The Origin of Power Laws in Internet Topologies Revisited

Authors: Q. Chen et al.

Presenter: Visa Holopainen

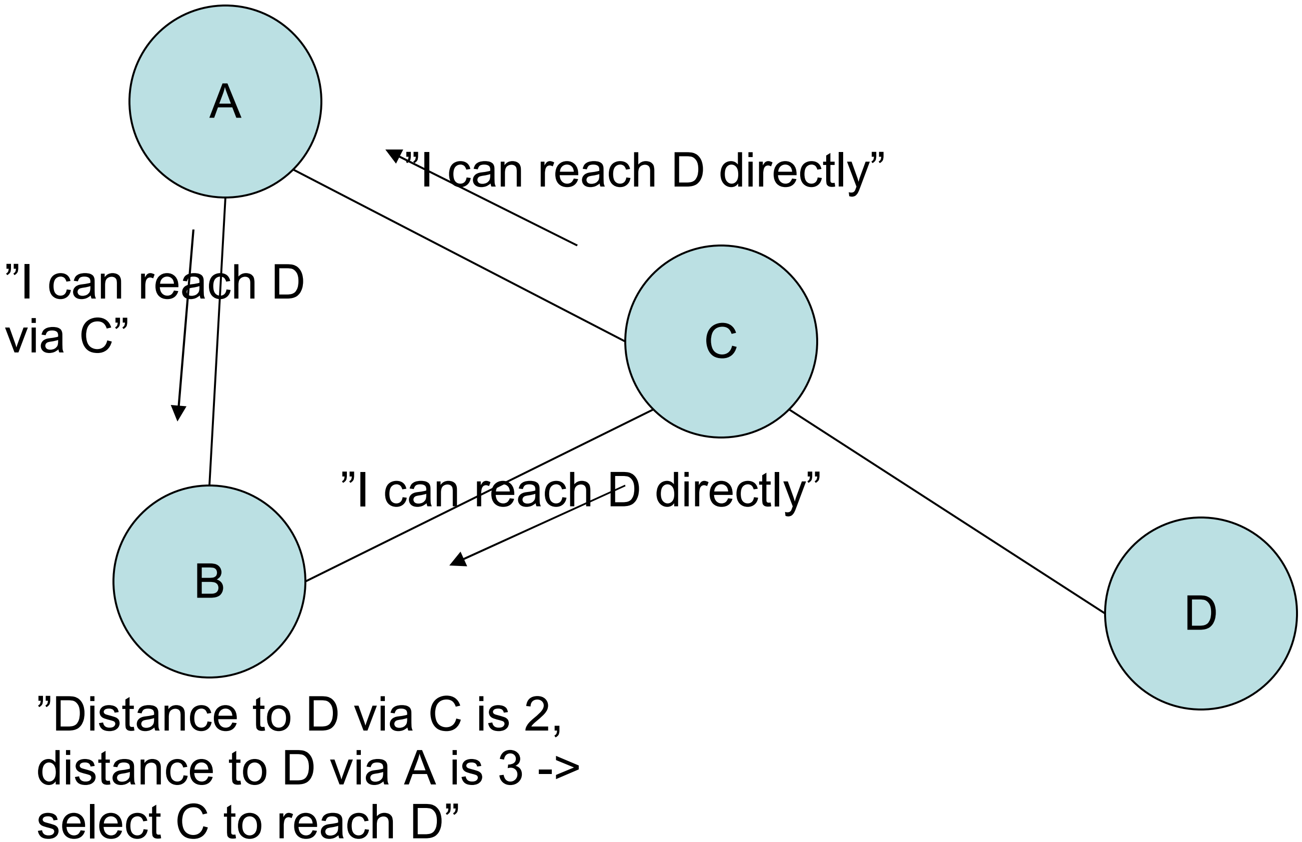
The actual topic:

How Internet evolves over time – models and explanations

Why Internet growth models?

- Predicting future from the past
- Simulations of new algorithms
- Anomaly detection – abnormal evolution
- Graph sampling – many real world graphs are too large to deal with

Border Gateway Protocol basics



Internet basics

- Internet is composed of Autonomous Systems (AS) that speak BGP to each other
  - Currently about 21000
- Examples: University of Jyväskylä (Commercial Internet Services), CSC, Oulu Telephone Company, Elisa Oyj, Akamai Technologies, AT&T

- *Question*: Can we reliably say something about AS-level connectivity based on BGP data?

Background – BA-model

- 3 mechanisms by which power-law networks evolve:

١. **Incremental growth.** Networks develop over time by adding nodes and links to existing graph.

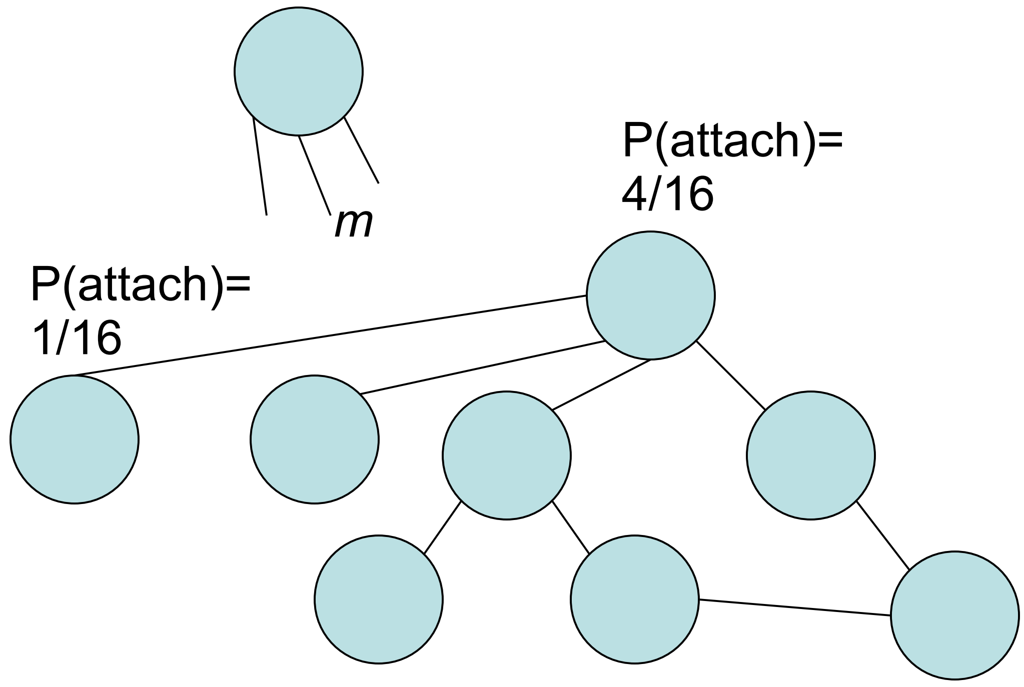
٢. **Preferential connectivity.** Probability Π that a new vertex will be connected to vertex *i* depends linearly on degree k_i

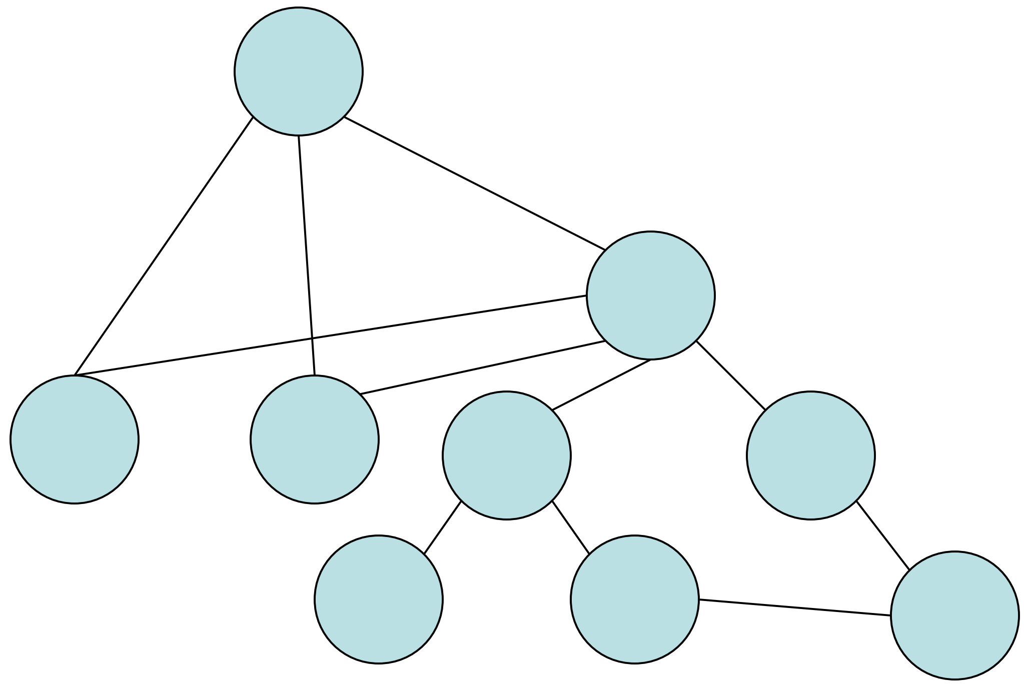
  - $\Pi(k_i) = k_i/\Sigma_j k_j$

٣. **Re-wiring.** At each iteration re-connect *m* existing links

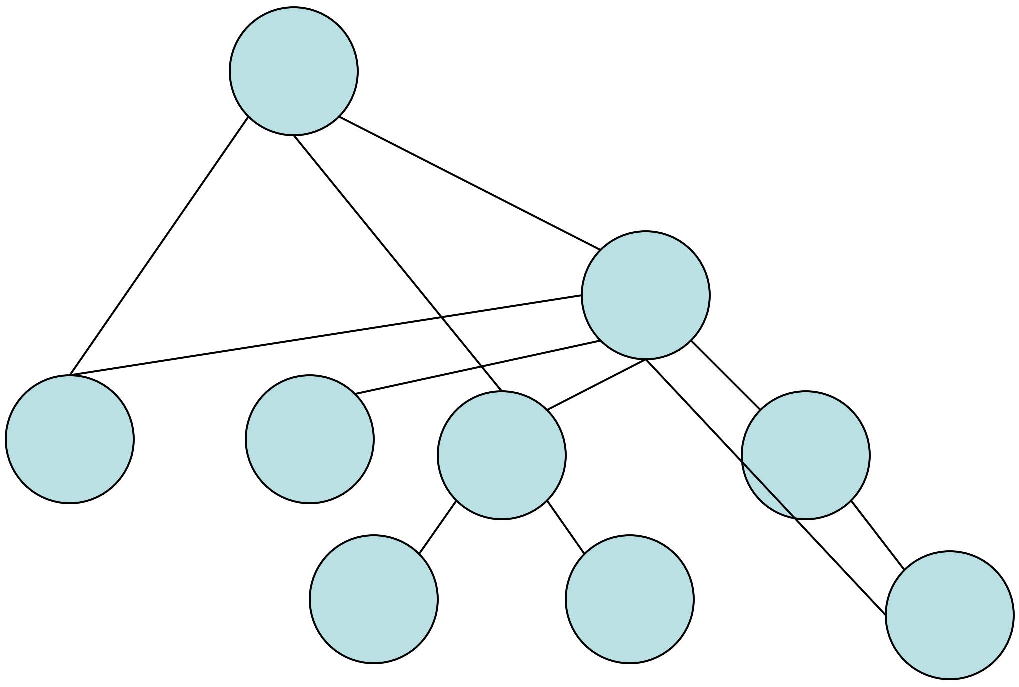
  - Find new destinations following rule 2

BA-model (preferential connectivity)





BA-model (re-wiring)



- In this model the rich get richer

    but…
- Maintains constant average degree

Background

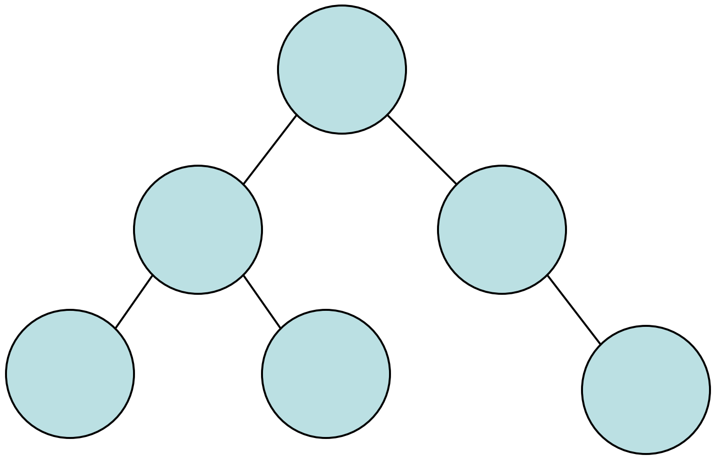
- Faloutsos brothers found surprisingly simple power laws of the Internet topology (On Power-Law Relatonships of the Internet Topology. ACM SIGCOMM 1999.)
  - Most significant: Autonomous System degree distribution
- Correlation coefficients at least 96%
- Implies very strict ”rich get richer” growth model for Internet (Barabàsi, Albert) aka BA-model
- Unexpected result in networking community

Starting point

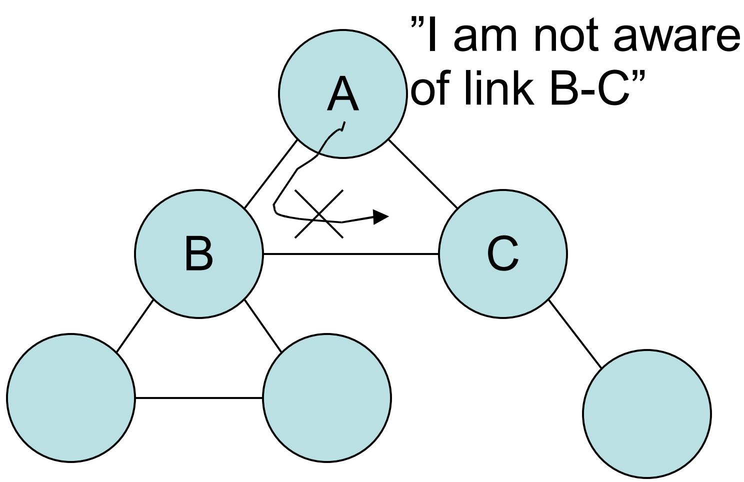
- Authors suspected that the BGP-data that was used by Faloutsos brothers did not represent Internet topology well enough, and hereby <u>should not have been used to make conclusions about Internet growth model</u> (in other words, Internet might not evolve according to BA-model)
  - The data was obtained from a route server located in Oregon

Why is it difficult to obtain Internet topology?

- A telecom network has strict hierarchy
  - ➢ Easy to obtain complete topology



- Internet has loose hierarchy
  - ➢ Important BGP detail: Upstream/downstrem nodes (ASes) are not usually aware of peering links!



Discovered problems in data

- The set of peers of Oregon route server changes over time -> some graph changes are cause by this
- Flapping links
- Truncated tables

➢ Difficult to identify real node/link births/deaths

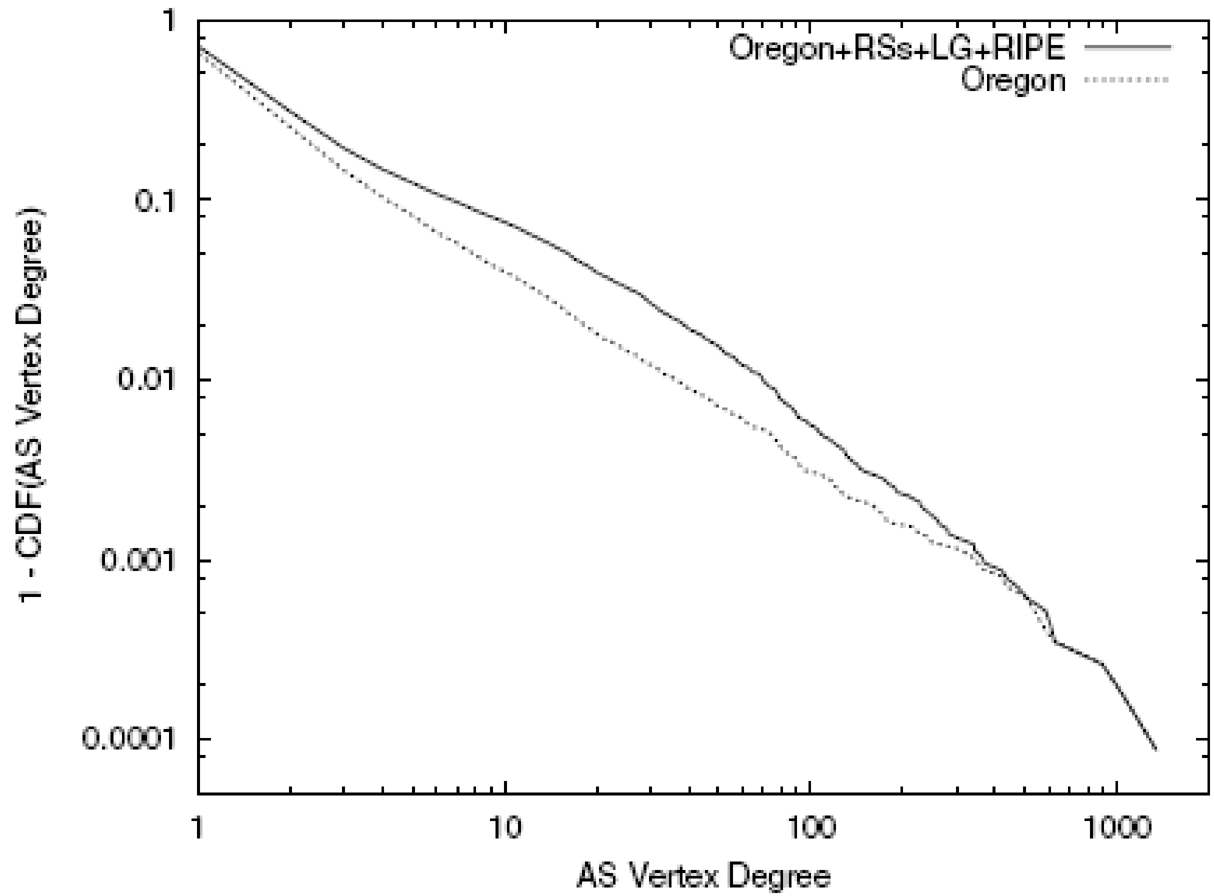
New Data Acquisition Method

- The authors used three additional methods to determine Internet topology (in addition to the snapshots of Oregon route server)
  ١. All publicly available BGP data
  ٢. Looking glass sites
  ٣. RIPE Internet Routing Registry

➢ Extended AS maps

Clear Finding

- While AS degree distribution is heavy-tailed, it does not follow a strict power-law
- Larger degree nodes are more popular than BA-model predicts

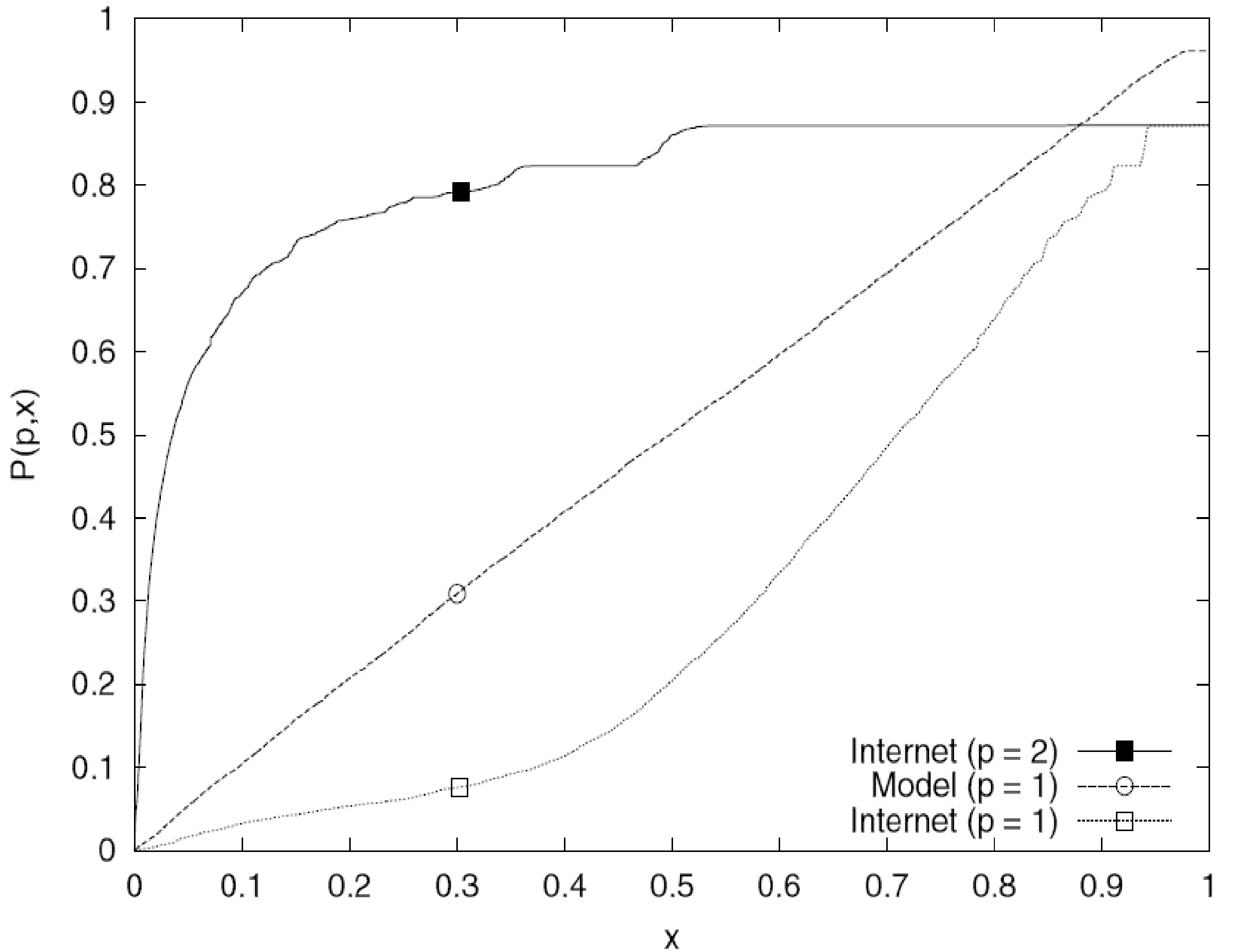


Unclear Findings

- ”We need a metric to analyze the evolution of connectivity graphs”
  - Why?
- The authors choose *cumulative degree ratio* to be this metric
- If we take the Internet connectivity graph, and run it through a non-linear ”conversion function”, we *may* get a cumulative degree ratio plot, that is close to the cumulative degree ratio plot produced by BA-model?
  - So what?

Unclear Finding

- Different cumulative degree ratio plots can be produced by running Internet graph through non-linear conversion function with varying exponents
- Some exponent *may* yield a plot close to BA-model?



Relevant Conclusions

- The probability of a new vertex forming a link with an existing vertex *seems* to follow a non-linear function of degree (BA-model suggests a linear function) – form of function unclear
- Existing vertices form new links – BA-model does not take this into account
- Vertex degree distribution and peering characretistics of Ases are more complex than what BA-model predicts
- What is the real growth logic/model?

Recent findings regarding growth models

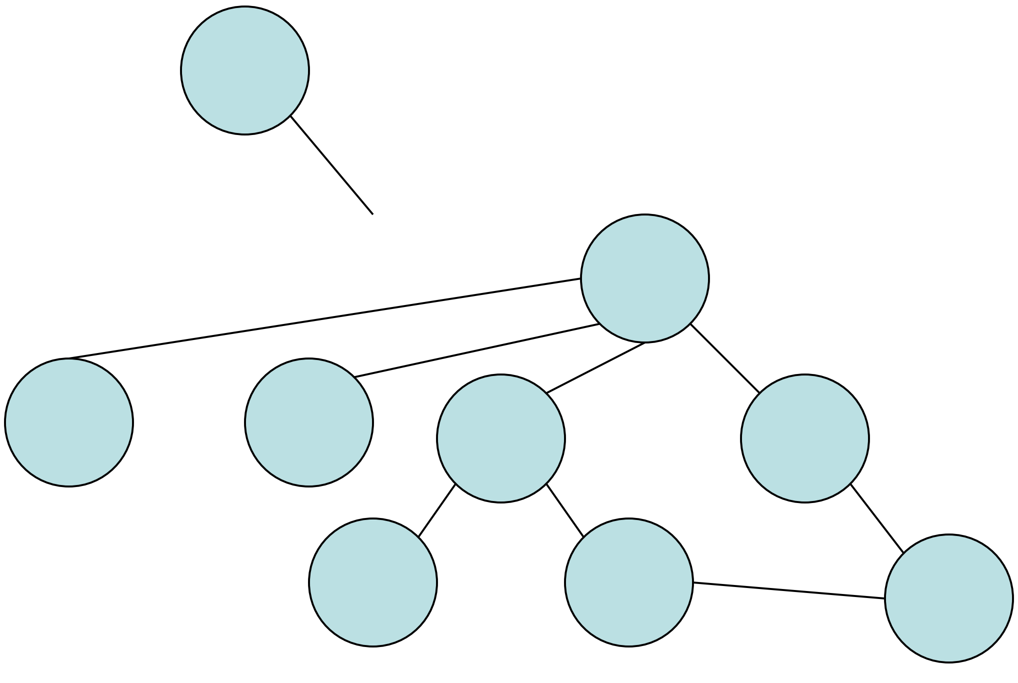
- J. Leskovec, J. Kleinberg, C. Faloutsos: (Graphs over time: densification laws, shrinking diameters and possible, 2005)

1. Many real networks become <u>*denser* over time</u> (average degree increases)
2. <u>Diameter</u> of real networks <u>*decreases* over time</u>
3. Degree distribution is heavy-tailed

➢ Two new growth models for real networks:

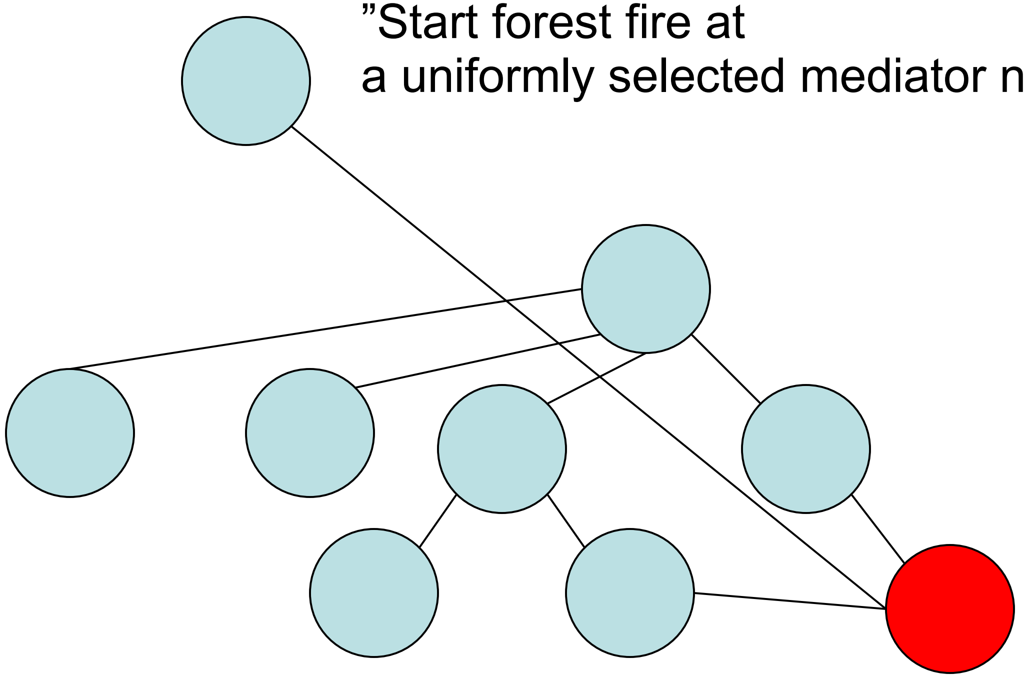
1. Community Guided Attachment
2. Forest Fire Model

Forest Fire Model

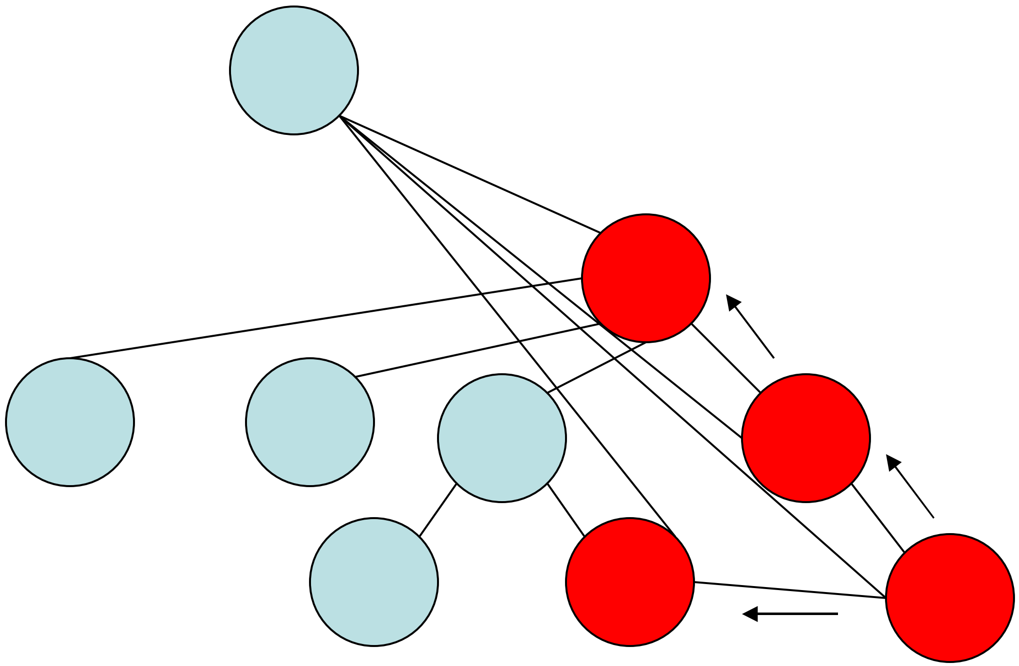


Forest Fire Model

”Start forest fire at
a uniformly selected mediator node”



Forest Fire spreads with exponentially decreasing probability



Properties of Forest Fire model

- ”Rich get richer”-attacment process (heavy-tailed degrees)
- Leads to community structure
- Densification
- Shrinking diameter

- Addition: orphan nodes

Problems in modeling Internet growth

- Different types of nodes (Universities, ISPs, Content distributors, BGP route servers, BGP beacons, …) - they are not distributed uniformly around the Internet
- ”Piecewise behavior” of nodes (dual-homing, some peer – some don’t)
- Nodes’ behavior may change arbitrarily (new business models, new BGP policies) - ISP peering is very much an art form
- ISP consolidation: Two nodes turn into one?
- Topology may look arbitrarily different at times *t* and *t+1* (route hijacking (misconfiguration), maintenance breaks, HW/SW faults/errors)

Open questions

- Are simplistic growth models of any use in the context of Internet?
- Has anyone ever actually used them for something (succesfully)?
- Averaging may lose much relevant data
- Would it be more reliable to identify *nodewise behavior* for Ases and extrapolate from that?
- How about in WWW?

Thanks!
Questions?