Microscopic Evolution of Social Networks

Jure Leskovec, Lars Backstrom, Ravi Kumar, Andrew Tomkins
Affiliations: Cornell and Yahoo!

Presenters: Vamsidhar Thummala and Nedyalko Borisov
CPS 296.3, Spring 2009

In brief..

- Study of network evolution in online social networks
  - LinkedIn, Yahoo! Answers, Delicious, and FlickR
- Study with emphasis on node arrival and edge creation (mircoscopic) in contrast to snapshot based approach
- Use model likelihood of edges for evaluation and comparison among different networks

Approach

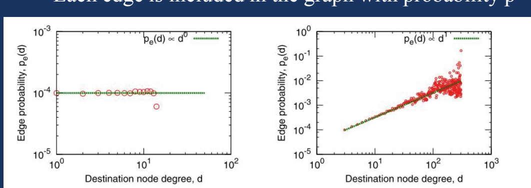
- Test individual edge attachment
  - Directly observe mechanisms leading to global network properties
    - E.g., What is really causing the power law distributions?
- Compare models via likelihood
  - Compare network models by likelihood (rather than summary statistics)
    - E.g., Is Preferential Attachment best model?

Preliminaries

- Preferential Attachment
  - Advocates the edge creation is governed by "popularity"
  - In theory, follows power law distribution
- Maximum likelihood principle
  - Way of tuning the free parameters for a good fit
  - General assumptions: iid and normal

Preliminaries

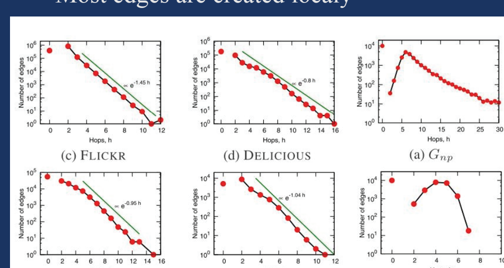
- G(n,p) model
  - A graph is constructed by connecting node randomly
  - Each edge is included in the graph with probability p



(a) G_{np} (b) PA

Locality of edge attachments

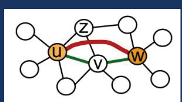
- Most edges are created localy



(c) FLICKR (d) DELICIOUS (a) G_{np}
(e) ANSWERS (f) LINKEDIN (b) PA

Triangle-closing models

- Various models
  - Random – uniformly random
  - Deg – proportional to the degree
  - Com – proportional to the common friends
  - Last – proportional to the last edge creation
  - Comlast – combination of com and last



FLICKR	random	$deg^{0.2}$	com	$last^{-0.4}$	$comlast^{-0.4}$
random	13.6	13.9	14.3	16.1	15.7
$deg^{0.1}$	13.5	14.2	13.7	16.0	15.6
$last^{0.2}$	14.7	15.6	15.0	17.2	**16.9**
com	11.2	11.6	11.9	13.9	13.4
$comlast^{0.1}$	11.0	11.4	11.7	13.6	13.2

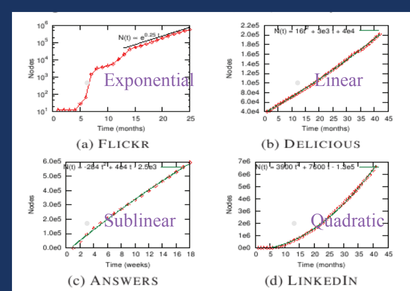
Edge-by-edge evolution model

- Three process that govern the evolution
  1. Node arrival process: nodes enter the network
  2. Edge initial process: each node decides when to initiate the edge
  3. Edge destination process: determines the destination after a node decides to initiate

Edge-by-edge evolution model

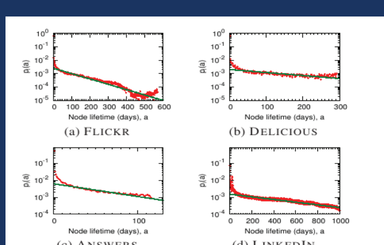
- Three process that govern the evolution
  1. Node arrival process: nodes enter the network
  2. Edge initial process: each node decides when to initiate the edge
  3. Edge destination process: determines the destination after a node decides to initiate

How fast the nodes are arriving?



- Node arrival process is dependent on the network

What is the node lifetime?



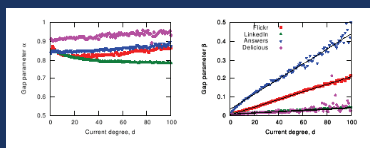
Node lifetimes are best modelled by exponential distribution

Edge-by-edge evolution model

- Three process that govern the evolution
  1. Node arrival process: nodes enter the network
  2. Edge initial process: each node decides when to initiate the edge
  3. Edge destination process: determines the destination after a node decides to initiate

Time gap between edges

- Edge gap: $\delta_u(d) = t_{d+1}(u) - t_d(u)$



- Power Law Distribution

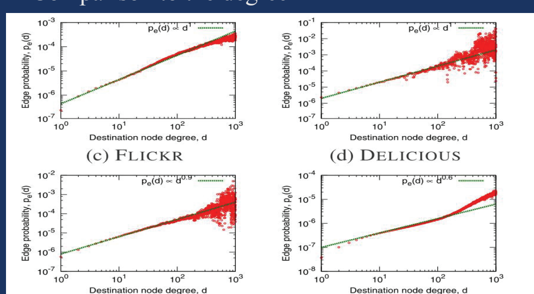
$$p_g(\delta(d);\alpha,\beta) \propto \delta(d)^{-\alpha}\exp(-\beta\delta(d))$$

Edge-by-edge evolution model

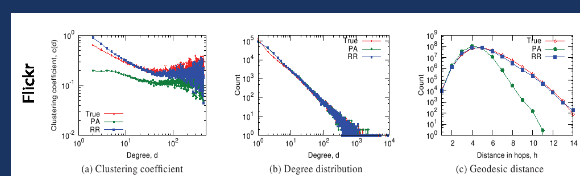
- Three process that govern the evolution
  1. Node arrival process: nodes enter the network
  2. Edge initial process: each node decides when to initiate the edge
  3. Edge destination process: determines the destination after a node decides to initiate

Edge destination

- Comparison to the degree



Model evaluation



Discussion

- What do we get out of it?
- Will this work for any kind of social network?
- What is the intuition behind the model parameters?
- Suggestions in the soc. networks follow random-random model, so it is expected to fit well.