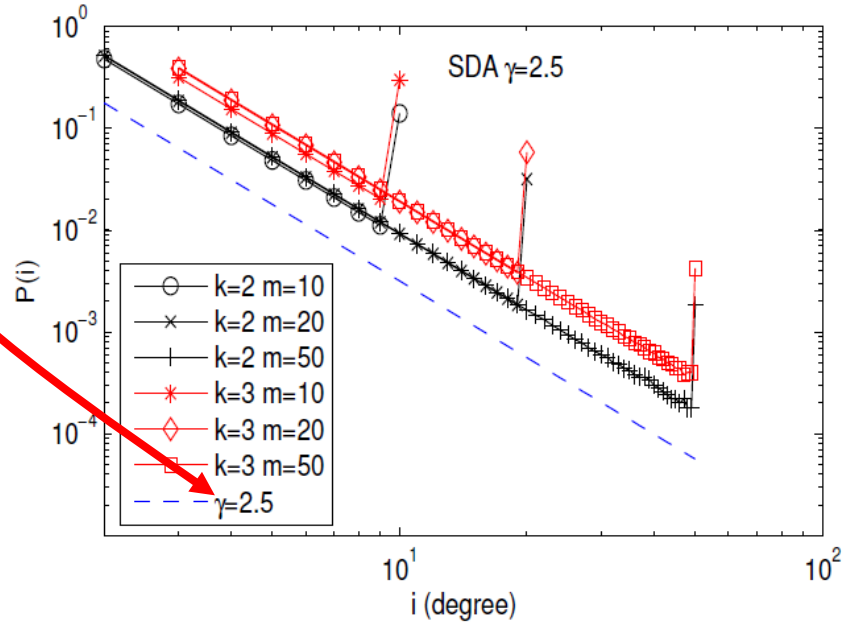
On Growth of Limited Scale-free Overlay Network Topologies

Eyuphan Bulut
bulute@cs.rpi.du, ebulut@cisco.com
Boleslaw Szymanski
szymansk@cs.rpi.edu

Scale-free networks

- Power-law degree distribution
  - Distribution is free of the system size N

$$P(k) \sim k^{-\gamma}$$



Scale-free networks

- Used in the design of overlay network topologies in multiple network types:
  - P2P networks
    - For better search efficiency and scalable networks
  - Wireless sensor networks
    - High degree overlay nodes are placed at the physical nodes with more power and often serve as the network's routers.
  - Database network (ex: GaianDB)
    - To reduce the cost of query retrieval from database nodes.

Scale-free networks

- Overlay topology features have significant effects on the search performance in unstructured P2P networks.

Search Efficiency vs. Exponent and Connectedness

	Diameter d	Exponent γ	Min degree k
Ultra-small	$O(\ln\ln N)$	(2,3)	≥1
	$O(\ln N/\ln\ln N)$	3	≥2
	$O(\ln N)$	3	1
Small-world	$O(\ln N)$	>3	≥1

Limited (or truncated) scale-free networks

- Nodes may not want or be able to cope with high connectivity in reality.
- Thus, some hard cutoffs are often imposed on the number of edges that each node can have.
  - As hard cutoff decreases, the diameter of the network increases
  - In [1], Guclu showed that hard cutoffs have positive impact on Normalized Flooding and Random Walk type search efficiency in unstructured P2P networks.

[1]: Guclu et. al. "Limited Scale-free Overlay topologies for unstructured p2p networks", IEEE TPDS 2009.

Motivation

How to construct a scale-free overlay topology with following properties:

- Evolutionary well-fitting to scale-freeness:
  - In each step of network growth it is as close to scale-free as possible to get more benefit from scale-free features for every network instance
- Cost-efficient:
  - Less communication overhead during construction
  - No global knowledge usage
- Limited:
  - No peer wants to take on the load – **hard cutoff** on the degree
- Parameterized:
  - As the desired post-construction parameters (i.e. γ) of the scale-free network can be predetermined.

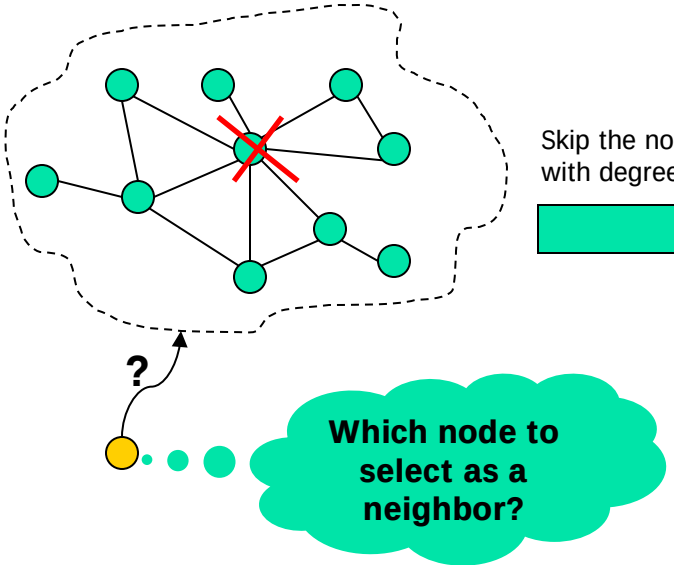
Previous Work

- Growth Algorithms for Limited Scale-free networks:
  1. Preferential attachment (Barabasi-Albert)
  2. Hop-and-Attempt-PA (HAPA)
  3. Gaian

1-Preferential Attachment (PA)

- **Preferential Attachment (PA)**
  - Connects an existing peer with probability proportional to its current degree (prefers the peers with larger degree)
  - Requires global info
  - Application on limited scale-free networks:
    - simply <u>skip</u> the existing peers already saturated their hard cutoffs
  - Ex: hard-cutoff (m) = 7

$$p_c(j) = \frac{d_j}{\sum_{i=1}^{n} d_i}$$



Skip the node with degree 7

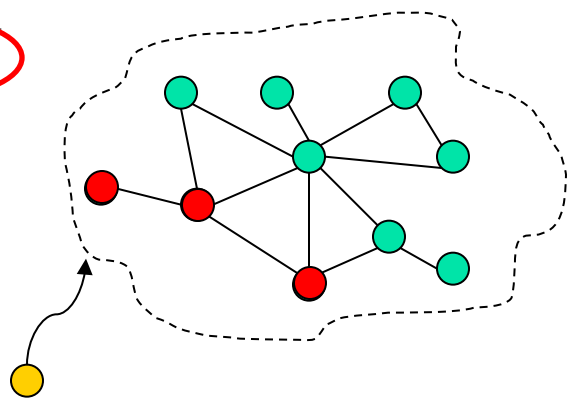
Degrees	Prob. of Attachment
4	0.21
3	0.15
3	0.15
2	0.10
2	0.10
2	0.10
1	0.05
1	0.05
+ 1	+ 0.05
19	1

2-Hop-and-attempt PA Model (HAPA)

- New node attempts to connect to a randomly chosen existing node A by using the preferential attachment rule
- Then it attempts to connect to a randomly chosen node B which is a neighbor of A
- The node repeats this procedure until it fills all its **k** edges (or the number of links it has reaches hard cutoff **m**)

$$p_c(j) = \frac{d_j}{\sum_{i=1}^{n} d_i}$$

2kn



Want to select k edges?

- Localized but may take too long
- Needs to know total node count (n) in the network to find denominator

3-Gaian Algorithm

- New node broadcasts a "Hello" message to announce its presence to the network.
- Other nodes response after a delay inversely proportional with their degree

  Response time = a random number in $[0, t_0/d_i]$ where t_0 is constant

- New node connects to first **k** responders
- Global information is not used by **computing with time**.

Previous Work Deficiencies

- Are not designed for limited scale-free networks (adjusted algorithms)
- Do not fit well to scale-free distribution
- Some need global or partially global information
- High communication overhead in construction
- Fixed γ

Analysis

- Degree distribution in limited scale-free networks:

$$P(i) = ci^{-\gamma} \text{ for nodes with degree } k \leq i < m$$
$$P(i) = 1 - \sum_{j=k}^{m-1} P(j) \text{ for } i = m$$

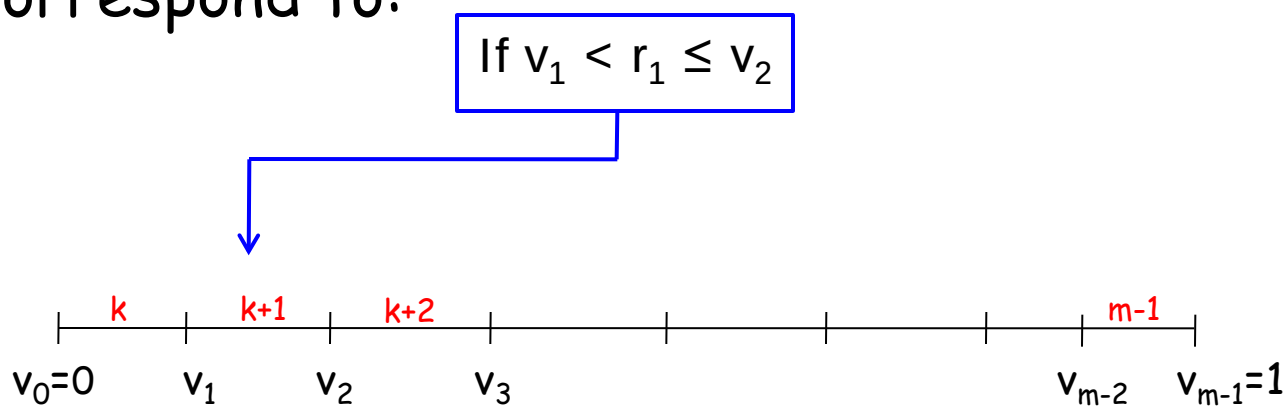
- By enumerating all node counts and edge counts:

$$n = \sum_{i=k}^{m} n_i \qquad 2kn = \sum_{i=k}^{m} in_i$$

- We found frequency (f_i) of nodes with each degree (i)

Proposed Algorithm

- Candidate Degree Selection based growth:
  - New node selects **k** random numbers ($r_1 \ldots r_k$) between [0,1] and finds the degrees they correspond to:



  - New node broadcasts a message with the list of **k** candidate degrees

Proposed Algorithm

- Existing nodes respond if their degrees are one of the sought ones.
  - To reduce the messaging overhead, a node which has forwarded k responses (of others or its own) towards new node, does not forward any more (suppressed reply mechanism).
- New node selects the first **k** responses.

Simulation Results

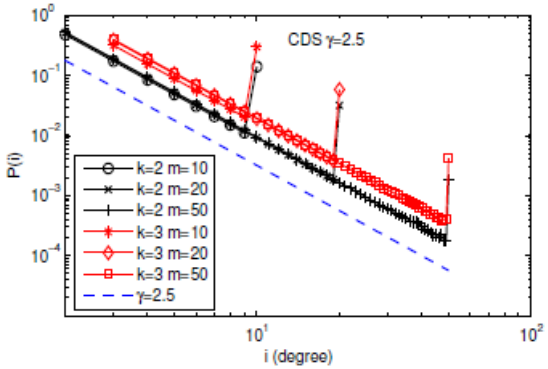
- Algorithms compared

Model	Global knowledge used	Flexible exponent (γ)
BA [12]	Degrees of all nodes	No
HAPA [16]	Total node count	No
Gaian [3]	None	No
CDS	None	Yes

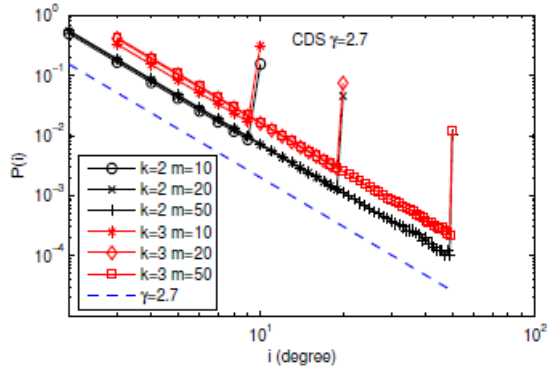
Simulation Results

(LS, MLS) results:
3x-20x better compliance
with fitting scale-free γ

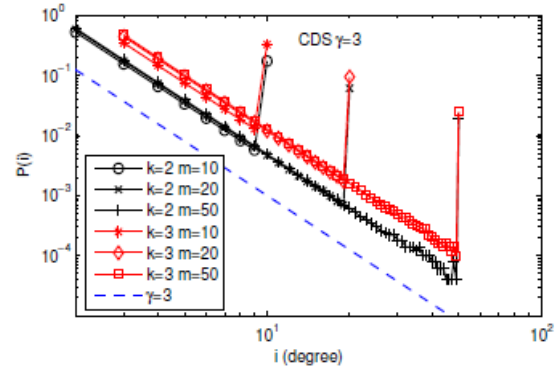
- Degree distributions



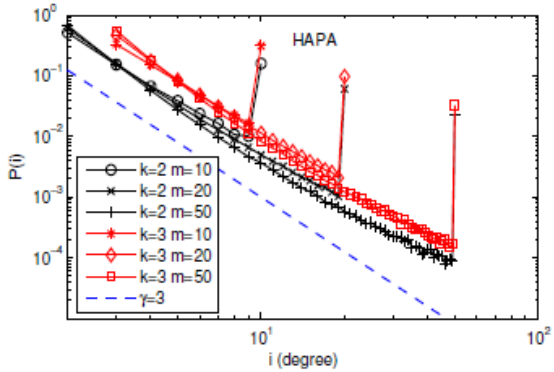
(a) CDS $\gamma = 2.5$



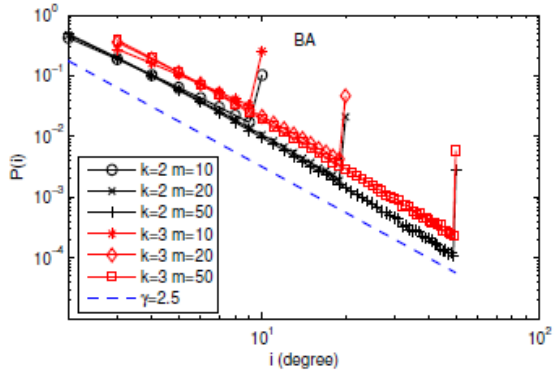
(b) CDS $\gamma = 2.7$



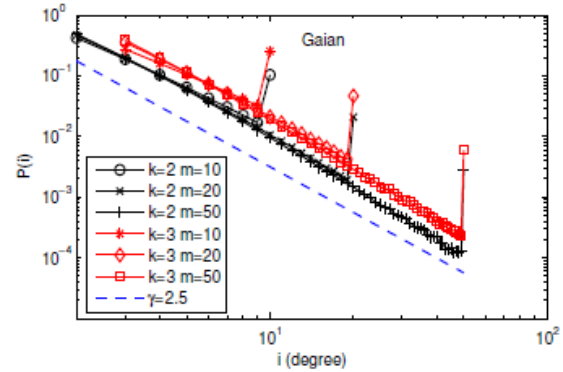
(c) CDS $\gamma = 3$



(d) HAPA



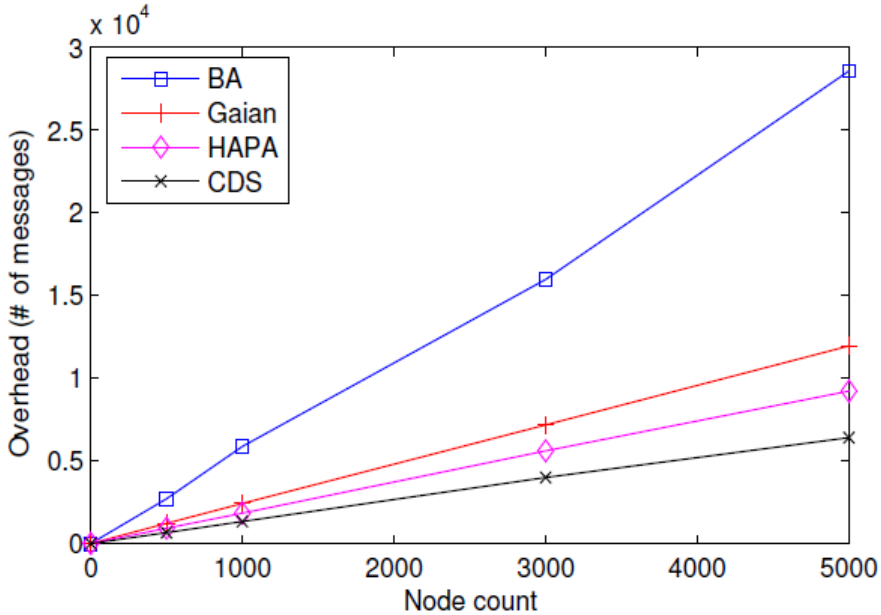
(e) BA



(f) Gaian

Simulation Results

- Communication overhead
  - Ex: when k=2, m=10



Hello broadcast +
reply of all nodes with their degrees

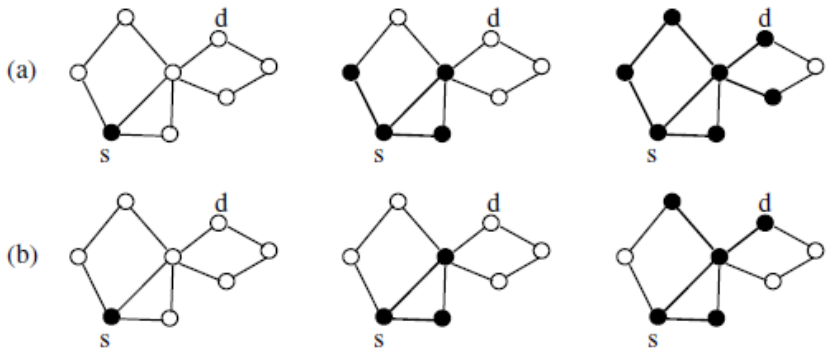
Hello broadcast +
suppressed reply of all nodes with delay inversely proportional to degrees

Random walk through neighbors

Hello broadcast +
Suppressed reply of selected-degree nodes

Search Methods

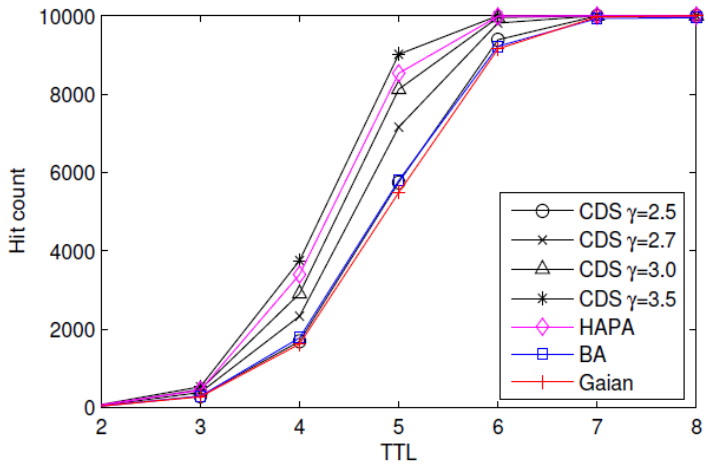
- a) Flooding
  - Source node sends a message to all its neighbors and every node which receives the message forwards it to all its neighbors until the target node receives the message
- b) Normalized flooding
  - The nodes send the messages to at most k (minimum edge count) neighbors



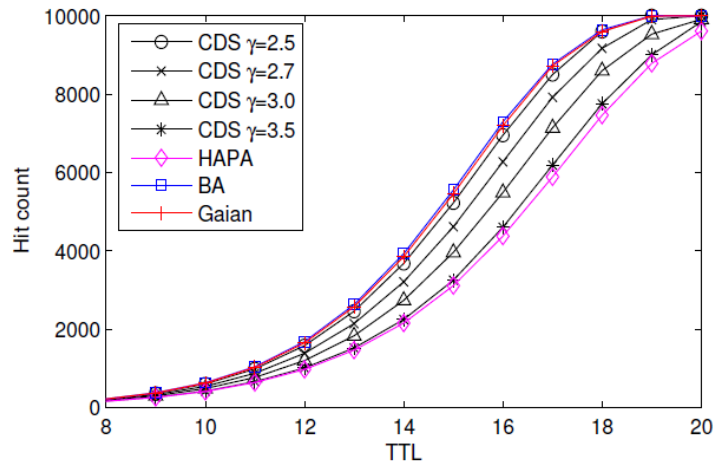
Search Efficiency Results

- In FL search, our algorithm with γ = 3.5 achieves the best hit ratios (HAPA algorithm is closest to it)
- In NF based search, our algorithm with γ = 2.5 together with BA and Gaian algorithms achieves the best hit ratios.

Best γ could be chosen to achieve best search efficiency with given search algorithm.



(b) FL search ($k = 2$, m=50)



(c) NF search ($k = 2$, m=50)

Summary of Contributions

- We have showed that without using global topology information, we can obtain well fitting scale-free networks with less overhead
- γ has significant impact on search efficiency of algorithms

- Future work:
  - How to consider the node leaves in this design?

Questions?

Thank you!