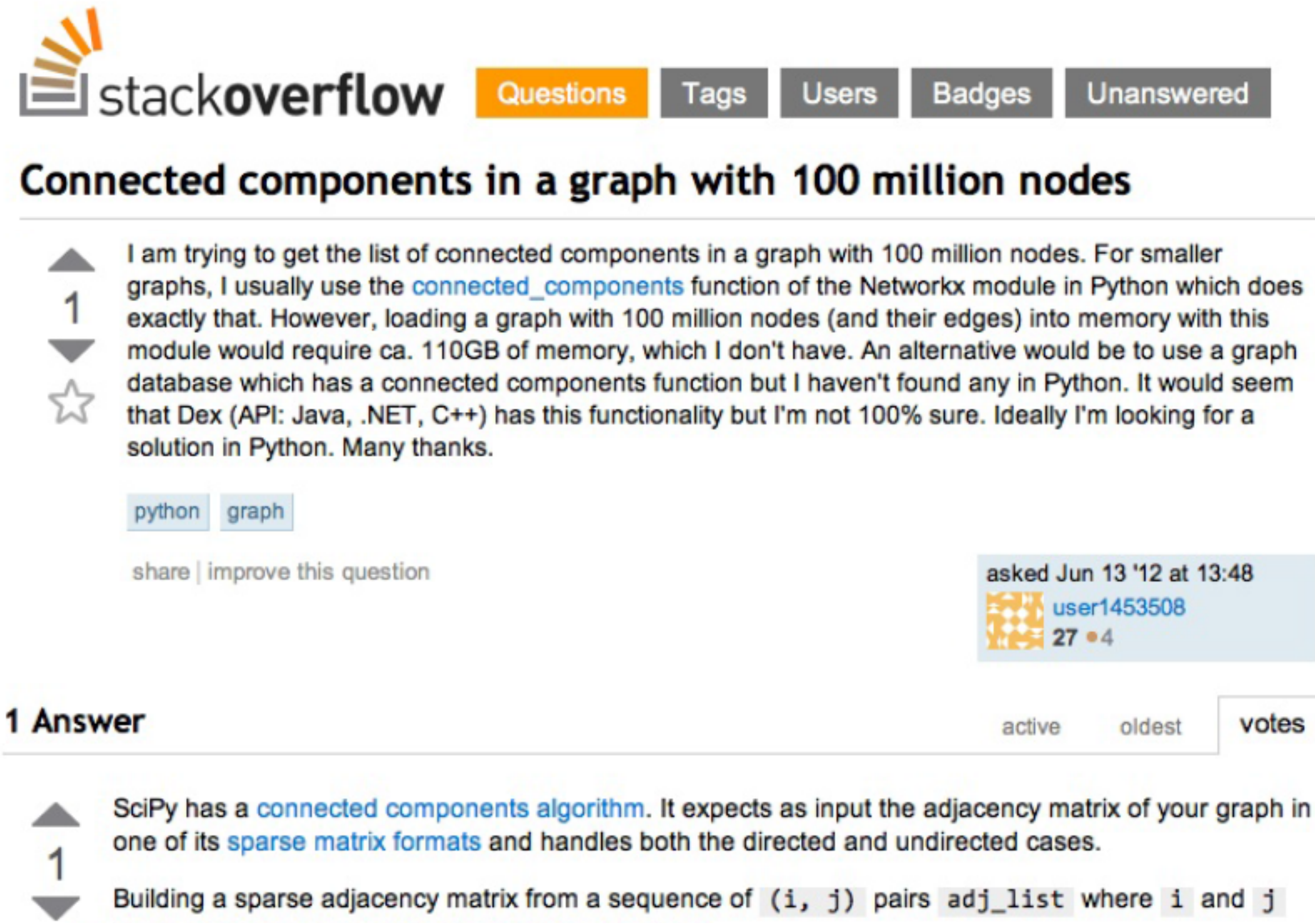
How do badges translate into effects on user behavior?

Need lots of data to tease out the effects and build a mathematical model

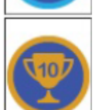
Badges on Stack Overflow Q&A site:
- 2M people
- 5M questions
- 10M votes



Badges on StackOverflow



Newbie: Congrats on your 1st answer

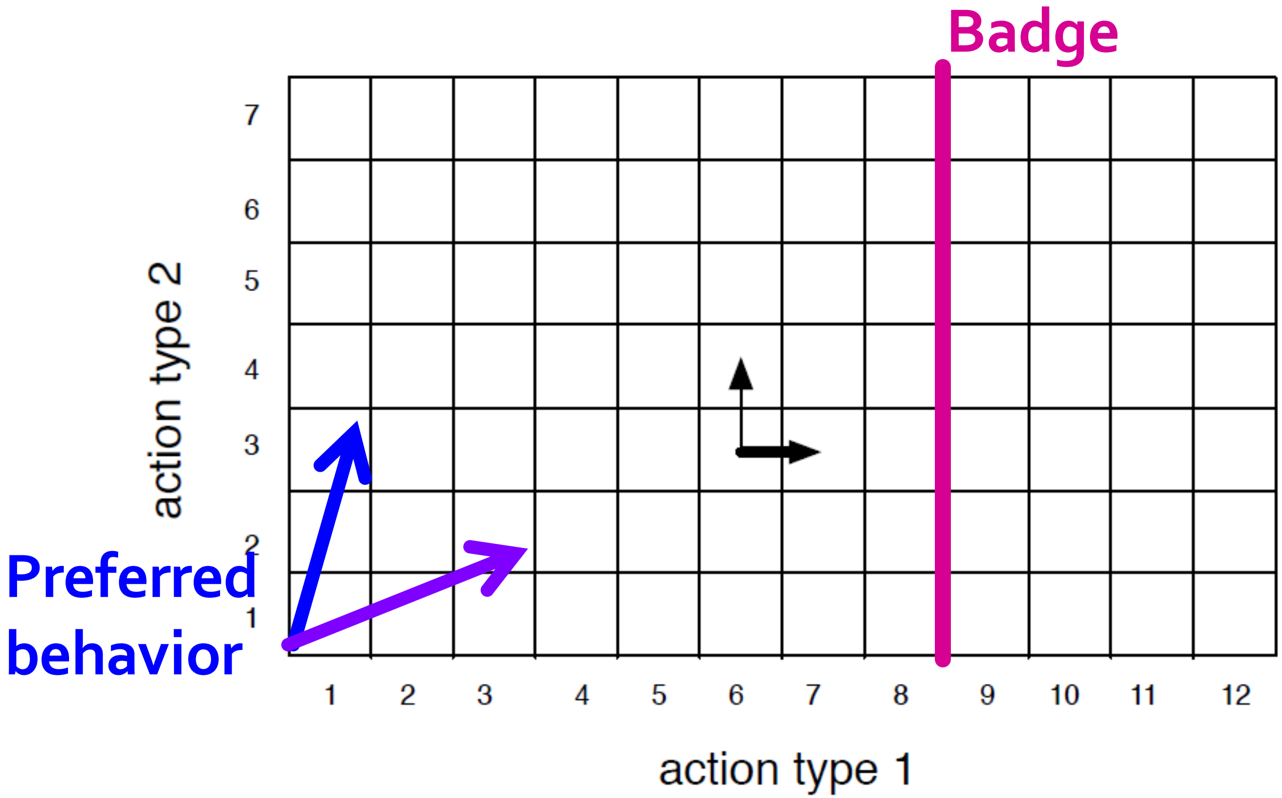


Superstar: You answered 10 questions

Model of Badges

- **Approach: Utility based model of badges**
  - User trades off between her preferred mix of activities and the goal of reaching the badge
  - Effects on both **engagement** and **steering**
    - **Engagement:** Increased user site activity
    - **Steering:** Users change the actions they do

Model of Badges



- 2 parts to the model:
  - User gains value from obtaining a badge
  - But it "costs" user to change behavior

Model of Badges

- **User's optimization problem:**
  Choose actions to maximize utility
  - User balances between achieving badges quickly and keeping cost low
- **At each point x_a in time:**
  - **Receive utility V_b** of **badges** already won
  - **Get penalized** for deviating from **preferred behavior p** according to loss function $g(p, x)$
  - Exit the site with probability θ

Model of Badges

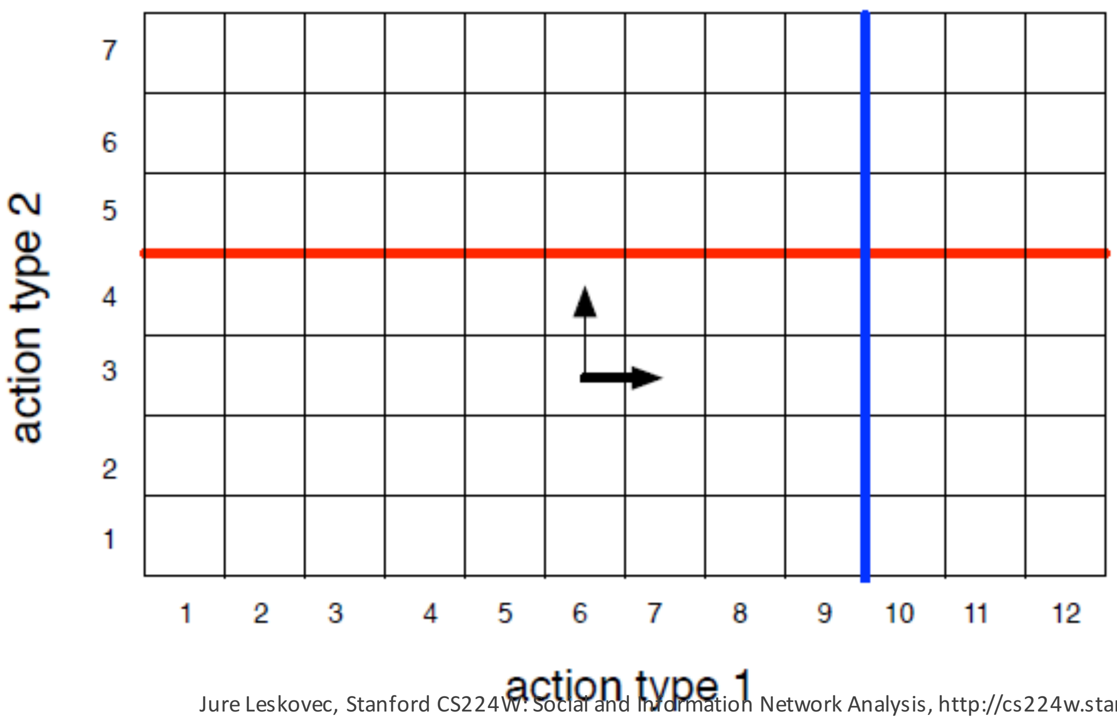
$$U(\mathbf{x_a}) = \sum_{b \in B} I_b(\mathbf{a}) V_b + \theta \sum_{i=1}^{n+1} \mathbf{x_a^i} \cdot U(\mathbf{x_{a+e_i}}) - g(\mathbf{x_a}, \mathbf{p})$$

Utility | Badges | Discounted future utility for action $\boldsymbol{e_i}$ | Cost for deviating from $\boldsymbol{p}$

- $\boldsymbol{U(x_a)}$... total utility for a user with action distribution $\boldsymbol{x_a}$
- $\boldsymbol{a}$ ... user's actions so far (count of actions)
- $\boldsymbol{I_b(a)}$ ... given user's actions $\boldsymbol{a}$, did she get badge $\boldsymbol{b}$ (0/1)
- $\boldsymbol{V_b}$ ... value of badge $\boldsymbol{b}$
- $\boldsymbol{\theta}$... prob. of exiting the site
- $\boldsymbol{x_a^i}$... user's prob. of taking action $\boldsymbol{i}$
- $\boldsymbol{U\left(x_{a+e_i}\right)}$... future utility after taking action $\boldsymbol{i}$
- $\boldsymbol{g(x_a, p) = \sum_i \left(x_a^i - p_i\right)^2}$ ... cost for deviating from $\boldsymbol{p}$
- $\boldsymbol{p}$... preferred action distribution

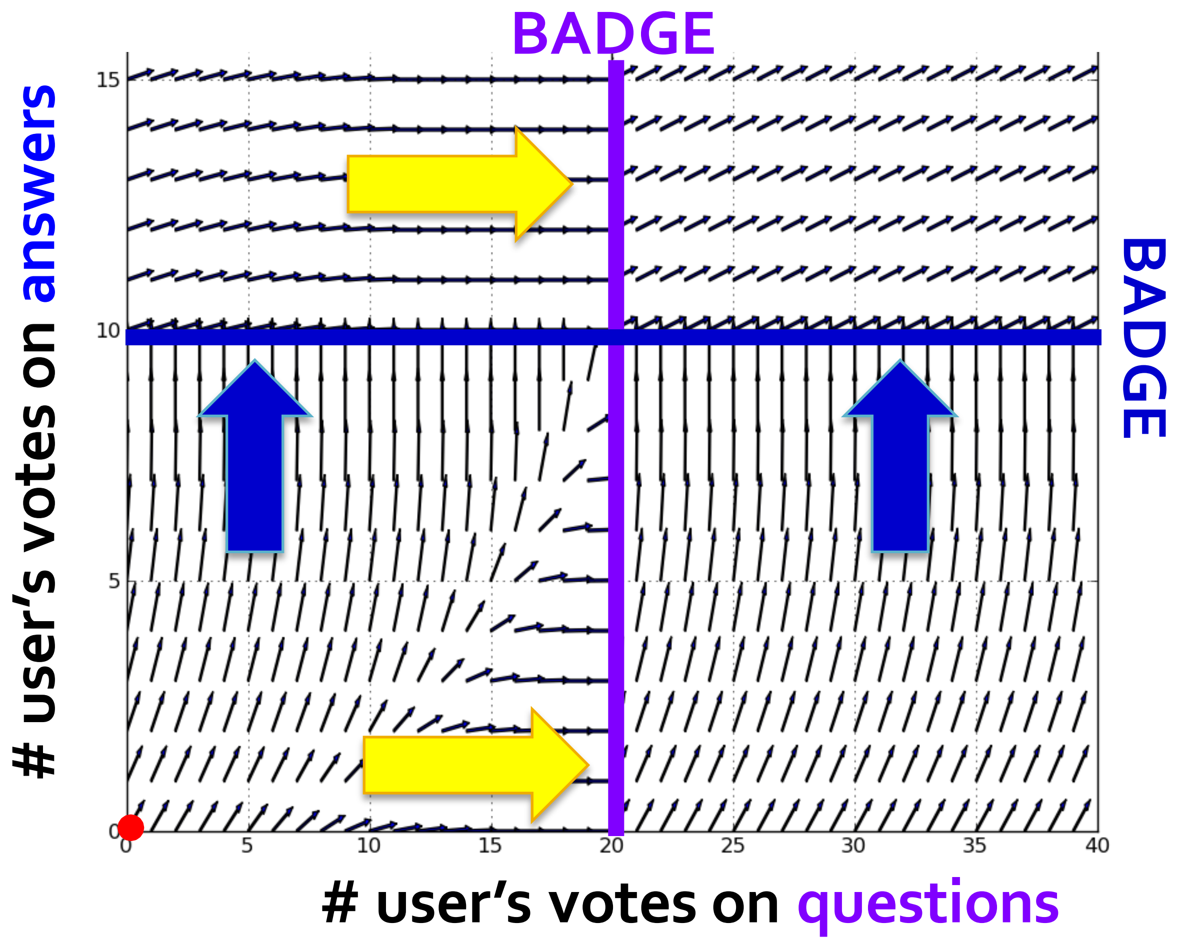
Solving the Model

- **Solve *arg max*** $U(\mathbf{x_a}) = \sum_{b \in B} I_b(\mathbf{a}) V_b + \theta \sum_{i=1}^{n+1} \mathbf{x_a^i} \cdot U(\mathbf{x_{a+e_i}}) - g(\mathbf{x_a}, \mathbf{p})$
  - Partition space based on badge boundaries
  - Inductively solve regions in order from large coordinates to small

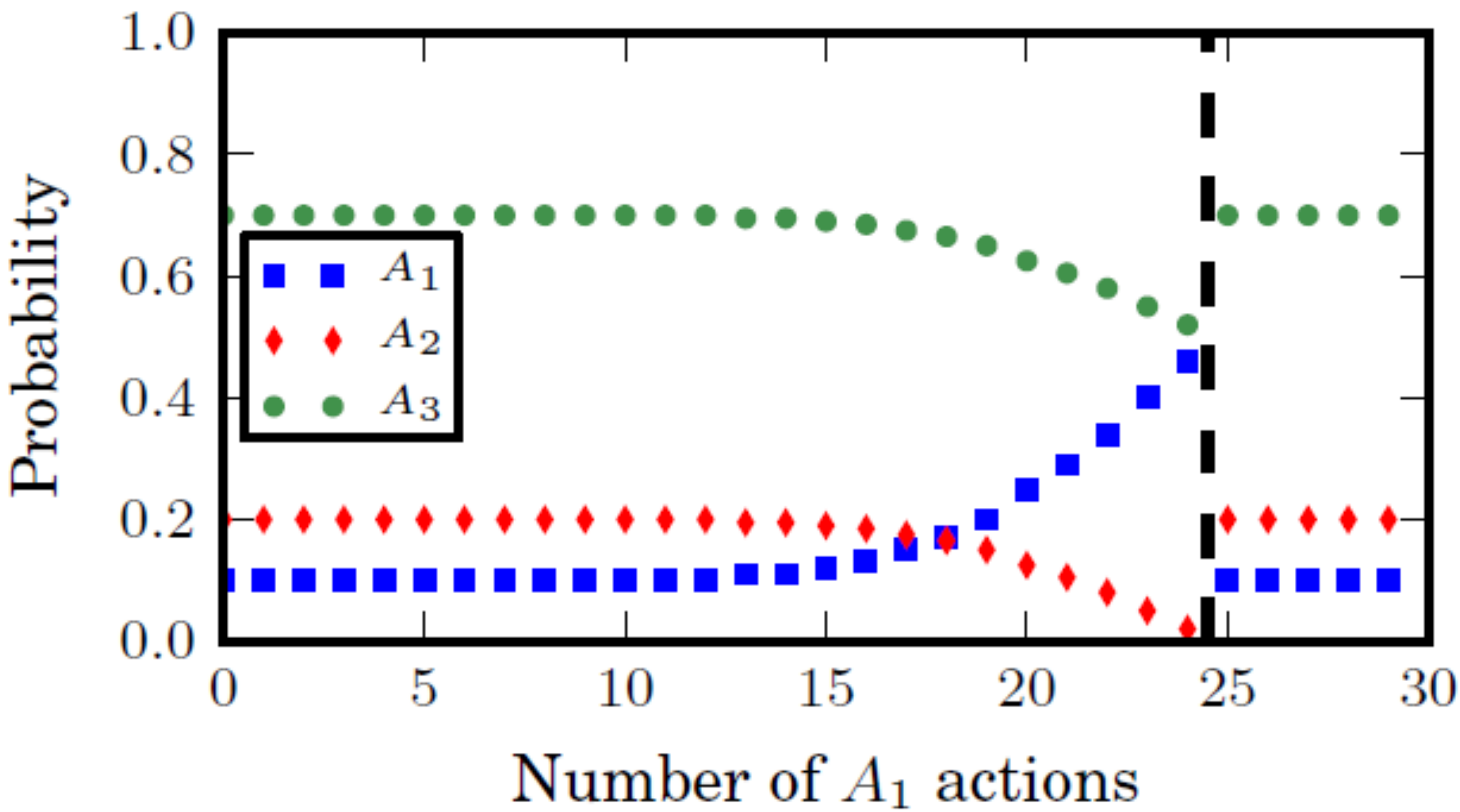


Example: 2 Badges

- **Influencing user behavior:**



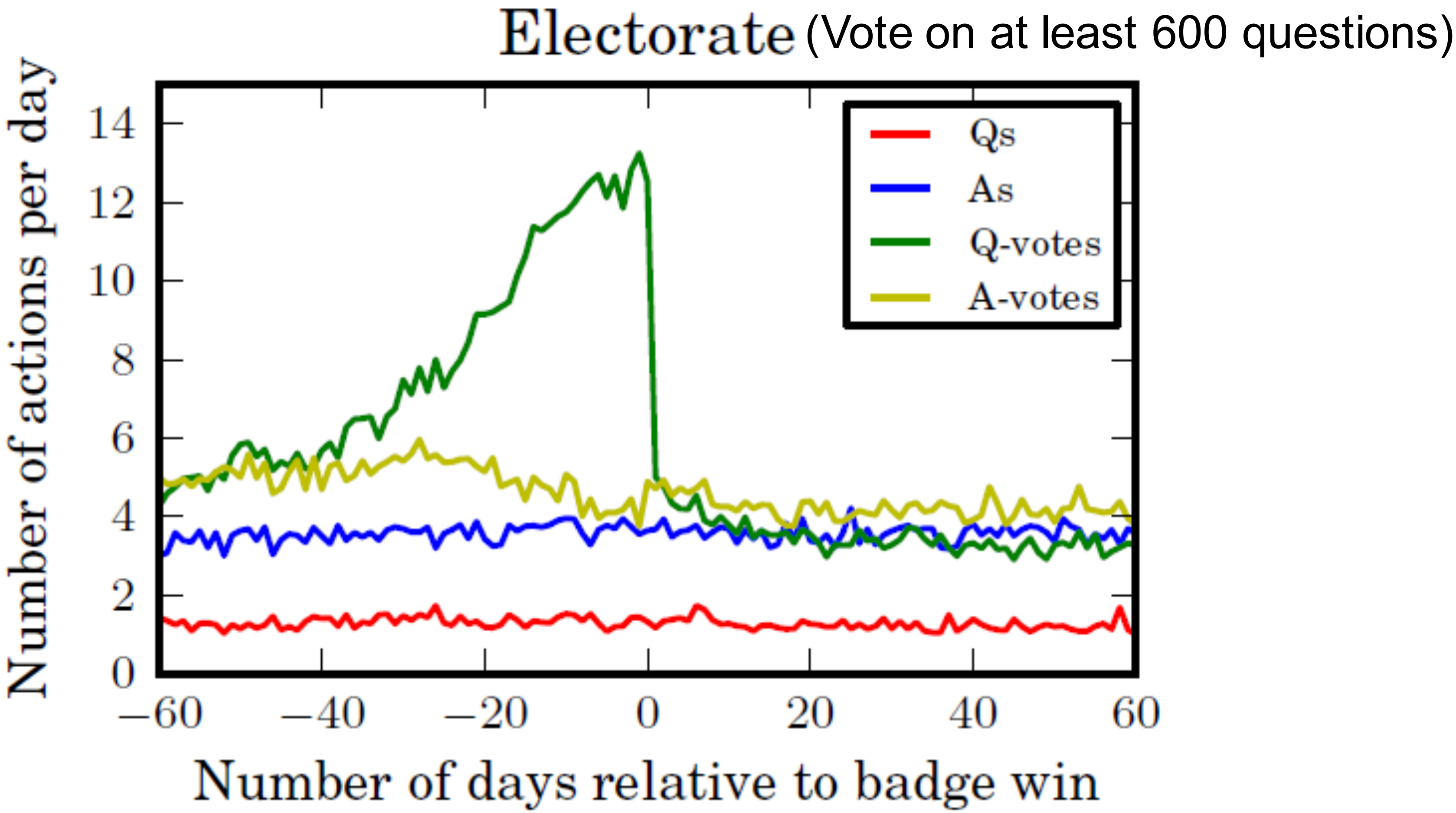
Example: 1 Badge



- **Example:** Badge at 25 actions of type $\mathbf{A_1}$
  - User **steers** in $\mathbf{A_1}$ direction as she approaches the badge boundary; then resets after receiving it

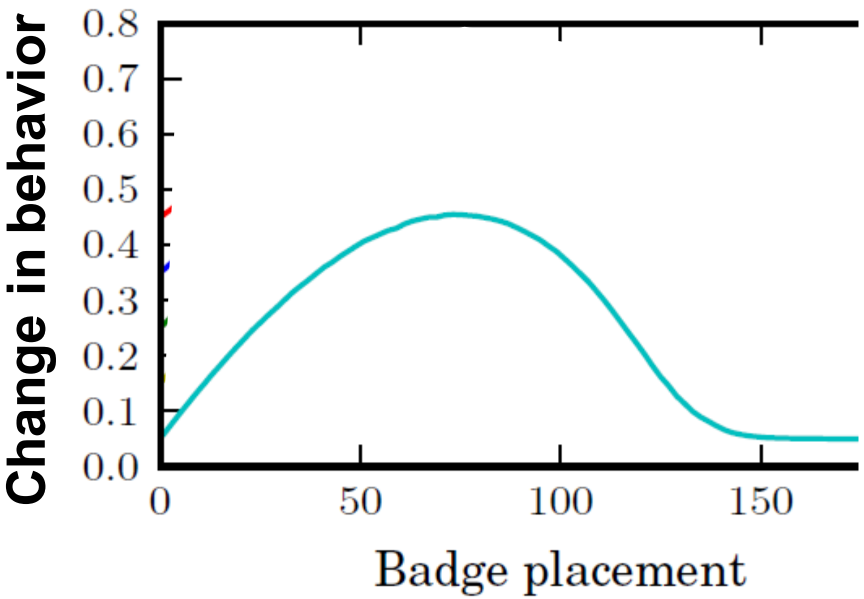
Model vs. Data

- **Model predicts qualitative behavior**



Badge Placement Problem

- **Question:** How should you "place" badges to achieve desired effects?
- **Our model allows for optimizing the badge placement for optimal behavior steering:**



Badges on Coursera

Badge Series (2 earned)

The Reader
To earn the next badge (Silver), you must read 30 threads from your classmates.

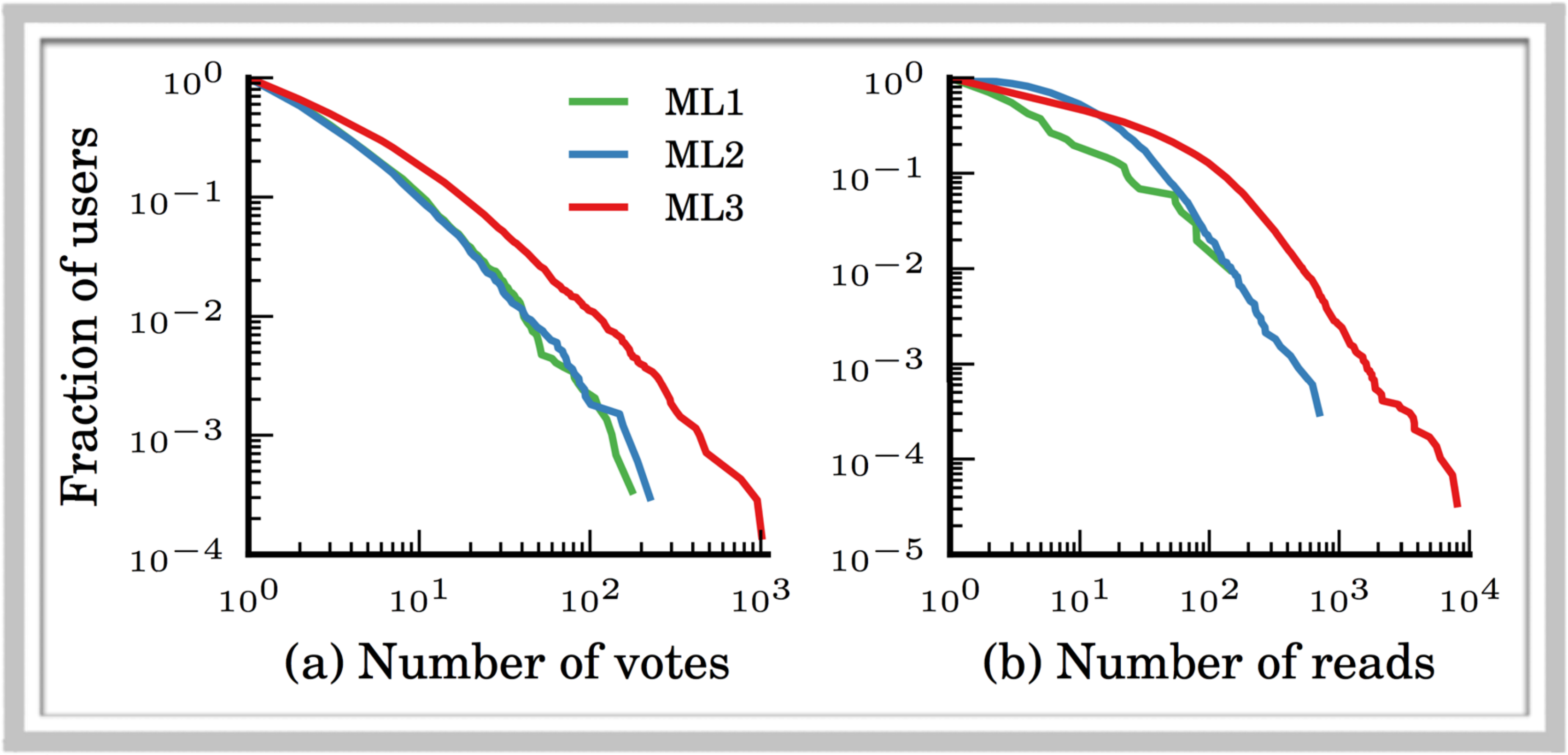
The Supporter
To earn the next badge (Silver), you must vote on 15 posts that you find interesting or useful.

The Contributor
To earn the next badge (Bronze), you must post 3 replies that your classmates find interesting.

The Conversation Starter
To earn the next badge (Bronze), you must start 3 threads that your classmates find interesting.

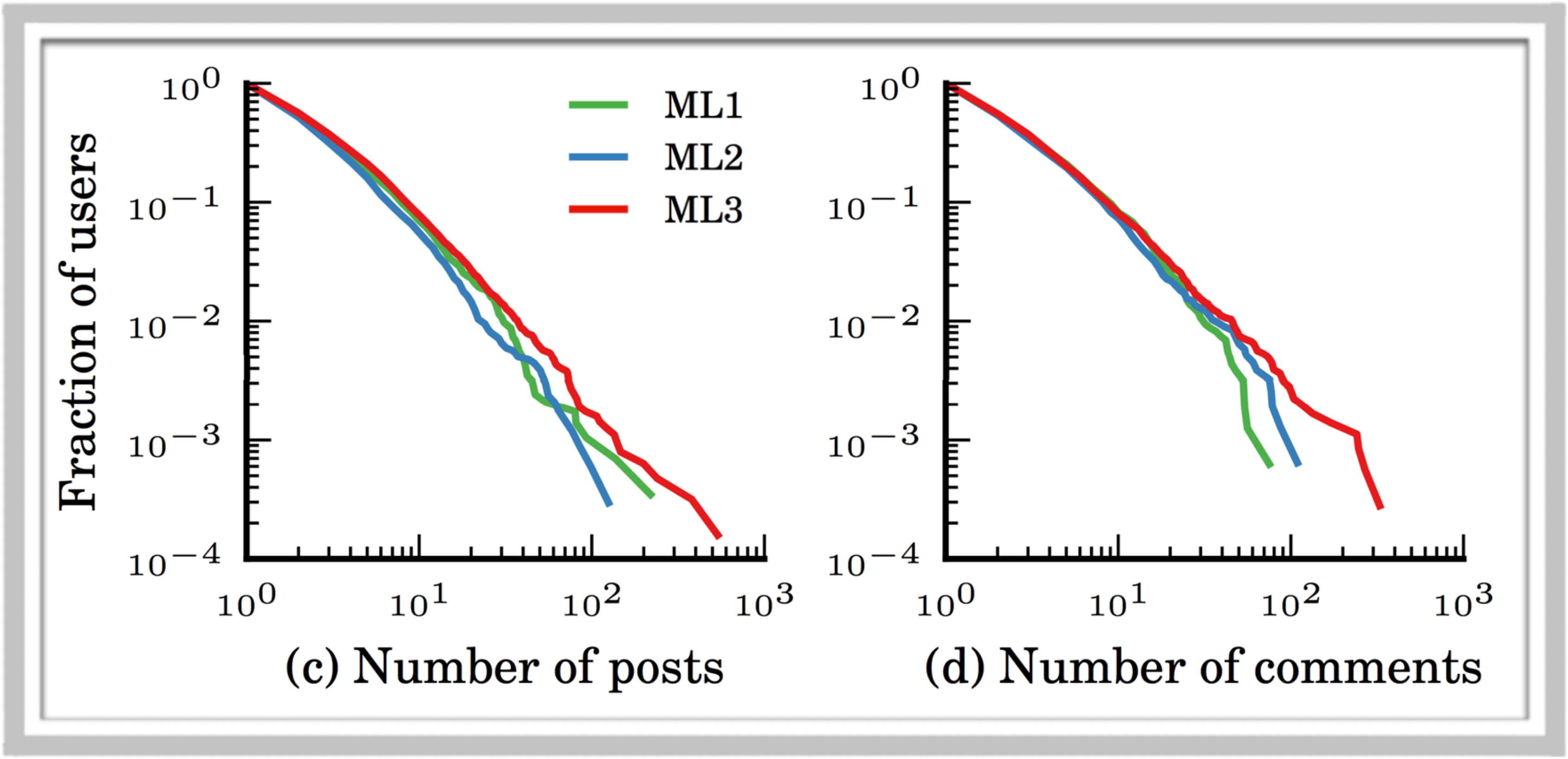
Top Posts
To earn the next badge (Bronze), you must write a post that gets 5 upvotes from your classmates.

Experiment: ML Class



5x more likely to get to 100 votes/reads!

Experiment: ML Class



No qualitative difference in posts/comments

No badges on these actions!

Badges: Further questions

- **Many questions, both within this framework and extending it:**
  - Where does the value of a badge come from? Internal, social, transactional, ...
  - How does achievement-seeking interact with competition and scarcity?
  - How far can we develop analogies with off-line domains?

Social and Information Networks:
Review of Key Concepts

Networks

How do we reason about networks?

Reasoning About Networks

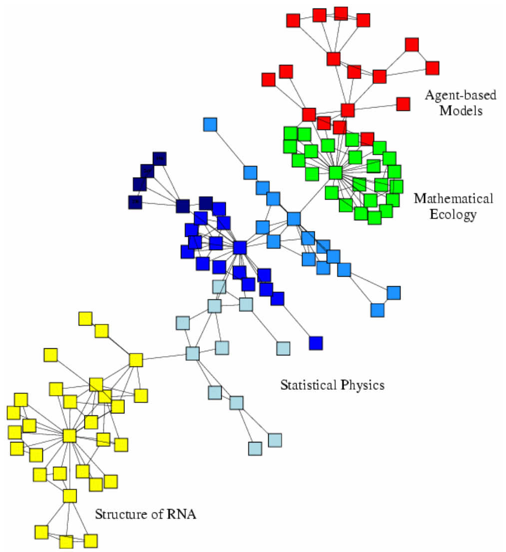
How do we reason about networks?

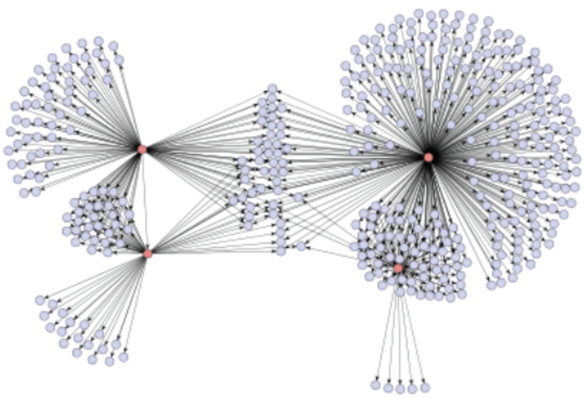
- **Empirical:** Study network data to find organizational principles
- **Mathematical models:** Probabilistic, graph theory
- **Algorithms:** Methods for analyzing graphs

Networks: Structure & Process

What do we study in networks?

- **Structure and evolution:**
  - What is the structure of a network?
  - Why and how did it become to have such structure?
- **Processes and dynamics:**
  - Networks provide "skeleton" for spreading of information, behavior, diseases





What We Have Covered

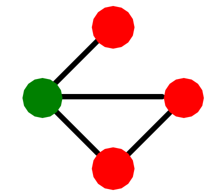
- Network diameter
- Edge clustering
- Scale-free networks
- Strength of weak ties
- Core-periphery structure
- Densification power law
- Shrinking diameters
- Structural Balance
- Status Theory
- Memetracking
- Small-world model
- Erdös-Renyi model
- Preferential attachment
- Network cascades
- Independent cascade model
- Decentralized search
- PageRank
- Hubs and authorities
- Girvan-Newman
- Modularity
- Clique percolation
- Supervised random walks
- Influence maximization
- Outbreak detection
- Linear Influence Model
- Network Inference
- Kronecker Graphs
- Bow-tie structure

How It All Fits Together

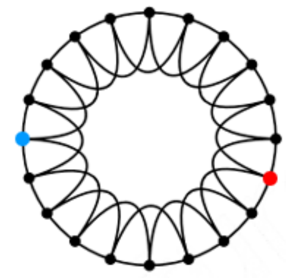
Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Kronecker Graphs	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Patterns of signed edge creation	Structural balance, Theory of status	Models for predicting edge signs
Information virality, Memetracking	Independent cascade model, Game theoretic model	Influence maximization, Outbreak detection, LIM

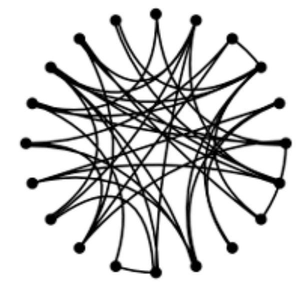
Small-World Phenomena

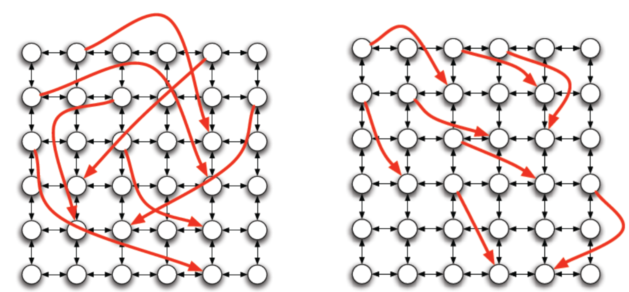
- **Properties:**
  - **Six degrees of separation**
    - Networks have small diameters
  - **Edges in the networks cluster**
    - Large clustering coefficient
- **Models:**
  - **Erdös-Renyi model**
    - Baseline model for networks
  - **The Small-World model**
    - Small diameter and clustered edges
- **Algorithms:**
  - **Decentralized search in networks**
    - Kleinberg's model and algorithm



C_i=1/3



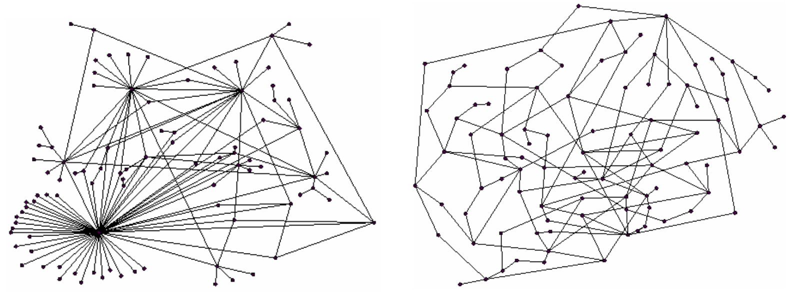


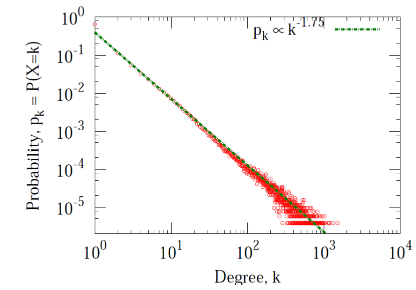


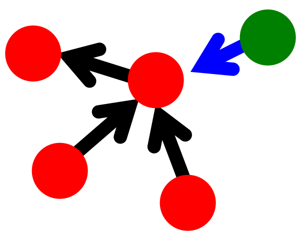
$$P(u \rightarrow v) \sim d(u,v)^{-\alpha}$$

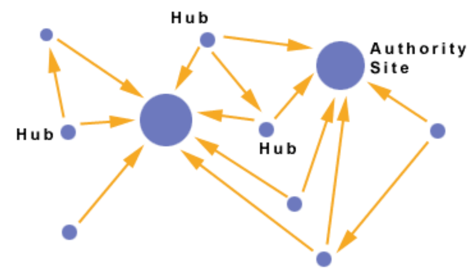
Scale-Free Networks

- **Properties:**
  - **Power-law degrees**
    - Degrees are heavily skewed
  - **Network resilience**
    - Networks are resilient to random attacks
- **Models:**
  - **Preferential attachment**
    - Rich get richer
- **Algorithms:**
  - **Hubs and Authorities**
    - Recursive: $a_i = \sum_{j \to i} h_j, h_i = \sum_{i \to j} a_j$
  - **PageRank**
    - Recursive formulation, Random jumps



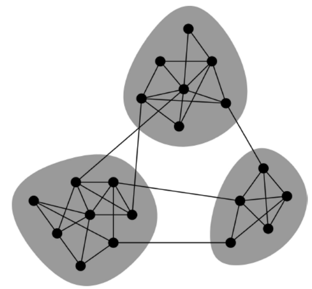


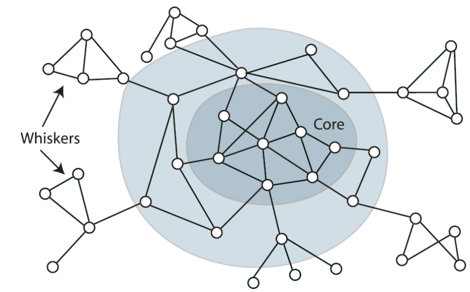




Community Detection

- **Properties:**
  - **Strength of weak ties**
  - **Core-periphery structure**

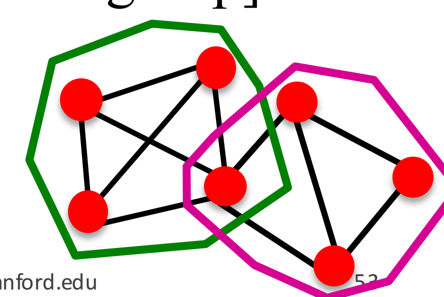




- **Models:**
  - **Kronecker graphs model**

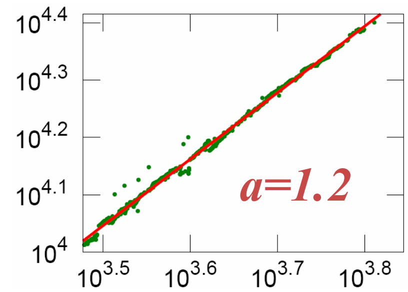
$$\begin{array}{|c|c|c|}\hline 1 & 1 & 0 \\ \hline 1 & 1 & 1 \\ \hline 0 & 1 & 1 \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|c|}\hline K_1 & K_1 & 0 \\ \hline K_1 & K_1 & K_1 \\ \hline 0 & K_1 & K_1 \\ \hline \end{array}$$

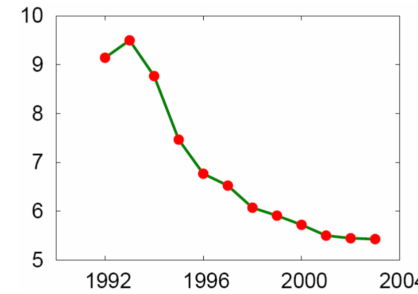
- **Algorithms:**
  - **Spectral Clustering**
  - **Girvan-Newman** (Betweeness centrality)
  - **Modularity:** #edges within group – E[#edges within group]
  - **Clique Percolation Method**
    - Overlapping communities

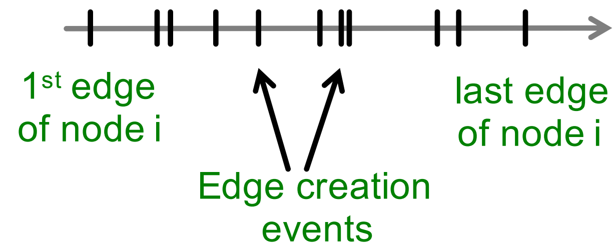


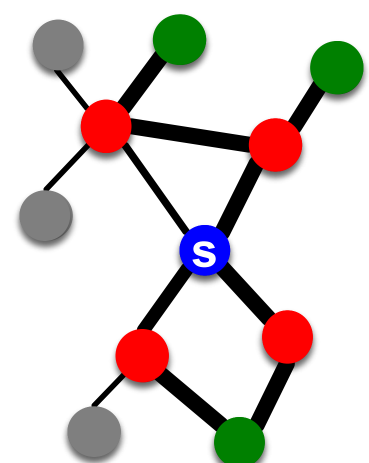
Network Evolution

- **Properties:**
  - **Densification Power Law**
    - $E(t) \propto N(t)^a$
  - **Shrinking Diameter**
- **Models:**
  - **Microscopic Network Evolution**
    - Exponential life-times, Evolving sleeping times
    - Random-Random edge attachment
- **Algorithms:**
  - **Link prediction**



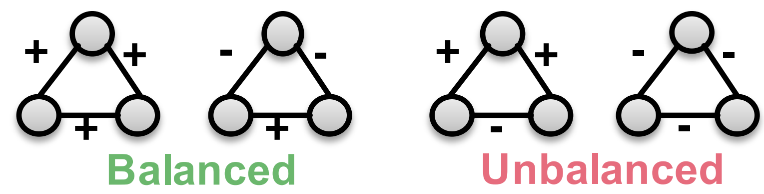


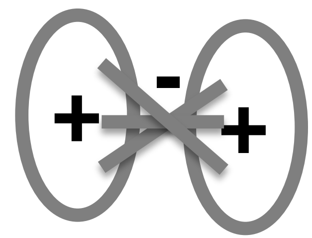


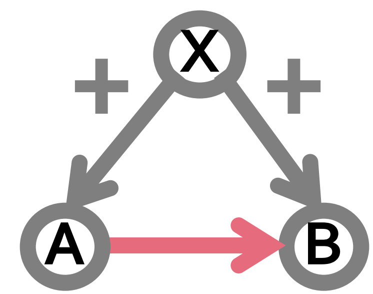


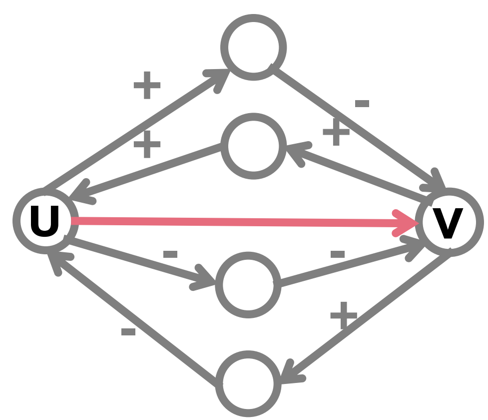
Signed Networks

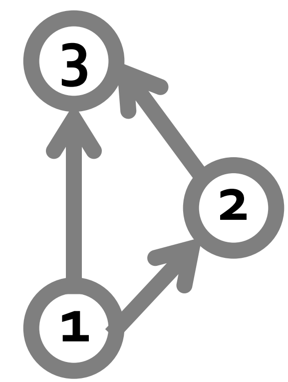
- **Properties:**
  - **Signed link creation**
  - **+links are more embedded**
- **Models:**
  - **Structural Balance**
    - Coalition structure of networks
  - **Status Theory**
    - Global node status ordering
- **Algorithms:**
  - **Predicting edge signs**





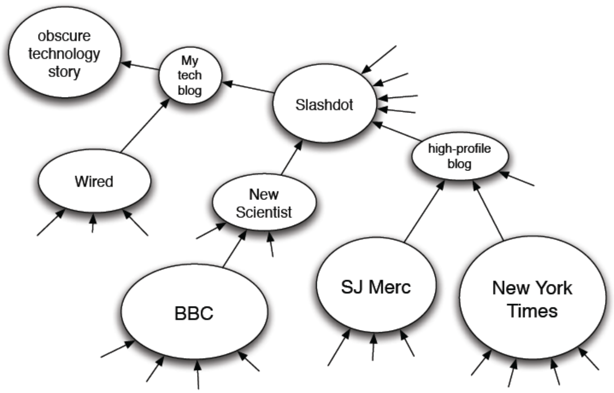


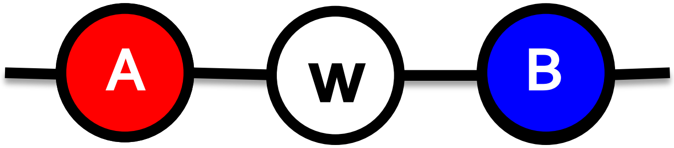




Network Diffusion (1)

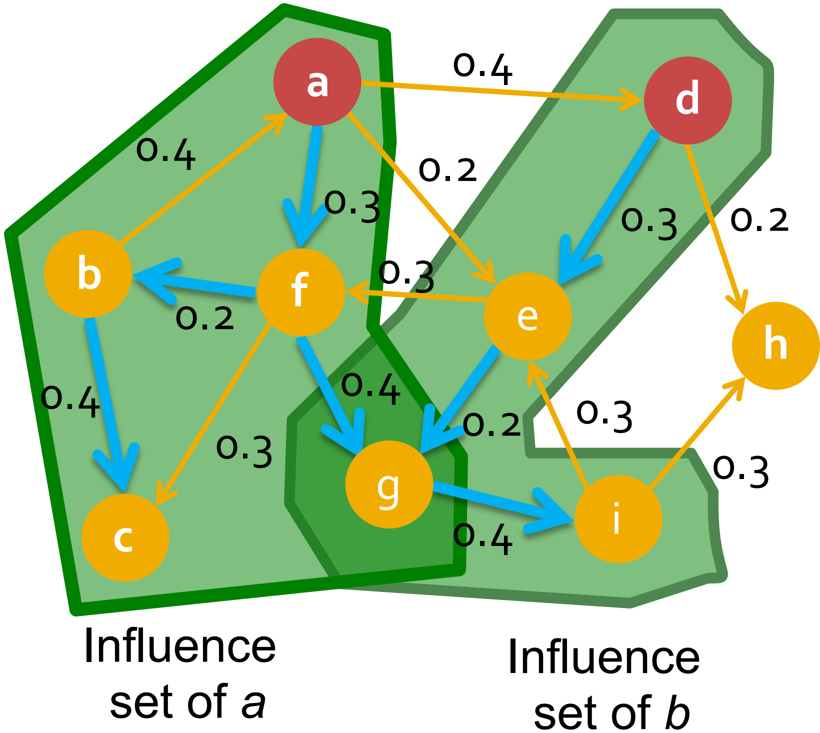
- **Properties:**
  - **Meme-tracking**
    - Blogs trail mass media
- **Models:**
  - **Game theoretic model:**
    - Payoffs, Competing products
  - **Independent Cascade Model**
    - Each node infects a neighbor with some probability





Network Diffusion (2)

- **Algorithms:**
  - **Influence Maximization**
    - Set *of k* nodes producing **largest expected cascade size** if activated
    - Submodularity
    - Greedy hill-climbing
  - **Outbreak Detection**
  - **Network Inference**
    - Infer networks based on information diffusion data



Map of Superpowers

Properties	Models	Algorithms
Small diameter, Edge clustering	Small-world model, Erdös-Renyi model	Decentralized search
Scale-free	Preferential attachment, Copying model	PageRank, Hubs and authorities
Strength of weak ties, Core-periphery	Kronecker Graphs	Community detection: Girvan-Newman, Modularity
Densification power law, Shrinking diameters	Microscopic model of evolving networks	Link prediction, Supervised random walks
Patterns of signed edge creation	Structural balance, Theory of status	Models for predicting edge signs
Viral Marketing, Blogosphere, Memetracking	Independent cascade model, Game theoretic model	Influence maximization, Outbreak detection, LIM

What Next?

- **Project write-ups:**
  - Sun Dec 11 Midnight **(11:59PM)** Pacific Time **No late days!**
    - **1 team member uploads PDF to Gradescope**
    - **See course website for more info**
- **Poster session:**
  - Tue Dec 13 from 3 - 6pm in **Gates**
    - All groups with at least one non-SCPD member must present
    - There should be 1 person at the poster at all times
    - **Prepare a 2-minute elevator pitch of your poster**
    - **More instructions to follow**

What Next? Seminars

- **Seminars:**
  - **InfoSeminar:** http://i.stanford.edu/infoseminar
    - Can be taken for credit (CS545, also on SCPD)
    - Great industry/academia speakers on Fridays
- **Conferences / Journals:**
  - **WWW**: ACM World Wide Web Conference
  - **WSDM**: ACM Web search and Data Mining
  - **ICWSM**: AAAI Int. Conf. on Web-blogs & Social Media
  - **KDD**: Conf. on Knowledge Discovery & Data Mining
  - **Journal of Network Science**
  - **Journal of Complex Networks**

What Next? Courses

- **CS246: Mining Massive Datasets (Winter 2017)**
  - Data Mining & Machine Learning for big data
    - (big==does' fit in memory/single machine), MapReduce
- **CS341: Project in Data Mining (Spring 2017)**
  - Groups do a research project on big data
  - We provide interesting data, projects and access to the Amazon computing infrastructure
  - Nice way to finish up CS224W project & publish it!

What Next? Courses

- **Other relevant courses:**
  - **MS&E 231:** Computational Social Science
  - **MS&E334:** The Structure of Social Data
  - **CS276**: Information Retrieval and Web Search
  - **CS229**: Machine Learning
  - **CS245**: Database System Principles
  - **CS347**: Transaction Processing & Databases

In Closing…

- **You Have Done a Lot!!!**
- **And (hopefully) learned a lot!!!**
  - Answered questions and proved many interesting results
  - Implemented a number of methods
  - **And are doing excellently on the class project!**

Thank You for the Hard Work!!!


Eve Online: Exodus developer CCP publisher CCP
© GameWallpapers.com hosted by JTLnet.com