Slide 1

Mathematics for Computer Science
MIT 6.042J/18.062J

PageRank
(by Google founder Larry Page)

Slide 2

Google Rankings

Which webpages are "more important?"

Model of internet:
- Users click random link on a page.
- Occasionally start over.

A page is "more important" if viewed a large fraction of time

Slide 3

Random Walk on the Web

View the entire web as digraph
- vertices are webpages
- edge (V,W) exists if link from page V to page W
- edges out of V equally likely
  $$\Pr[(V,W)] = 1/\text{outdeg}(V)$$

Slide 4

Random Walk on the Web
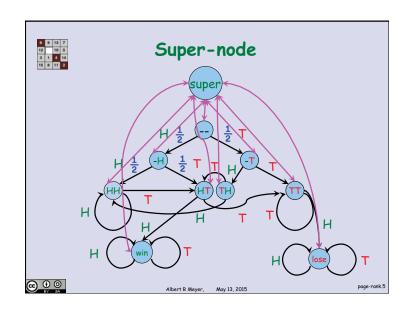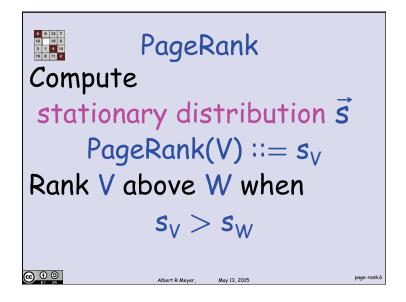
To model starting over:
* add a "super-node" to the graph
* an edge from super-node to each other node
* edges from each other node back to super-node
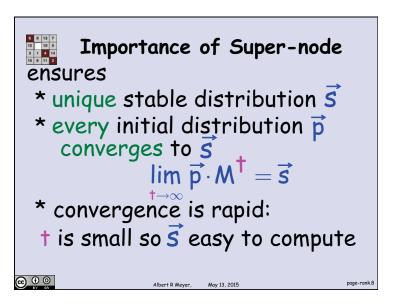 -may get customized probabilities

1

Super-node

PageRank

Compute
stationary distribution $\vec{s}$

$$\text{PageRank}(V) ::= s_V$$

Rank **V** above **W** when

$$s_V > s_W$$

Resistance to scamming

* Creating fake nodes
  pointing to self
* Adding links to important
  nodes

won't improve PageRank

Importance of Super-node

ensures
* unique stable distribution $\vec{s}$
* every initial distribution $\vec{p}$
  converges to $\vec{s}$

$$\lim_{t \to \infty} \vec{p} \cdot M^t = \vec{s}$$

* convergence is rapid:
  t is small so $\vec{s}$ easy to compute

Actual Google Rank

Google rank rules are a closely held trade secret using text, location, payment, and other criteria that have evolved for 15 years.

But PageRank continues to play a significant role.

3

6.042J / 18.062J Mathematics for Computer Science
Spring 2015