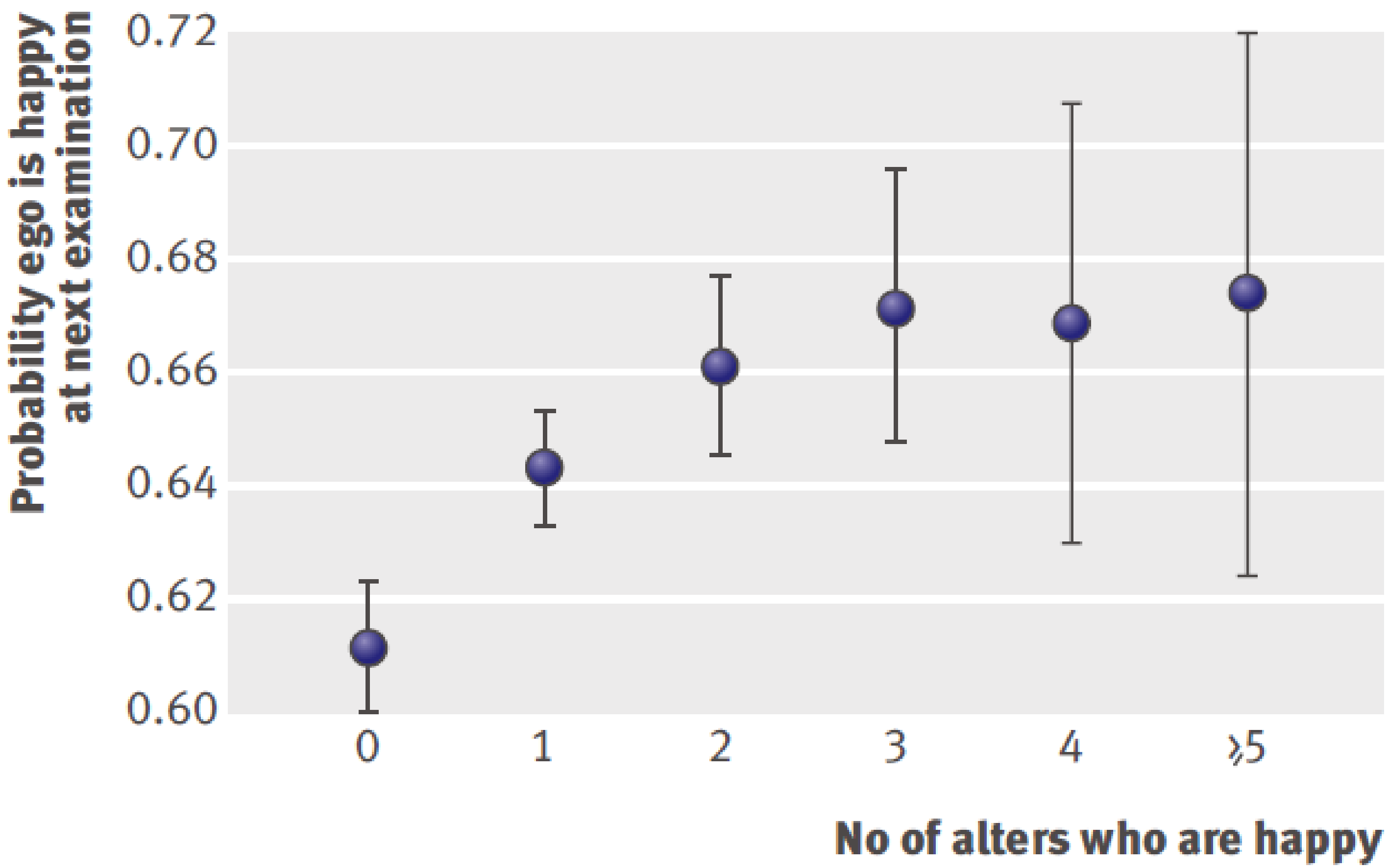
Influence and Correlation in Social Networks

Social Correlation

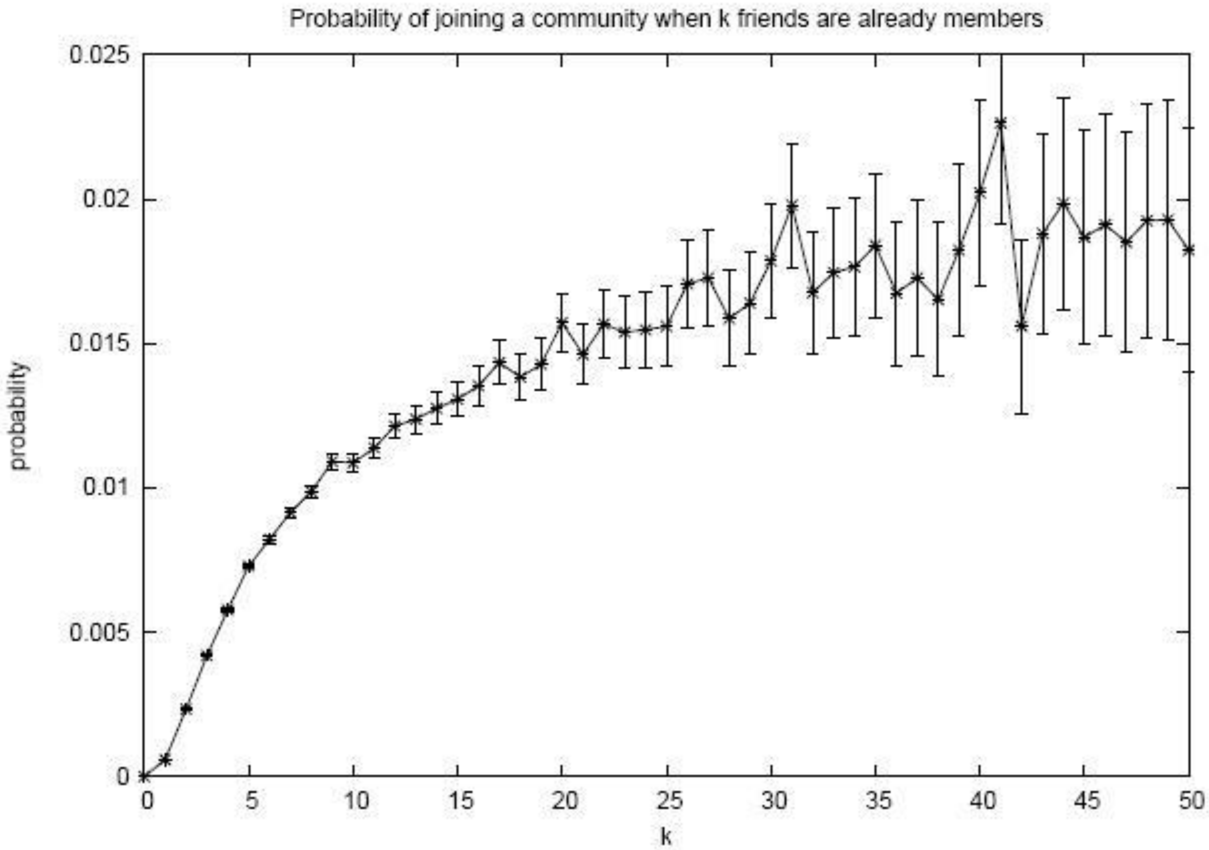
- How similar is the behavior of connected users.

- Previous studies:
  - Offline behavior
    - Fashion
    - Happiness
    - Publishing in conferences [Backstrom et al.]
  - Online behavior
    - Joining online communities [Backstrom et al.]
    - Tagging vocabulary on Flickr [Marlow et al.]
    - Using a VoIP service

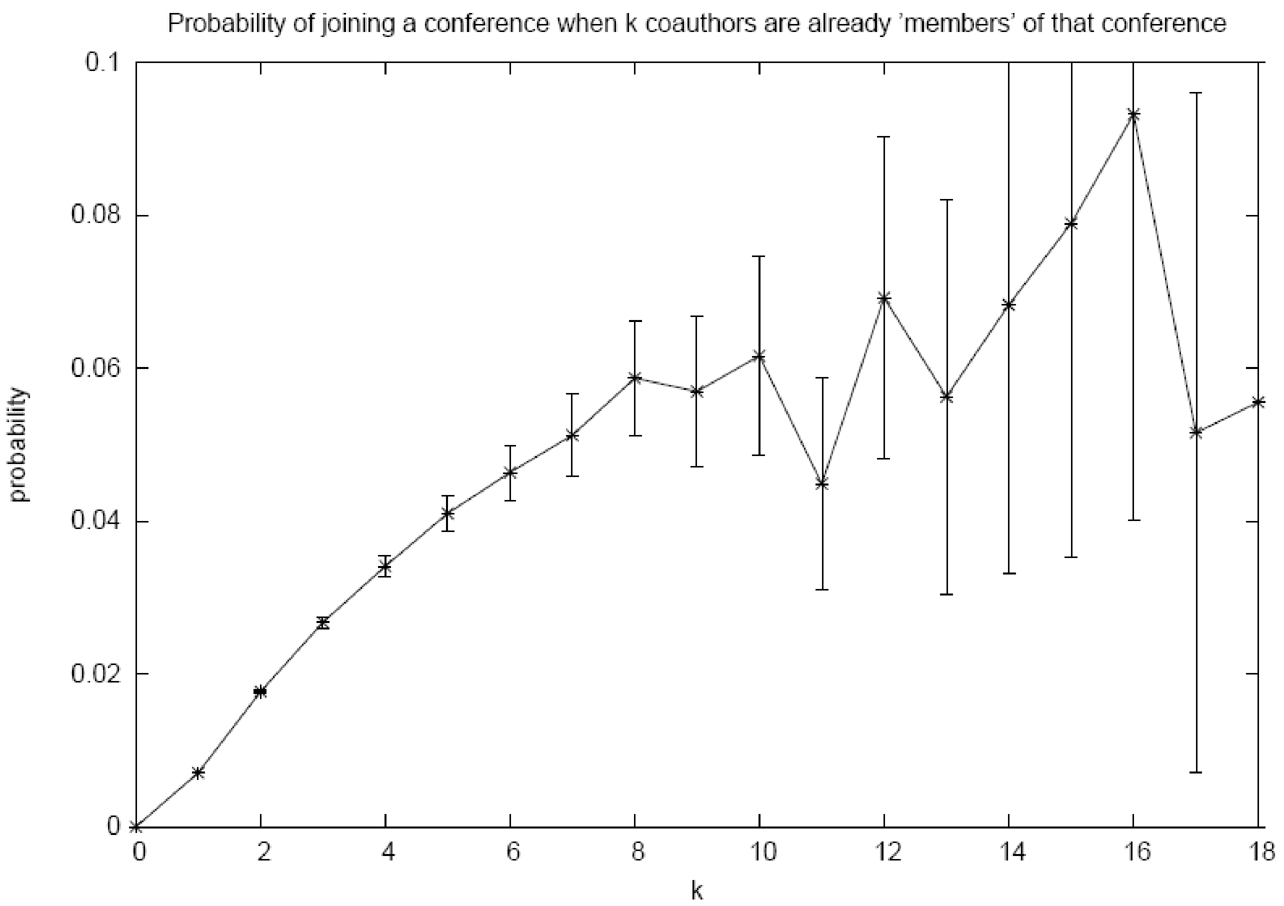
Happiness [Fowler and Christakis]



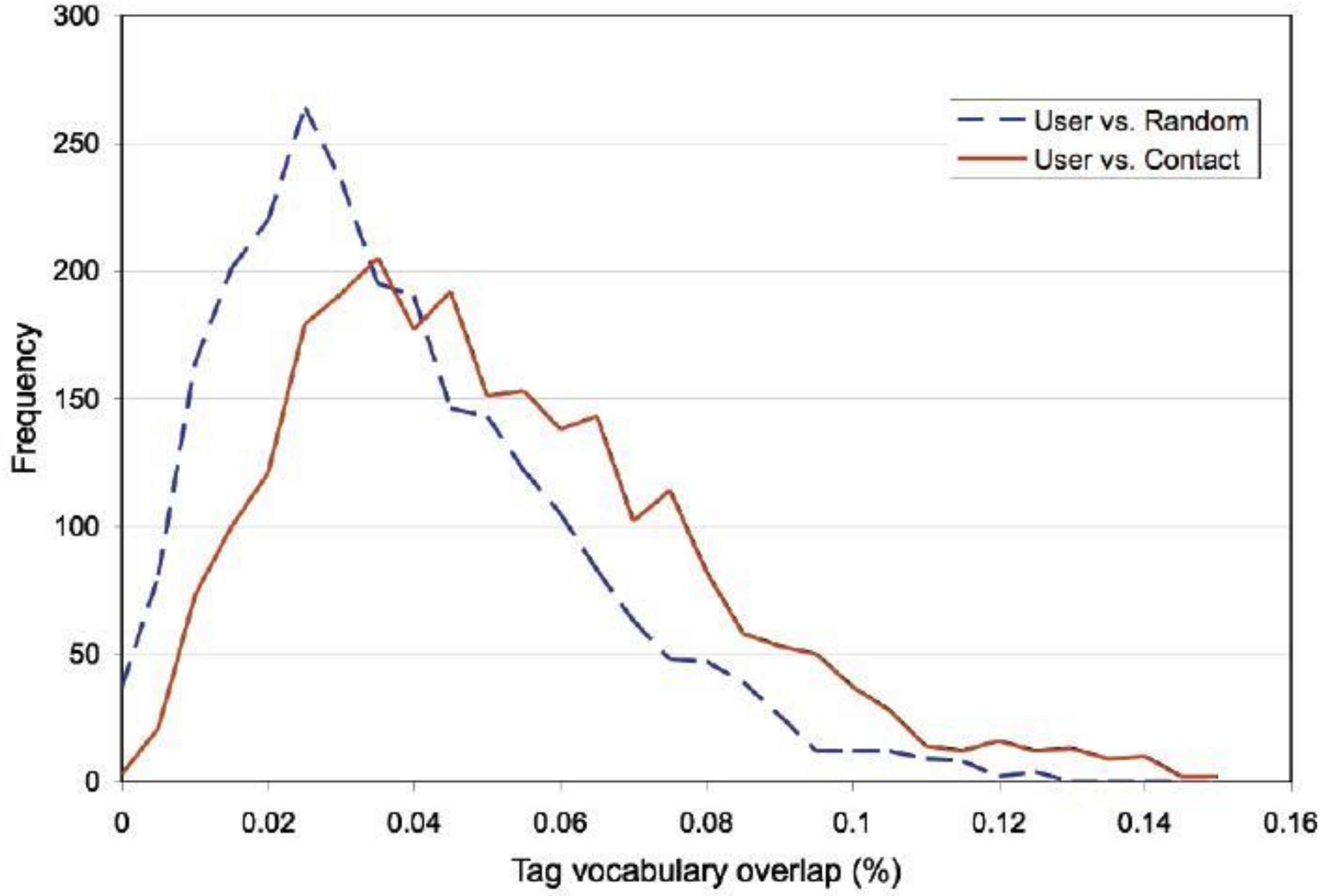
Joining communities [Backstrom et al]

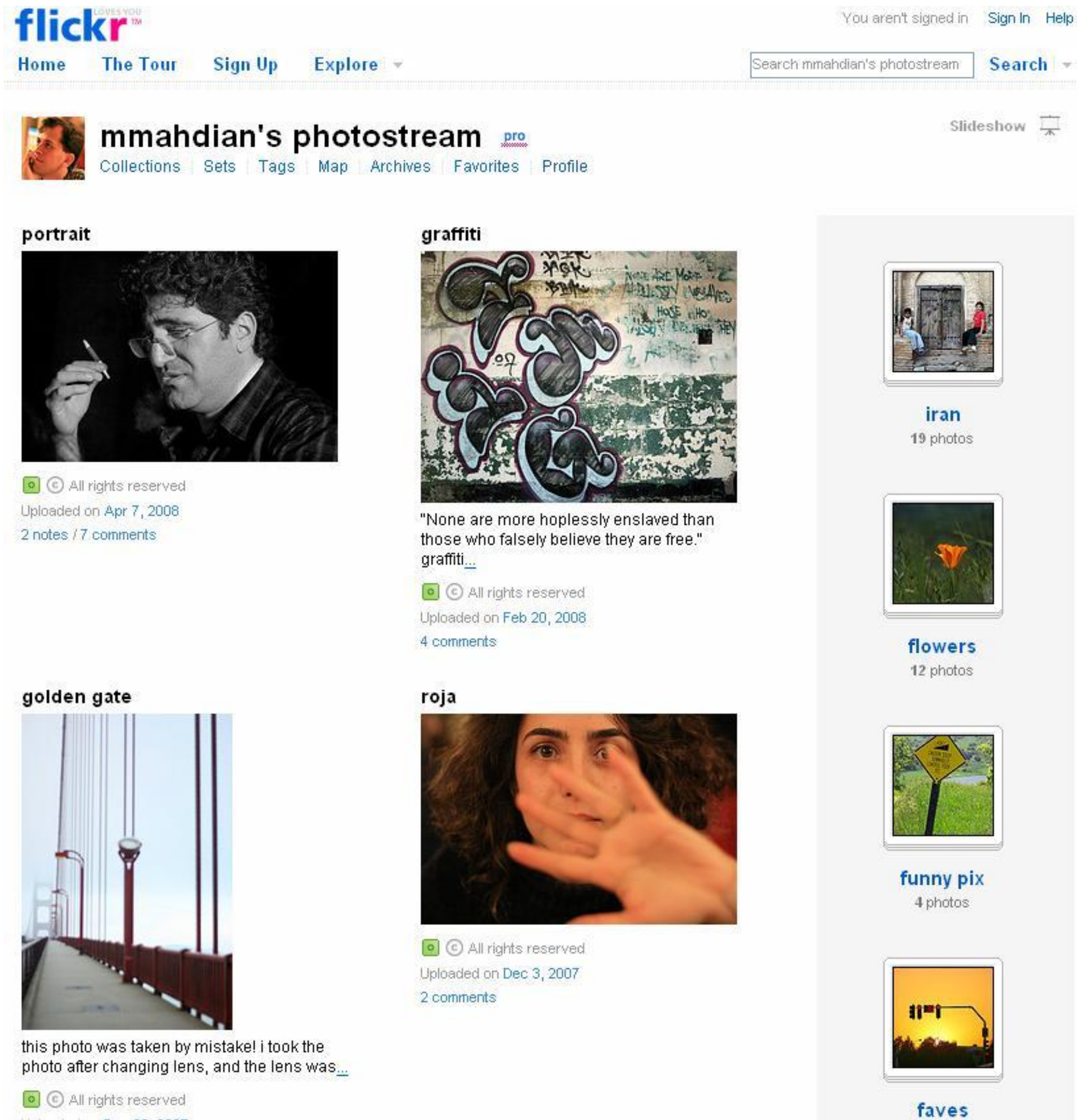


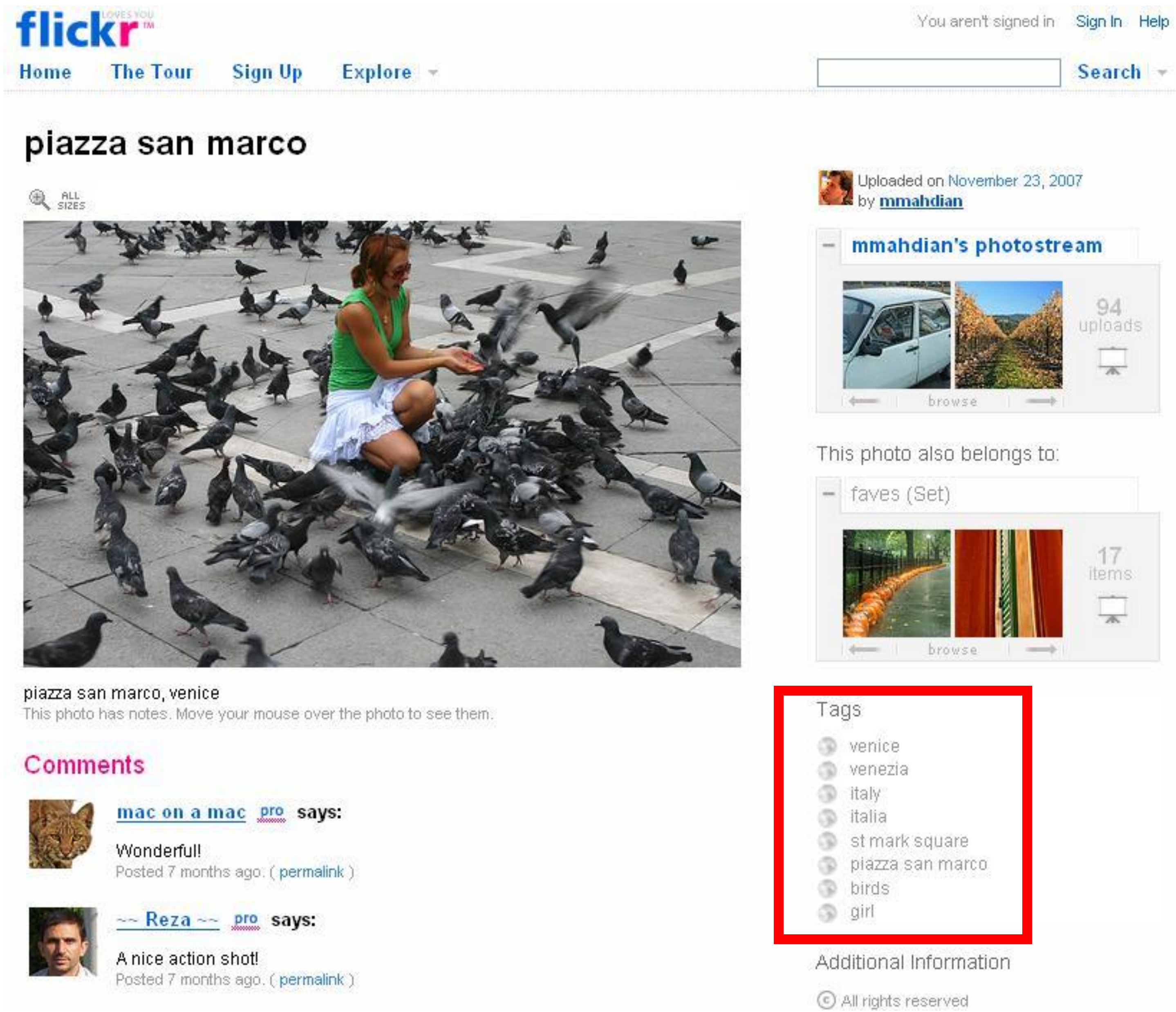
Publishing in conferences



Flickr tag vocabulary [Marlow et al.]




flickr
You aren't signed in
Sign In
Help
Home
The Tour
Sign Up
Explore
Search mmahdian's photostream
Search
mmahdian's photostream
pro
Slideshow
Collections
Sets
Tags
Map
Archives
Favorites
Profile
portrait
All rights reserved
Uploaded on Apr 7, 2008
2 notes / 7 comments
graffiti
"None are more hoplessly enslaved than those who falsely believe they are free." graffiti...
All rights reserved
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...
All rights reserved
roja
All rights reserved
Uploaded on Dec 3, 2007
2 comments
iran
19 photos
flowers
12 photos
funny pix
4 photos
faves


flickr
You aren't signed in Sign In Help
Home The Tour Sign Up Explore
Search
piazza san marco
ALL SIZES
Uploaded on November 23, 2007
by mmahdian
mmahdian's photostream
94 uploads
browse
This photo also belongs to:
faves (Set)
17 items
browse
piazza san marco, venice
This photo has notes. Move your mouse over the photo to see them.
Tags
venice
venezia
italy
italia
st mark square
piazza san marco
birds
girl
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)
Additional Information
All rights reserved

Sources of Correlation

- Social influence (induction):
  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 (selection):
  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

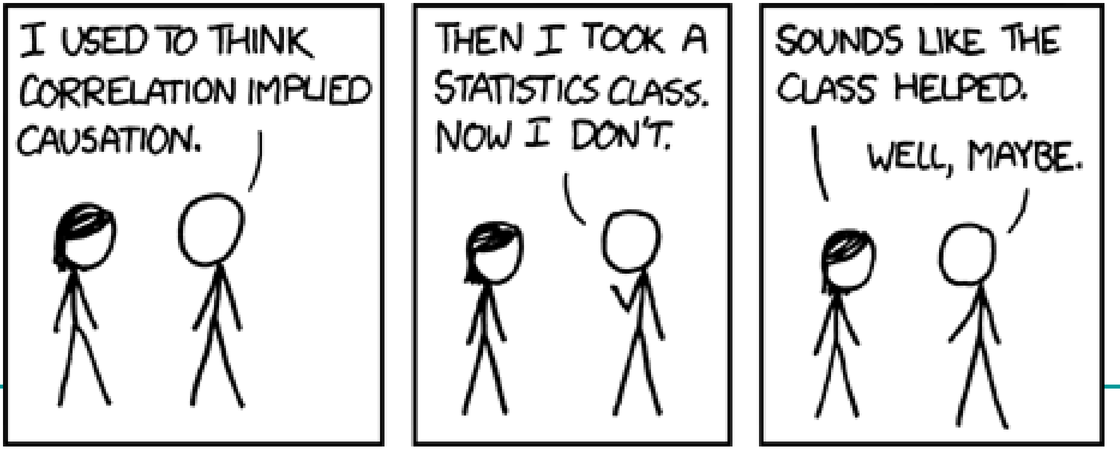
Social Influence

- 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“
  - Ice cream sales and drowning



Identifying social influence

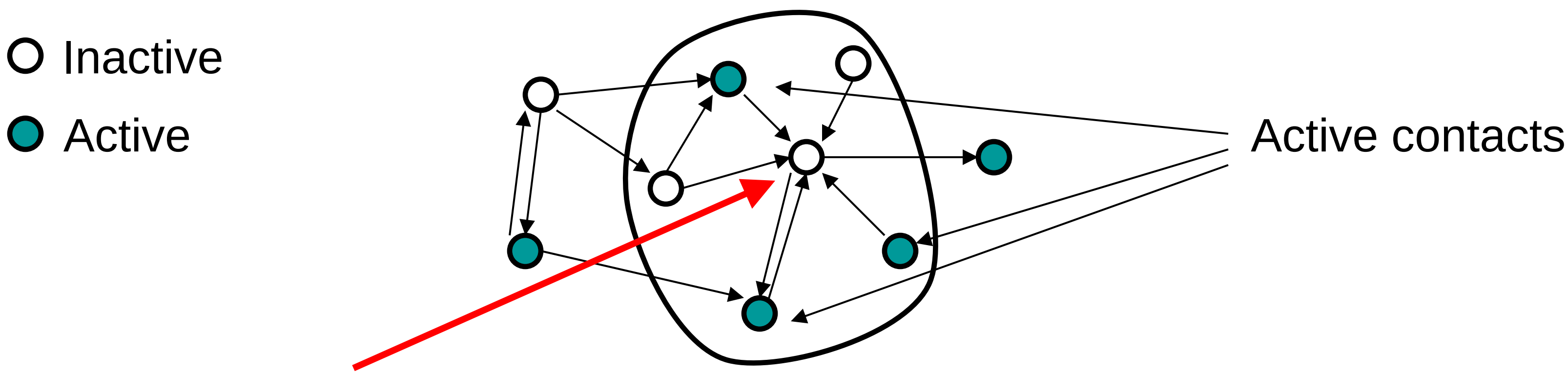
- Why is it important?
- Analysis: predicting the dynamics of the system. Whether a new norm of behavior, technology, or idea can diffuse like an epidemic
- Design: designing a system to induce a particular behavior, e.g.:
  - vaccination strategies (random, targeting a demographic group, random acquaintances,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

Approach

- Measure correlation
- Models for influence and correlation
- Tests for distinguishing influence from correlation
- Theoretical results
- Apply tests on synthetic data
- Apply tests on real data (Flickr)

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(x), where x is the # active contacts



Model – Influence Probability

- Natural choice for $p(x)$: logistic regression function:

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

with x (# active contacts) is the explanatory variable. I.e.,

$$p(x) = \frac{e^{\alpha \cdot x+\beta}}{1+e^{\alpha \cdot x+\beta}}$$

- Given data, can estimate α with Maximum Likelihood
- Coefficient α measures social correlation.

Measuring social correlation

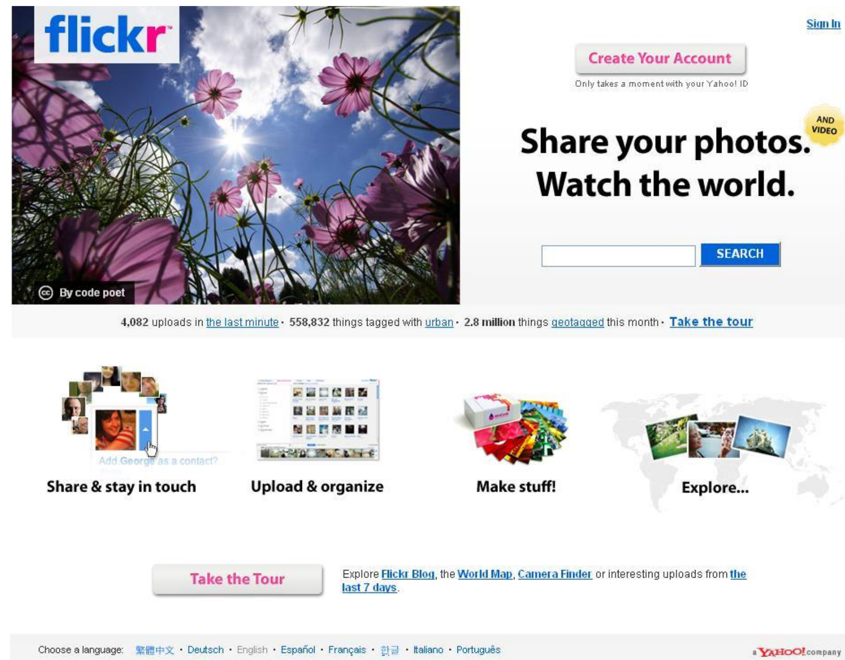
- Given data, we compute the maximum likelihood estimate for parameters α and β.
- Compute values Y_0, N_0, Y_1, N_1, Y_2, N_2, …
  - Y_x = # pairs (user u, time t) where at beginning of time step t, user u is not active and has x active friends and becomes active in this step.
  - N_x = …… does not become active in this step.
- Find α, β to maximize the likelihood function:

$$f(\alpha, \beta, \mathbf{Y_x}, \mathbf{N_x}) = \prod_x p(x)^{Y_x}(1-p(x))^{N_x}$$

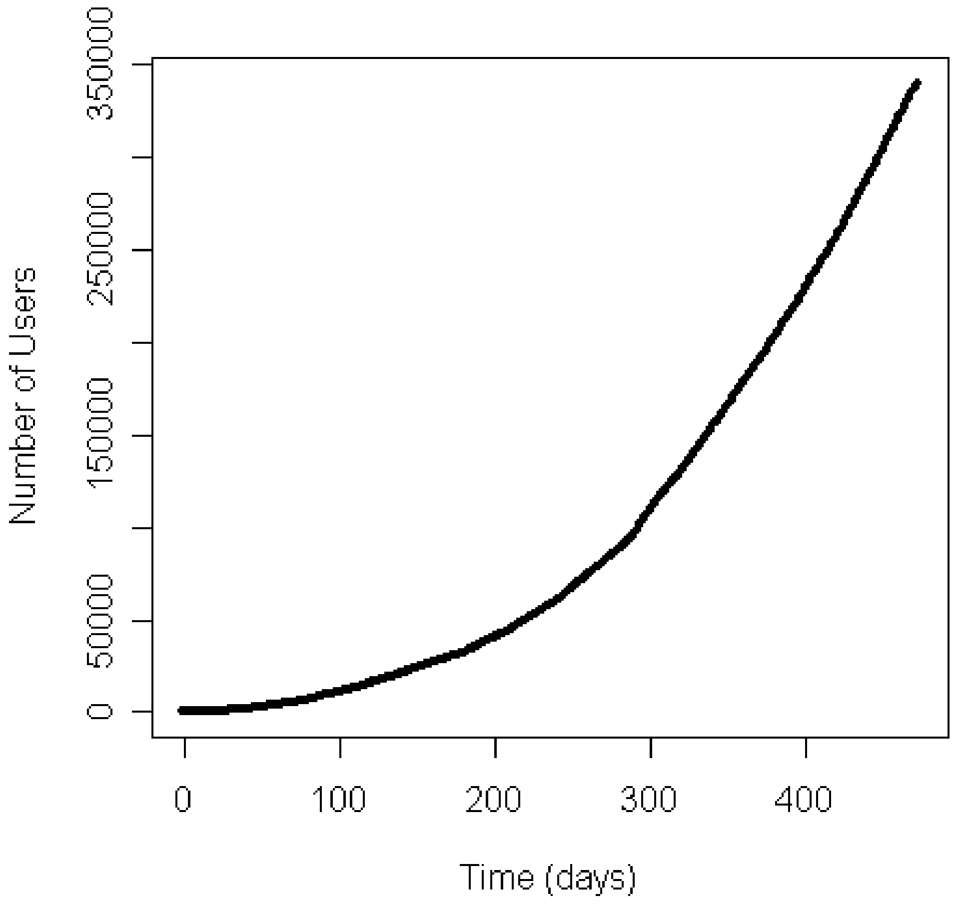
- For convenience, we cap x at a value R.

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.



Flickr data set, growth

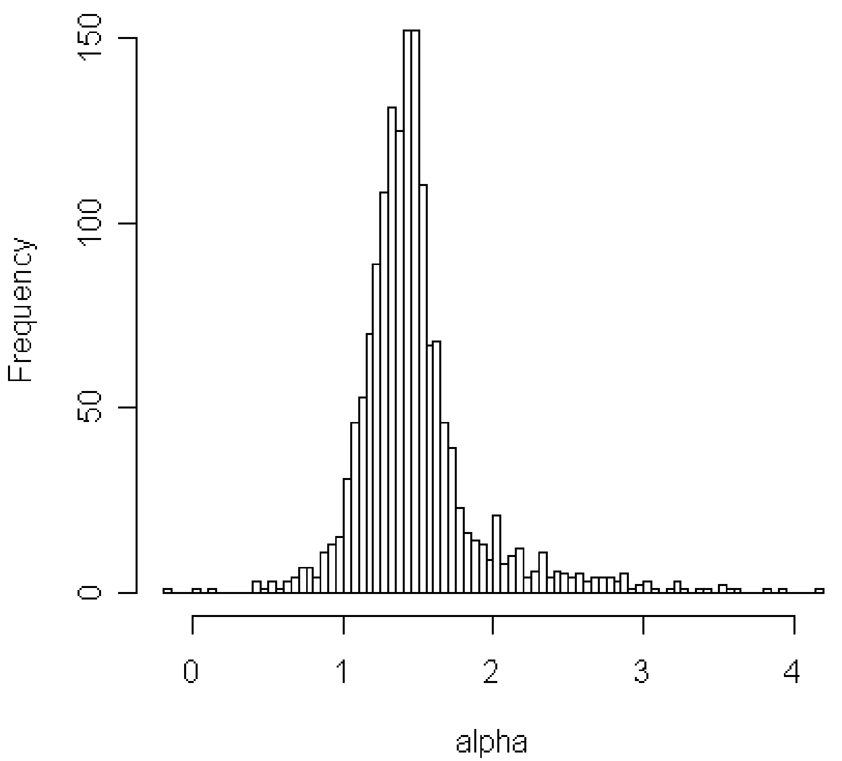


Flickr tags

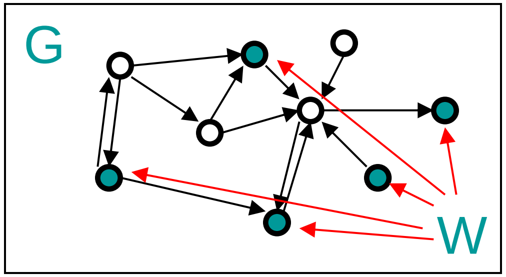
- ~10K tags
- We focus on a set of 1700
- Different growth patterns:
    - bursty (“halloween” or “katrina”)
    - smooth (“landscape” or “bw”)
    - periodic (“moon”)
- For each tag, define an action corresponding to using the tag for the first time.

Social correlation in flickr

- Distribution of α values estimated using maximum likelihood:

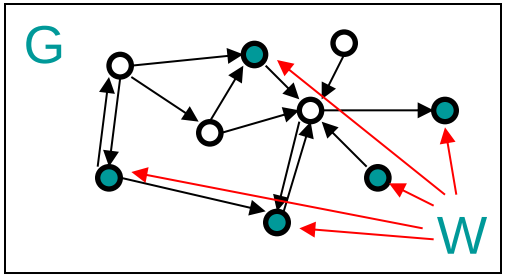


Distinguishing influence



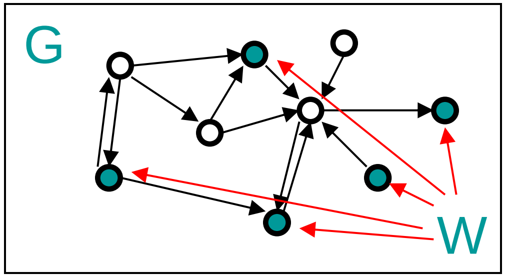
- Recall: graph G, set W of active nodes
- Influence model
  - First G is selected
  - Then W is picked from a distribution depending on G

Correlation Models



- 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

Correlation Model



- 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

- Shuffle Test:
  - **Simple Idea:** In non-influence model, even though an agent's probability of activation can depend on friends, her timing of activation is independent
  - Compute coefficient α
  - Shuffle time-stamp of all actions, and re-estimate coefficient α'
  - If $\alpha \approx \alpha'$, social influence is ruled out.
  - If $\alpha \neq \alpha'$, social influence can't be ruled out.
- Edge-Reversal Test:
  - Reverse direction of all edges, and re-estimate α.

Testing for Influence

Edge-Reversal Test:

- **Simple Idea:**
  - Main idea: assume edge (u → v), where u, v become active
  - If we have influence u is expected to become active before v
  - If there is no influence, each is equally likely to become active first
- **Test:**
  - Reverse direction of all edges, and re-estimate α.

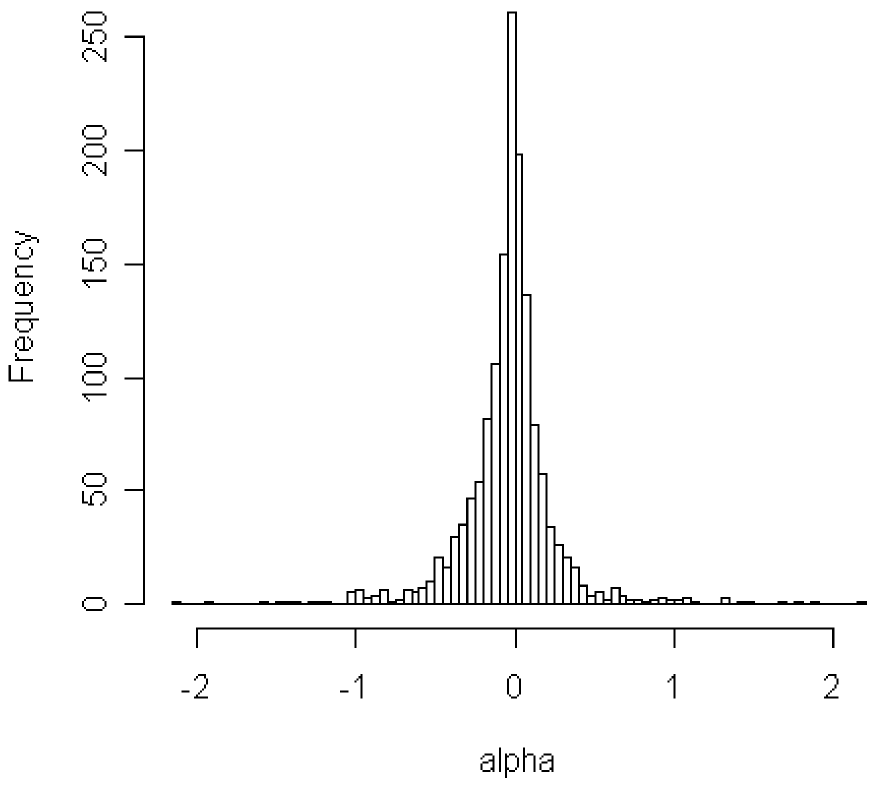
Shuffle test, theoretical justification

- Theorem. If the graph is large enough, the shuffle test rules out the general model of correlation.

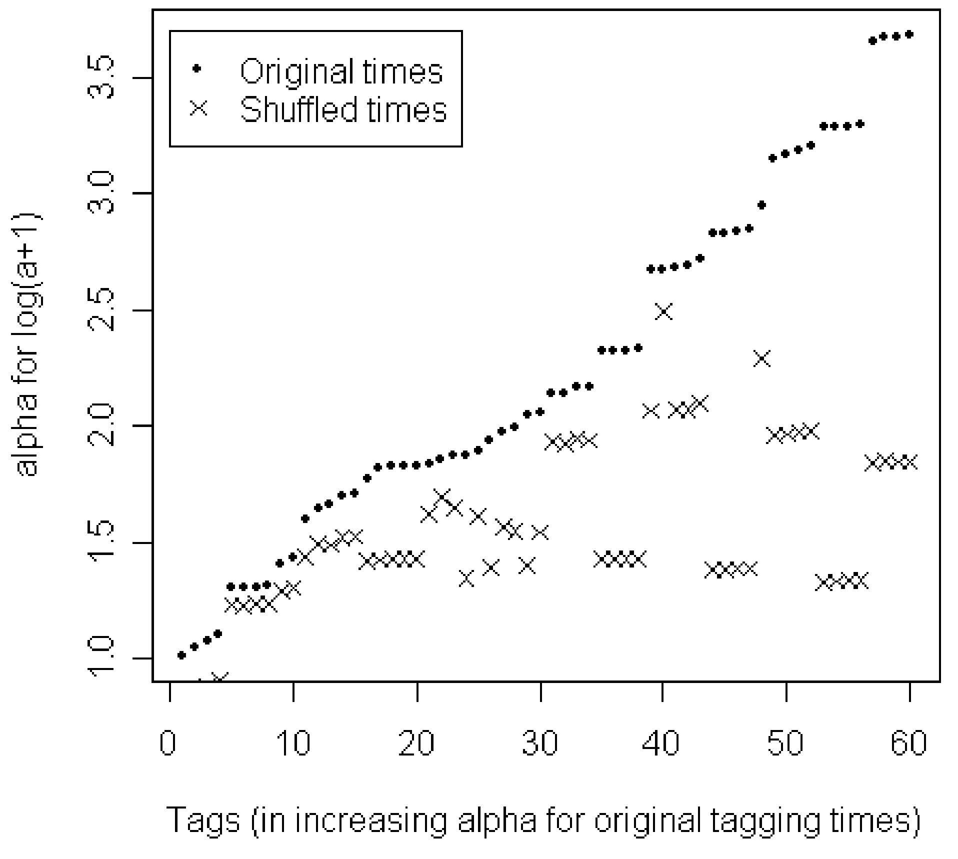
Simulations

- 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.

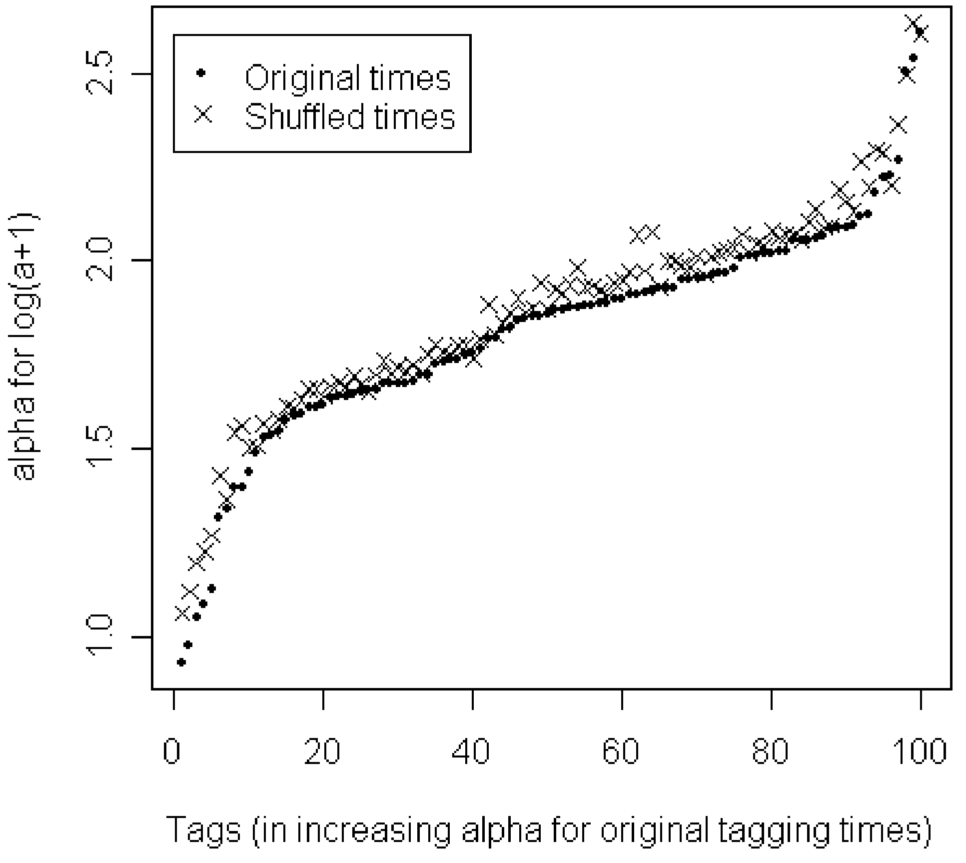
Simulation Results, Baseline



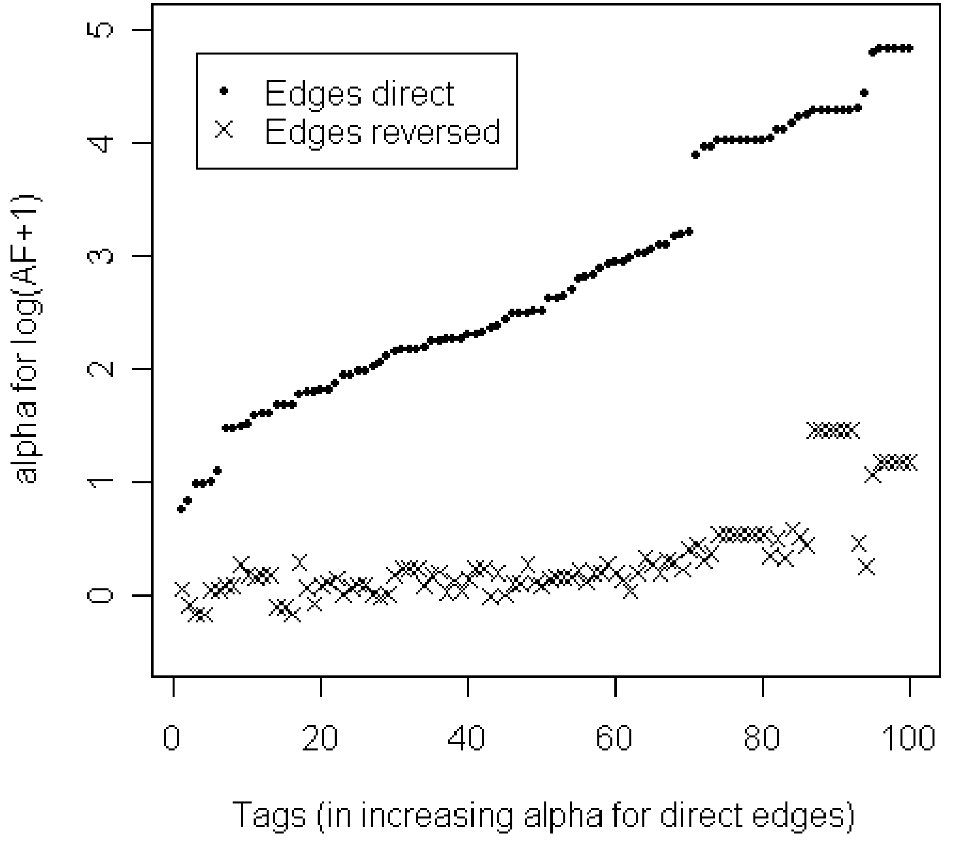
Shuffle Test, Influence Model



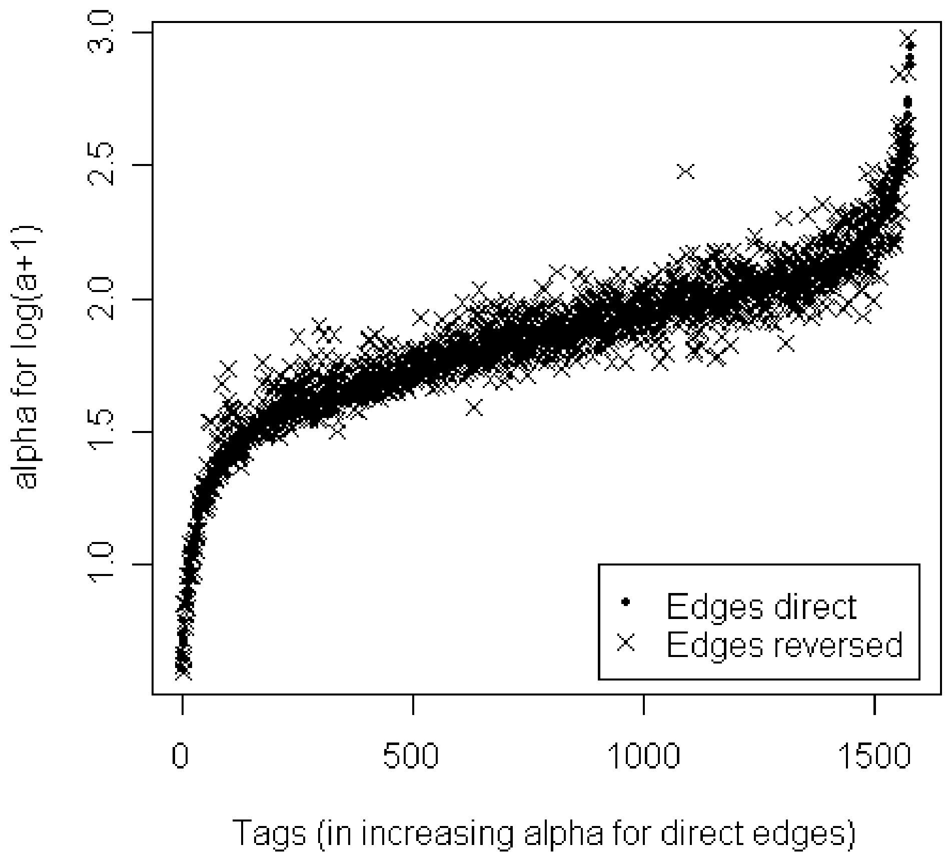
Shuffle Test, Correlation Model



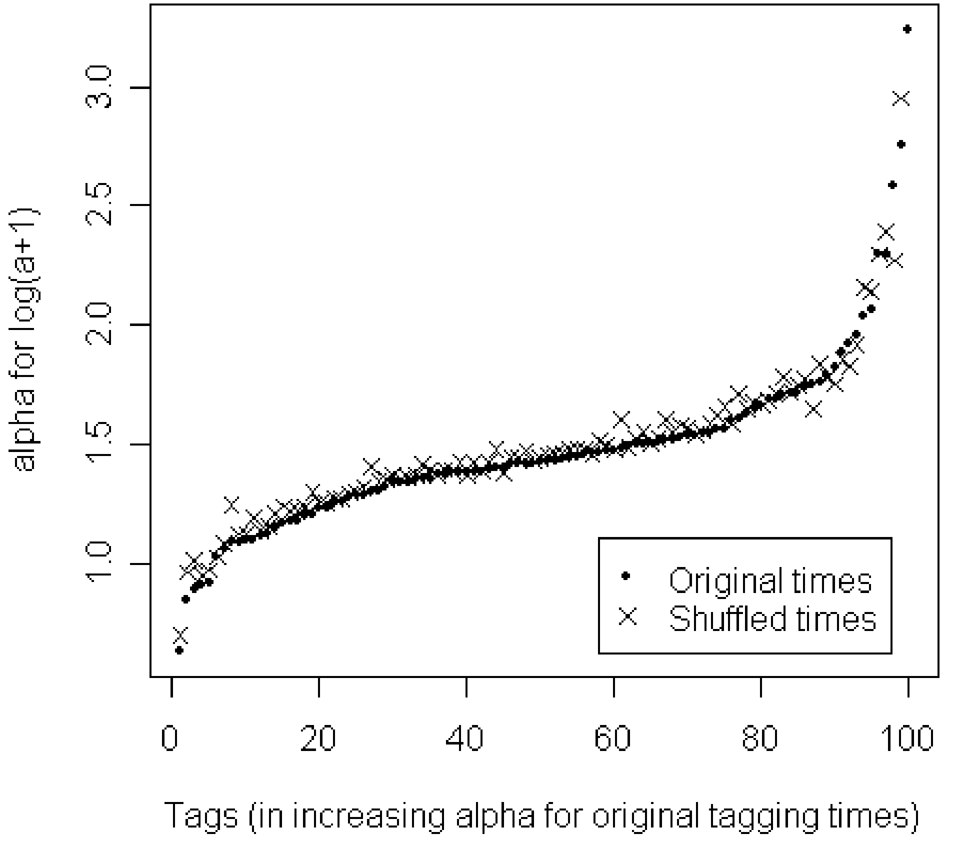
Edge-Reversal Test, Influence Model



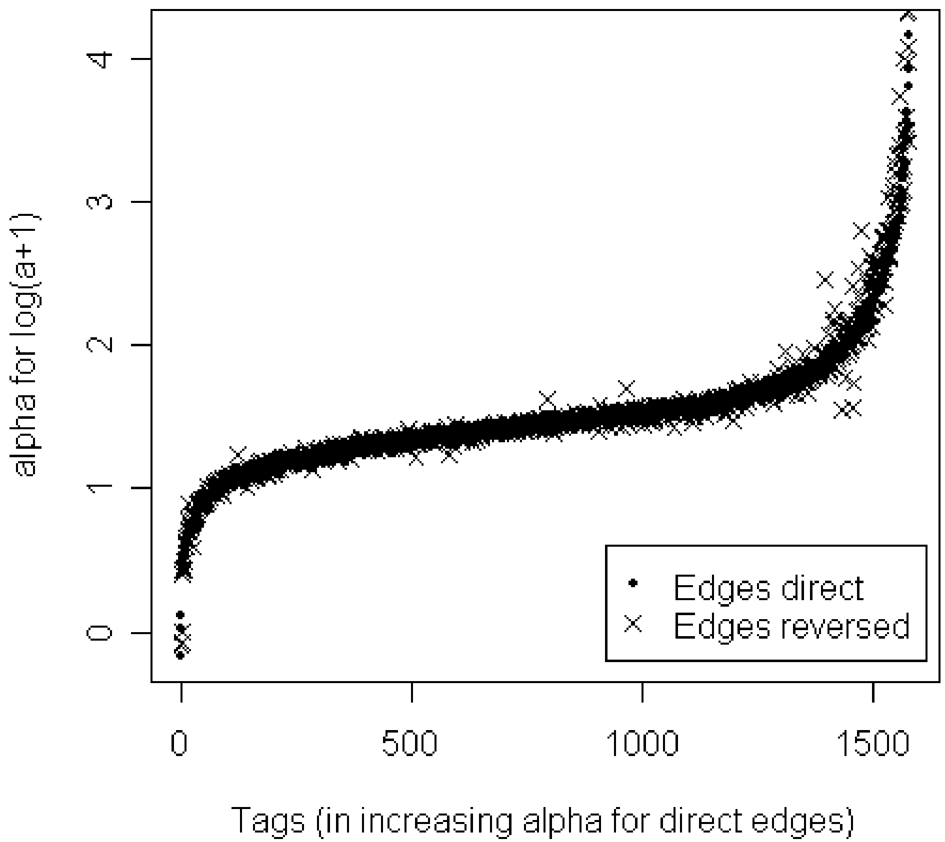
Edge-Reversal Test, Correlation Model



Shuffle Test on Flickr Data



Edge-Reversal Test on Flickr Data



Conclusions

- Our contributions
  - Defined two models that exhibit correlation, one with and the other without social influence
  - Developed statistical tests to distinguish the two
  - Theoretical justification for one of the tests
  - Simulations suggest that the tests “work” in practice
  - On Flickr, we conclude that despite considerable correlation, no social influence can be detected
- Discussion
  - cannot conclusively say there is influence without controlled experiments (example: flu treatment)
  - still can rule out potential candidates
  - Open: develop algorithms to find “influential” nodes/communities given a pattern of spread