Deciding on the type of a graph from a BFS

Reporter: Wang Xiaomin
Joint work with Matthieu Latapy and Michèle Soria

6 mars 2011

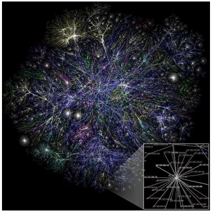
Aléa 2011

Outline

Outline

The difficulty of the measurement of the Internet

1. Internet: Complex
2. Measurement: Sampling
3. Problem: Partial and Biased

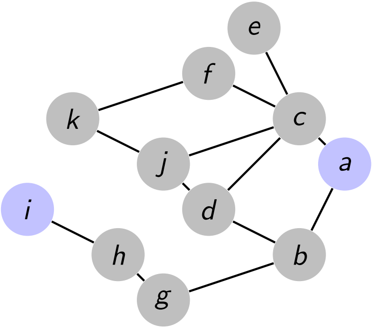


Definition

Degree distribution $\triangleq$ the fraction P_k of nodes with k links.
Type of distribution: Poisson, Power-law, Regular...

Traceroute

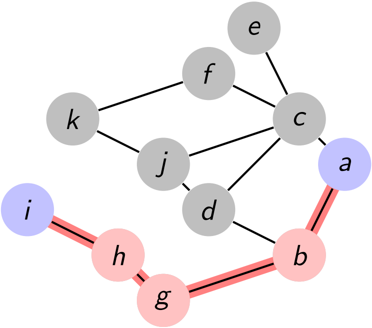
Traceroute $\rightarrow$ route from a monitor to a destination.



route $\sim$ shortest path

Traceroute

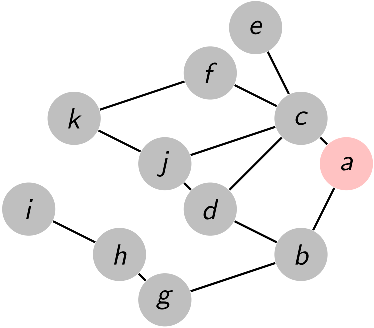
Traceroute → route from a monitor to a destination.



route ∼ shortest path

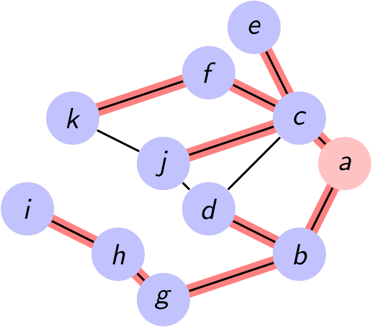
BFS Tree

1 monitor, many destinations
$\rightarrow$ BFS: Breadth First Search Tree



BFS Tree

1 monitor, many destinations
→ BFS: Breadth First Search Tree



BFS Tree: Power-law degree distribution

[*Achliotas*, *Clauset*, *Kempe*, *Moore*, *JACM*, 2005]

$$a_{m+1}^{obs} = \sum_i a_i \left[\int_0^1 i t^{i-1} \binom{i-1}{m} p_{vis}(t)^m (1 - p_{vis}(t))^{i-1-m} \, dt \right] \quad (1)$$

$$p_{vis}(t) = \frac{1}{\sum_j j a_j t^j} \sum_k k a_k t^k \left(\frac{\sum_j j a_j t^j}{\delta t^2} \right)^k \quad (2)$$

$$g^{obs}(z) = z \int_0^1 g' \left[t - \frac{(1-z)}{g'(1)} g' \left(\frac{g'(t)}{g'(1)} \right) \right] dt \quad (3)$$

BFS Tree: Power-law degree distribution

Degree distribution of the BFS is always Power-law.

1. Poisson: $a_m = \frac{\lambda^m e^{-\lambda}}{m!} \rightarrow a^{obs}_{m+1} \sim m^{-1}$
2. Regular: $a_r = 1 \rightarrow a^{obs}_{m+1} \sim \frac{1}{rm}$
3. Power-law: $a_m \sim m^{-\alpha} \rightarrow$ underestimate α

Problem: How to get information on the type of the graph from a BFS tree (always Power-law)?

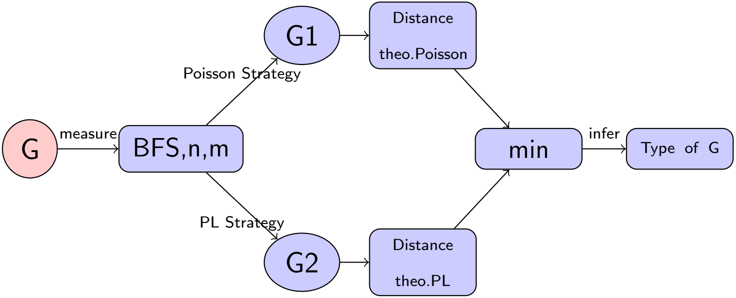
1. **Current approach**: collect samples as large as possible $\rightarrow$ still biased?
2. **Our approach**: infer the properties of the graph from a BFS.

Outline

Methodology

Distance:

1. Distribution 1: from the reconstructed graph (G1 or G2).
2. Distribution 2: calculated with (n,m,type).



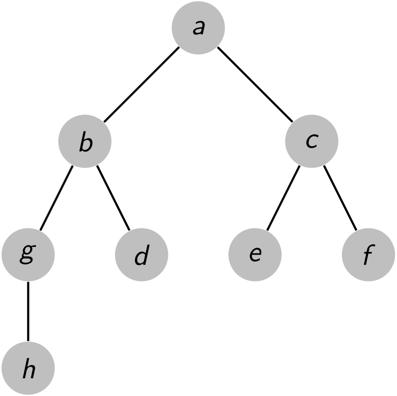
Step 1: decide on the type of a graph from **(n,m,BFS)**
Step 2: decide on the type of a graph from **(n,BFS)**

Rebuilding: Methodology

Unknown connected $G \rightarrow$ BFS tree T, n, $m \rightarrow G'$
Method: add $m - n + 1$ links to T
How to rebuild:

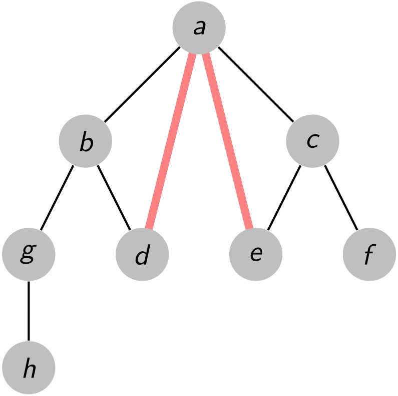
1. Forbidden positions
2. RR, PP and RP strategies

Rebuilding: Forbidden positions



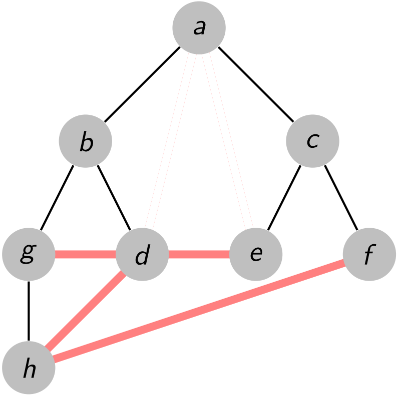
Any link of G is necessarily between two nodes in consecutive levels of T, or in the same level of T.

Rebuilding: Forbidden positions



Any link of G is necessarily between two nodes in consecutive levels of T, or in the same level of T.

Rebuilding: Forbidden positions



Any link of G is necessarily between two nodes in consecutive levels of T, or in the same level of T.

Rebuilding strategies: RR and PP

Among all allowed positions:

1. RR: Inspired from Erdös-Rényi Model, the two extremities are chosen with uniform probability.

$$E\left(d_{G'}(v)=l\right)=\frac{1}{n}\sum_{j>0}\sum_{k=1}^{l}n_{jk}P\left(k\rightarrow l,j\right) \tag{4}$$

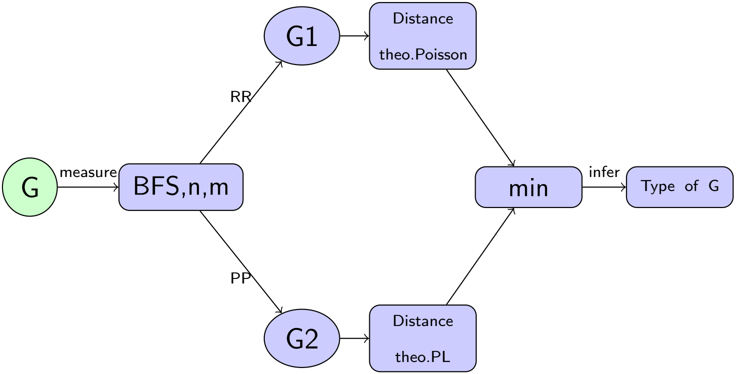
2. PP: Inspired from Barabási-Albert Model, the two extremities are chosen with probability proportional to their degree.

$$E\left(d_{G'}(v)=l\right)=\frac{1}{n}\sum_{j>0}\sum_{k=1}^{l}n_{jk}P(k\rightarrow l,j,m') \tag{5}$$

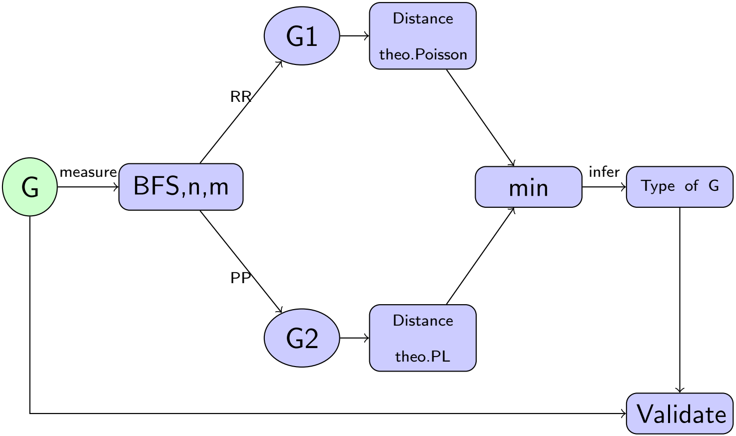
3. Other strategies

Outline

Validation: process



Validation: process



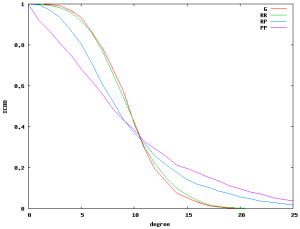
Validation: datasets

Model graphs

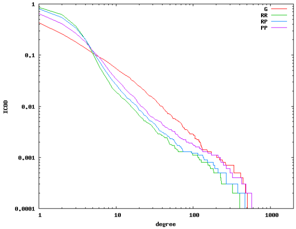
1. Simple, connected
   $[F.Viger, M.Latapy, 11thICCC, 2005]$
2. Poisson: 3 to 10
   Power-law: 2.1 to 2.5
3. Size: 1000 to 100000 nodes
4. Sample: 10 each

Graphic Validation on model graphs

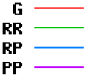
Poisson 10



Power-law 2.2



Poisson 10: RR is best.
Power-law 2.2: PP is best.

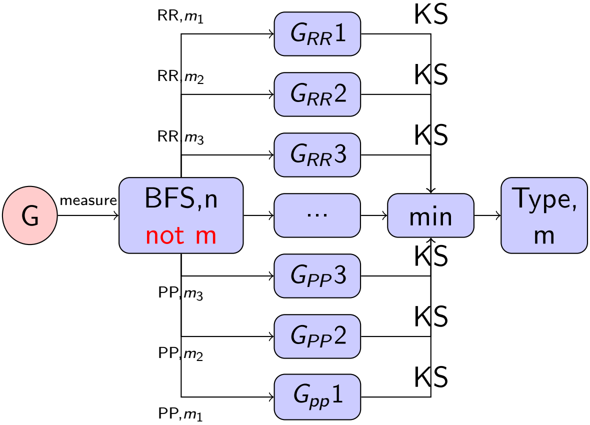


Our strategies work on model graphs

Outline

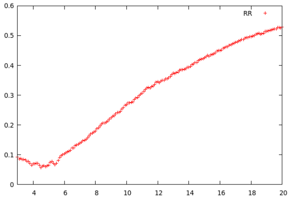
Deciding without m

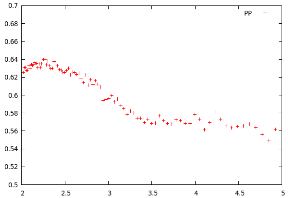
Hypothesis : n, a complete BFS, not m
Try wide range of m, RR: $m \in (2, 50]$; PP: $m \in (2, 10]$
Kolmogorov-Smirnov distance: $KS = max_k \frac{1}{2} \sum_{i=0}^{k} (p_i - q_i)$.



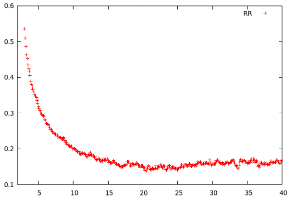
Deciding the type without m: one BFS tree

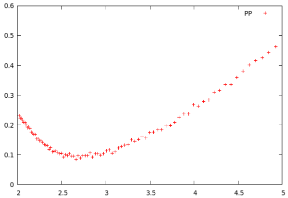
Poisson 3 (m=3)





Power-law 2.2 (m=3.14)





Deciding the type without m: multi-BFS trees

Nodes: 10000

Table: Result of multi-BFS: Poisson 3

BFS number	min KS
1 BFS	RR at 4.5 = 0.0207
2 BFSs	RR at 4 = 0.0342
5 BFSs	RR at 3 = 0.0402
10 BFSs	RR at 3 = 0.0177
20 BFSs	RR at 3 = 0.0169

Deciding the type without *m*: multi-BFS trees

Skitter-AS graph is more likely to a power-law graph, with 5776 nodes and 12822 links.

Table: Result of multi-BFS: Skitter-AS

BFS number	RR λ	m	PP α	m
1 BFS	21	60648	2.28	7683
2 BFSs	20	57760	2.15	10228
5 BFSs	16.5	47652	2.08	12388

Outline

Conclusion and perspectives

Conclusion: A new approach succeeds in distinguishing between Poisson and power-law, but needs a complete BFS.

Future work:

1. Use the profile of BFS.
   $\sim$ a diminishing urn model M=(-1,-1;0,-2).
2. Partial BFS or BFS limited by number of hop.
3. Several BFSs.
   1. How many BFS?
   2. How to choose?

Thank you!