## **Google and the Random Surfer**



An important page is one that lots of pages point to?
No …

## **Google and the Random Surfer**



• An important page is one that lots of *important* pages point to.

# **Google and the Random Surfer**



- An important page is one that lots of important pages point to.
- Start at any web page and follow links at random. Forever.
- You' II see "important" pages more often than unimportant ones.

### Analyzing the Web with graphs and matrices



- Graph nodes are web pages
- Arrows between nodes are links between web pages
- Matrix entries are links from "column" pages to "row" pages
- The Page Rank comes from algebra on the matrix

### Analyzing the Web with graphs and matrices



- Graph nodes are web pages
- Arrows between nodes are links between web pages
- Matrix entries are links from "column" pages to "row" pages
- The Page Rank comes from algebra on the matrix
- The matrix has over 30,000,000,000 rows & columns

### **Random Surfer Rule**



An important page is one that many important pages point to.

- If there are no links out of this page, choose a page at random.
- Otherwise, with probability p (= .85), follow a random link out of this page.
- Or, with probability m = 1 p (= .15), choose a page at random.

# A Page Rank Matrix

- Importance ranking of web pages
- Stationary distribution of a Markov chain
- Power method: matvec and vector arithmetic
- Page rank matrix from a web crawl of mit.edu (170,000 pages)

