Macroscopic Evolution of Networks

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

Macroscopic 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"

Three questions

- What is the relation between the number of nodes $n(t)$ and number of edges $e(t)$ over time t?
- How does diameter change as the network grows?
- How does degree distribution evolve as the network grows?

[Leskovec et al. KDD 05]

Network Evolution

- $N(t)$ … nodes at time t
- $E(t)$ … edges at time t
- Suppose that

$$N(t+1) = 2 * N(t)$$

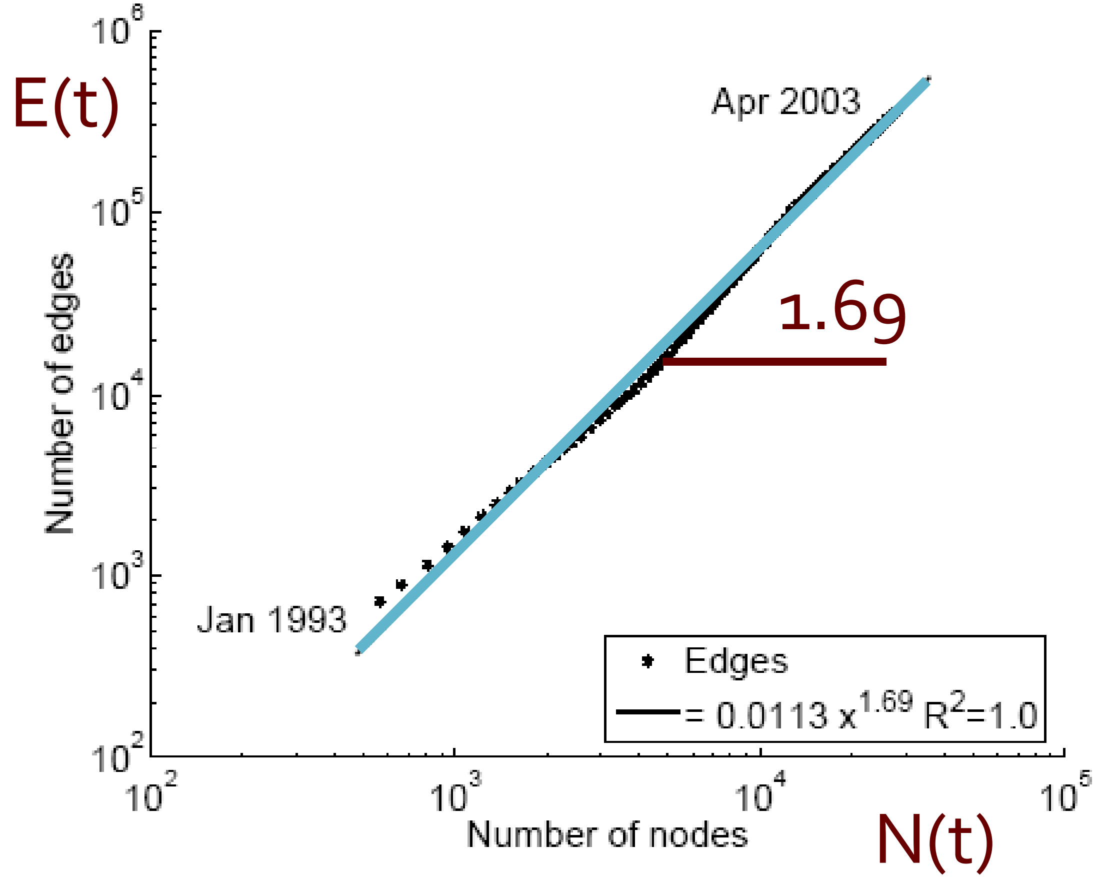
- Q: what is

$$E(t+1) = 2\ E(t)$$

- A: over-doubled!
  - But obeying the Densification Power Law

Densification – Physics Citations

- Citations among physics papers
- 1992:
  - 1,293 papers, 2,717 citations
- 2003:
  - 29,555 papers, 352,807 citations
- For each month *M*, create a graph of all citations up to month *M*



Densification Power Law

- Densification Power Law
  - the number of edges grows faster than the number of nodes – average degree is increasing

$$E(t) \propto N(t)^a$$

or equivalently

$$\frac{\log(E(t))}{\log(N(t))} = const$$

a … densification exponent: $1 \leq a \leq 2$:

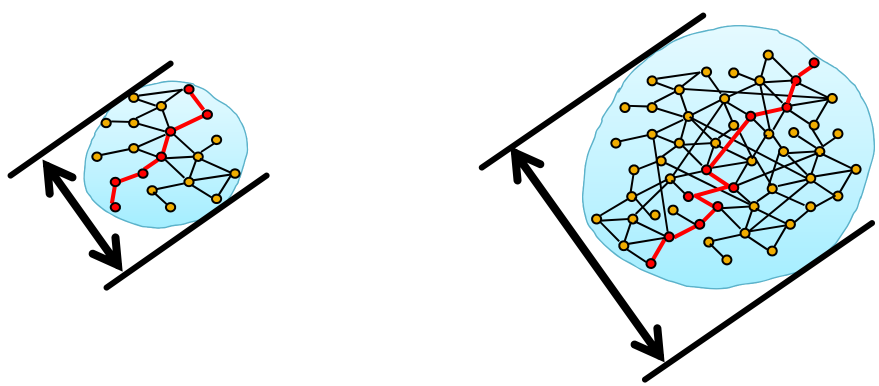
- a=1: linear growth – constant out-degree (traditionally assumed)
- a=2: quadratic growth – clique

$a \approx 1.67$

[Leskovec et al. KDD 05]

Evolution of the Diameter

- Prior work on Power Law graphs hints at Slowly growing diameter:
  - diameter ~ O(log N)
  - diameter ~ O(log log N)



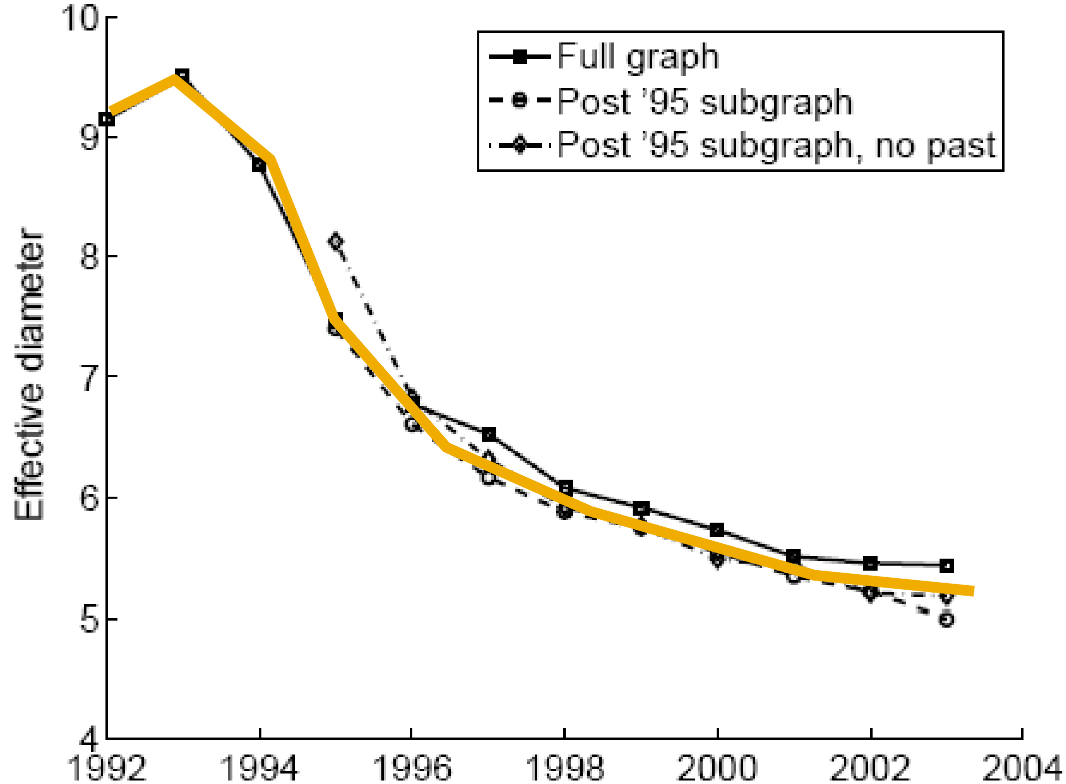
- What is happening in real data?

[Leskovec et al. KDD 05]

Diameter – ArXiv citation graph

- Citations among physics papers
- 1992 –2003
- One graph per year

diameter

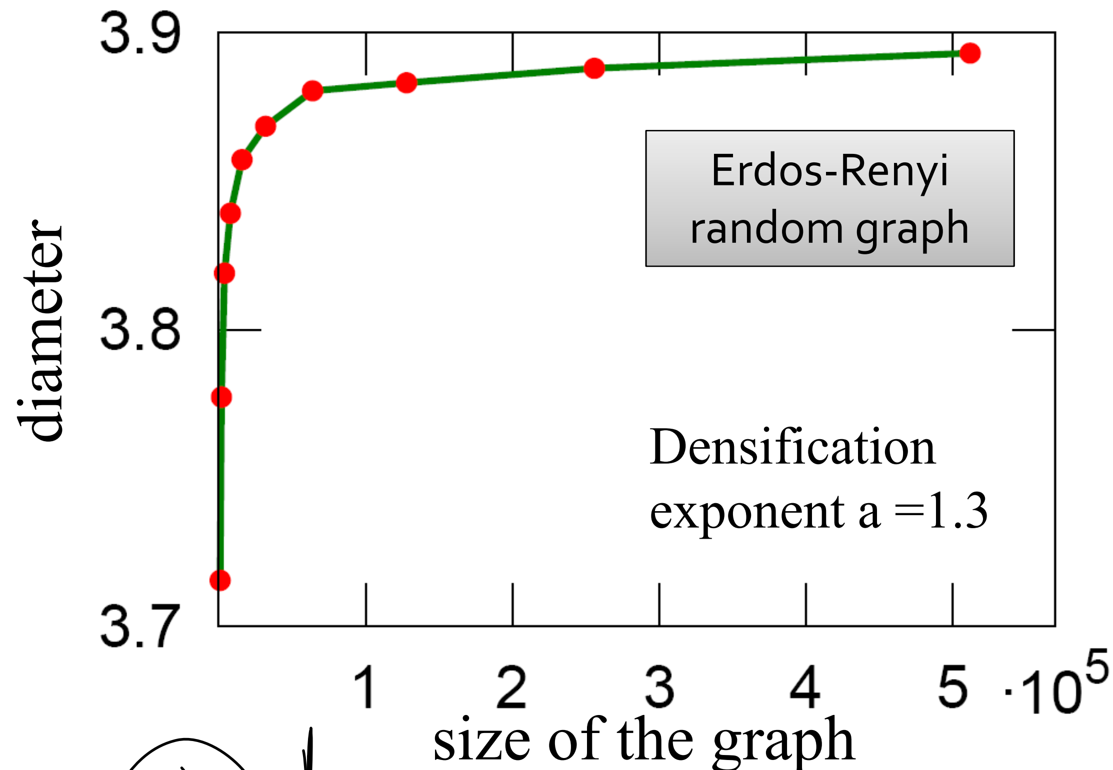


time [years]

[Leskovec et al. TKDD 07]

Diameter of a densifying G_{np}

Is shrinking diameter just a consequence of densification?



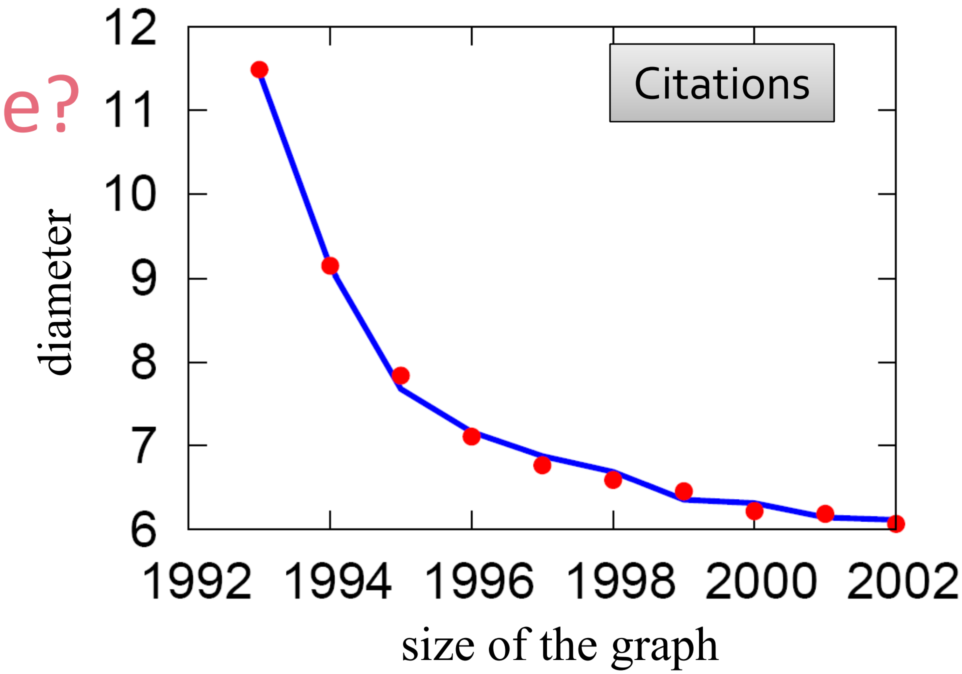
Densifying random graph has increasing diameter ⇒ There is more to shrinking diameter than just densification

Diameter of a rewired network

Is it the degree sequence?

Compare diameter of a:

- True network (red)
- Random network with the same degree distribution (blue)



Densification + degree sequence give shrinking diameter

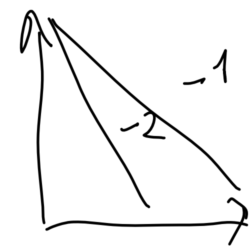
[Leskovec et al. TKDD 07]

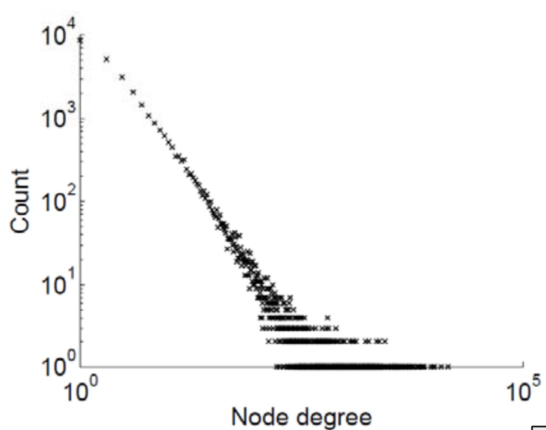
Connecting Degrees & Densification

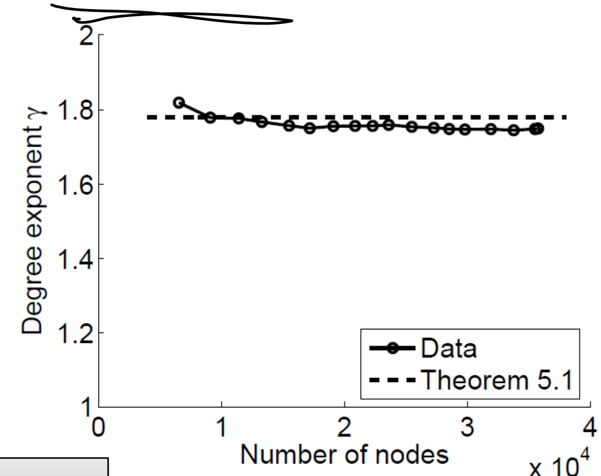
- How does degree distribution evolve to allow for densification?
- 2 Options:
- 1) Degree exponent γ is constant:
  - Fact 1: For degree exponent $1<\gamma<2$: $a = 2/\gamma$

$\gamma < k \quad k \in \mathbb{N}$

k-1st moment won't exist





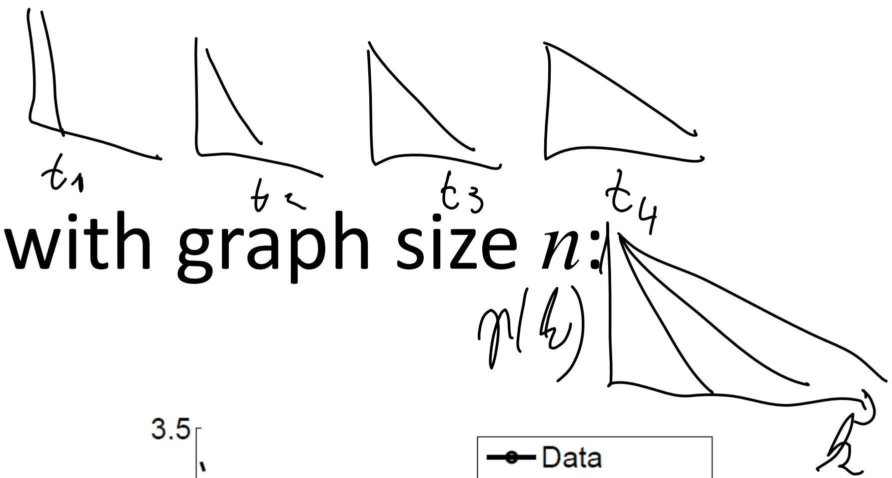


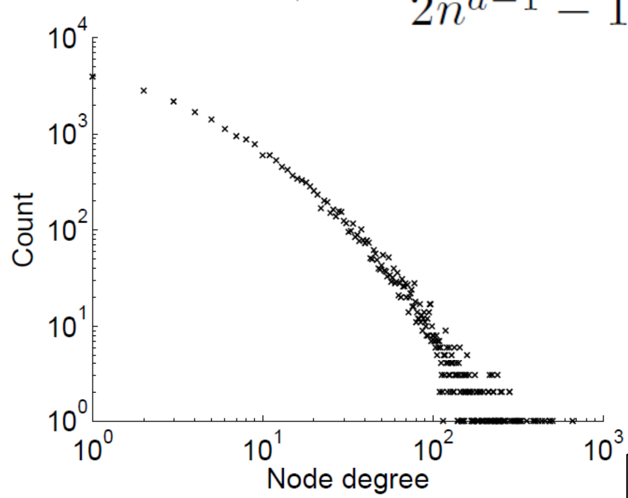
Email network

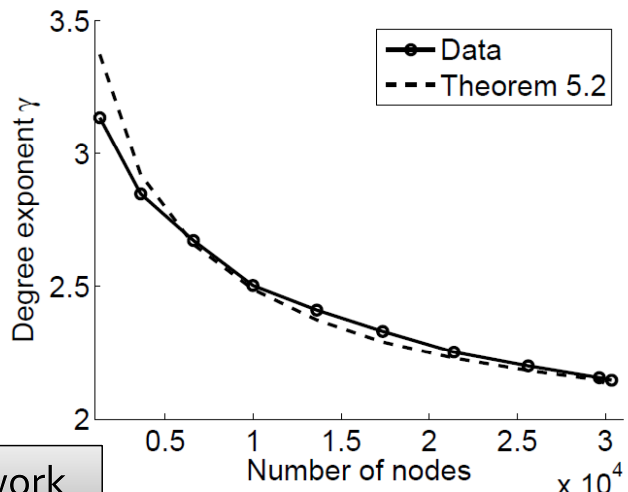
[Leskovec et al. TKDD 07]

Connecting Degrees & Densification

- How does degree distribution evolve to allow for densification?
- 2 Options:
- 2) Exponent γ_n evolves with graph size n:
  - Fact 2: $\gamma_n = \frac{4n^{a-1} - 1}{2n^{a-1} - 1}$







Citation network

[Leskovec et al. TKDD 07]

Patterns Hold in Many Graphs

- Densification and Shrinking diameter can be observed in many real life graphs:

Dataset	Nodes	Edges	Time	DPL exponent
Arxiv HEP–PH	30,501	347,268	124 months	1.56
Arxiv HEP–TH	29,555	352,807	124 months	1.68
Patents	3,923,922	16,522,438	37 years	1.66
AS	6,474	26,467	785 days	1.18
Affiliation ASTRO–PH	57,381	133,179	10 years	1.15
Affiliation COND–MAT	62,085	108,182	10 years	1.10
Affiliation GR-QC	19,309	26,169	10 years	1.08
Affiliation HEP–PH	51,037	89,163	10 years	1.08
Affiliation HEP–TH	45,280	68,695	10 years	1.08
Email	35,756	123,254	18 months	1.12
IMDB	1,230,276	3,790,667	114 years	1.11
Recommendations	3,943,084	15,656,121	710 days	1.26

Densification – Possible Explanation

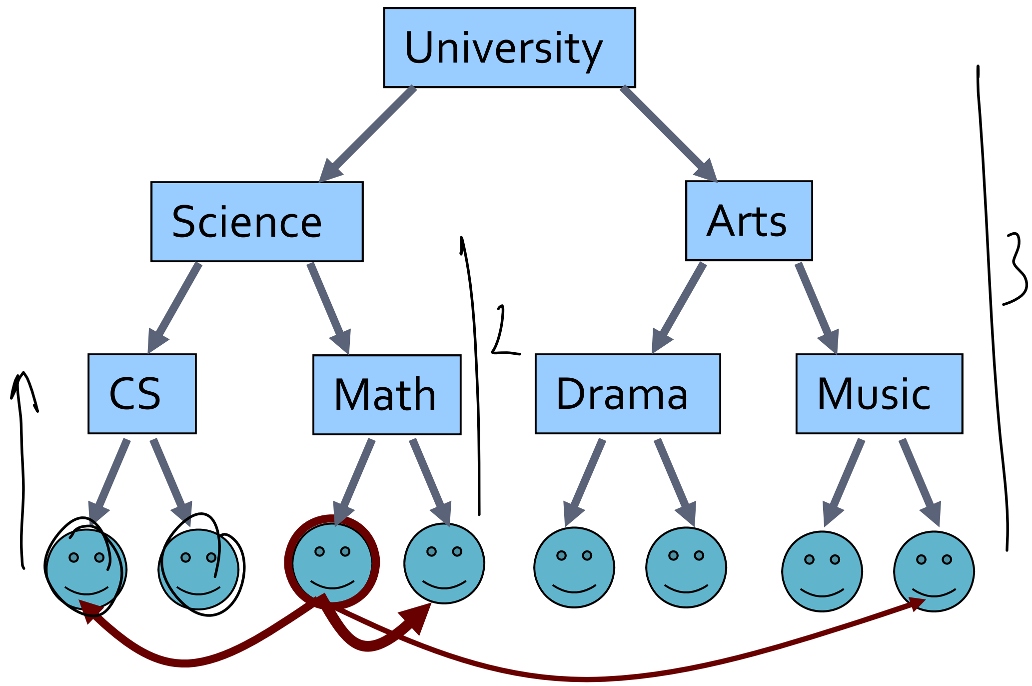
- What explains the Densification Power Law and Shrinking diameters
- Can we find a simple model of local behavior, which naturally leads to the observed phenomena?
- Yes! Community Guided Attachment

[Leskovec et al. TKDD 07]

Community structure

- Let's assume the community structure
- One expects many within-group friendships and fewer cross-group ones

- How hard is it to cross communities?



Self-similar university community structure

Main Assumption

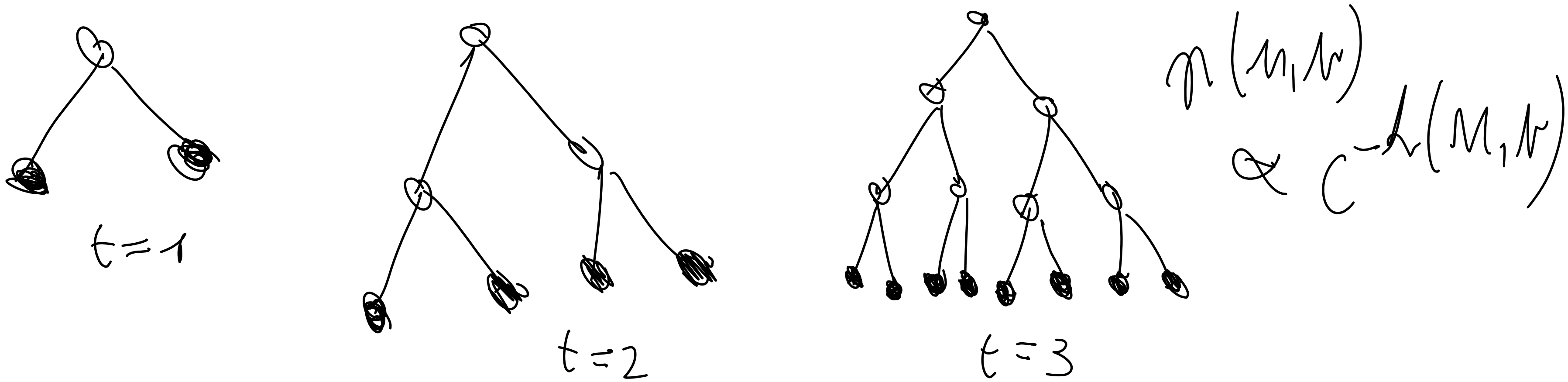
- Assume the cross-community linking probability of nodes at tree-distance h is scale-free
- Then the cross-community linking probability is:

$$f(h) = c^{-h}$$

where: $c \geq 1$ … the *Difficulty constant*
h … tree-distance

Community guided attachment

- $n = 2^k$ nodes reside in the leaves of the b-way community hierarchy (assume $b=2$)



- Each node then independently creates edges based the community hierarchy, $f(h) = c^{-h}$
- How many edges m are in a graph of n nodes?

Densification Power Law

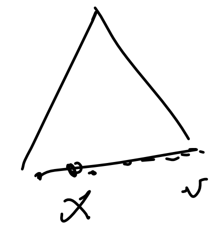
[Leskovec et al. TKDD 07]

- Claim: Community Guided Attachment random graph model, the expected out-degree of a node is proportional to

$$1.\quad n^{1-\log_b(c)} \quad \textit{if } 1 \le c < b$$
$$2.\quad \log_b(n) \quad \textit{if } c = b$$
$$3.\quad \textit{constant} \quad \textit{if } c > b$$

Proof

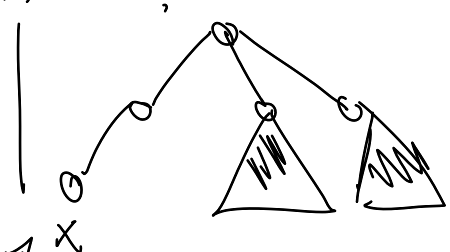
- What is expected out-degree of a node x?



$$\bar{d}_x = \sum_{v \neq x} f(h(v,x)) = \sum_h \#\text{nodes at } h \cdot \text{prob of link to } h$$

How many nodes are at dist. h?

$(b-1)\, b^{h-1}$ prob link c^{-h}



$$= \sum_{h=1}^{\log_b n} (b-1)\, b^{h-1} \cdot c^{-h} = \frac{b-1}{c} \sum_{h}^{\log_b n} \left(\frac{b}{c}\right)^{h-1}$$

Proof

IF $c > b$ then $\sum_{i}^{\log n} \left(\frac{b}{c}\right)^i$ converges

$c < b$ then -||- $\propto$ largest element

$\propto \left(\frac{b}{c}\right)^{\log n}$

FACT:
$x^{\log y} = y^{\log x}$

OPLEXP:

$\bar{d} \propto n^{\log_b \frac{b}{c}} = n^{1-\log_b c} = n^{\alpha}$

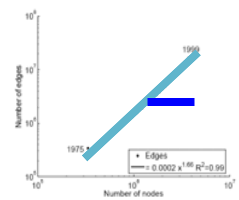
$E = n \cdot n^{1-\log_b c} = n^{2-\log_b c}$

Densification Power Law (2)

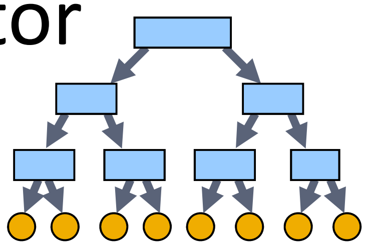
- Claim: The Community Guided Attachment leads to Densification Power Law with exponent

$$a = 2 - \log_b(c)$$

- a … densification exponent $E(t) \propto N(t)^a$



- b … community tree branching factor
- c … difficulty constant, $1 \leq c \leq b$



Difficulty Constant

- DPL:

$$a = 2 - \log_b(c)$$

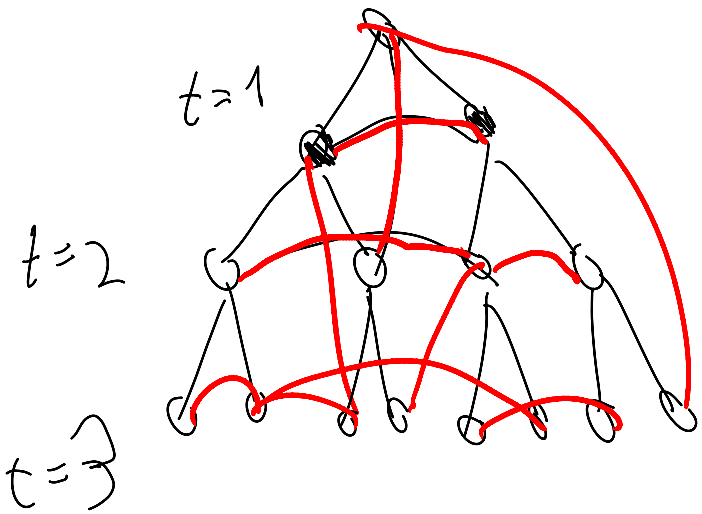
- Gives any non-integer Densification exponent
- If $c = 1$: easy to cross communities
  - Then: $a=2$, quadratic growth of edges – near clique
- If $c = b$: hard to cross communities
  - Then: $a=1$, linear growth of edges – constant out-degree

$p(v, u) = p^{-h}$

[Leskovec et al. TKDD 07]

Extension of the model

- Community tree evolves by a complete new level of nodes being added in each time step
  - Now nodes also reside in non-leafs of the tree

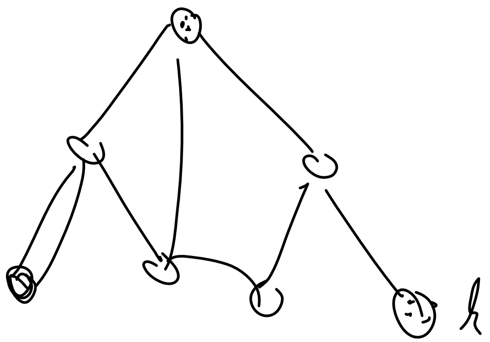


- Can show:
  - Densification + power-law degree distribution

[Leskovec et al. TKDD 07]

B-way Branching Tree

- Topics of web pages:
  - Each node represents k pages
  - Nodes everywhere in the tree
  - Link up h levels with prob. ~ γ^{-h}

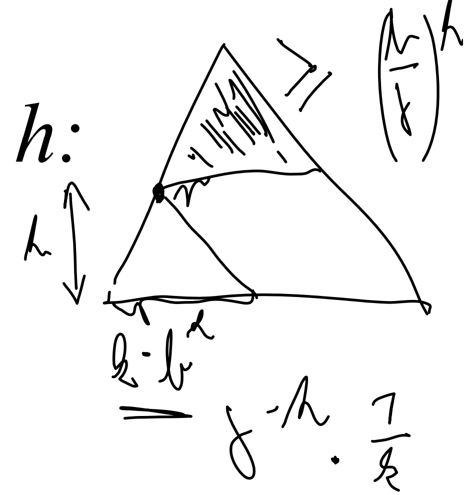


- Simplification: only bottom level creates links

[Leskovec et al. TKDD 07]

Expected degree at height h

- Expected degree of a node at height h:



- Fraction of nodes with degree $>d$ as a function of d?

Power-laws vs. Log-normal

- Normal $N(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - Limit of avg. of indep. draws from a random var from any distribution
- X is lognormally distributed if $Y = \ln X$ is normally distributed
- Log-normal density
  $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} d(\ln x) = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{x} \cdot e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$

Lognormal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-(\ln x - \mu)^2 / 2\sigma^2}$$

- Properties:
  - Finite mean/variance.
  - Skewed: mean > median > mode
  - Multiplicative:
    - X_1 lognormal, X_2 lognormal implies $X_1 X_2$ lognormal.

[Mitzenmacher]

Power-law – Lognormal: similarities

- Looks similar to power-law on log-log plot:
- Power-law has linear log-density

$$\ln f(x) = -\alpha \ln x + \ln c$$

- For large σ, lognormal has nearly linear log-density:

$$\ln f(x) = -\ln x - \ln \sqrt{2\pi}\sigma - \frac{(\ln x - \mu)^2}{2\sigma^2}$$

- Similarly, both have near linear log-ccdfs.
- Question: how to differentiate them empirically?

Lognormal vs. Power Law

- Question: Is this distribution lognormal or a power law?
  - Reasonable follow-up: Does it matter?
- Generative model:
  - Size-independent + growth
  - Size (e.g., popularity, degree) start at one and changes in each step by a random multiplicative factor F

Generative Models: Lognormal

- Start with an organism of size X_0.
- At each time step, size changes by a random multiplicative factor. $$X_t = F_{t-1} X_{t-1}$$
- <u>Claim:</u> If F_t is taken from a lognormal distribution, each X_t is lognormal.

Generative Models: Lognormal

- <u>Claim:</u> If F_t is taken from a lognormal distribution, each X_t is lognormal.
- If F_t are independent, identically distributed then (by CLT) X_t converges to lognormal distribution.

$$X_t = X_{t-1} F_{t-1} = \prod_i F_i = \sum_i \log F_i$$

$$X_0 = 1$$

- $\ln X_t$ = size at time t = $\ln F_1 + \ln F_2 + \ldots + \ln F_t$
- $\ln X_t = \sum \ln F_j$ ~ normal dist → X_t is lognormal

[Mitzenmacher]

BUT!

- If there exists a lower bound:

$$X_t = \max(\varepsilon, F_{t-1} X_{t-1})$$

- then X_t converges to a power law distribution [Champernowne, 1953]

- Lognormal model easily pushed to a power law model

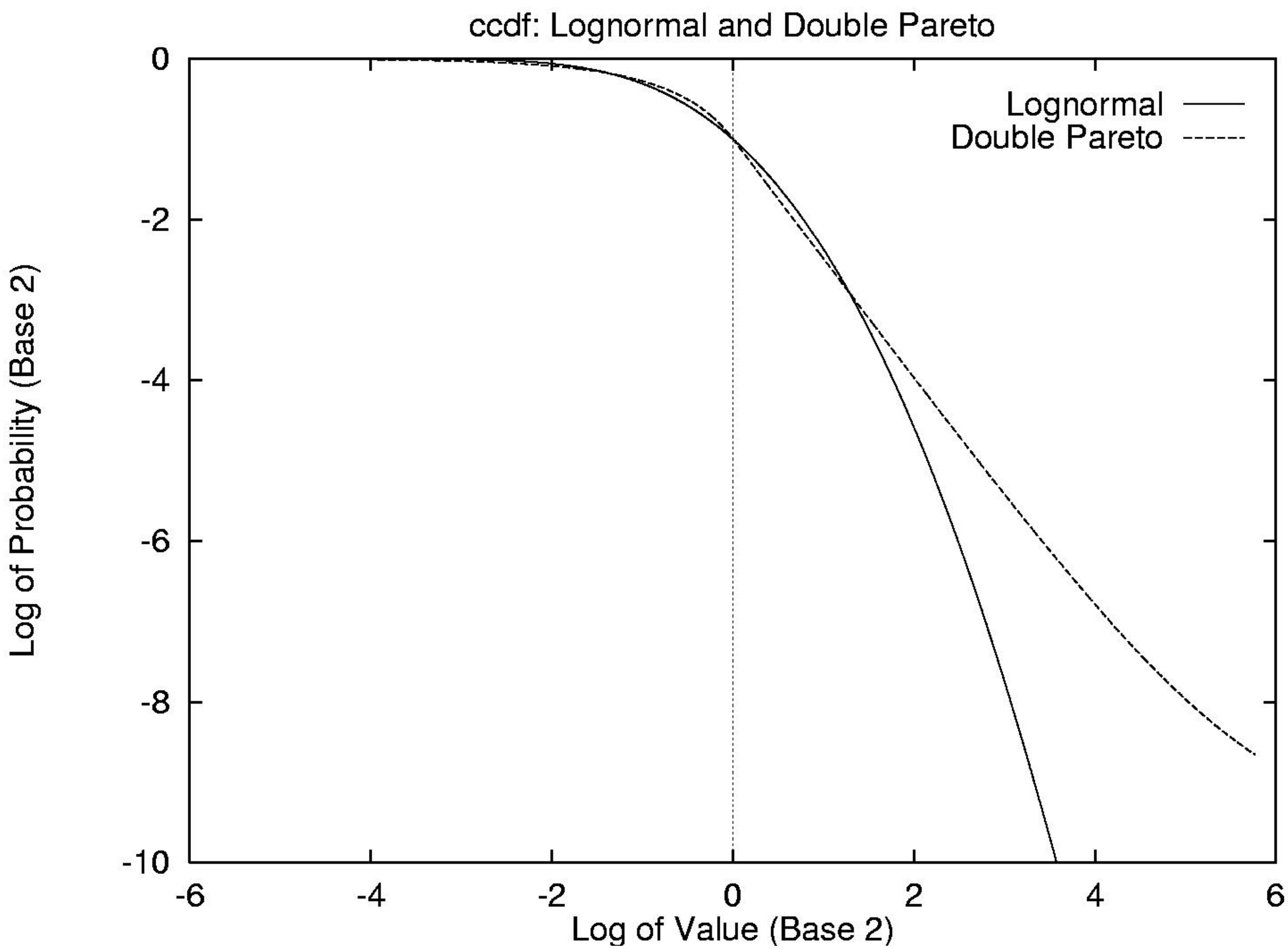
[Mitzenmacher]

Double Pareto Distributions

- Consider continuous version of lognormal generative model:
  - At time t, log X_t is normal with mean μt and variance $\sigma^2 t$
- Suppose observation time is distributed exponentially
  - E.g., When Web size doubles every year.
- Resulting distribution is Double Pareto.
  - Between lognormal and Pareto.
  - Linear tail on a log-log chart, but a lognormal body.

[Mitzenmacher]

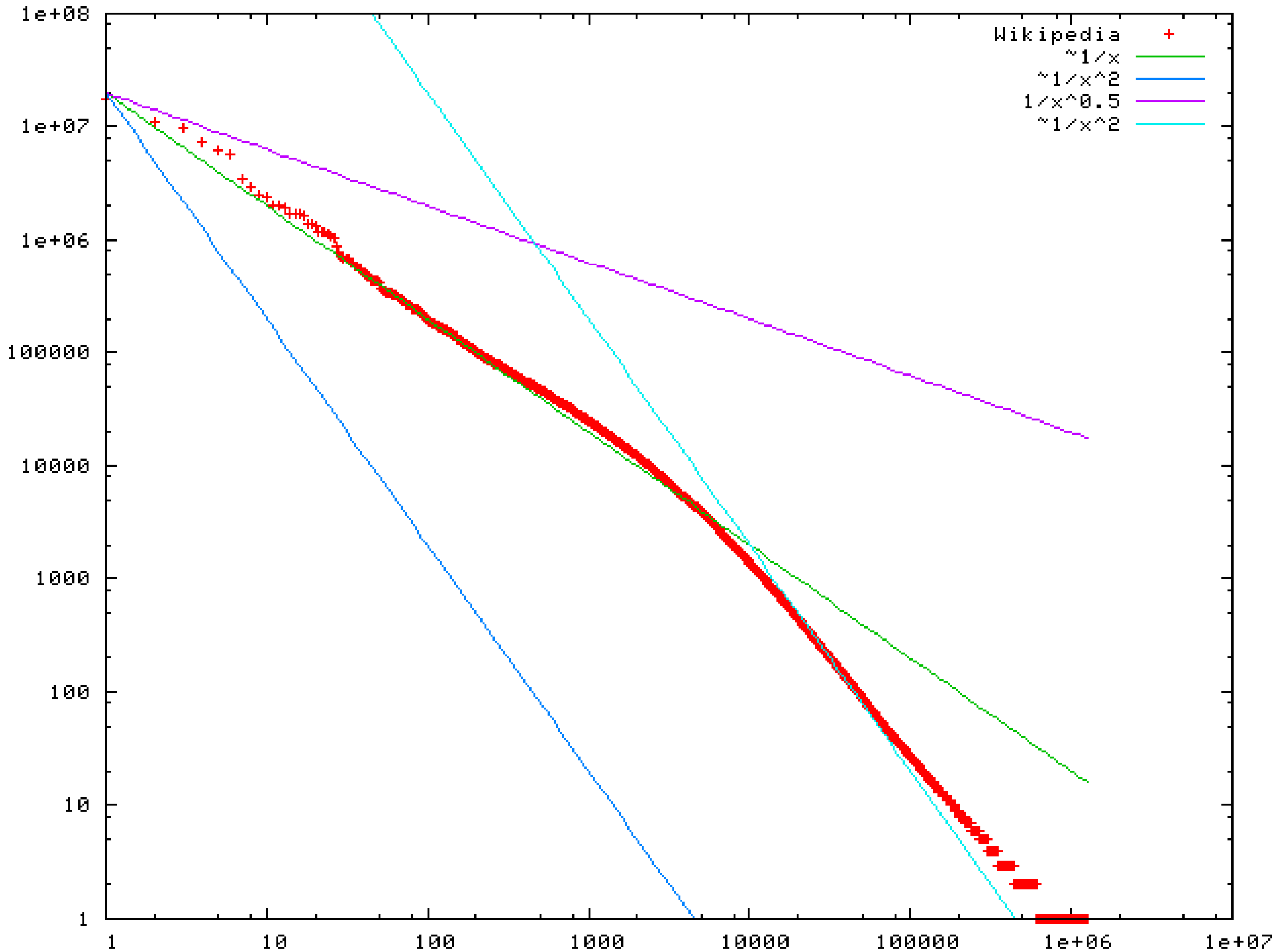
Lognormal vs. Double Pareto



More of power-laws

- Hierarchical model that produces power-laws
- Zipf's law – how does the frequency of a word depend on its rank?
- Empirically: $\text{rank} \sim 1/\text{freq}$

Rank vs. freq of words on Wikipedia



Optimization Model: Power Law

- Mandelbrot experiment: design a language over a d-ary alphabet to optimize information per character.
  - Probability of j^{th} most frequently used word is p_j.
  - Length of j^{th} most frequently used word is c_j.
- Average information per word:

$$H = -\sum_j p_j \log_2 p_j$$

- Average characters per word:

$$C = \sum_j p_j c_j$$

- Optimization leads to power law.

Explanation: Miller '57

- Alphabet of c letters plus spacebar
- Monkey sits at a typewriter and randomly pushes keys
  - Uniformly random sequence of c+1 symbols – spaces mark word boundaries
- Word ranks
  - 1, 2, …, c — 1-letter words
  - c+1, …, c^2+c+1 — 2-letter words
  - c^2+c+1+1, … — 3-letter words

Explanation

- Each j-letter word has frequency $(c+1)^{-(j+1)}$

- Pick representative point c^j for j-letter word
- Rank c^j is a j-letter word with freq $(c+1)^{-(j+1)}$

Explanation

- Write $(c+1)^{-(j+1)}$ as a function of rank