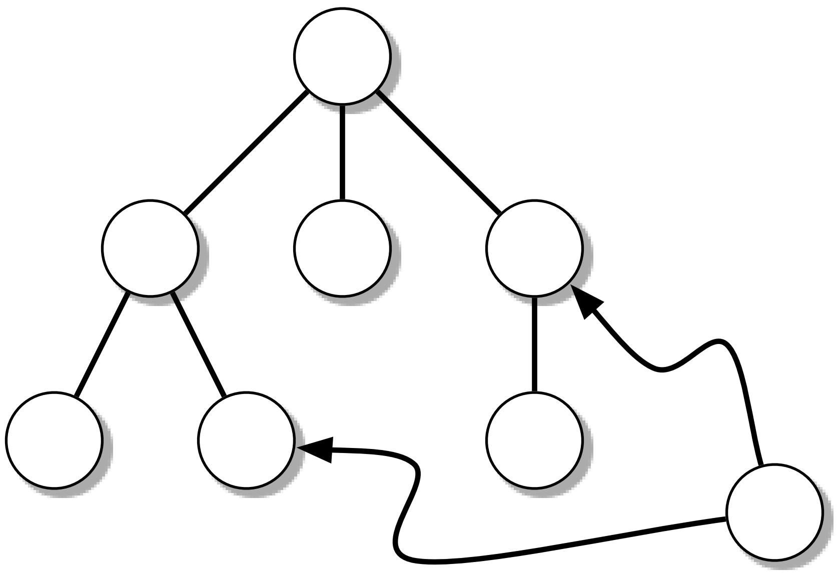
The Power of Choice in Social Networks

Raissa D'Souza, UC Davis
Paul Krapivsky, Boston
Cristopher Moore, UNM/Santa Fe

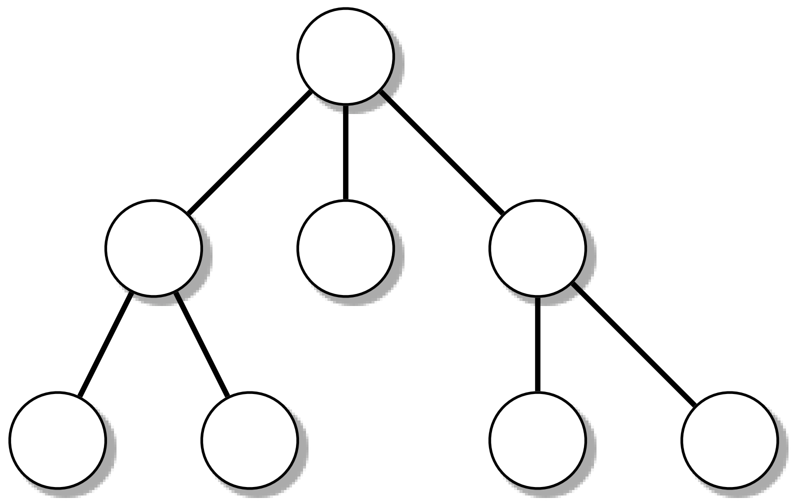
A growing organization

- NetCorp has a hierarchical organization; a tree with the CEO at the root
- You have k contacts, randomly chosen (where k is a fixed constant)
- You go to work for the one closest to the root

A growing organization



A growing organization

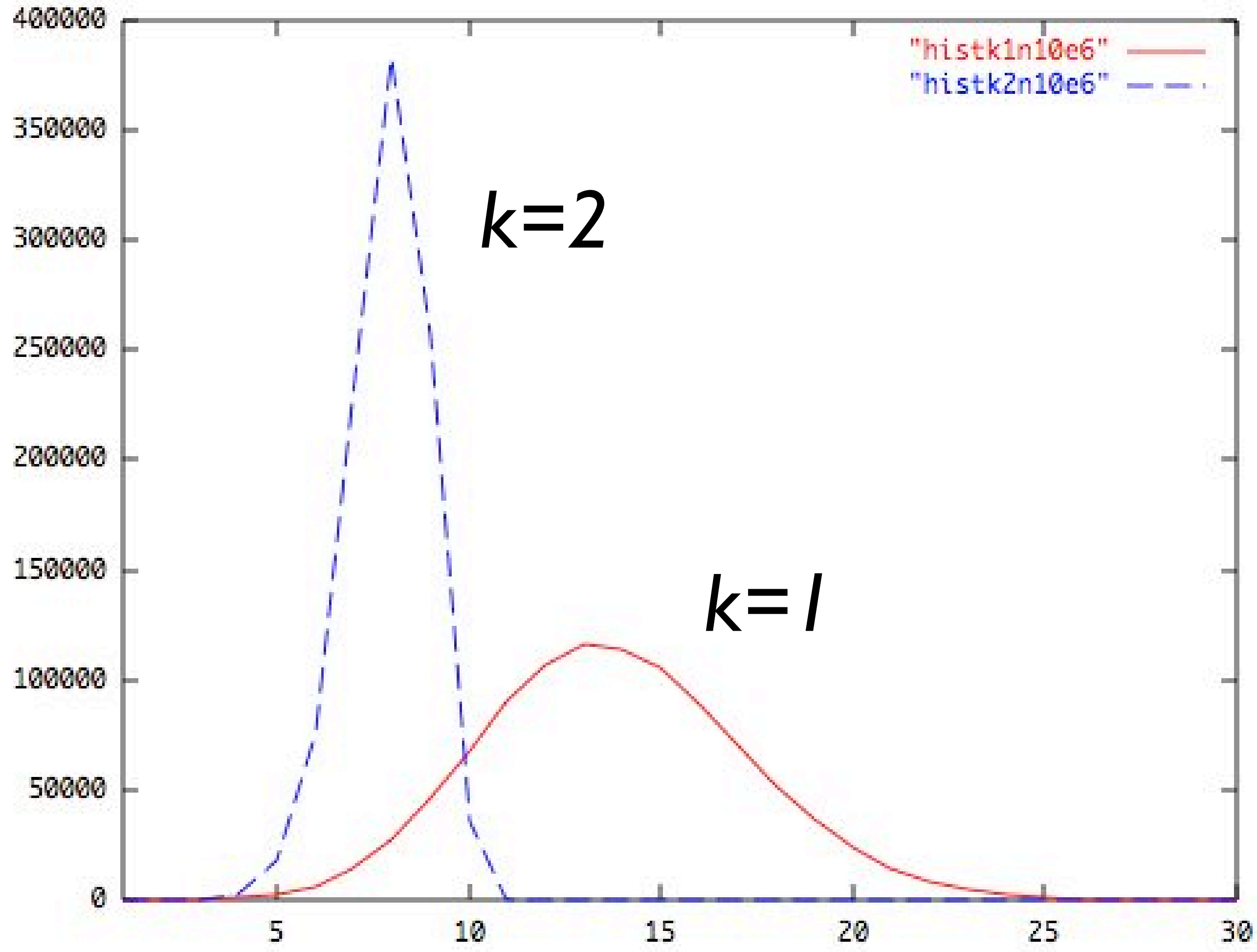


- What does the tree look like for large *n*? What is the *depth distribution* of nodes or leaves?

Random trees

- When $k=1$, you become a daughter of a random node
- The distribution of depths is Poisson, with mean $\ln n$
- With high probability the greatest depth of any leaf is $O(\ln n)$
- But for $k>1$, we expect a shallower tree. How much?

A first glance



Balls in Bins

- Problem in computer science: assign tasks to processors
- If I throw m balls in n bins randomly, the number of balls in a particular bin is Poisson-distributed with mean m/n
- If $m=O(n)$, the largest number of balls in any one bin is $O(\log n / \log \log n)$

Balls in Bins, with Choice

- What if I choose k random bins, and throw the new ball into the bin with the fewest balls?
- Now, for any $k \geq 2$, the largest number of any one bin becomes $O(\log \log n)$ — the distribution is much closer to uniform!
- Does something similar happen here?

[ABKU '98, Mitzenmacher '01]

The Depth Distribution

- Let q_j be the fraction of vertices at depth j or greater (the cumulative distribution)
- The new node only has depth $\geq j$ if all k of its contacts had depth $\geq j-1$, so

$$\frac{d}{dn} n q_j = q_{j-1}^k$$

- Changing variables to $t = \ln n$ gives

$$\frac{dq_j}{dt} = -q_j + q_{j-1}^k$$

A Traveling Wave

- Let's try a solution of the form

$$q_j(t) = q(j - vt)$$

- This gives us a differential equation in a single variable,

$$v\frac{dq}{dx} = q(x) - q(x-1)^k$$

- The "wave front" (average depth) moves at velocity *v*. But what is *v*?

Tails of the Wave

- When $x \to -\infty$ the tail is exponential,

$$1 - q(x) \sim e^{\lambda x}$$

- Plugging this in gives the velocity

$$v = \frac{1 - ke^{-\lambda}}{\lambda}$$

- The "selection principle" says that *v* is the *largest possible.* Maximizing over λ (when k=2) gives *v*=0.373365...

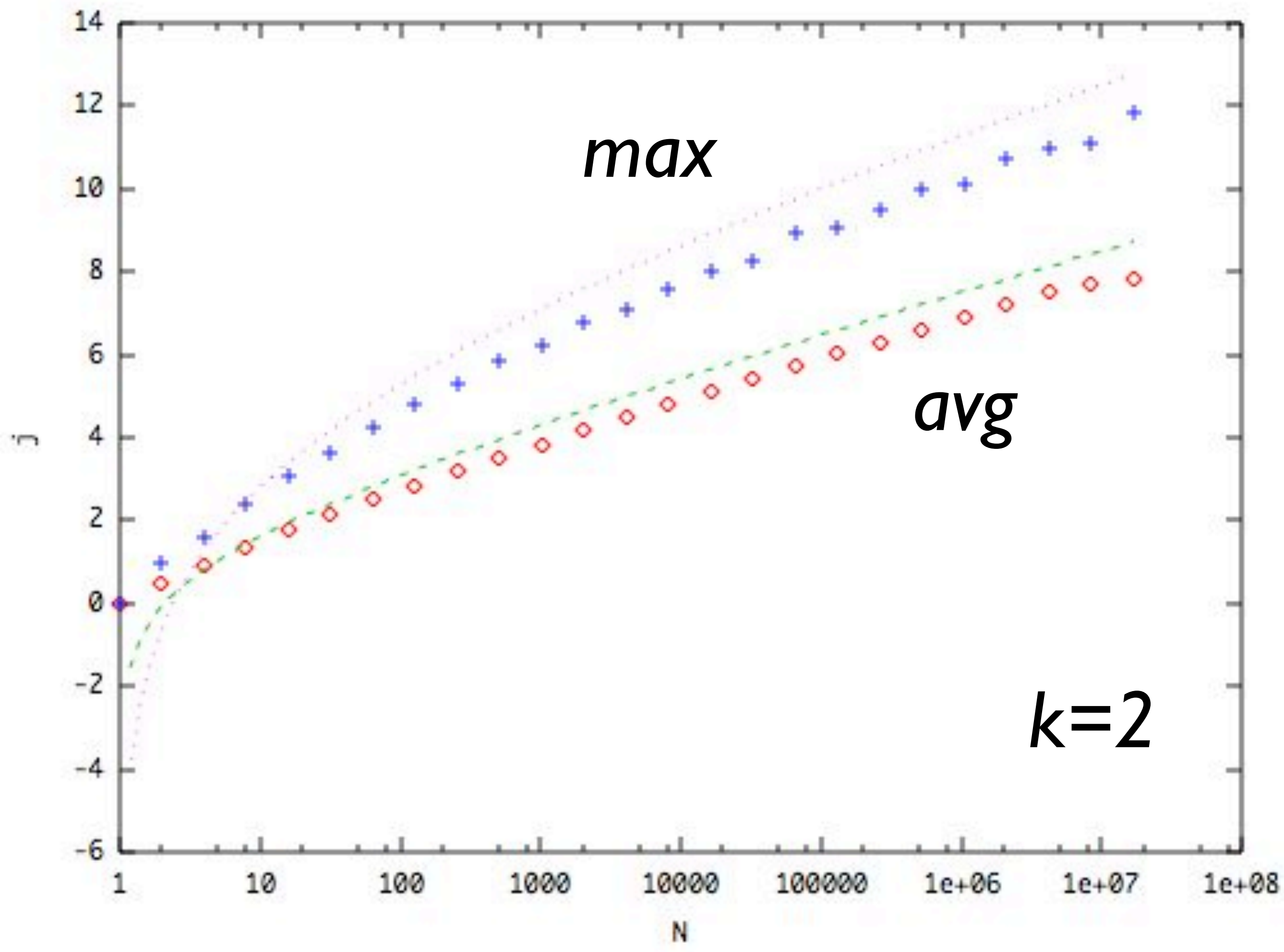
A Narrow Wavefront

- When *k*=1, we get a Poisson distribution of mean $t = \ln n$, and width $\sqrt{\ln n}$
- In contrast, here the width of the wavefront stays constant as *t* increases.
- And, since $q(x) \leq q(x-1)^k$, the tail drops off doubly-exponentially:

$$q(x) \sim e^{-Ak^x}$$

- Setting this to 1/*n*, the max depth is *O(log log n)* greater than the average.

Simulation



max
avg
k=2
j
N
14
12
10
8
6
4
2
0
-2
-4
-6
1
10
100
1000
10000
100000
1e+06
1e+07
1e+08

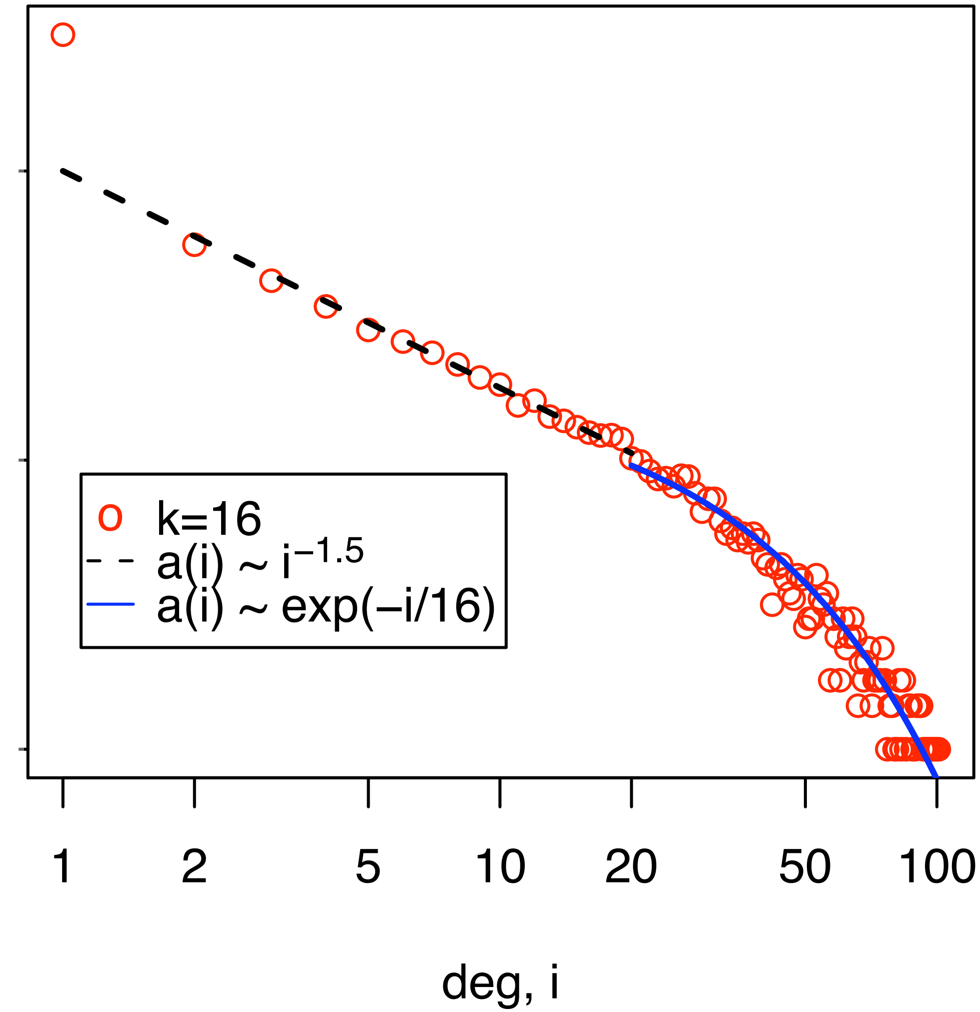
Highest Degree

- We have k random contacts; but now we join the one with highest degree. (Like preferential attachment, but not really.)
- Let c_j be the fraction of vertices with degree j or less. This decreases if the highest degree of the k contacts is j, so at a steady state we have

$$c_j = 1 - (c_j^k - c_{j-1}^k)$$

A Pudgy Tail

- When *k* is large (but still constant), this gives a power law up to *j*~*k* (with an exponent tending to −1) followed by an exponential cutoff.



And Now, Lowest Degree

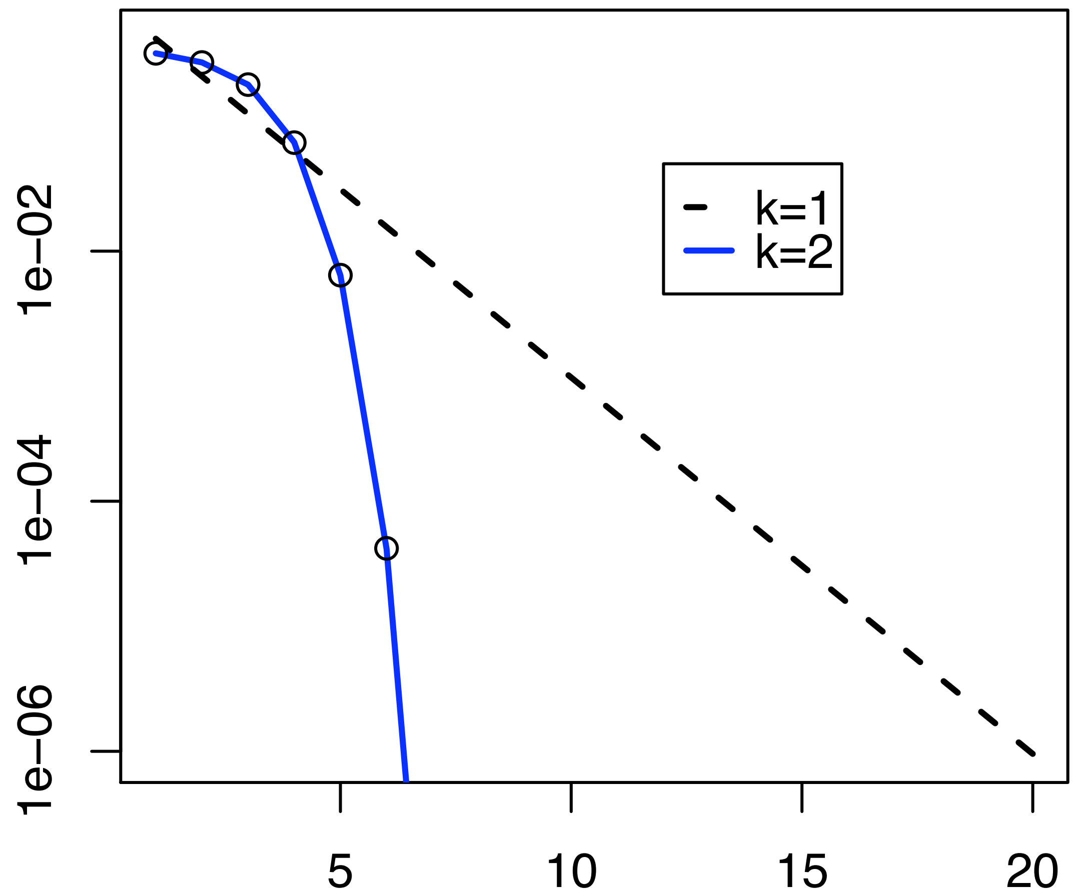
- We have *k* random contacts; but now we join the one with *lowest* degree.
- Now let c_j be the fraction of vertices with degree *j* or *more*. This increases if the lowest degree of the *k* contacts is *j*–1, so

$$c_j = c_{j-1}^k - c_j^k$$

- But then $c_j \leq c_{j-1}^k$, so $c_j \sim e^{-Ak^j}$ is doubly-exponential; setting to 1/*n* gives a max degree *O(log log n)*.

A Very Skinny Tail

- Degree distribution is nearly uniform.
- Useful for peer-to-peer networks?



Shameless Plug

Oxford, 2007/8

Acknowledgments