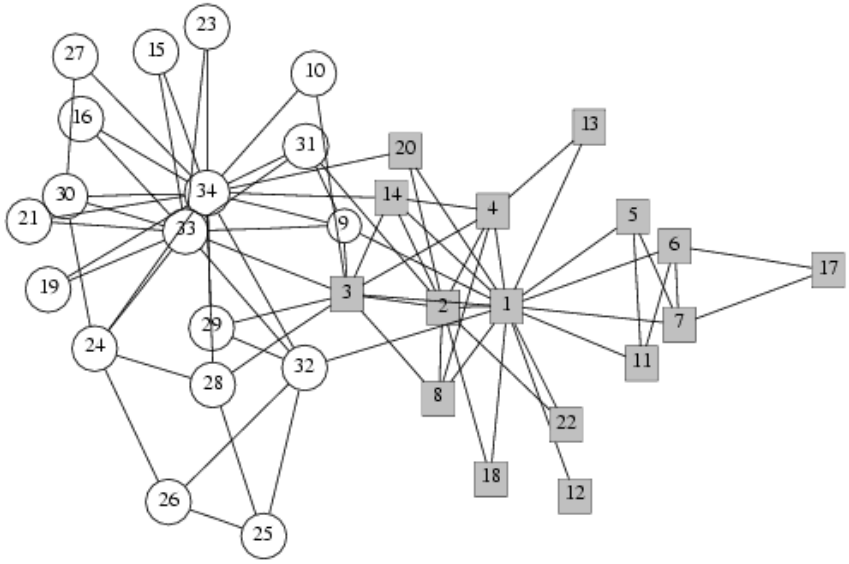
Online Social Networks

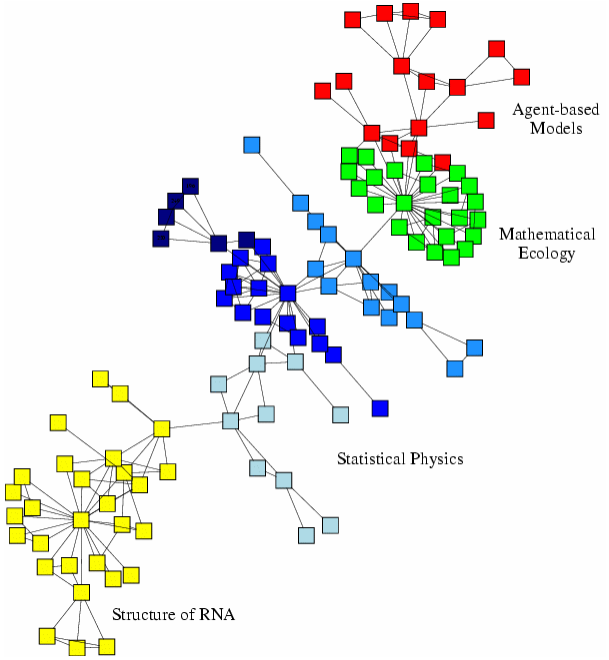
Daniel Huttenlocher

John P. and Rilla Neafsey Professor
of Computing, Information Science and Business

Social Network Models



[Zachary77]



- Individuals and relationships between them
  - E.g., karate club friendships, paper co-authors
- Density and patterns of connectedness as relate to behavior

Characterizing Social Networks

- Six degrees of separation
  - Popular lore (Kevin Bacon game)
  - Deliveries to strangers [Milgram67, Watts01]
    - Experiments where delivery averaged < 6 hops
- Rule of 150
  - "Genuine social network" limited in size
    - E.g., size of village
- Network properties
  - Density, clustering, etc.
    - E.g., open vs. closed triads, "social capital"

Social Networks: Past and Present

- Network models of social interactions have 50 year history in academia
  - Been difficult to study except on small scale
  - Now have large-scale data, but often asking simple questions
- Social networking web sites date back several years (e.g., Classmates ’99)
  - Varying levels of popularity, boom-bust
  - Recently become mass phenomenon
    - MySpace surpassed Google page views in 2005
    - Lots of visibility, sometimes negative

Social Software on Web

- Enables people to connect, rendezvous, collaborate, form communities
  - Content: blogs, wikis, media sharing (video, photo, audio)
    - E.g., Blogger, Wikipedia, YouTube, Flickr
  - Connections: job networking, friend networking
    - E.g., LinkedIn, Facebook
  - Hybrids
    - E.g., MySpace, LiveJournal, Orkut, Xanga
- Replacing other forms? Enhancing?

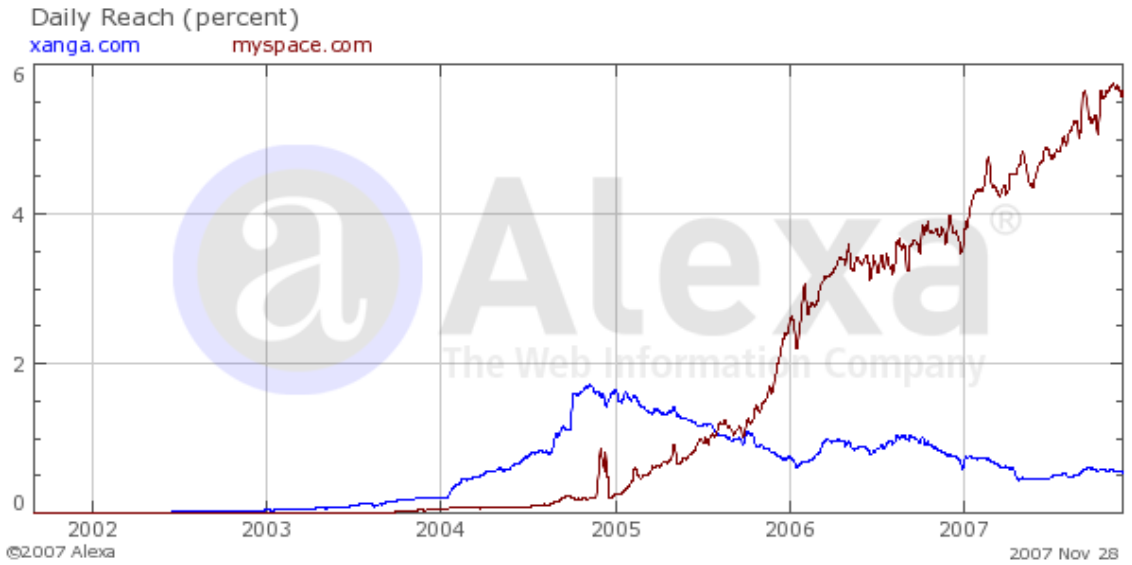
Social Software Web Site Usage

- MySpace
  - #3 overall rank in US, same as year ago
    - Daily reach 30K/M, page views 20K/M
    - Vs. #1 Yahoo: reach 300K/M, views 70K/M
- YouTube
  - #4 vs. #8 year ago
- Facebook
  - #5 vs. #15 year ago
- Wikipedia
  - #10 vs. #9 year ago

Source: Alexa.com

Fickleness of Internet Users

- Staying power remains to be seen – former leaders can quickly fall from grace
  - E.g., evolution of Xanga vs. MySpace



Recent Studies at Larger Scales

- Studying formation and evolution of groups, online and offline
  - Large scale, hundred thousand or more people
  - Work with Jon Kleinberg, students and postdocs
  - Informed by social science questions
- Part of Cornell wide focus on Social and Information Networks
  - 2006-07 Institute for Social Sciences Theme
    - 10 faculty from across Cornell, visiting scholars
    - New courses at undergrad and grad level
  - 2006-08 NSF Cyberinfrastructure Tools Project

Studies of Evolving Social Groups

- What factors influence a person's decision to join a group?
- What factors indicate that a group will grow in membership?
- What causes groups to change their focus?
- How do changes in focus correlate to changes in group membership?
- Online vs. offline?
  - Political group vs. MySpace group?

Propensity to Join Groups

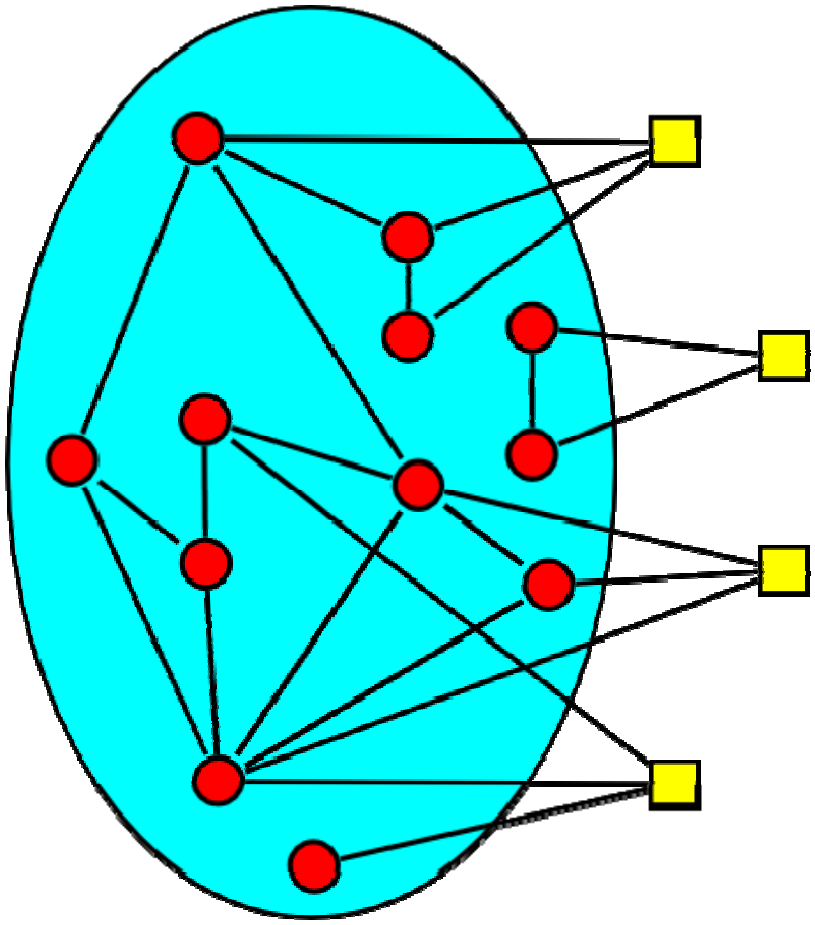
- Groups where members explicitly affiliate
  - Political, religious, service, professional, athletic, …
- Extent to which influenced by presence of friends in (connections to) group
  - Observational study, not causal model
  - Number of friends, relationships between those friends
- Closely related topic of academic inquiry
  - Diffusion of innovation – patterns of adoption

Studied Two Kinds of Groups

- Groups on LiveJournal web site
  - About 250K groups, 2M "active" users of 12M
  - People join groups to participate and to gain more ready access to content
  - People explicitly declare friendship with others
    - About 80% reciprocal
- Computer Science conferences (DBLP)
  - About 100 conferences, 100K authors
  - Authors "join" by publishing paper in conference
  - Friend relationships based on previous co-authorship

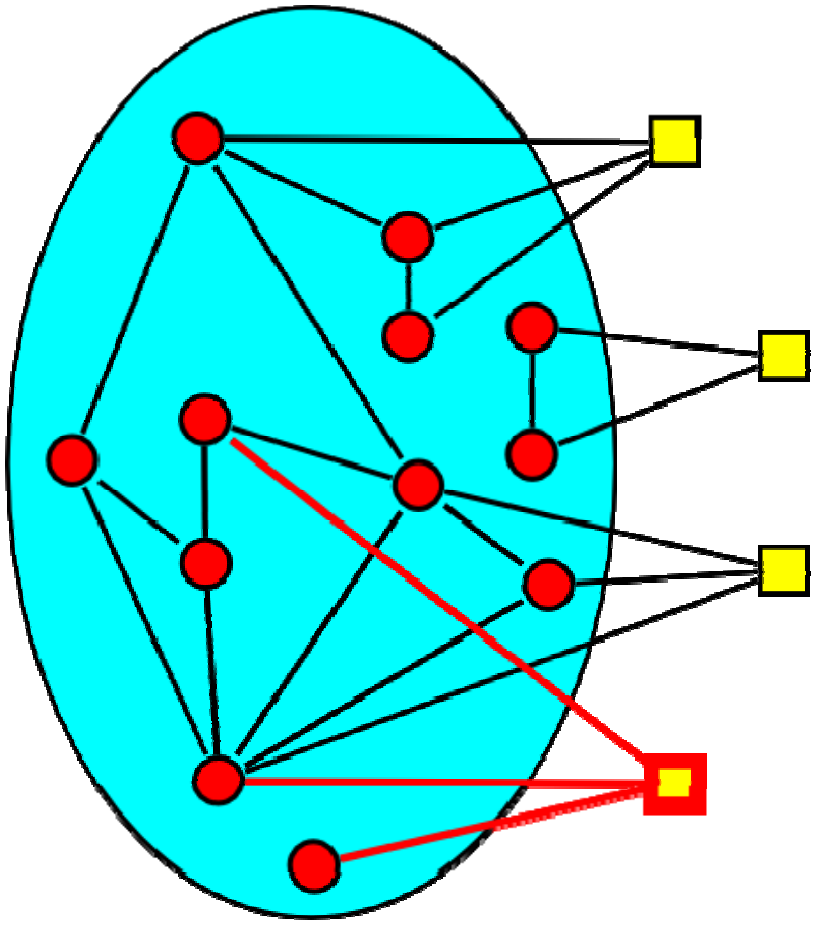
Joining a group

- Model as diffusion of innovation
  - Analogous to adopting new technology, acquiring disease
  - In theoretical models, probability, *p*, depends on *k*, number of friends
  - Large datasets allow comparison of theoretical models
  - Other structural factors also important



Example: Joining a group

- Who most likely to join?
- Red circles represent those in group, yellow squares might join
  - 3 friends vs. 2 friends
- Other structural features
  - E.g. how connected are your friends?
    - 3 friends, 3 pairs
    - 3 friends, no pairs

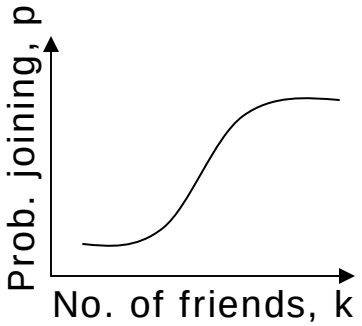


Probability of Joining

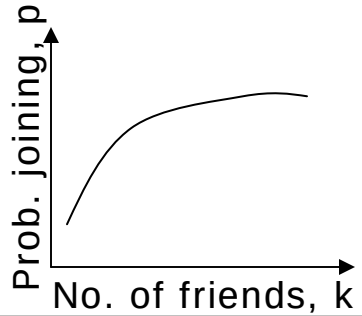
- Probability of joining as function of number of friends already in group – p vs. k
  - For each person who joined, k is how many friends in group at time that they joined
  - Probability p is fraction of people who joined (for each number of friends)
- Similar or different for online vs. offline?
  - Online groups are often casual in nature, join with a mouse click
  - Conference papers require expertise and strict review process to be accepted in group

P vs. k Curve

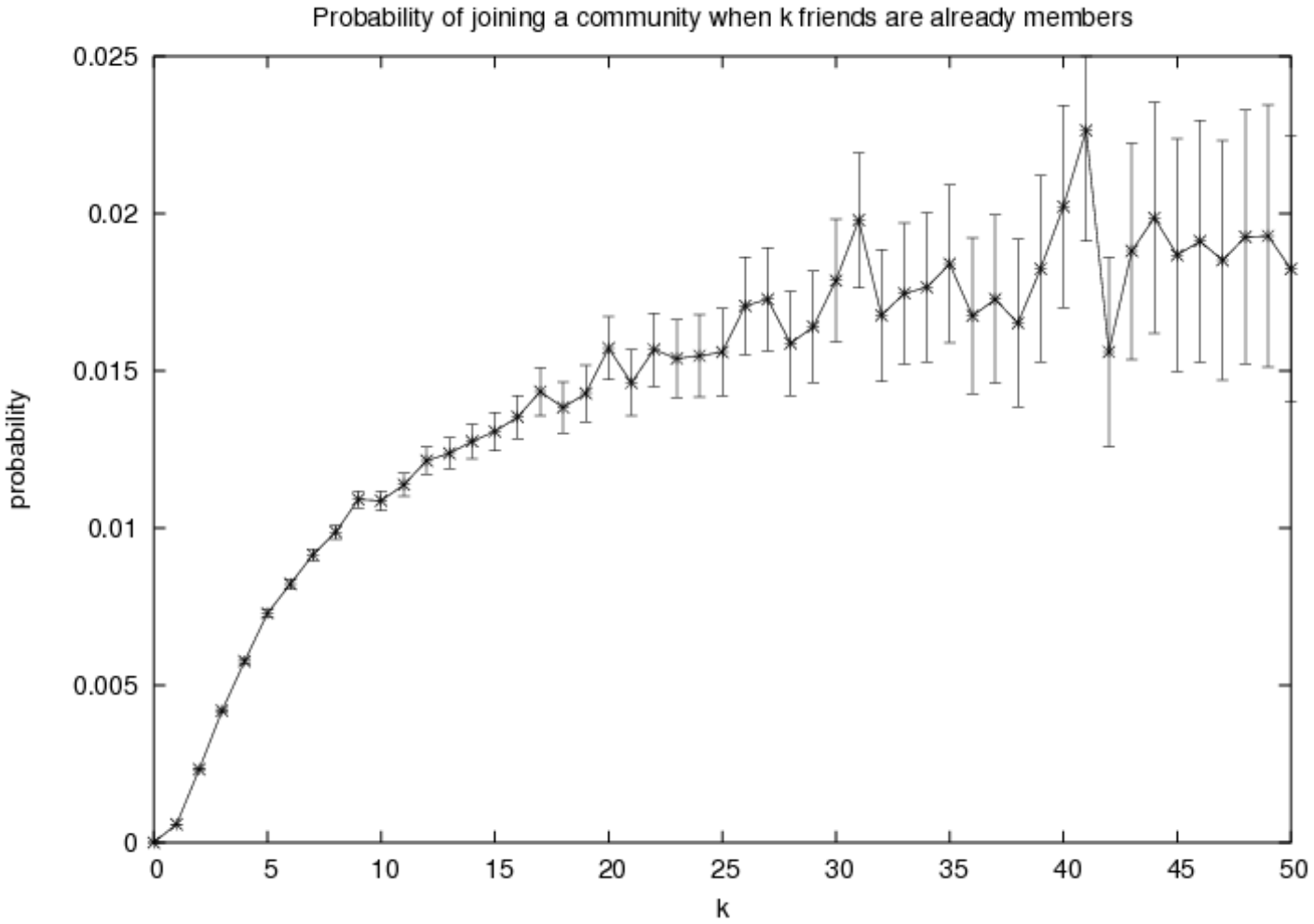
- Shape of these curves?
  - S-shaped? Critical mass effect
    - Time-based adoption curves tend to be this shape



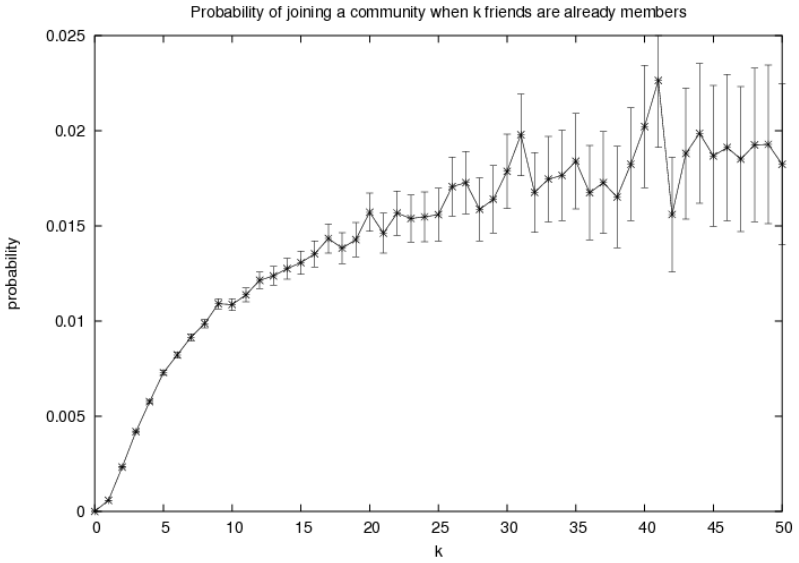
  - Convex? Diminishing returns

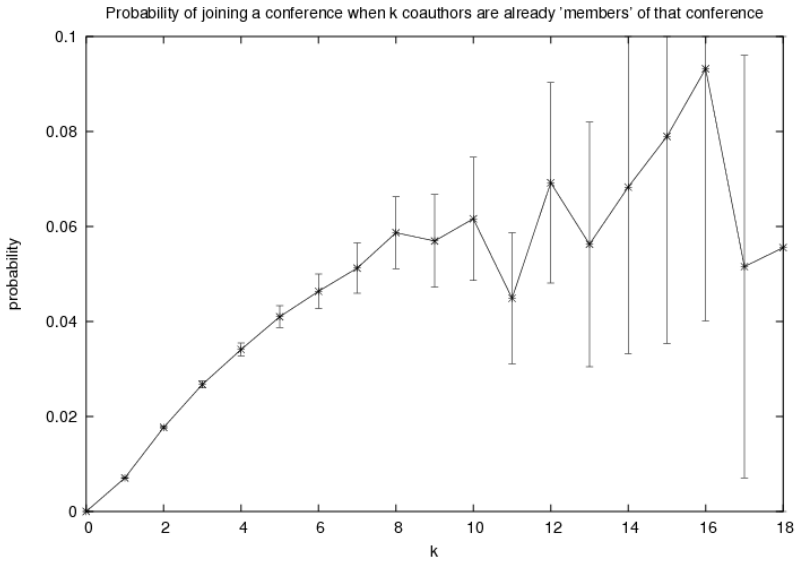


P vs. k for LiveJournal



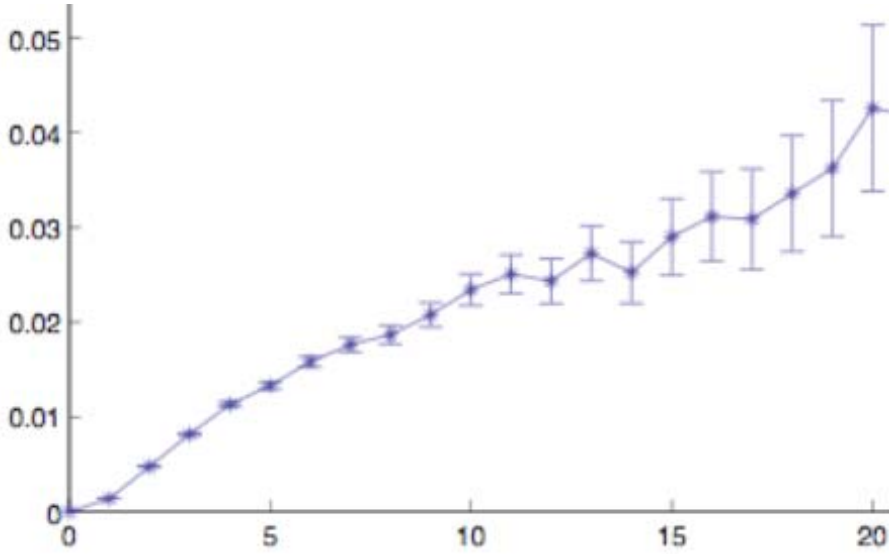
LiveJournal vs. Conferences

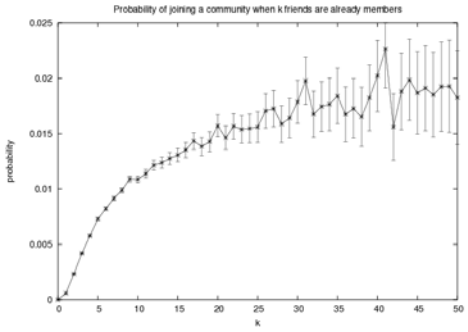




- Similar shape for both
  - Billions of data points, smoother curve
- Supra-linear for 0-1-2, then diminishing returns even for large number of friends
  - Information effect, but not “critical mass”

Wikipedia Page Editing

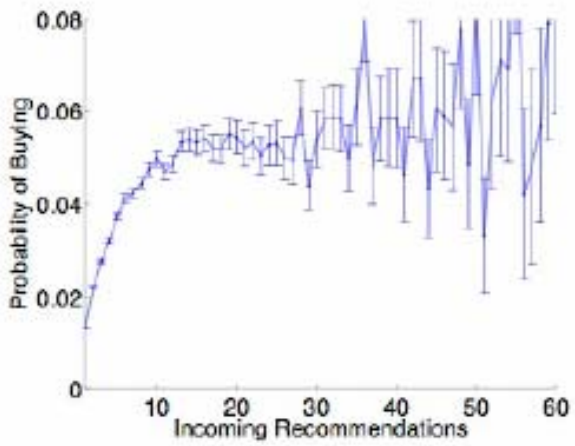




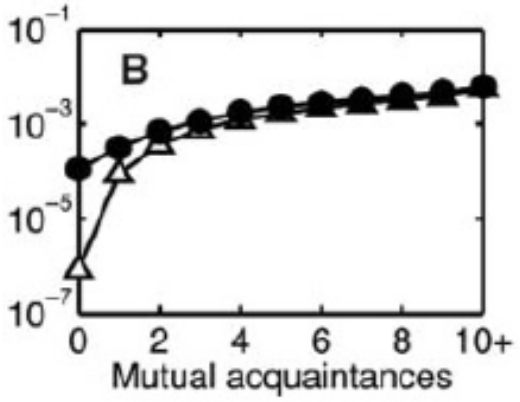
- "group" = editors of an article
- "friendship" = posting on user talk page
- Intermediate effort level
- Still has 0-1-2 "information" effect

Recent Related Work

- Buying DVD from large online retailer
  - As function of recommendations [Leskovec et al '06]



- First engaging in email correspondence
  - As function of friends in common [Kossinets-Watts '06]



Connectivity Among Friends

- How does degree of connectivity among friends affect propensity to join?
  - More connected friends have more "social capital", increasing probability of joining [Coleman88]
  - Less connected friends provide more independent sources of information, increasing probability of joining [Granovetter73]
- Only found reliable effect for LiveJournal, CS conferences not enough data
  - Even with 100K authors

Connectivity Among Friends

- For LiveJournal groups
  - More highly connected friends in group increased probability of joining
  - Consistent with social capital rather than independent sources model
- Measuring connectedness
  - Proportion of friends in group who themselves are friends
- Probability of joining increases as connectedness proportion increases
  - Control for effect of number of friends

Evolution of Groups - Growth

- How does group growth relate to structural properties such as clustering?

  - Groups with high clustering tend to grow slower
    - Clustering defined as # triangles/#open triads

  - Yet just saw individuals more likely to join when friends know one another
  - Discover other properties by learning predictors

Changes in a Community

- Topics of discussion and interest change over time, as does membership
- How are changes in membership related to changes in topic?
  - Do newly shared interests attract members?
  - Do new members bring new interests to a group?
- Things always changing, challenge to study relations between types of change
  - New methodology based on “bursts”

Preliminary Studies of Group Change

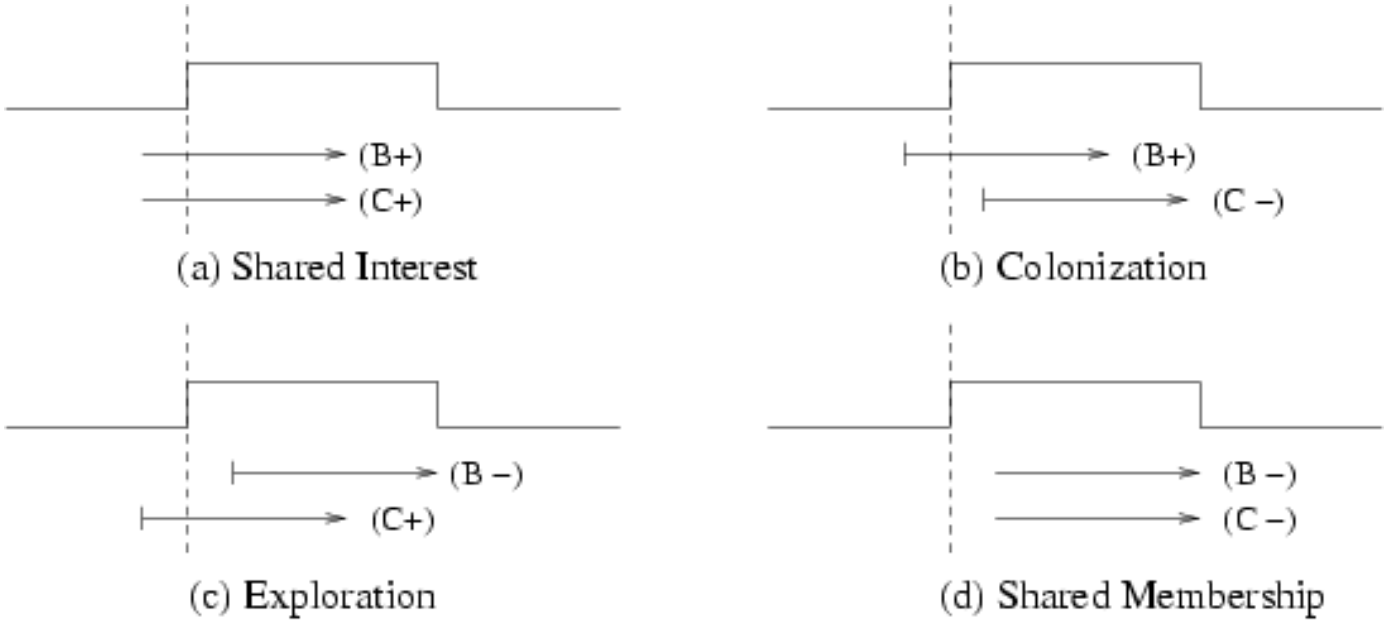
- Use paper titles to indicate topics
  - Bypass need for more sophisticated natural language processing
- Hot topics detected by bursts of title terms
  - E.g., Google's "In the News"

- How do hot topics in two conferences related to people moving between them?
  - Relative timing of bursts
- Topics draw people or vice versa?

Author or Topic Movement First?

- Define hot topic in terms of burst in paper title words
- Define big motion in members in terms of burst in first-time authors in conference
- Interactions between the two
  - Are movers hot?
  - Do hot topics inspire movement?
  - Does movement bring hot topics with it?
  - Does movement occur after a topic is hot everywhere?

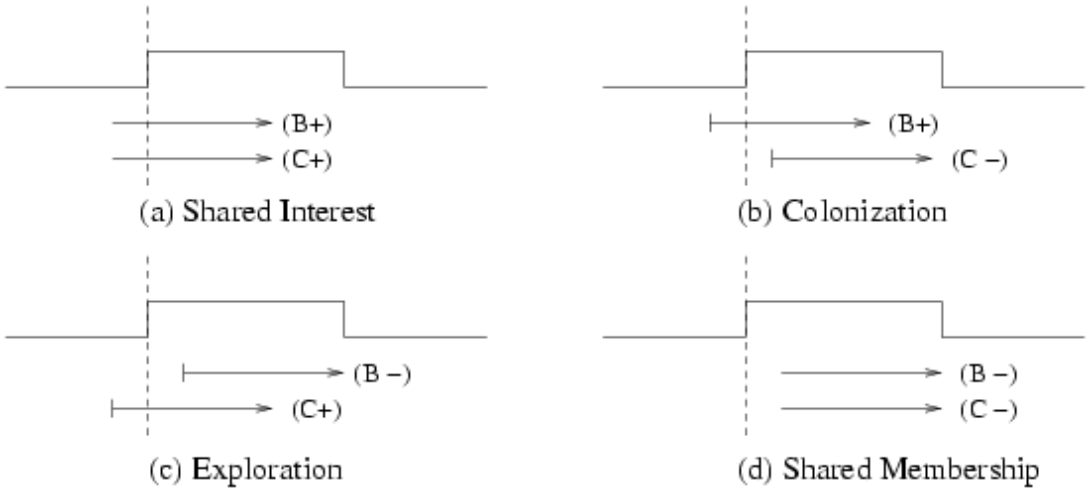
Topic diffusion Among Conferences



- Shared Interest: Hot topic, then movement
- Colonization: Hot at B, movement to C, hot at C
- Exploration: Hot at C, movement to C, hot at B
- Shared membership: movement, then hot topic

Topic diffusion Among Conferences

	C+	C-
B+	(a) 0.602	(b) 0.099
B-	(c) 0.109	(d) 0.189



- Dominant pattern shared interest
  - Topic becomes hot in both conferences
  - Then movement occurs

Wrapping Up

- Social software on web is a mass phenomenon
  - Hundreds of millions of users, among the most active sites
  - Particular sites changing fairly rapidly
- Many parallels between online world and physical world
  - Studies suggesting similarity of behaviors
    - Diminishing returns
    - 0-1-2 effect: 1 an accident 2 a pattern
  - But certainly differences as well