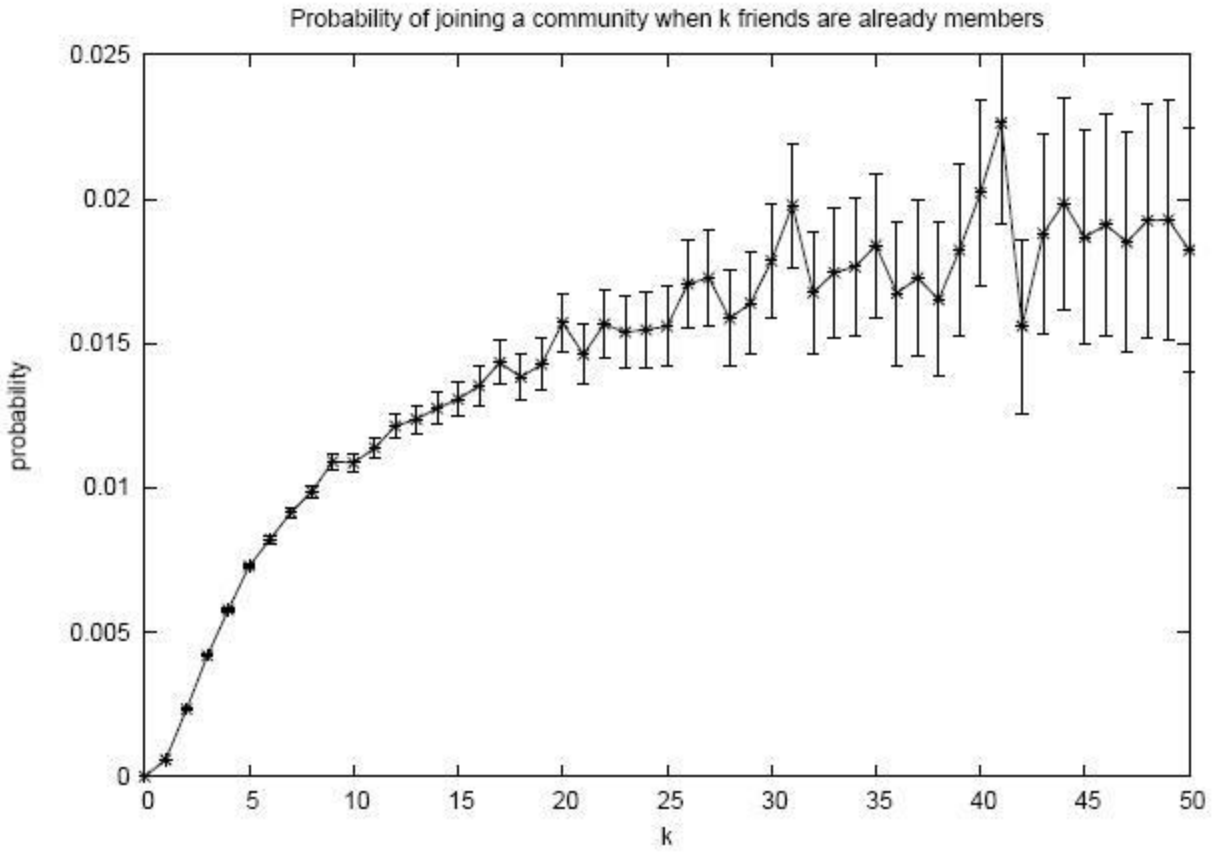
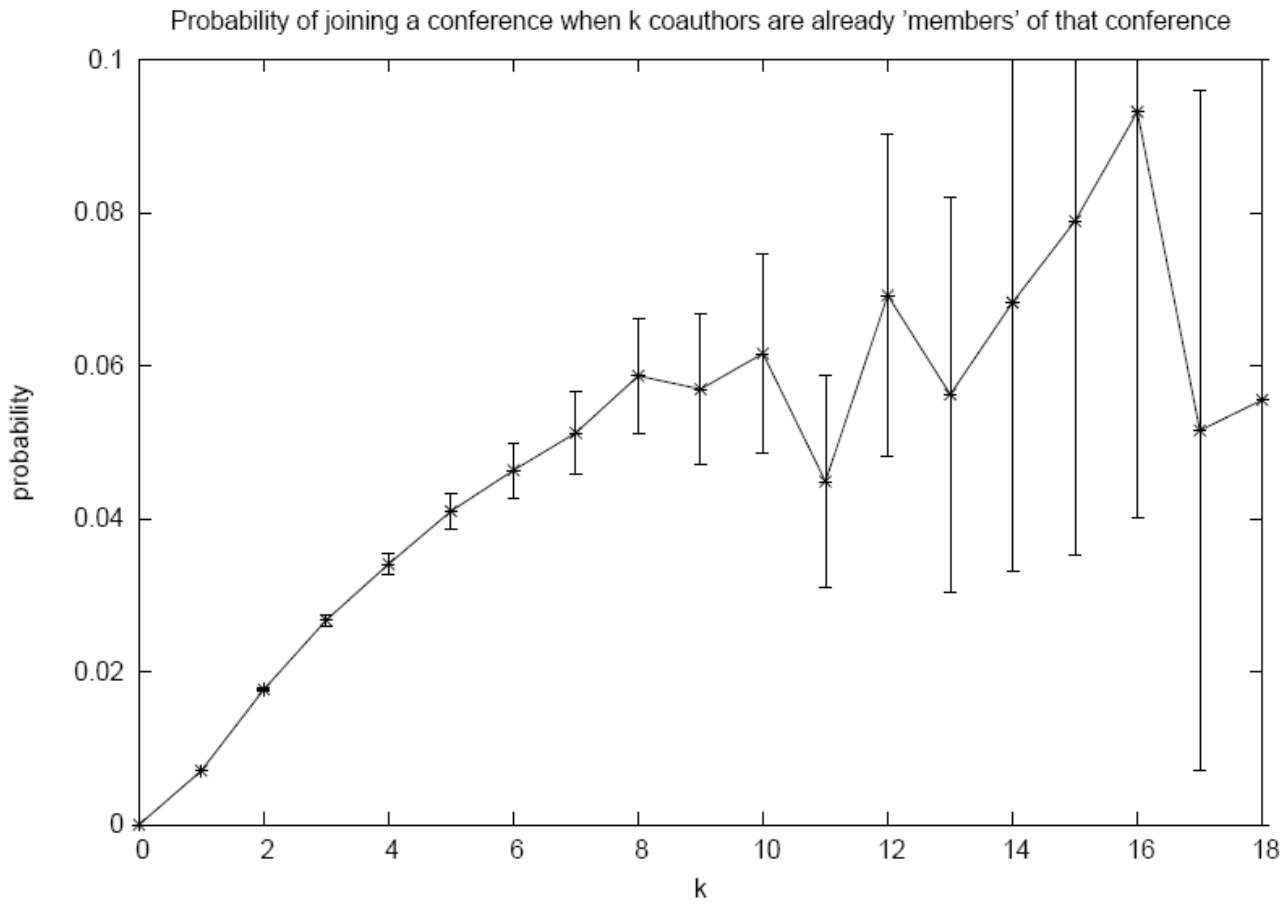
Social correlation

- How similar is the behavior of connected users.
- Previous studies:
  - Joining LiveJournal communities [Backstrom et al.]
  - Publishing in conferences [Backstrom et al.]
  - Tagging vocabulary on flickr [Marlow et al.]
  - Adoption of paid VoIP service in IM
  - Offline: Smoking habits of teenagers
  - …

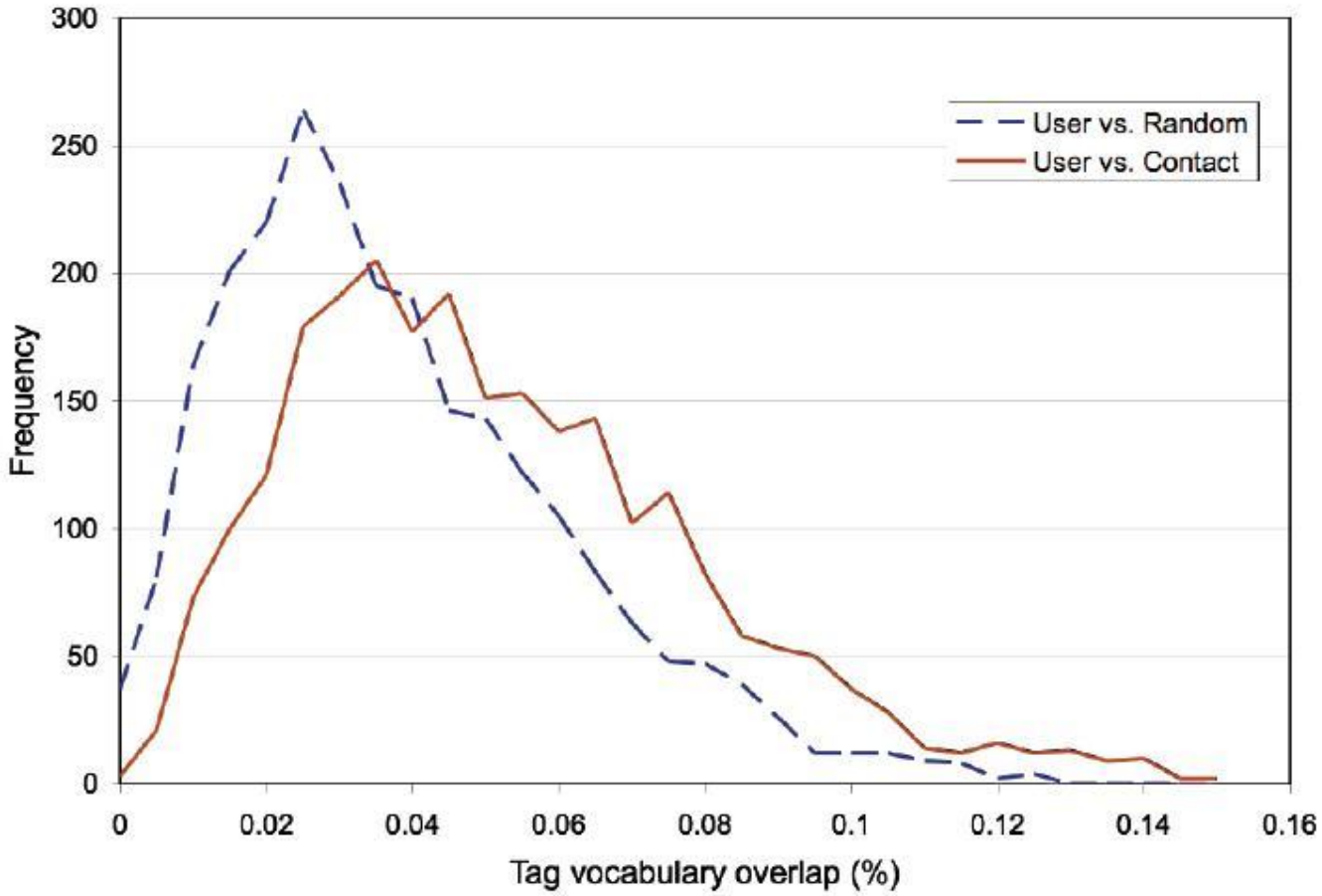
Joining communities [Backstrom et al]



Publishing in conferences



Flickr tag vocabulary [Marlow et al.]



Sources of correlation

- Social influence: One person performing an action can cause her contacts to do the same.
  - by providing information
  - by increasing the value of the action to them
- Homophily: Similar individuals are more likely to become friends
  - Example: two mathematicians are more likely to become friends
- Confounding factors: External influence from elements in the environment
  - Example: friends are more likely to live in the same area, thus attend and take pictures of similar events, and tag them with similar tags

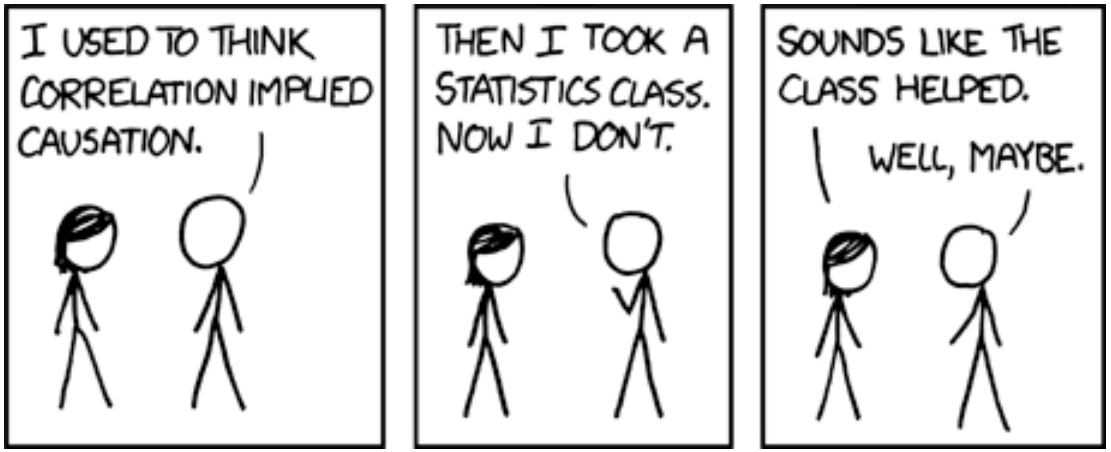
Social influence

- Focus on a particular "action" A.
  - E.g.: buying a product, joining a community, publishing in a conference, using a particular tag, using the VoIP service, …
- An agent who performs A is called "active"
- x has influence over y if x performing A increases the likelihood that y performs A.

- Distinguishing factor: causality relationship

Causation vs. Correlation

- What we try to do is essentially distinguish causation from correlation.
- Common mistake, especially by journalists:
  - People who drink more coffee live longer
  - People who drive red cars create more accidents
  - Eating pizza "cuts cancer risk“
  - People who go to school, live longer

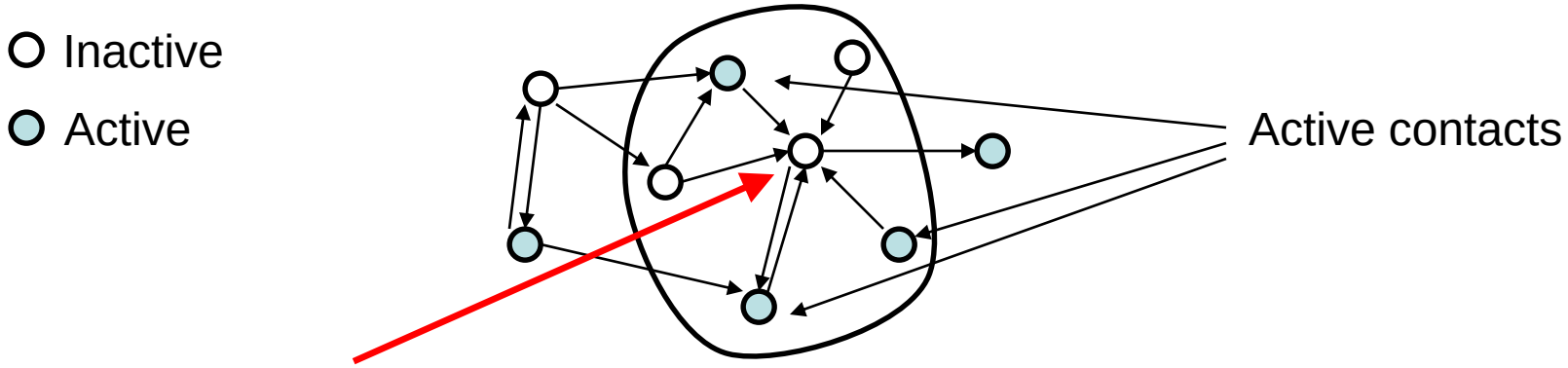


Identifying social influence

- Why is it important?
- Analysis: predicting the dynamics of the system. Whether a new norm of behavior, technology, or idea can diffuse like an epidemic
- Design: for designing a system to induce a particular behavior, e.g.:
  - vaccination strategies (random, targeting a demographic group, random acquaintances, etc.)
  - viral marketing campaigns

Influence Model

- Graph (static or dynamic)
- Edge (u,v): Node u can influence node v
- Discrete time: t = 0, 1, 2, …, T
- For each t, every inactive node becomes active with probability p(a), where a is the # active contacts



Model – Influence probability

- Natural choice for p(a): logistic regression function:

$$\ln\left(\frac{p(a)}{1-p(a)}\right) = \alpha \ln(a+1) + \beta$$

with ln(a+1) as the explanatory variable. I.e.,

$$p(a) = \frac{e^{\alpha \ln(a+1)+\beta}}{1+e^{\alpha \ln(a+1)+\beta}}$$

- Coefficient α measures social correlation.

Measuring social correlation

- Given data, we compute the maximum likelihood estimate for parameters α and β.

- Let Y_a = # pairs (user u, time t) where u is not active and has a active friends at the beginning of time step t, and becomes active in this step.
- Let N_a = …… does not become active in this step.

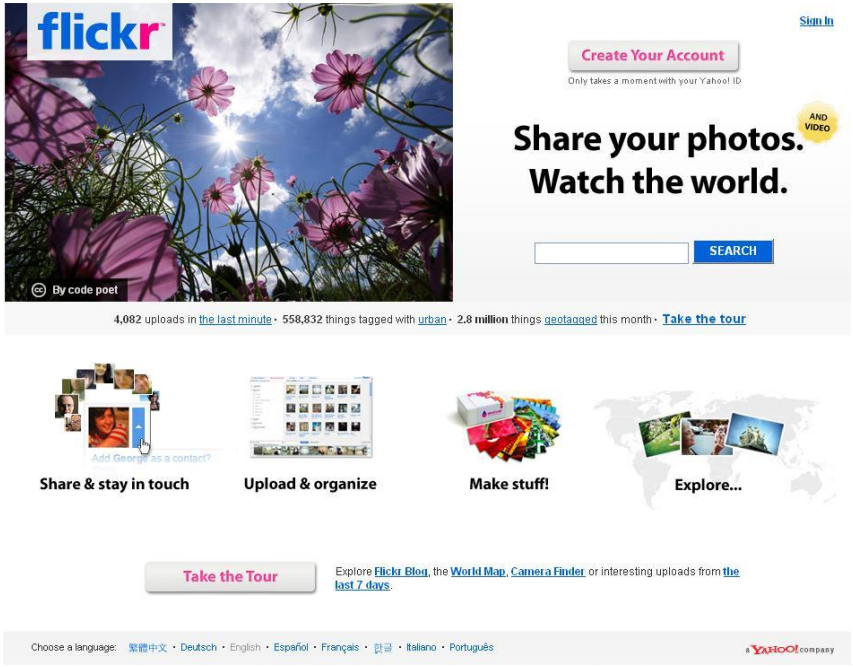
- Find α, β to maximize the likelihood function:

$$f(\alpha, \beta, \mathbf{Y_a}, \mathbf{N_a}) = \prod_a p(a)^{Y_a}(1 - p(a))^{N_a}$$

- For convenience, we cap a at a value R.

Flickr data set

- Photo sharing website
- 16 month period
- Growing # of users, final number ~800K
- ~340K users who have used the tagging feature
- Social network:
  - Users can specify “contacts”.
  - 2.8M directed edges, 28.5% of edges not mutual.
  - Size of giant component ~160K



mmahdian's photostream pro

portrait

Uploaded on Apr 7, 2008

2 notes / 7 comments

graffiti

"None are more hoplessly enslaved than those who falsely believe they are free." graffiti...

Uploaded on Feb 20, 2008

4 comments

golden gate

this photo was taken by mistake! i took the photo after changing lens, and the lens was...

roja

Uploaded on Dec 3, 2007

2 comments

iran

19 photos

flowers

12 photos

funny pix

4 photos

faves

piazza san marco

ALL SIZES

piazza san marco, venice

This photo has notes. Move your mouse over the photo to see them.

Comments

mac on a mac pro says:

Wonderful!

Posted 7 months ago. (permalink)

~~ Reza ~~ pro says:

A nice action shot!

Posted 7 months ago. (permalink)

Uploaded on November 23, 2007
by mmahdian

mmahdian's photostream

94 uploads

browse

This photo also belongs to:

faves (Set)

17 items

browse

Tags

- venice
- venezia
- italy
- italia
- st mark square
- piazza san marco
- birds
- girl

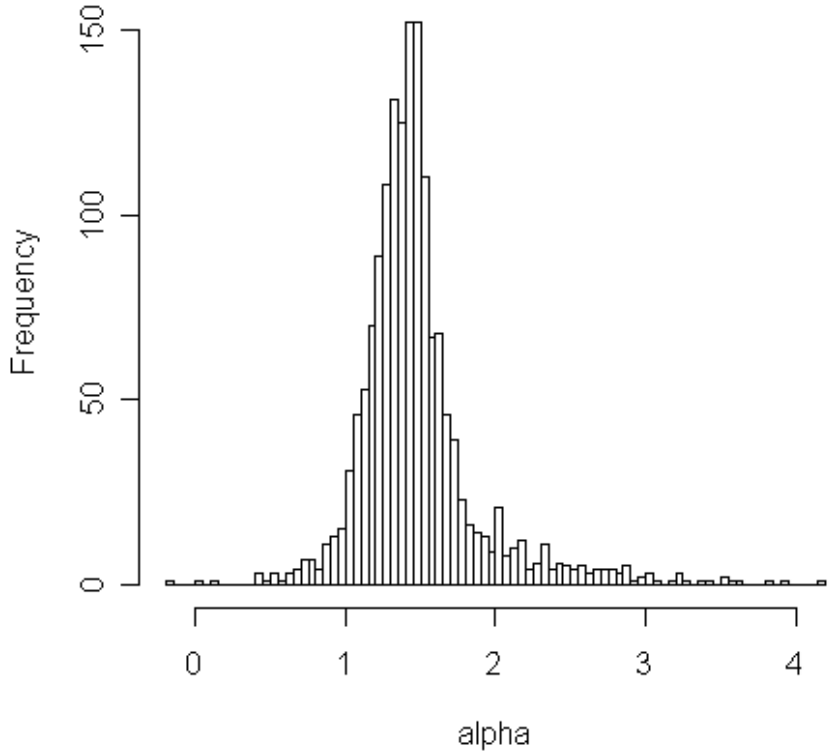
Additional Information

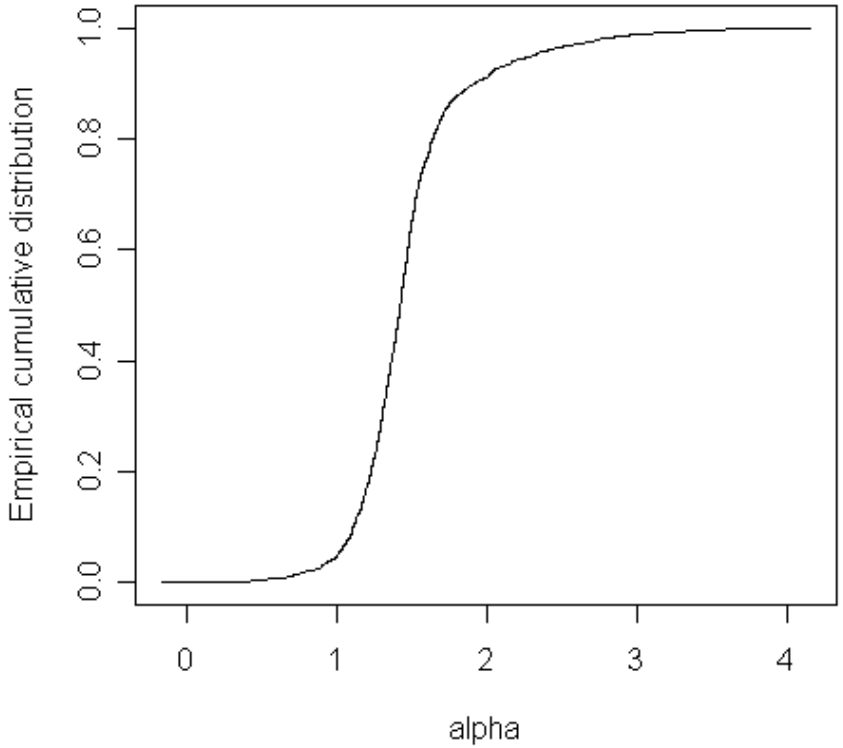
Flickr tags

- ~10K tags
- We focus on a set of 1700
- Different growth patterns:
  - bursty (“halloween” or “katrina”)
  - smooth (“landscape” or “bw”)
  - periodic (“moon”)
- For each tag, define an action corresponding to using the tag for the first time.

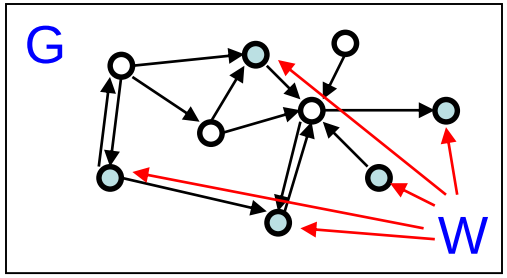
Social correlation in flickr

- Distribution of α values estimated using maximum likelihood:



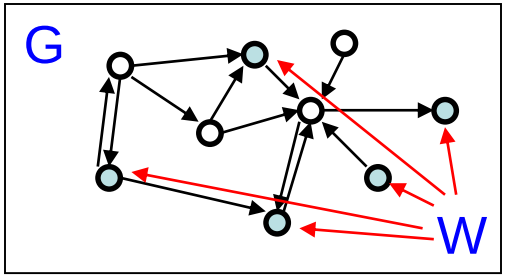


Distinguishing influence



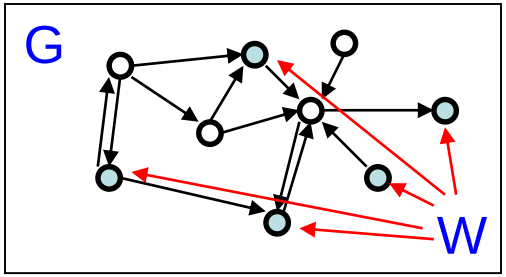
- Recall: graph G, set W of active nodes
- Influence model
  - First G is selected
  - Then W is picked from a distribution depending on G

Distinguishing influence



- Noninfluence models
  - Homophily (Similar individuals are more likely to become friends):
    - First W is picked, then G is picked from a distribution that depends on W
  - Confounding factors (External influence from elements in the environment):
    - Both G and W are picked from distributions that depend on another var X

Distinguishing influence



- Generally, we consider this correlation model:
  - (G,W) are selected from a joint distribution
  - Each agent in W picks an activation time i.i.d. from a distribution on [0,T]

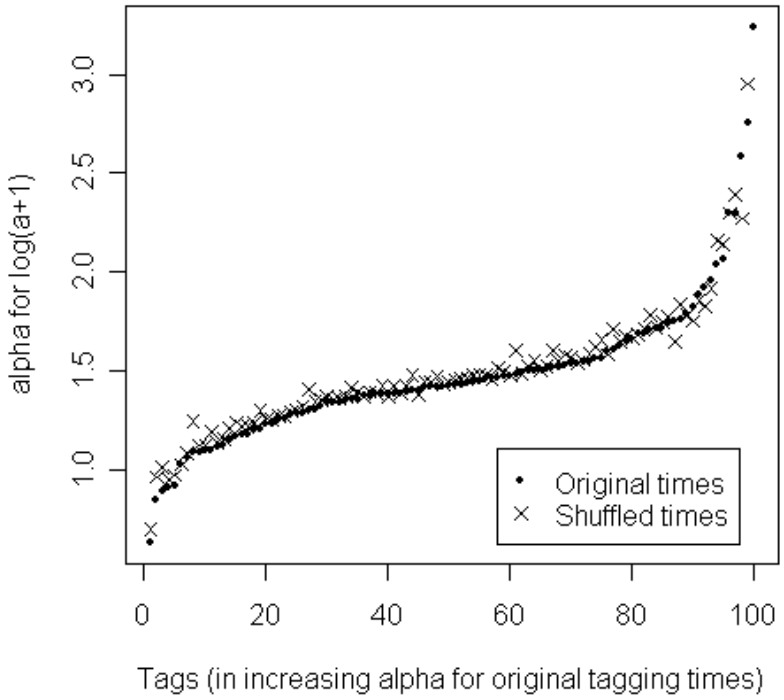
Testing for influence

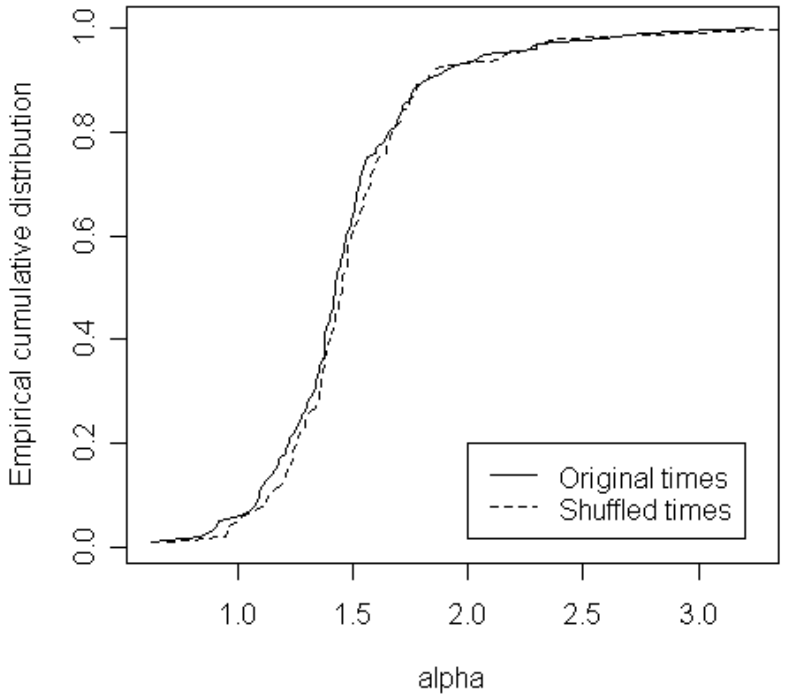
- **Simple idea:** even though an agent's probability of activation can depend on friends, her timing of activation is independent
- Shuffle Test: re-shuffle the time-stamp of all actions, and re-estimate the coefficient α. If different from original α, social influence can't be ruled out.

Testing for influence

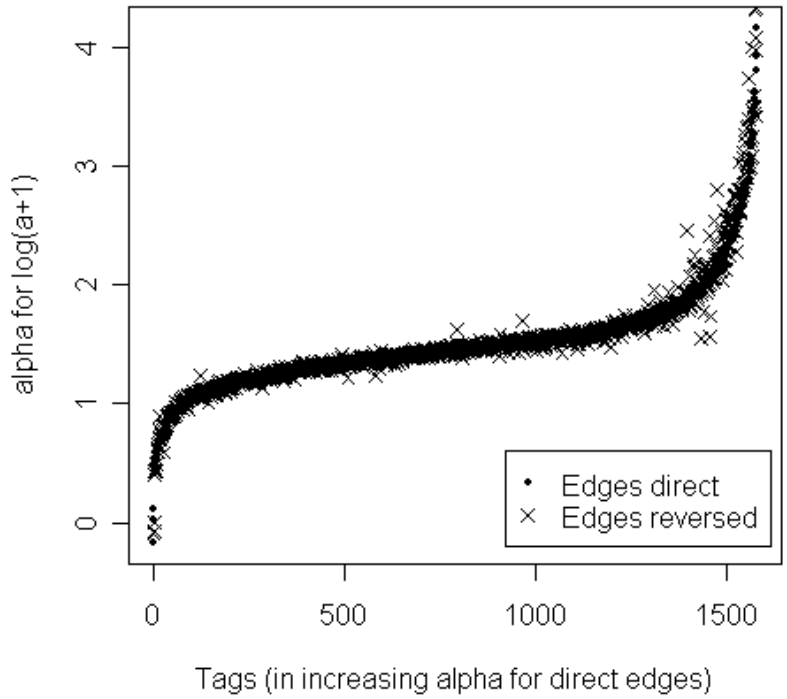
- **Simple idea:** even though an agent's probability of activation can depend on friends, her timing of activation is independent
- Shuffle Test: re-shuffle the time-stamp of all actions, and re-estimate the coefficient α. If different from original α, social influence can't be ruled out.
- Edge-Reversal Test: reverse the direction of all edges, and re-estimate α.

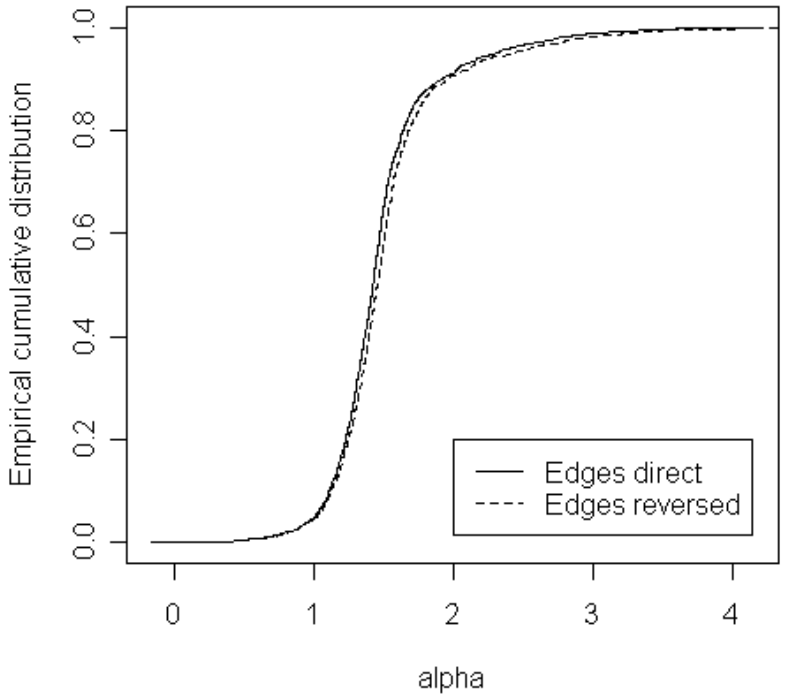
Shuffle test on Flickr data





Edge-reversal test on Flickr data

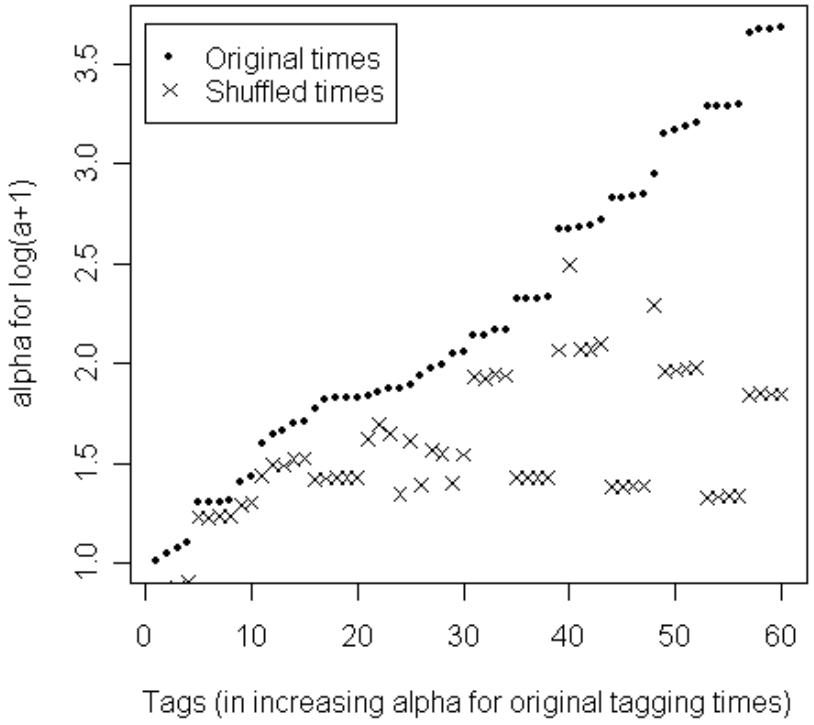


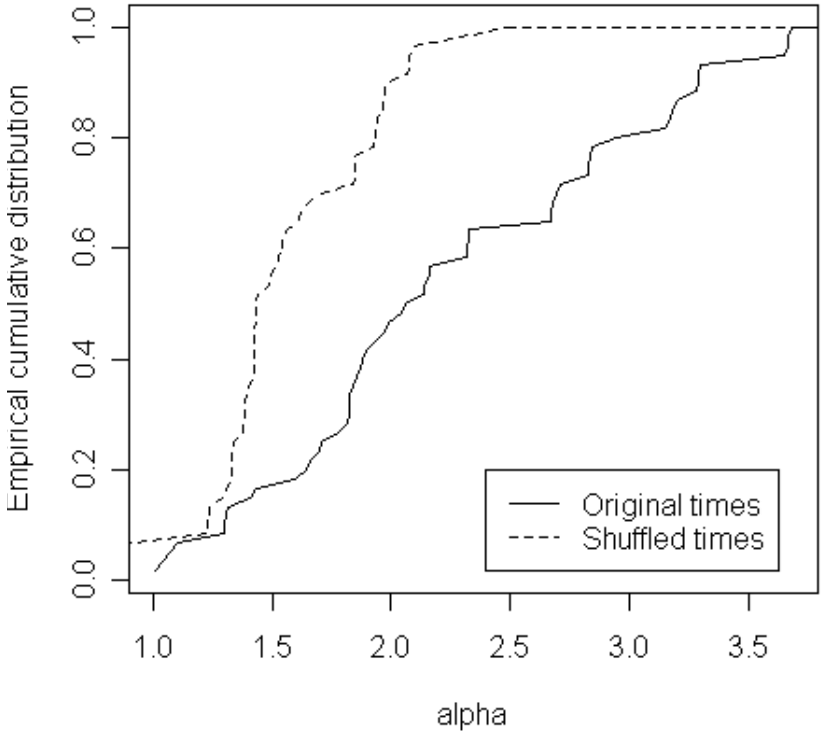


Simulations

- Run the tests on randomly generated action data on flickr network.
- Baseline: no-correlation model, actions generated randomly to follow the pattern of one of the real tags, but ignoring network
- Influence model: same as described, with a variety of (α,β) values
- Correlation model: pick a # of random centers, let W be the union of balls of radius 2 around these centers.

Shuffle test, influence model





Edge-reversal test, influence model