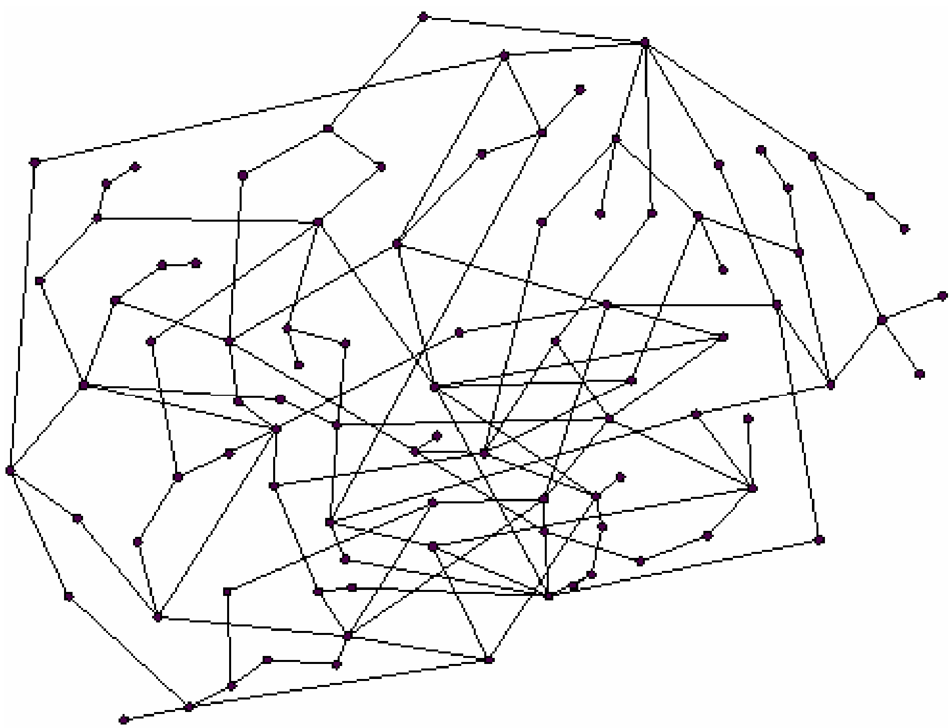
Models of evolving networks

CS 322: (Social and Information) Network Analysis
Jure Leskovec
Stanford University

Project proposal

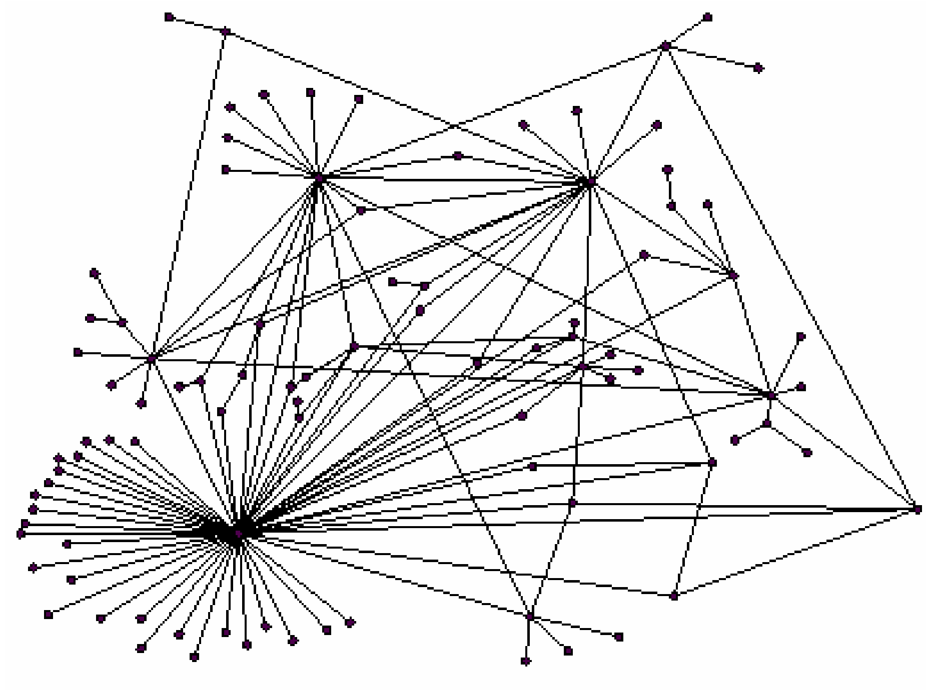
- 3 pages per group
  - Should include both
    - Reaction part
    - Proposed work part

Recap from last time



Random network

(Erdos-Renyi random graph)



Scale-free (power-law) network

Model: Preferential attachment

- Preferential attachment [Price 1965, Albert-Barabasi 1999]:
  - Nodes arrive in order
  - A new node j creates m out-links
  - Prob. of linking to a previous node i is proportional to its degree d_i

$$P(j \rightarrow i) = \frac{d_i}{\sum_k d_k}$$

[Mitzenmacher, '03]

Rich-get-richer

- Pages are created in order $1,2,\ldots,n$
- When node j is created it produces a single link to some earlier node i chosen:

1) With prob. p, j links to i chosen uniformly at random

2) With prob. $1\text{-}p$, j links to i with prob. proportional to d_i (degree of i)

Continuous approximation (1)

- How does degree d_i of node *i* grow over time?

$$\frac{\mathrm{d}\, d_i}{\mathrm{d}\, t} = p\frac{1}{t} + q\frac{d_i}{t}$$

$q=1-p$

- What is the degree $d_i(t)$ of node *i* at time *t*?

$$d_i(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right]$$

Continuous approximation (2)

- Fraction of nodes with degree $>d$ at time t

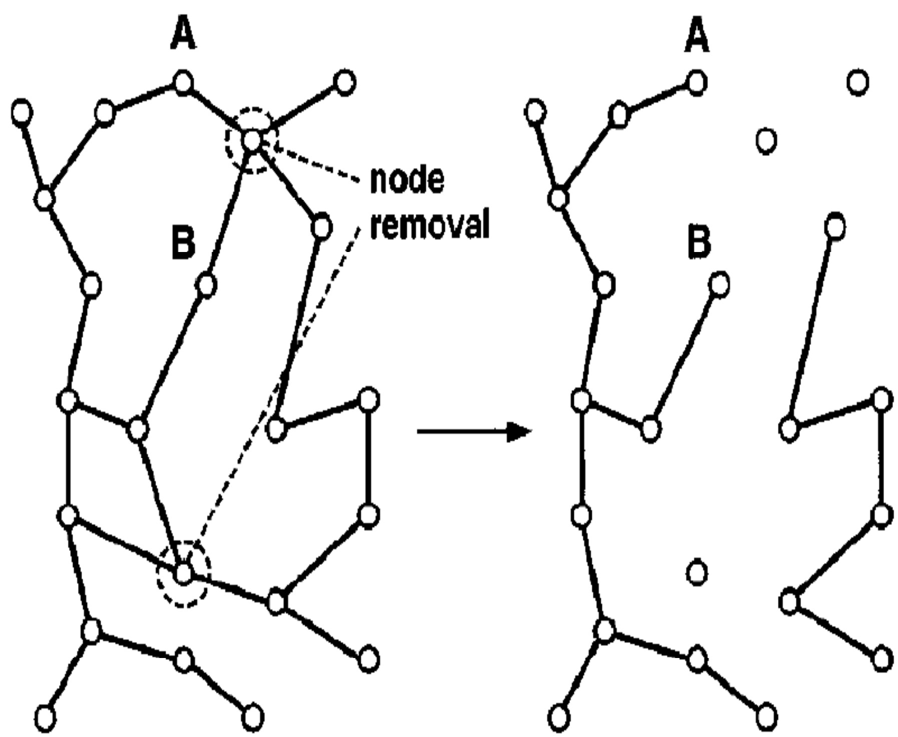
$$d_i(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^{q} - 1\right] > d \Rightarrow i < t\left[\frac{q}{p}d + 1\right]^{-\frac{1}{q}}$$

- Fraction of nodes with degree d?

$$= \frac{1}{p}\left[\frac{q}{p}d + 1\right]^{-1-\frac{1}{q}} \Rightarrow \alpha = 1 + \frac{1}{q} = 1 + \frac{1}{1-p}$$

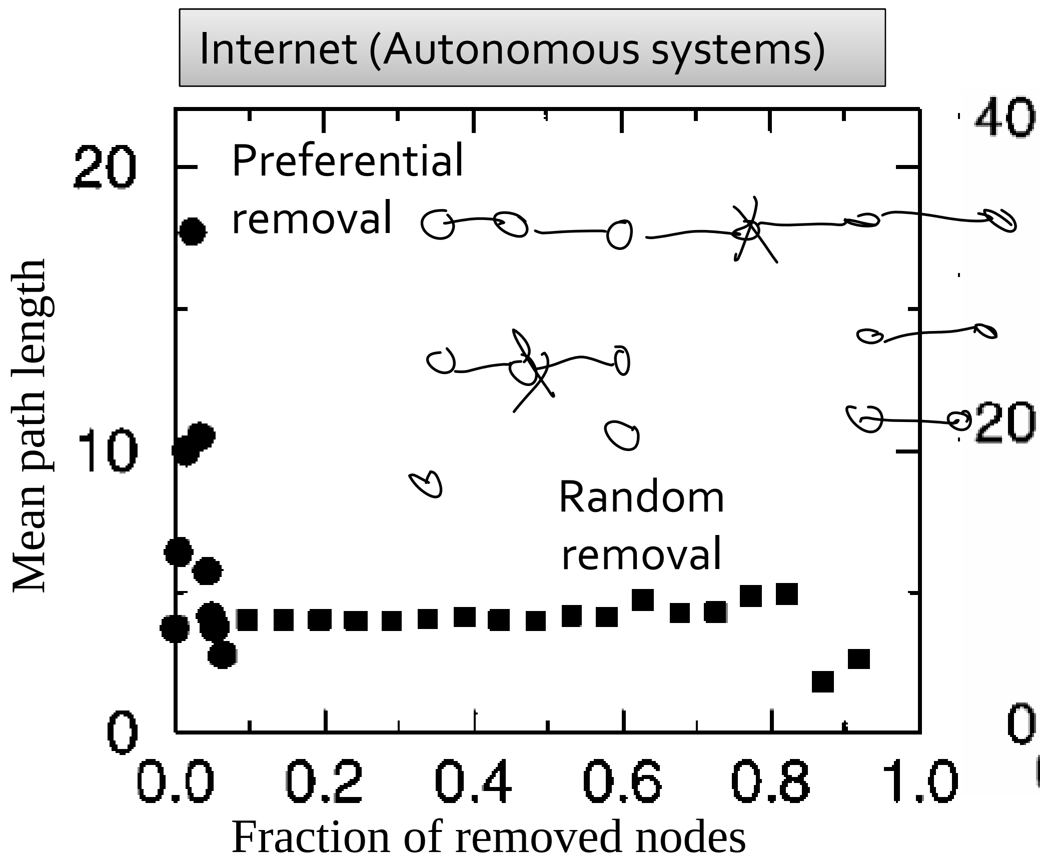
Network resilience (1)

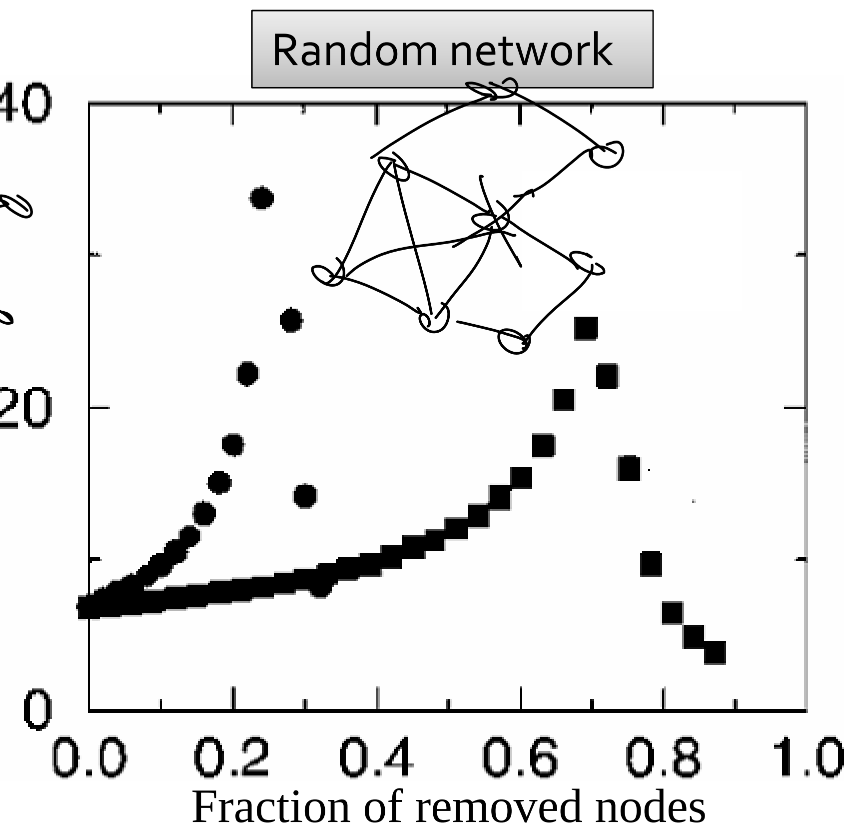
- We observe how the connectivity (length of the paths) of the network changes as the vertices get removed [Albert et al. 00; Palmer et al. 01]
- Vertices can be removed:
  - Uniformly at random
  - In order of decreasing degree
- It is important for epidemiology
  - Removal of vertices corresponds to vaccination



Network resilience (2)

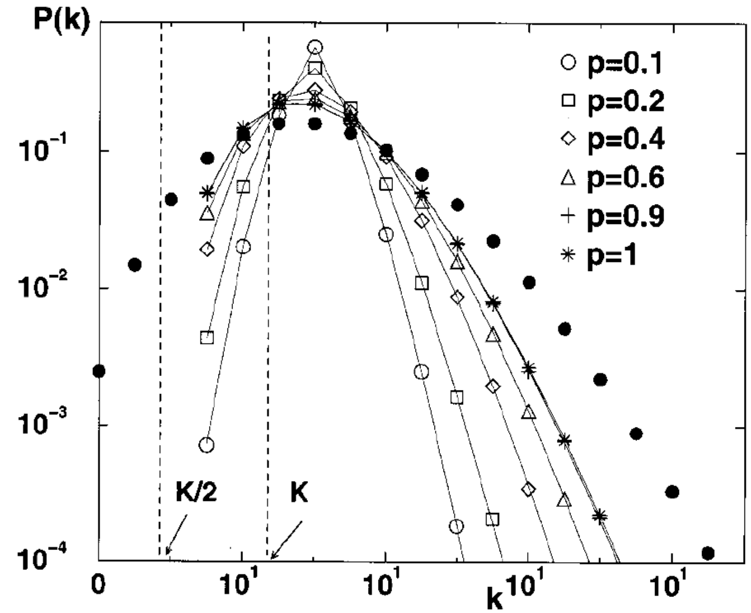
- Real-world networks are resilient to random attacks
  - One has to remove all web-pages of degree > 5 to disconnect the web
  - But this is a very small percentage of web pages
- Random network has better resilience to targeted attacks





Recap: Models so far (1)

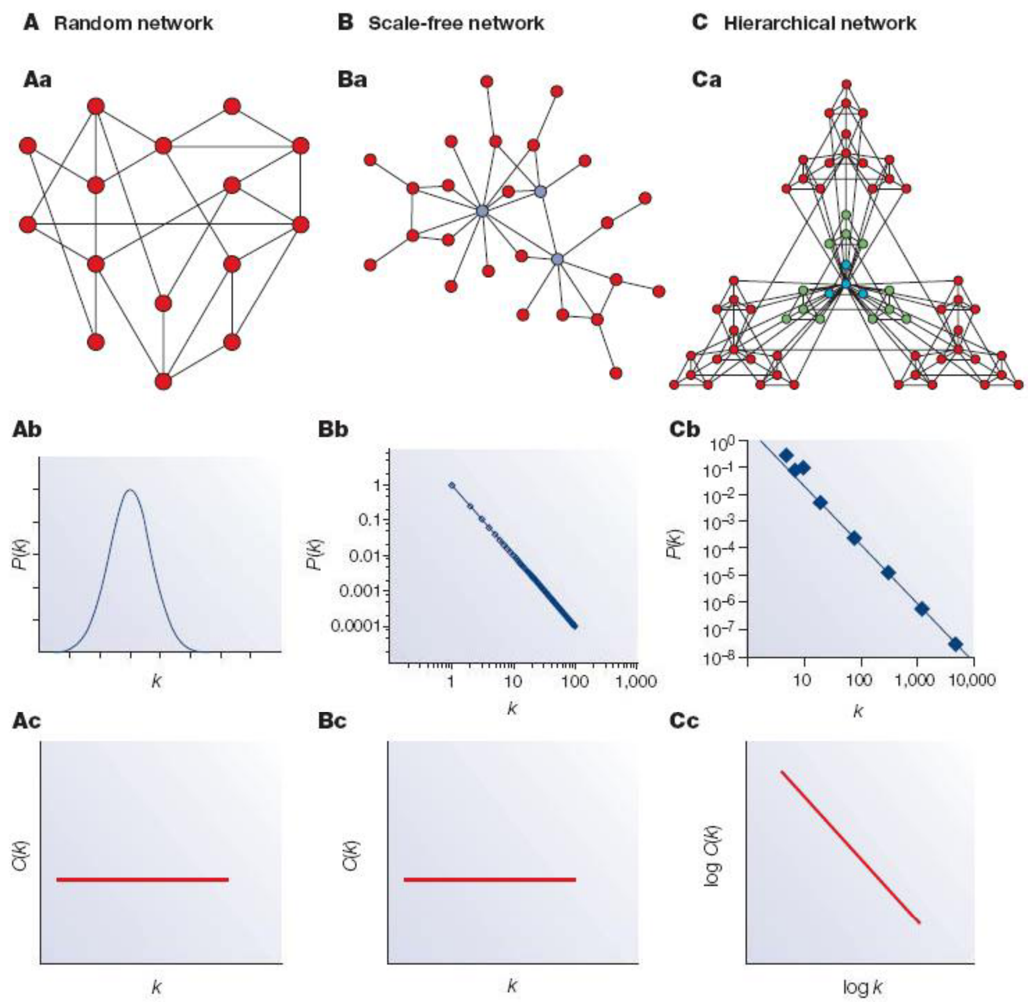
- G_{np}:
  - not realistic
  - But gives small diameters
- Small world:
  - small diameter + local structure
  - But no power-law degrees



Degree distribution of the Small world model

Recap: Models so far (2)

- Preferential attachment:
  - Power law degree distributions
  - But no local clustering
- Can we get all of them?



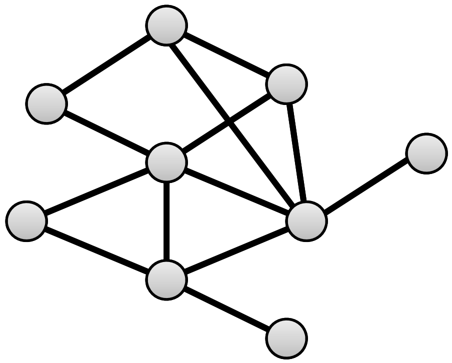
Observation

- Preferential attachment is a network growth model
- What governs the network growth and evolution?
  - P1) Node arrival process: nodes enter the network
  - P2) Edge initiation process: each node decides when to initiate an edge
  - P3) Edge destination process: determines destination after a node decides to initiate

[Leskovec et al. KDD 08]

Let's look at the data

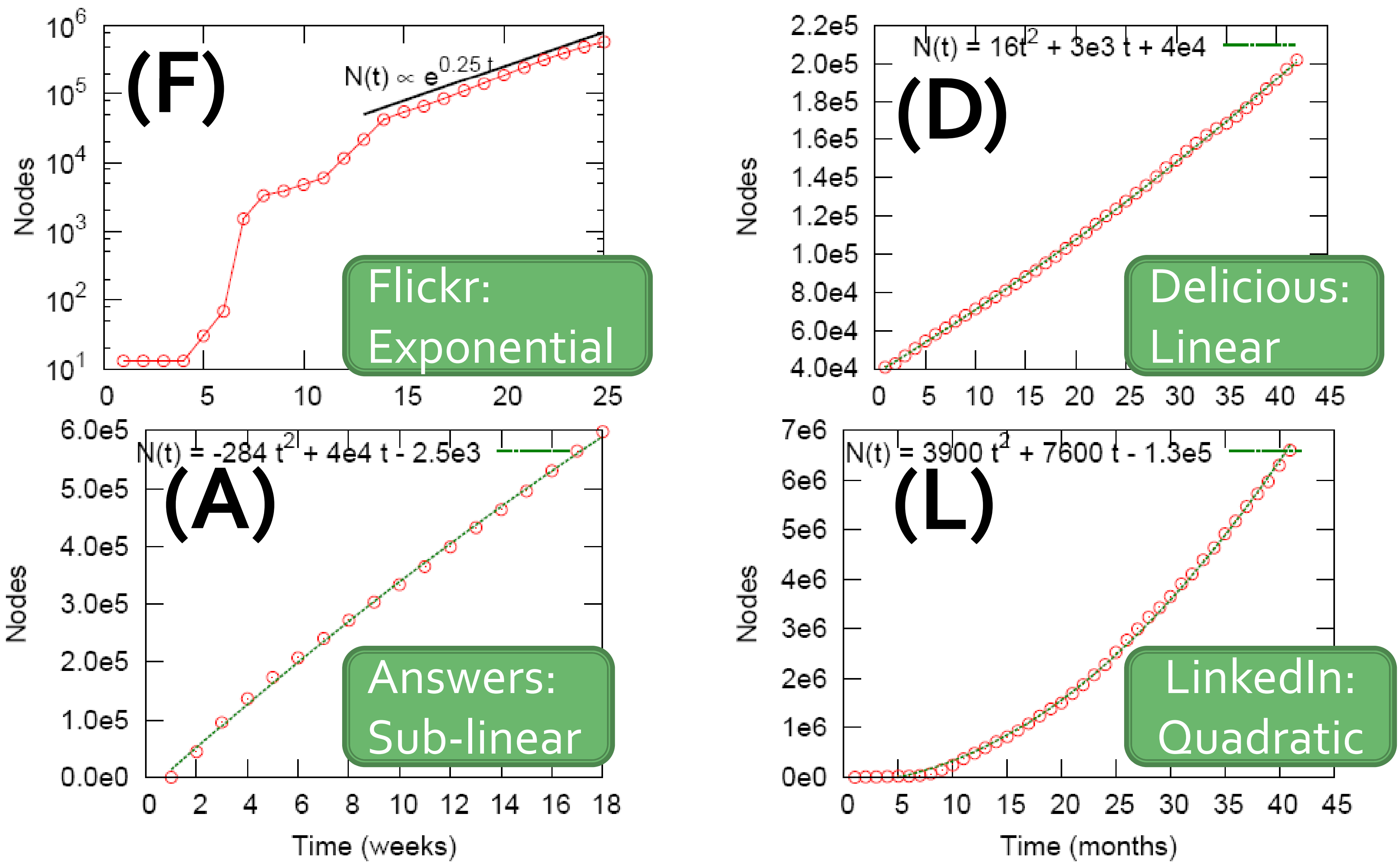
- 4 online social networks with exact edge arrival sequence
  - For every edge *(u,v)* we know exact time of the appearance t_{uv}
- Directly observe mechanisms leading to global network properties



and so on for millions…

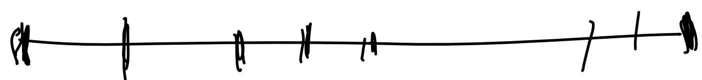
	Network	T	N	E
(F)	FLICKR (03/2003–09/2005)	621	584,207	3,554,130
(D)	DELICIOUS (05/2006–02/2007)	292	203,234	430,707
(A)	ANSWERS (03/2007–06/2007)	121	598,314	1,834,217
(L)	LINKEDIN (05/2003–10/2006)	1294	7,550,955	30,682,028

P1) How fast are nodes arriving?

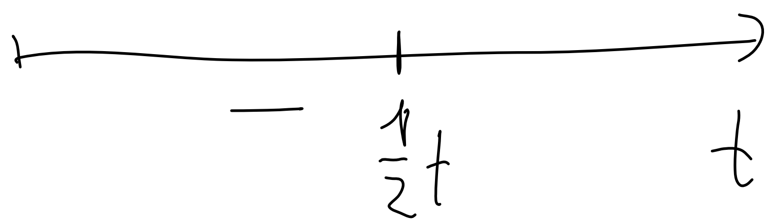


P2) When are nodes creating edges?

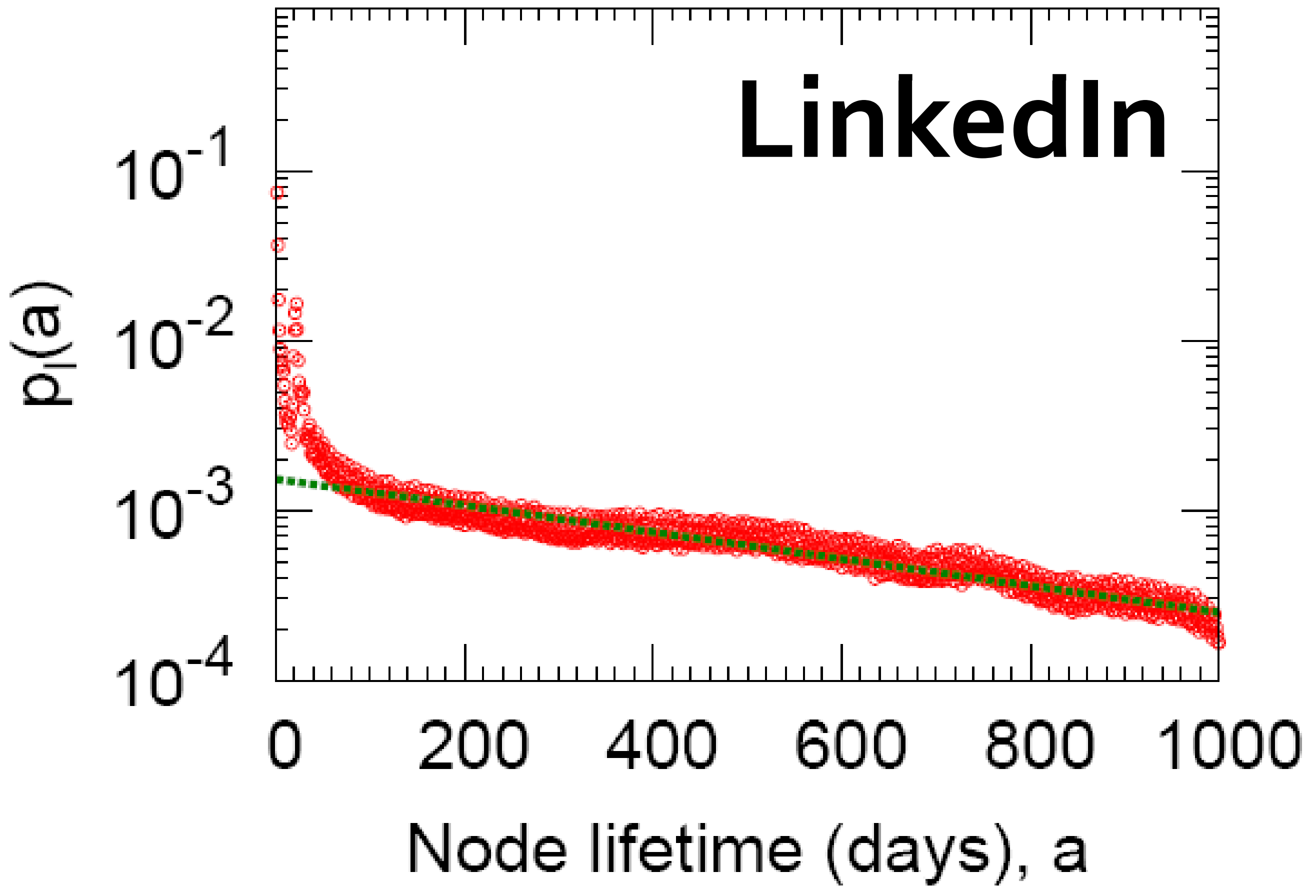
- How long do nodes live?
- How often nodes "wake up" to create edges?



- How long do nodes live?
  - Node life-time is the time between the 1st and the last edge of a node



P2) What is node lifetime?

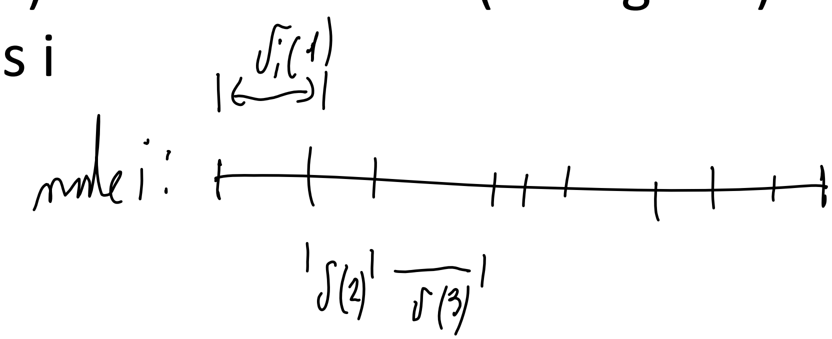


- **Lifetime a**: time between node's first and last edge

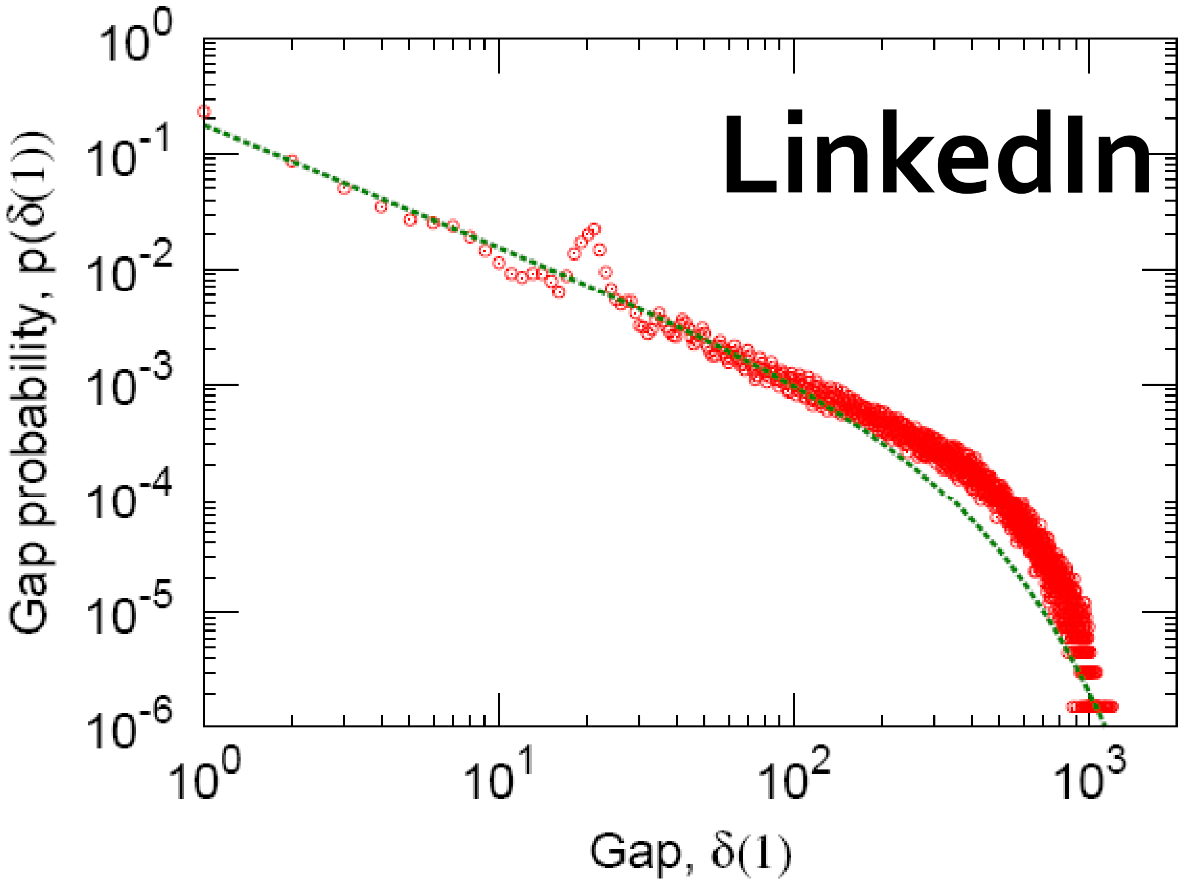
Node lifetime is **exponential**: $p(a) = \lambda \exp(-\lambda a)$

P2) When are nodes creating edges?

- How long do nodes live?
- How often nodes "wake up" to create edges?
  - **Edge gap δ(d)**: time between *d*th and *d+1*st edge of a node:
    - Let $t_i(d)$ be the creation time of d-th edge of node i
    - $\delta_i(d) = t_i(d+1) - t_i(d)$
    - Then δ(d) is a distribution (histogram) of $\delta_i(d)$ over all nodes i



2) How are edges initiated?



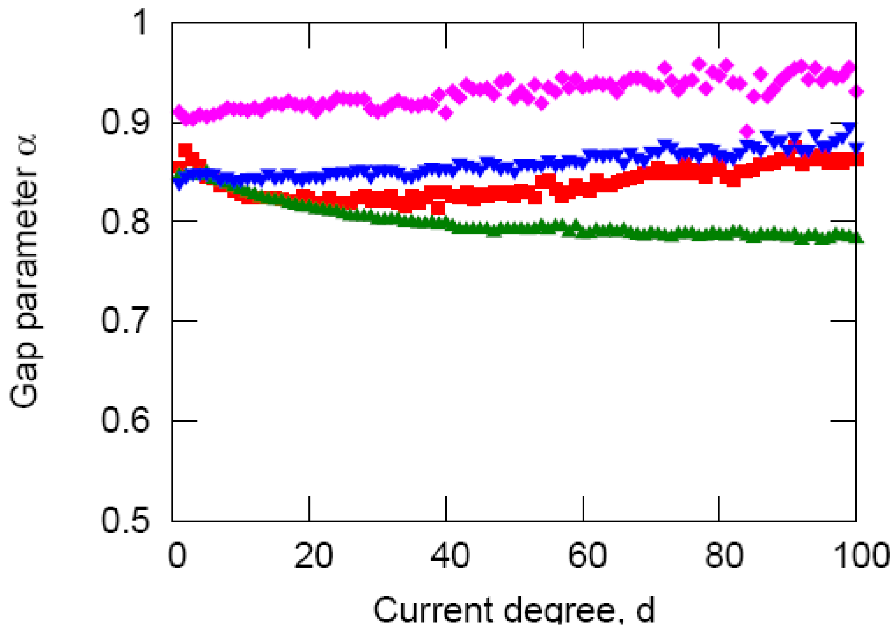
Edge gap δ(d): inter-arrival time between *d*th and *d+1*st edge

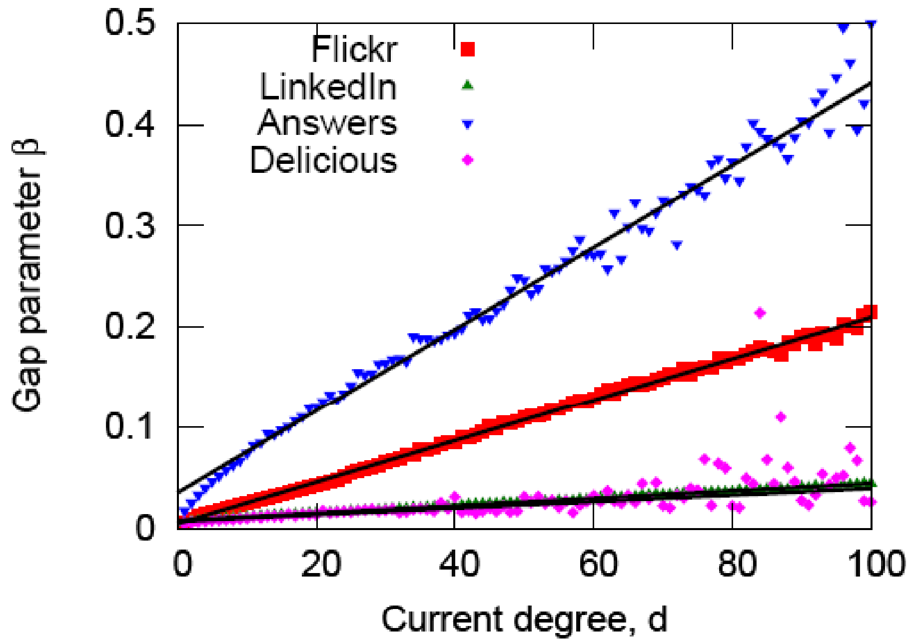
For every *d* we get a different plot

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

P2) How do α & β evolve with d?

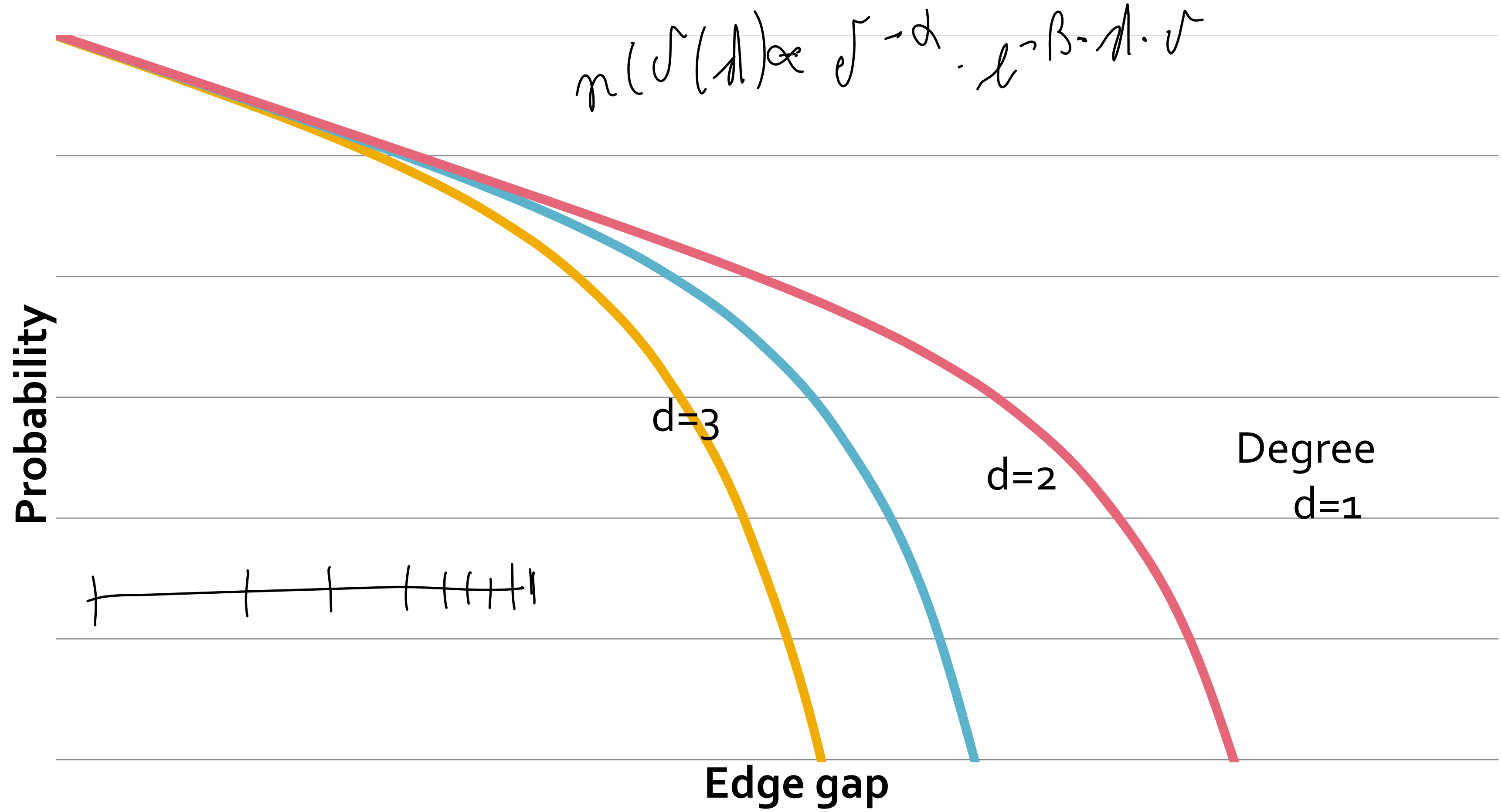
- As the degree of the node degree increases, how α and β change?





P2) Evolution of edge gap

- α is const, β linear in d – gaps get smaller with d



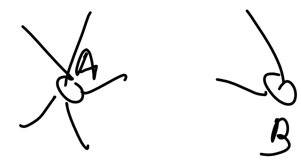
The model so far …

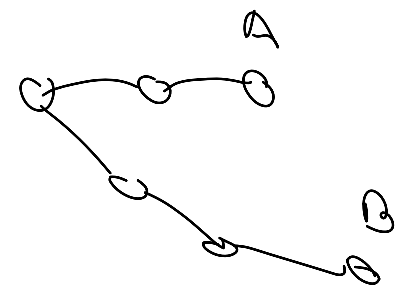
- What do we know so far?

Process	Our finding
P1) Node arrival	• Node arrival function is given
P2) Edge initiation	• Node lifetime is exponential • Edge gaps get smaller as the degree increases
P3) Edge destination	

How is destination node selected?

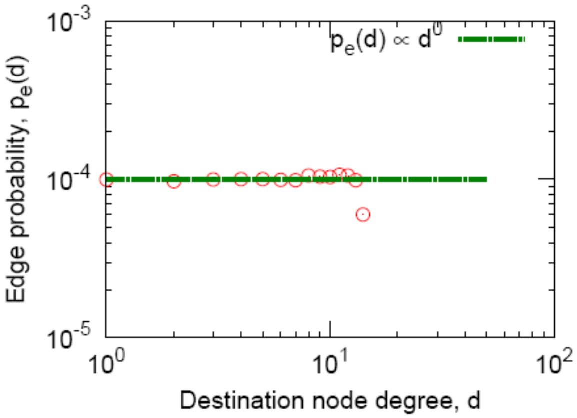
- Source node wakes up and creates an edge
- How does it find a target node?
  - What is the degree of the target?
    - Do preferential attachment really hold?
  - How many hops away if the target?
    - Are edges really local?



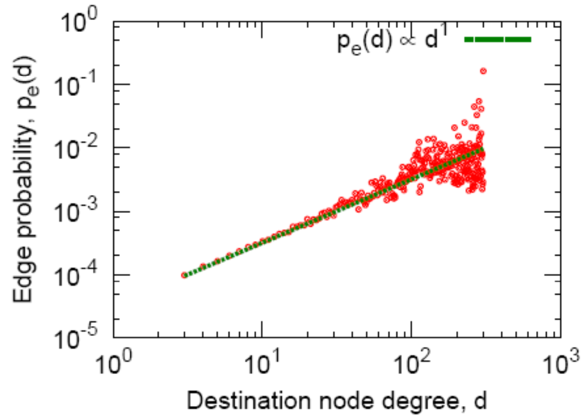


What is the degree of the target?

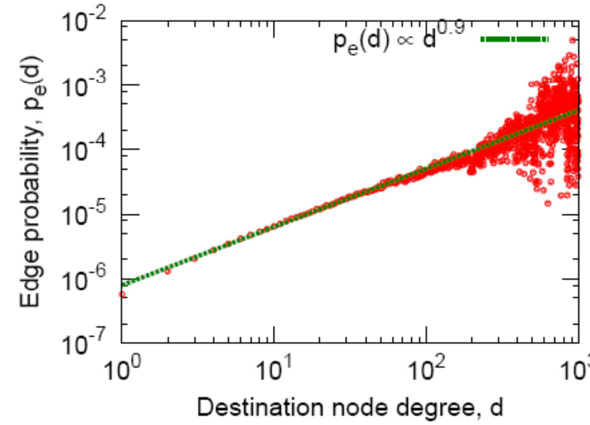
Degree of Preferential Attachment



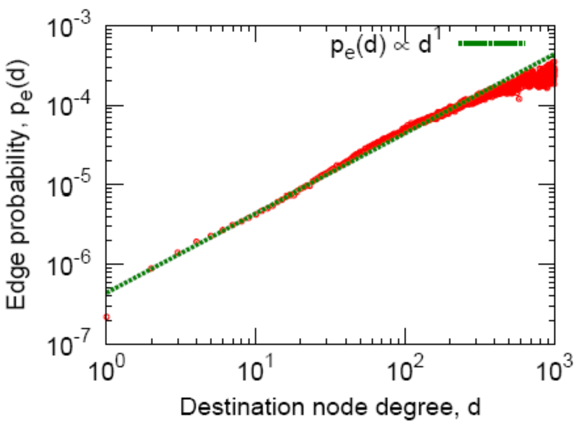
(a) G_{np}



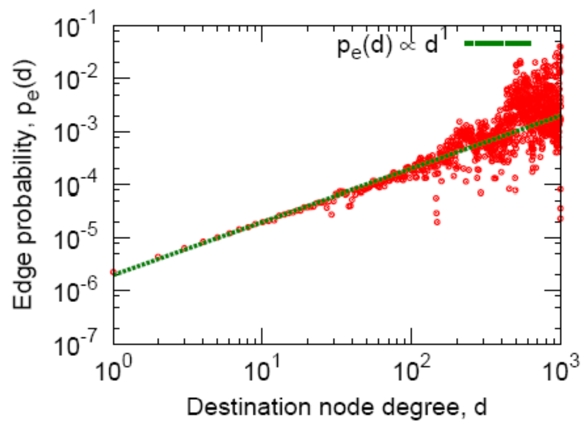
(b) PA



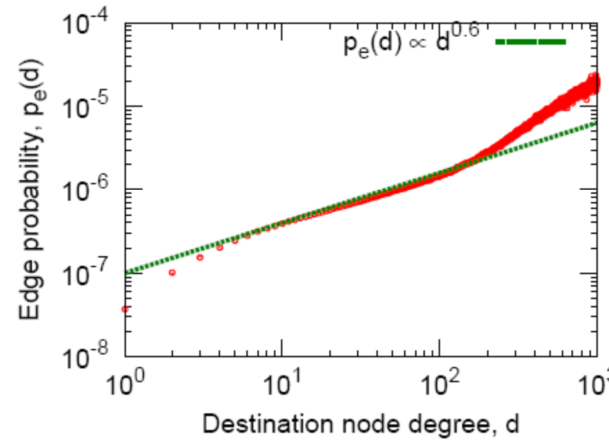
(e) ANSWERS



(c) FLICKR



(d) DELICIOUS

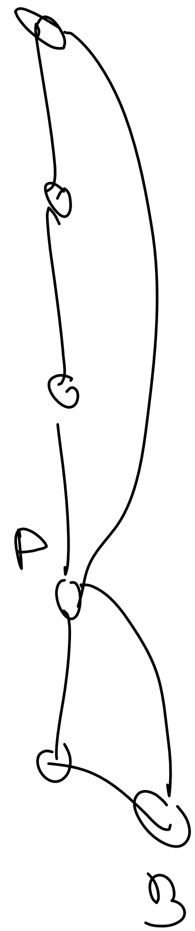


(f) LINKEDIN

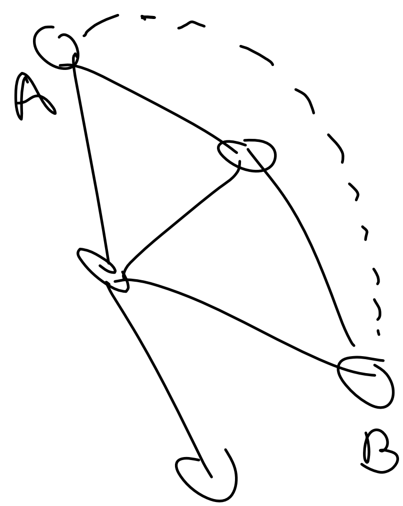
Degree of PA

$$p_e(k) \propto k^{\tau}$$

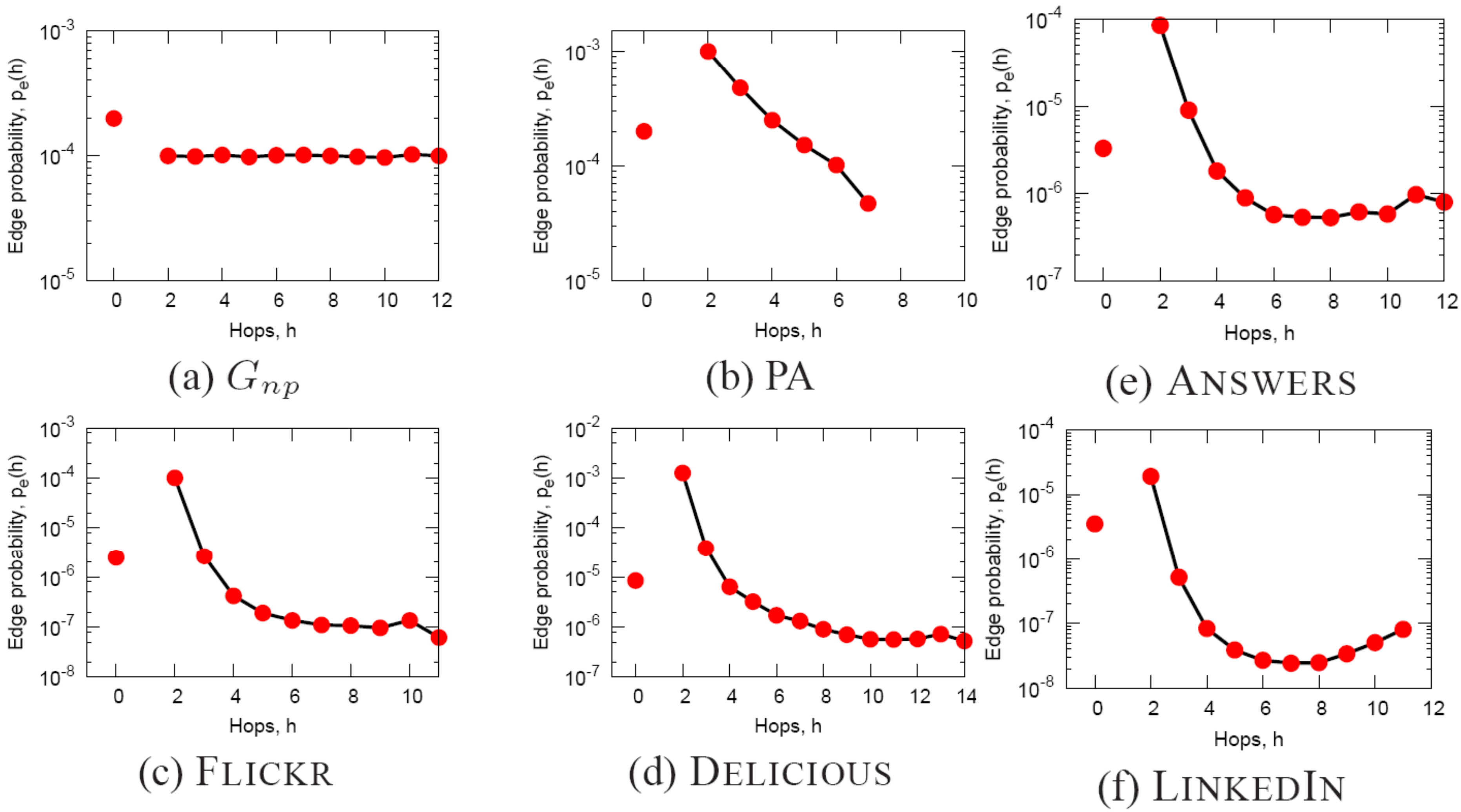
Network	τ
G_{nm}	0
PA	1
F	1
D	1
A	0.9
L	0.6



How far do edges go?



How "far" do edges go?



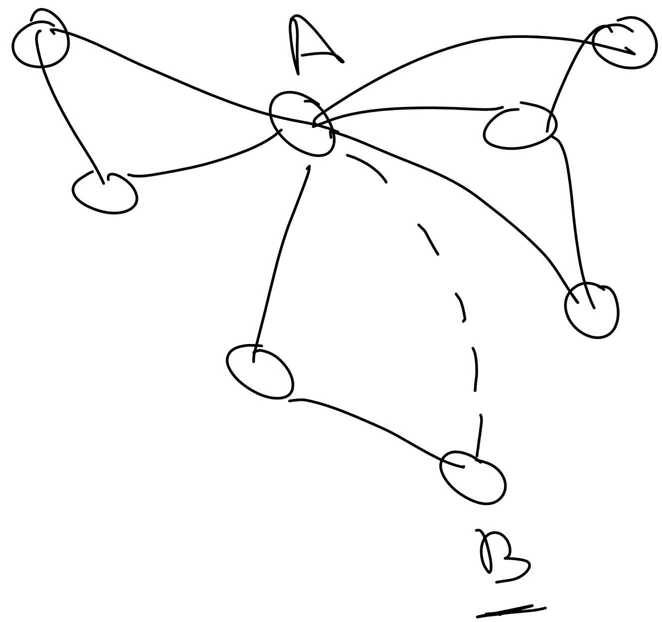
(a) G_{np} (b) PA (e) ANSWERS

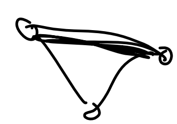
(c) FLICKR (d) DELICIOUS (f) LINKEDIN

It is all about closing triangles

Fraction of triad closing edges

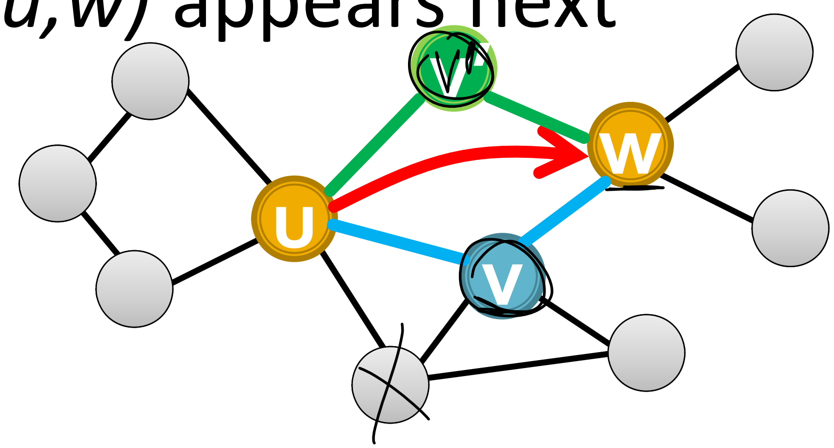
Network	% Δ
F	66%
D	28%
A	23%
L	50%





How to close triangles?

- New triad-closing edge *(u,w)* appears next
- We model this as:
  1. Choose *u*'s neighbor *v*
  2. Choose *v*'s neighbor *w*
  3. Connect *(u,w)*



- Compute edge prob. under Random-Random: *p(u,w)* = $\frac{1}{5} \cdot \frac{1}{2} + \frac{1}{5} \cdot 1$

Score of a graph = $\prod$p(u,w)

Triad closing strategies

- Improvement over the baseline:

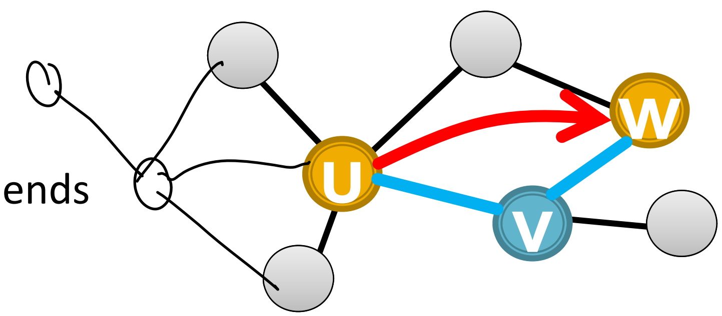
Strategy to select v (1st node)

Select w (2nd node)

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

Strategies to pick a neighbor:

- **random**: uniformly at random
- **deg**: proportional to its degree
- **com**: prop. to the number of common friends
- **last**: prop. to time since last activity
- **comlast**: prop. to **com*last**



Analysis of our model

- **Theorem:** node lifetimes and edge gaps lead to power law degree distribution
- Interesting as **temporal behavior predicts structural network property**

Idea of the proof

- Node lifetime: $p_l(a) = \lambda e^{-\lambda a}$
- Let a node have life-time a, what will its final degree D be? $\sum_{i=1}^{D} \delta(i) \le a \Rightarrow D = e^{f(\alpha,\beta) a}$

  $D = r(a)$

- $D \sim p_l(r^{-1}(D)) \left| \frac{dr^{-1}}{dD} \right|$

  $\sim D^{-(1+\lambda/f(\alpha,\beta))}$

- The two exponential functions "cancel out" and the power law part remains

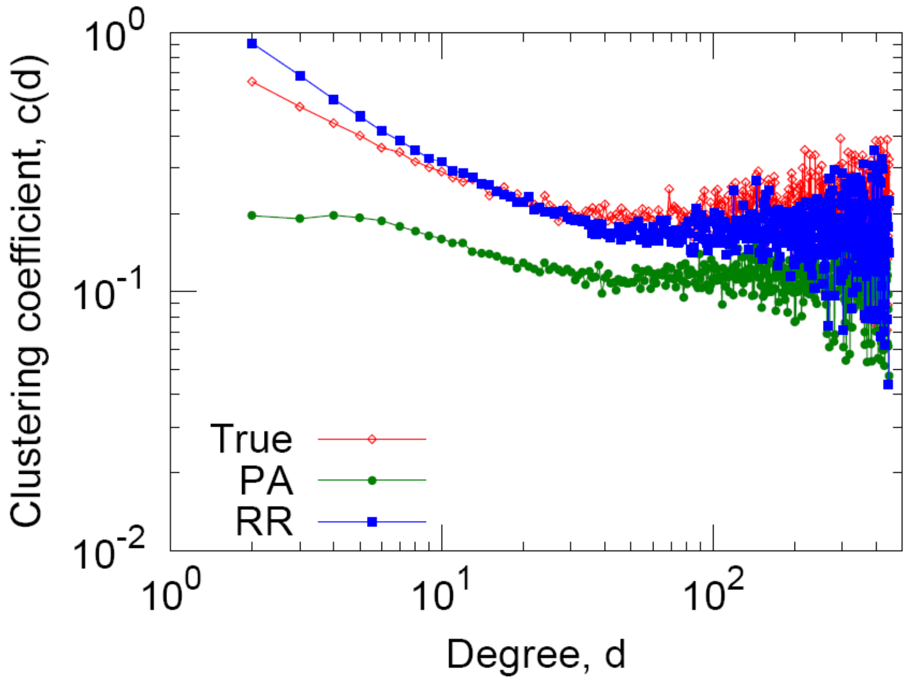
Evolving the networks

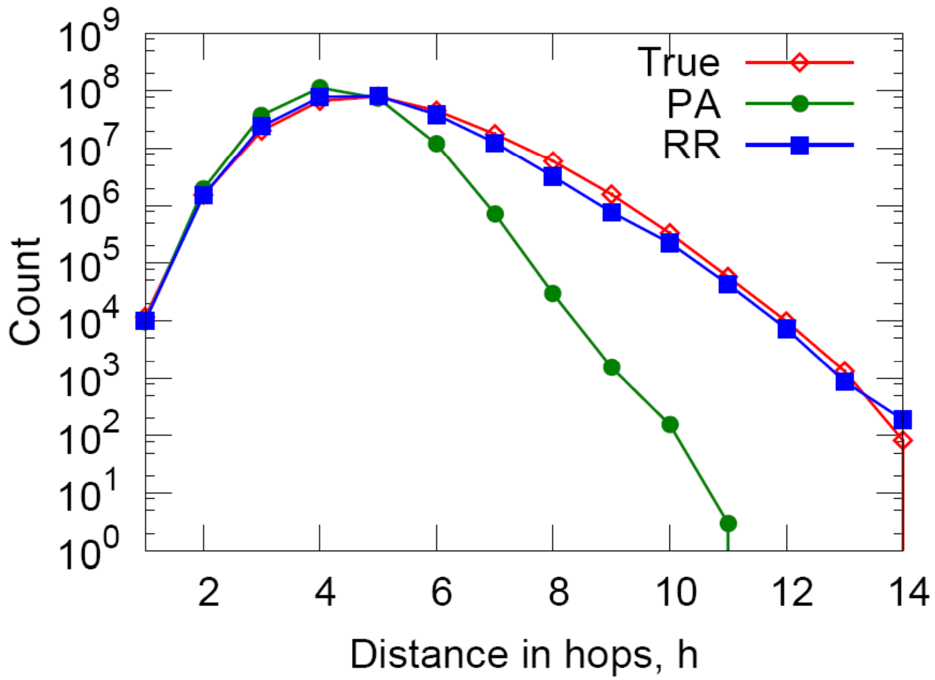
- Given our model one can take an existing network continue its evolution

	FLICKR	DELICIOUS	ANSWERS	LINKEDIN
λ	0.0092	0.0052	0.019	0.0018
α	0.84	0.92	0.85	0.78
β	0.0020	0.00032	0.0038	0.00036
true	1.73	2.38	1.90	2.11
predicted	1.74	2.30	1.75	2.08

Comparison to PA

- Take Flickr at time T/2 and then further evolve it continue evolving it using PA and our model.





Macro evolution

- How do networks evolve at the macroscopic level?
- Are there global phenomena of network evolution?
- What are analogs of the “small world” and “power-law degrees”