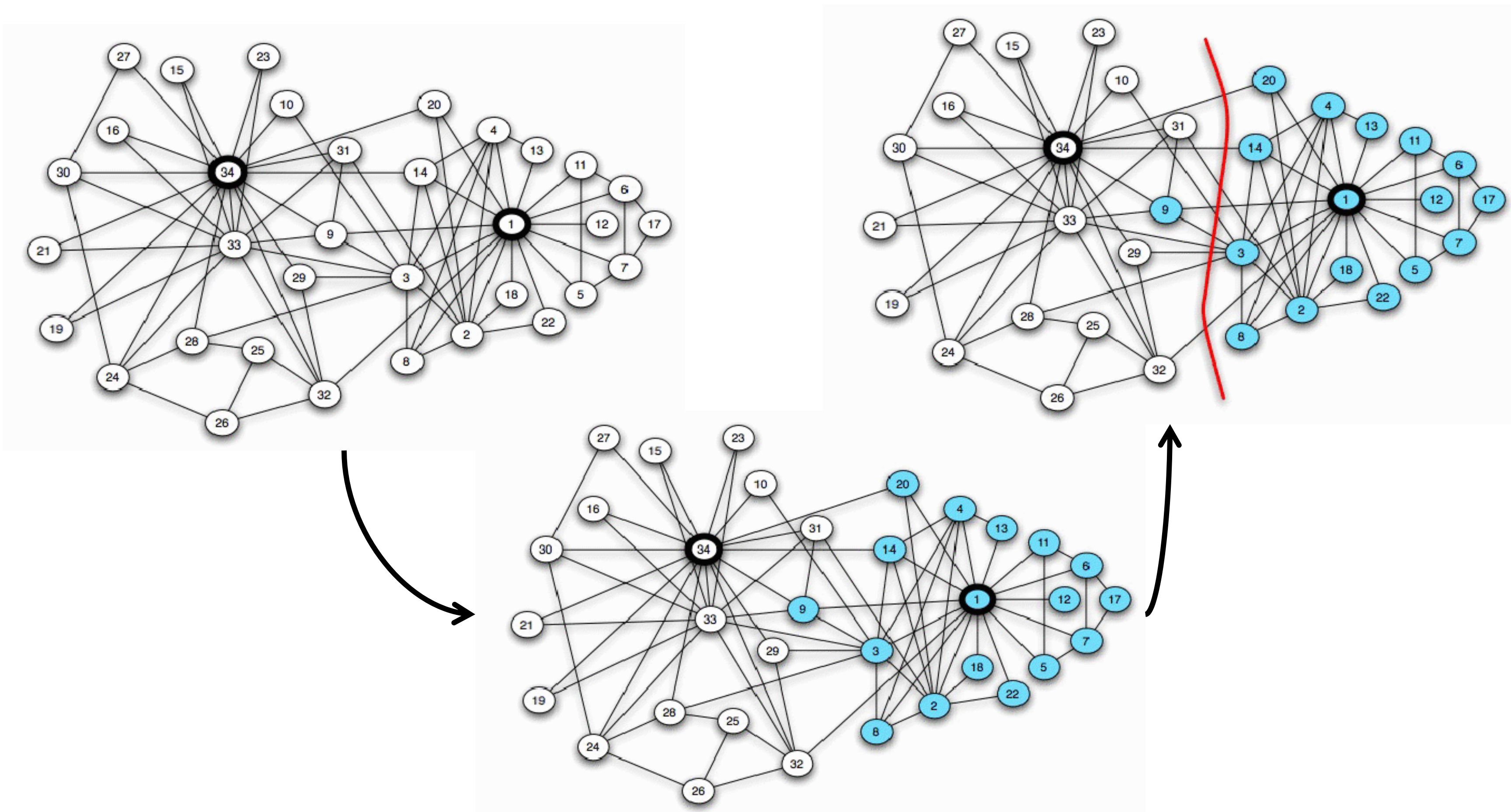
Micro and macroevolution of social networks

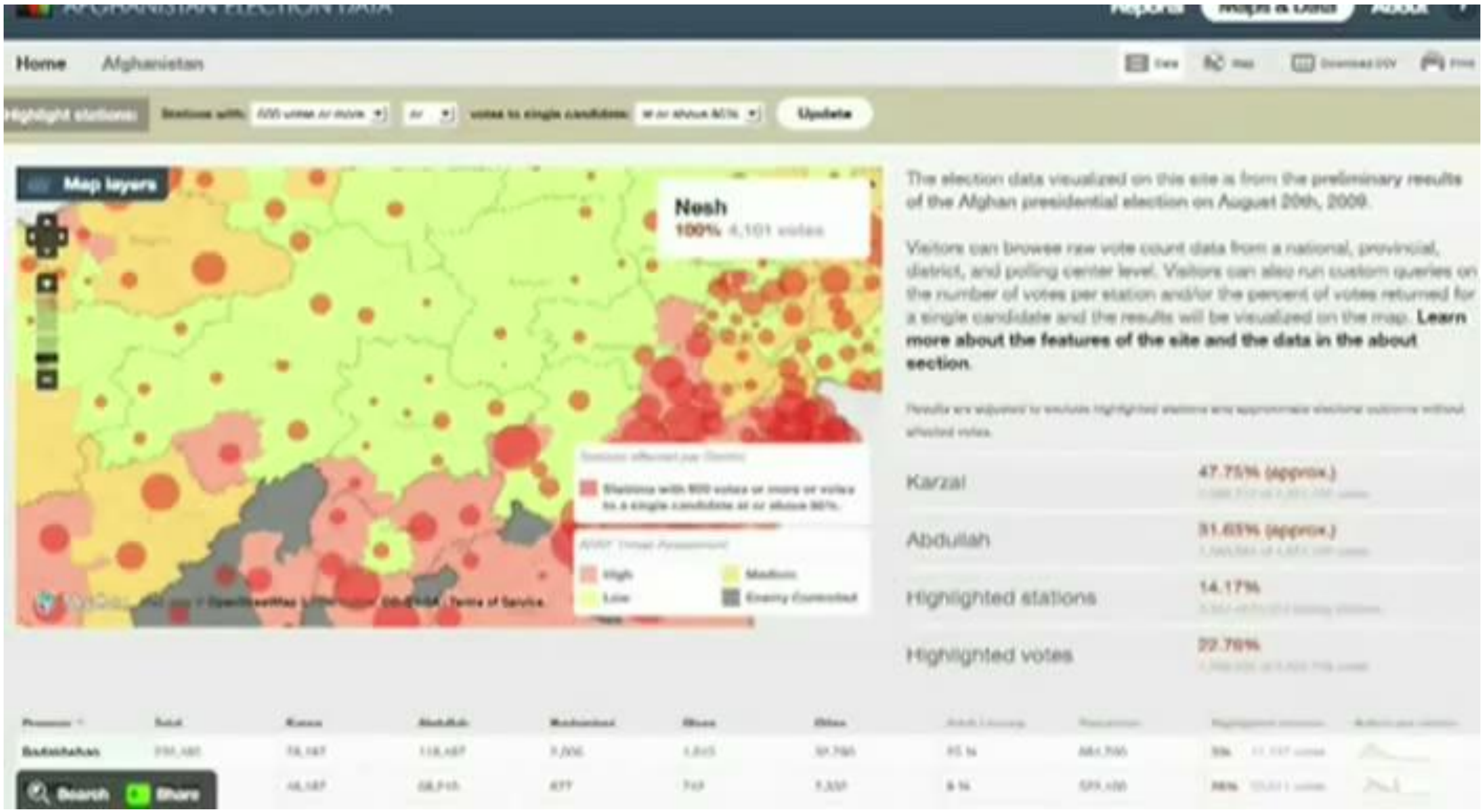
Mikolaj Morzy, Politechnika Poznanska

Example of a SNA

- Analysis of an academic karate club (Zachary, 1972)



Bursts of activities



http://www.youtube.com/watch?v=3YcZ3Zqk0a8

Can bursts of activities be predicted?

[…] by studying the communication and movement of millions of individuals through the electronic records they left behind, like mobile phone records, we have found a huge degree of predictability of individual behavior. The measurements told us that to those familiar with our past, our future acts should rarely be a surprise.

Albert-Barabasi model

- Algorithm for generating artificial networks
  - generates network with power-law degree distribution $P(k) \propto k^3$
  - assumes constant growth of the network
  - uses preferential attachement (autocatalysis)
  - average distance between nodes $l \propto \frac{\ln n}{\ln \ln n}$
  - clustering coefficient $C \propto n^{-0.75}$
  - nodes are added sequentially
  - probability of linking to node v_i given by $p(v_i) = \frac{\deg(v_i)}{\sum_j \deg(v_j)}$

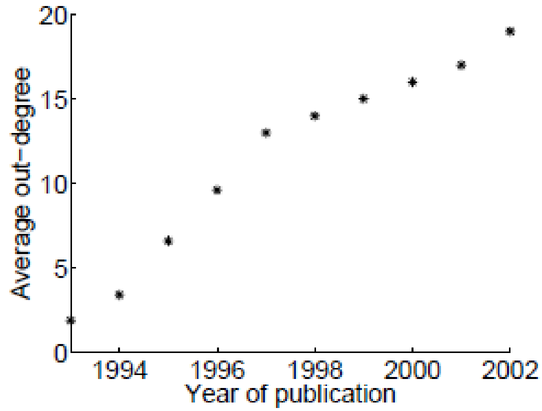
Macroevolution of social networks

- Wrong assumptions of the Barabási-Albert model
  - assumption about constant average degree of nodes
  - assumption on slow network diameter increase
- In reality… (Leskovec et al., 2005)
  - social networks become denser over time, i.e., the number of edges grows over-linearly w.r.t. the number of nodes

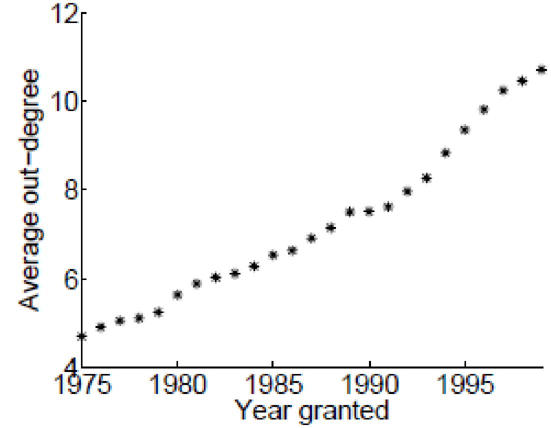
$$e(t) \propto n(t)^{\alpha} \qquad \alpha \in \langle 1,2 \rangle$$

    - $\alpha = 1$: constant average node degree
    - $\alpha = 2$: constant percentage of nodes linking to a given node
  - network diameter shrinks as more nodes join the netwok

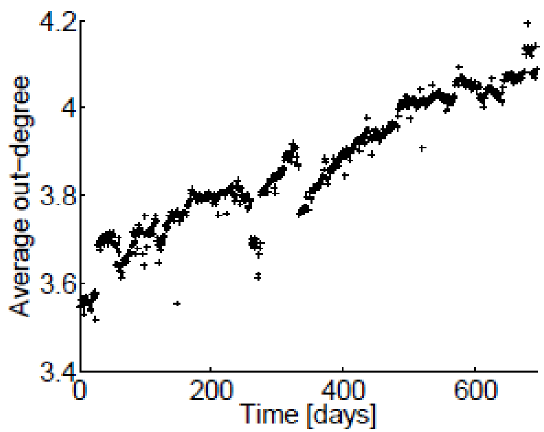
Empirical data : outdegree



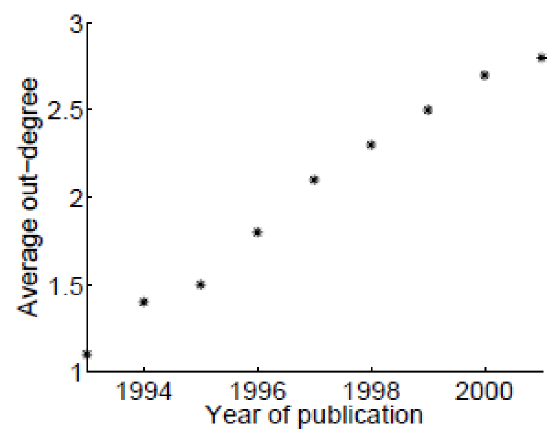
(a) arXiv



(b) Patents



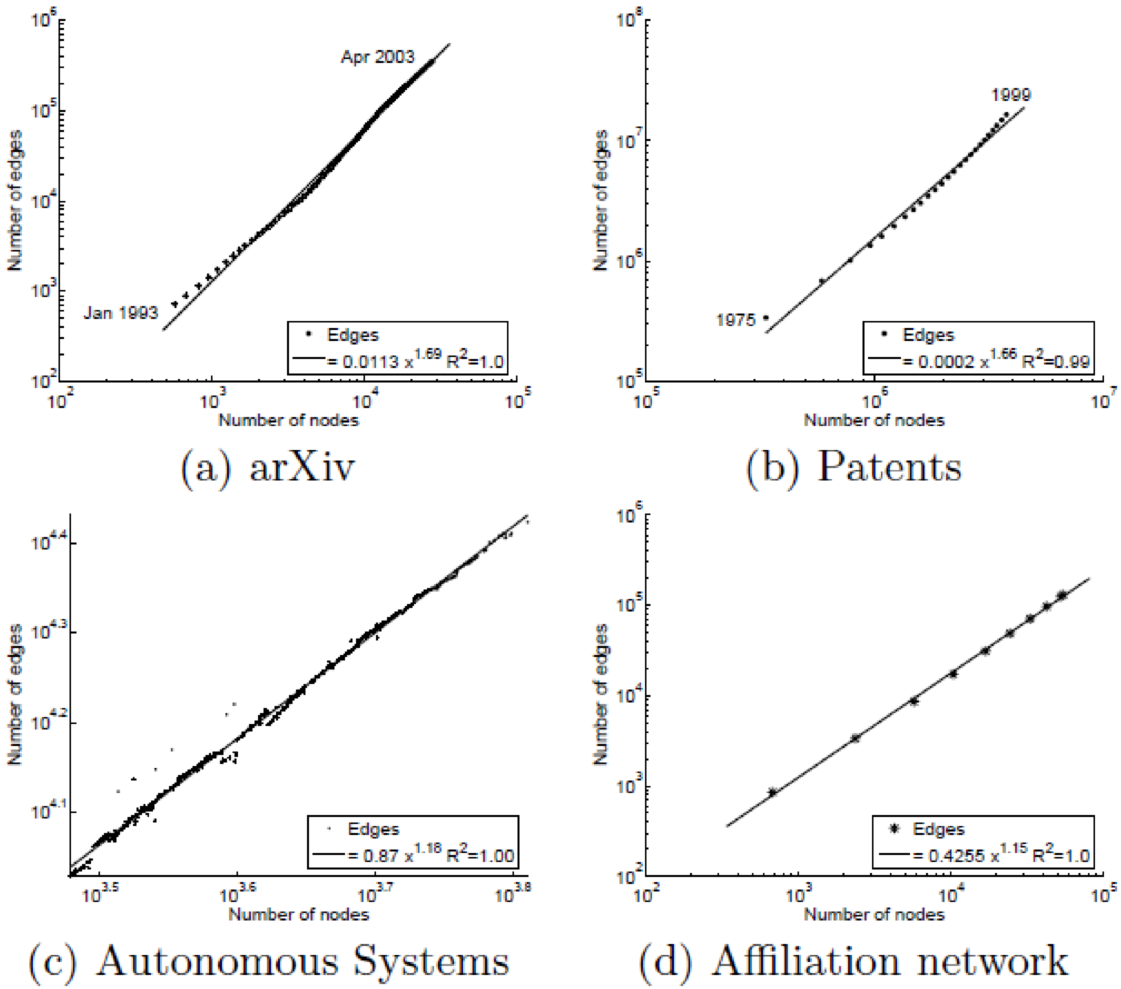
(c) Autonomous Systems



(d) Affiliation network

source: "Graphs over time...", Leskovec et al., KDD'05

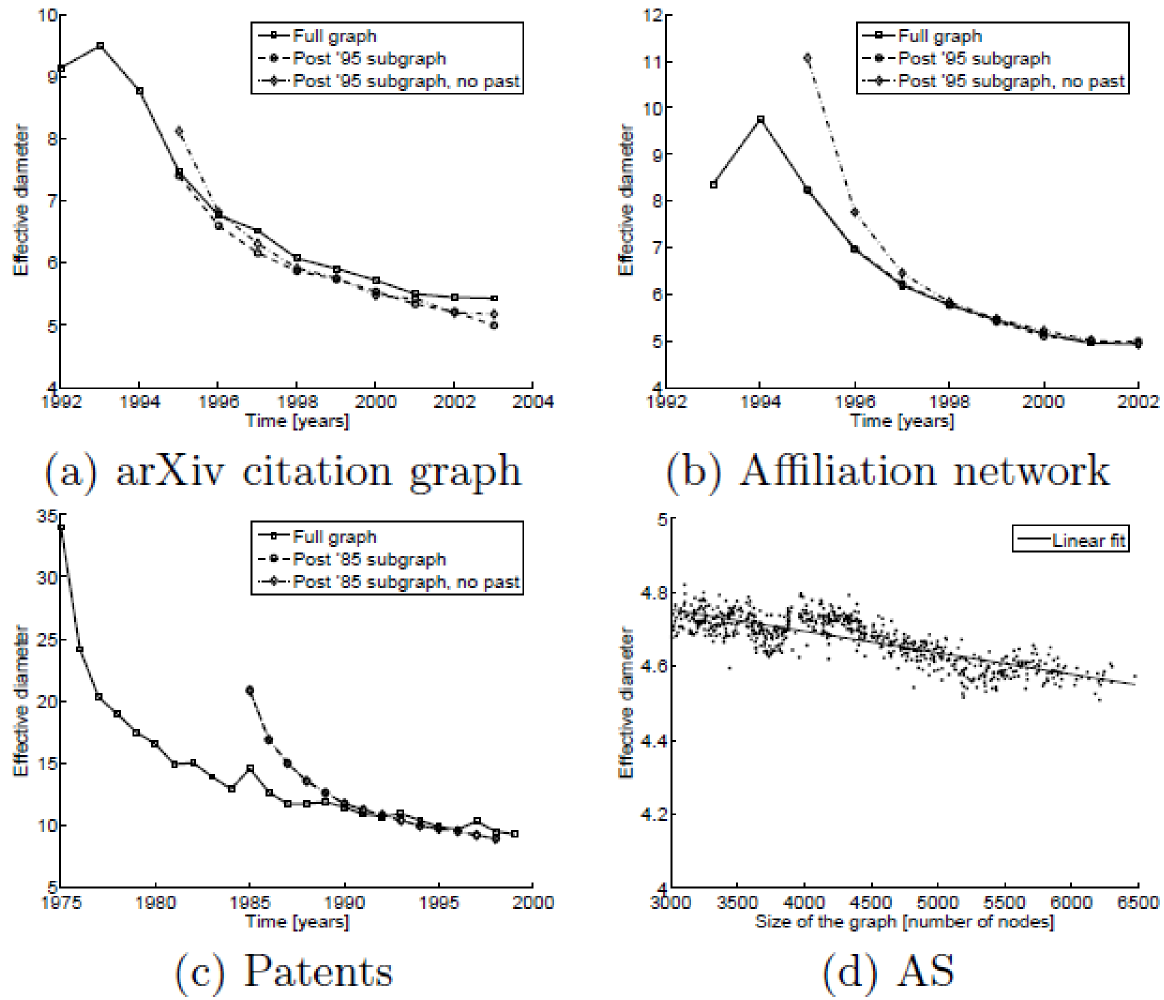
Empirical data : number of edges



(a) arXiv (b) Patents (c) Autonomous Systems (d) Affiliation network

source: "Graphs over time...", Leskovec et al., KDD'05

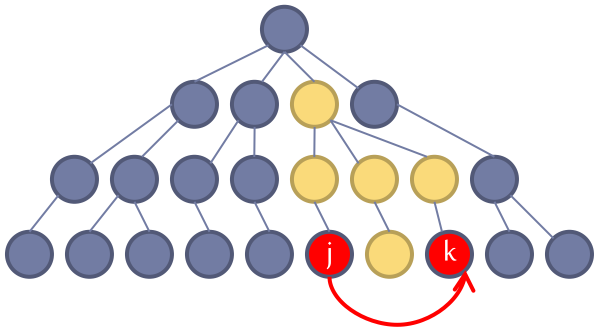
Empirical data : diameter



source: "Graphs over time...", Leskovec et al., KDD'05

Models for social network evolution (1/2)

- Community-guided attachement
  - probability of making a new edge $$p(j,k) \propto c^{-h}$$
  - average node degree $$\overline{\deg}(n) = n^{1-\log_b c} \quad 1 \le c < b$$
  - edges and nodes $$e(t) \propto n(t)^{2-\log_b c}$$



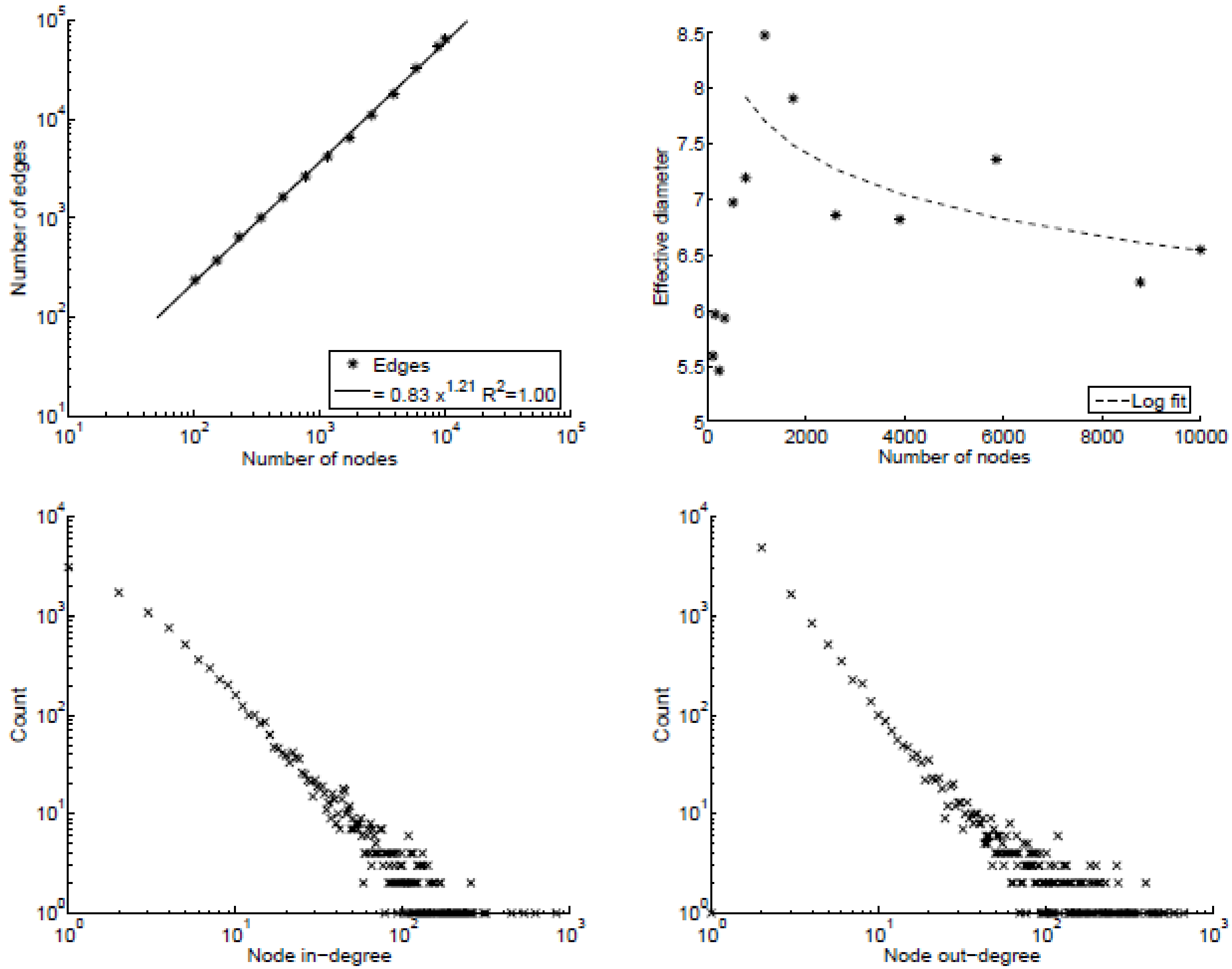
- What is wrong with CGA model?
  - no shrinking diameter property
  - no power-law distribution of node degrees

Models for social network evolution (2/2)

- Forest fire model
  - analogies: forest fire, citations, professional associations
  - parameters
    - forward burning probability p
    - backward burning ratio r

1. node v randomly chooses an ambassador a and forms an edge
2. node v chooses x edges going in and out of a with probability 1/r and forms edges to these nodes, where x is chosen from the binomial dist. with the average $(1-p)^{-1}$
3. node v repeats step (2) for nodes $a_1, a_2, \ldots, a_x$ chosen in step (2)

Properties of networks generated using forest fire model



In-degree

Out-degree

Microevolution of social networks

- How do individual nodes influence global properties of the network?
  - process of new node creation
  - process of edge initialization
  - process of choosing destination node
  - timespan of node's life and activity

Example datasets

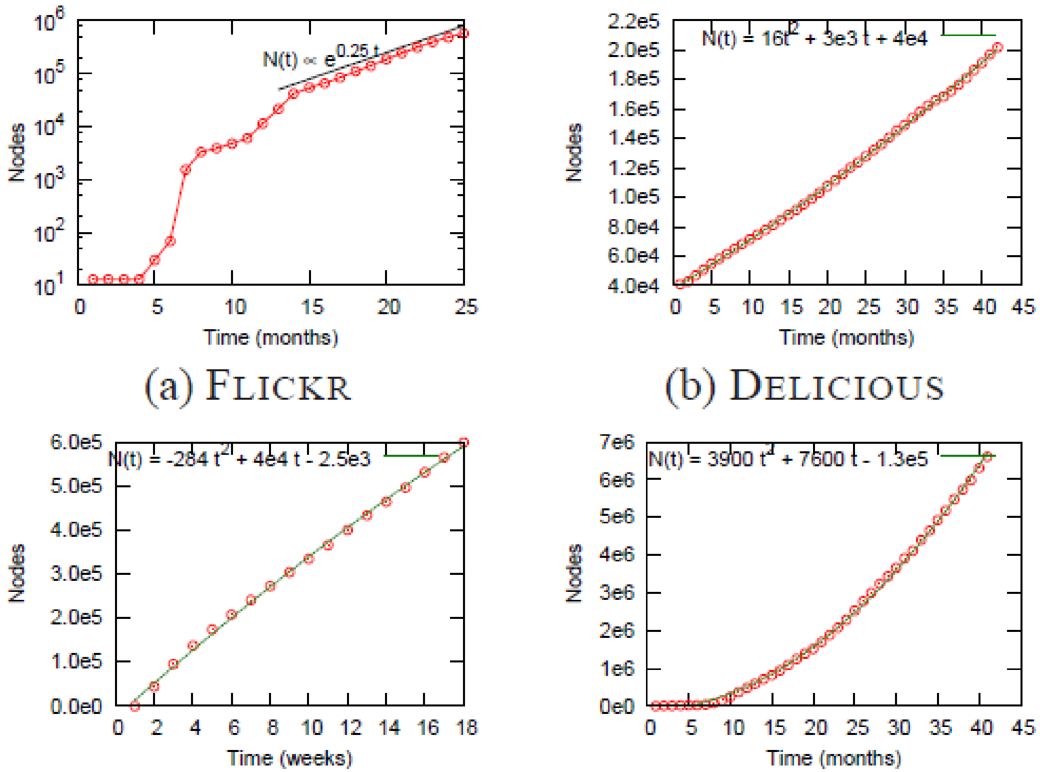
sieć	T	N	E	E_b	E_Δ	%	ρ	κ
Flickr	621	584207	3554130	2594078	1475345	65.63	1.32	1.44
Delicious	292	203234	430707	348437	96387	27.66	1.15	0.81
Answers	121	598314	1834217	1067021	303858	23.36	1.25	0.92
LinkedIn	1294	7550955	30682028	30682028	15201596	49.55	1.14	1.04

- T: number of time intervals
- N: number of nodes
- E: number of edges
- E_b: number of reciprocal edges
- E_Δ: number of triangle-closing edges
- %: percent of edges that close triangles
- ρ: exponent of network densification
- κ: exponent of distant edge creation

$$E(t) \propto N(t)^\rho$$
$$E_h \propto \exp(-\kappa h)$$

How do nodes appear on the scene?

- Function $N(\cdot)$ of new node creation strongly depends on the social process responsible for network creation

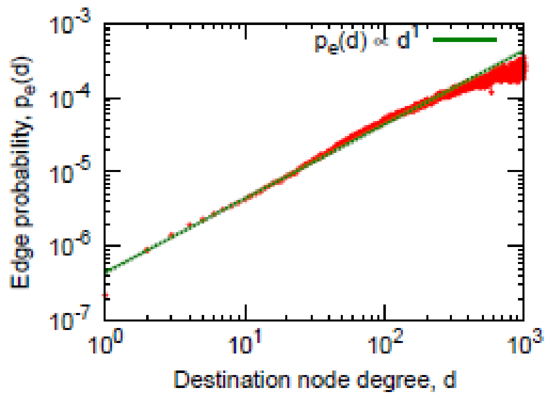


(a) FLICKR (b) DELICIOUS (c) ANSWERS (d) LINKEDIN

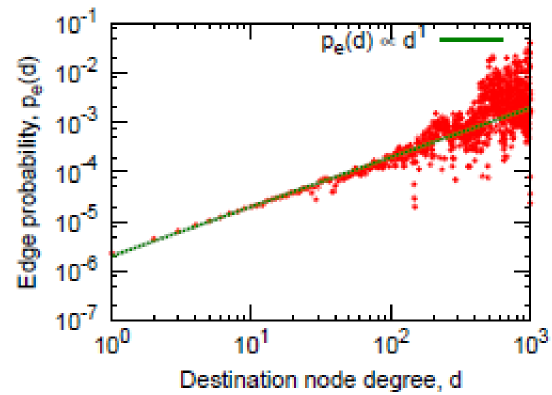
network	N(t)
Flickr	exp(0.25t)
Delicious	$16t^2$+3000t+40000
Answers	$-28t^2$+40000t-2500
LinkedIn	$3900t^2$+76000t-130000

How important is the popularity?

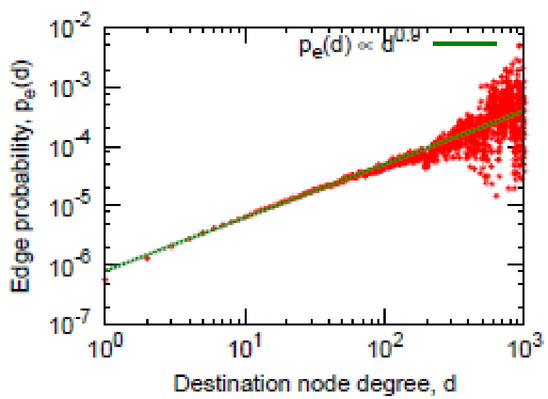
- Probability of choosing a node v with $\deg(v) = d$



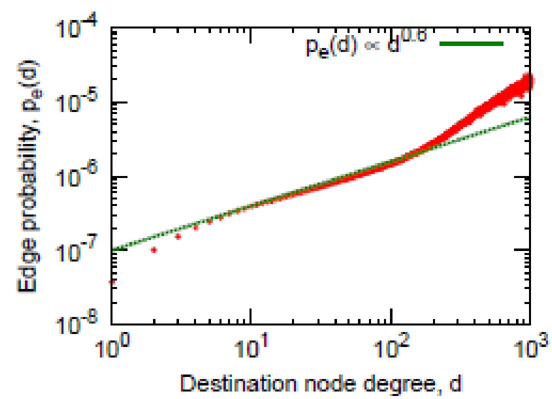
(c) FLICKR



(d) DELICIOUS



(e) ANSWERS



(f) LINKEDIN

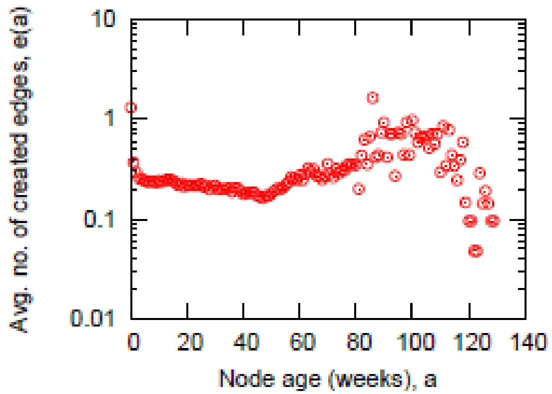
Models for destination node selection

- Probability of choosing node v with $\deg(v) = d$
  - model D $p(v) \propto d_t(v)^{\tau}$
  - model DR $p(v) \propto \tau d_t(v) + (1-\tau)\dfrac{1}{N(t)}$
  - model A $p(v) \propto a_t(v)^{\tau}$
  - model DA $p(v) \propto d_t(v) a_t(v)^{\tau}$

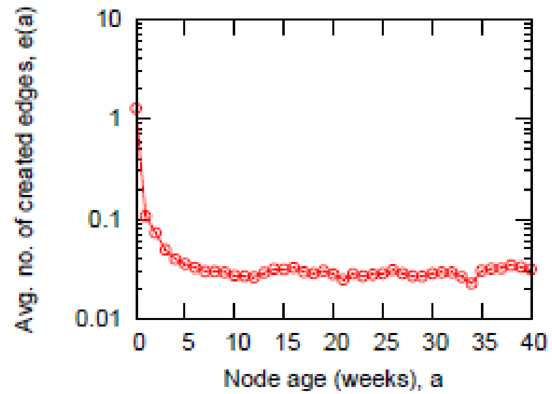
- Fitting models to real world networks
  - Flickr: preferential attachment, model D $(\tau=1)$
  - Delicious: random preferential attachment, model DR $(\tau=0.5)$
  - Answers: aging preferential attachment, model DA $(\tau=0.4)$
  - LinkedIn: random preferential attachment, model DR $(\tau=0.9)$

How does age influence activity?

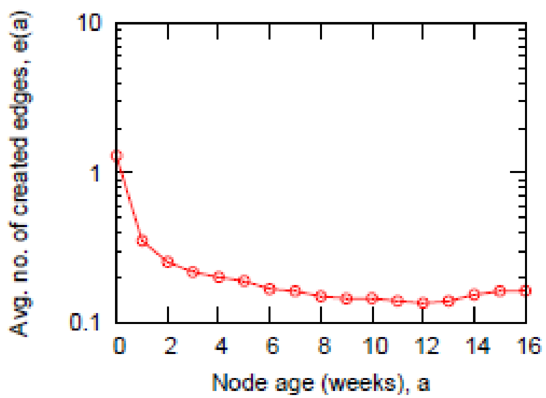
- Activity of a node v with age a



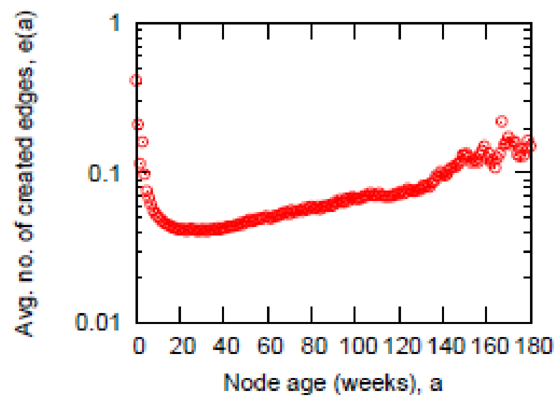
(a) FLICKR



(b) DELICIOUS



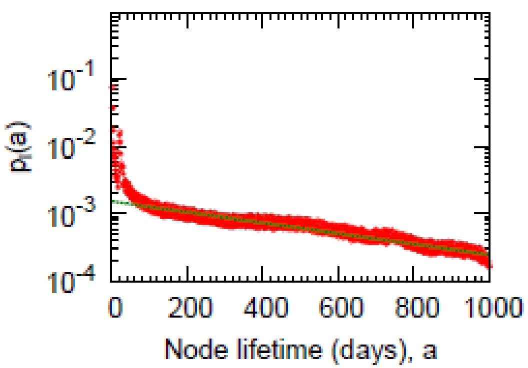
(c) ANSWERS



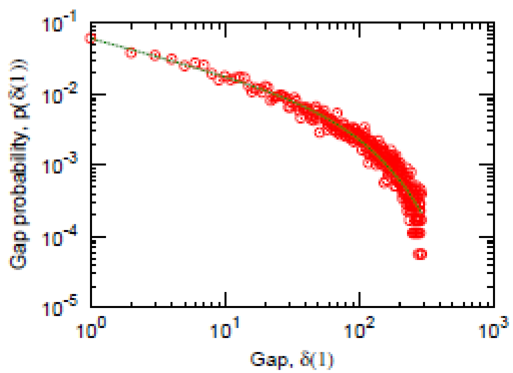
(d) LINKEDIN

Node age and edge creation

- Lifespan of nodes
  - best fitting for $p_l(a) \propto \exp(-\lambda a)$
  - no fitting for nodes with very short lifespan
- Edge creation activity
  - best fitting for $p_g(\delta(d)) \propto \delta(d)^{-\alpha} \exp(-\beta\delta(d))$
  - constant element $\alpha(d)$ linear element $\beta(d)$
  - no correlation between $\delta(1)$ and final node degree



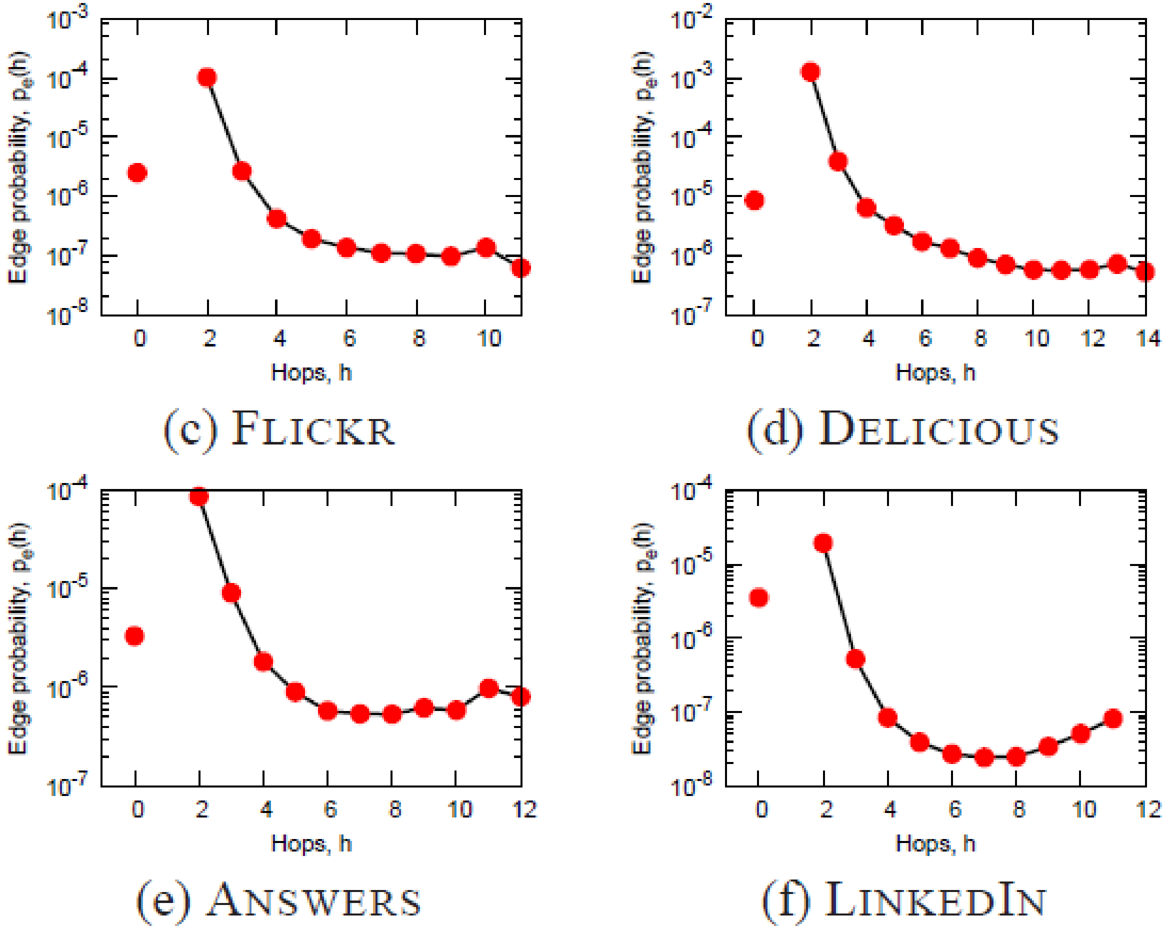
(d) LinkedIn



(b) Delicious

How often neighbours are chosen?

- Number of edges created to nodes with distance *h*



(c) FLICKR

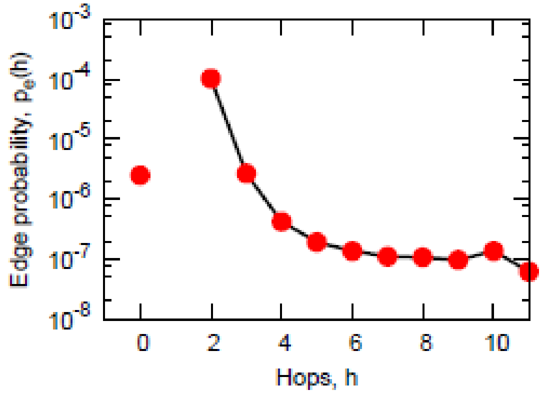
(d) DELICIOUS

(e) ANSWERS

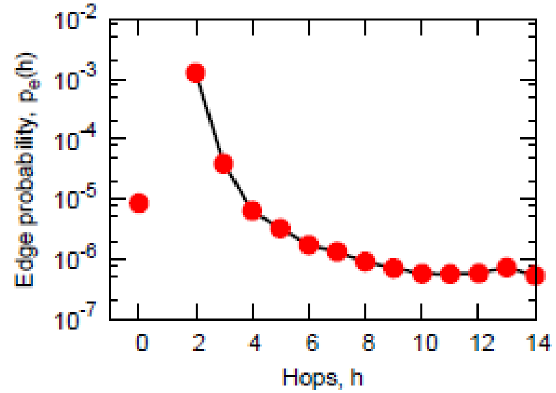
(f) LINKEDIN

Is triangle-closing model correct?

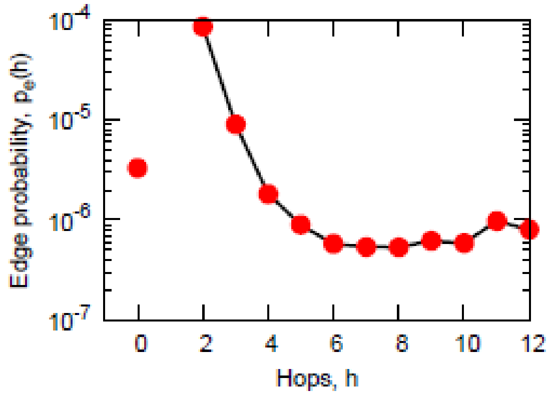
- Probability of linking to nodes with distance *h*



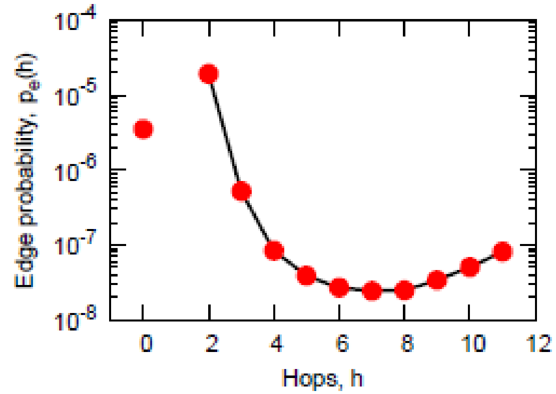
(c) FLICKR



(d) DELICIOUS



(e) ANSWERS



(f) LINKEDIN

General model for microevolution of social networks

1. nodes appear following the function $N(\cdot)$
2. lifespan a is chosen from the distribution $p_l(a) = \lambda \exp(-\lambda a)$
3. node links to a node v using preferential attachment
4. node chooses a gap δ from the distribution $p_g(\delta \mid d, \alpha, \beta) = \delta^{-\alpha} \exp(-\beta d \delta)$ and becomes inactive for δ steps
5. after waking up, if the lifespan has not expired, the node creates a triangle-closing edge
6. node loops to (4)

Summary

- Evolution of many social network reveals similarities
- New models regenerate network evolution with great precision
  - modeling local and global phenomena in the network
  - modeling processes that are contextually dependant on the characteristics of social network
- Open problems
  - information diffusion in social networks
  - maximization of influence in social networks
  - formation and evolution of microgroups in social networks