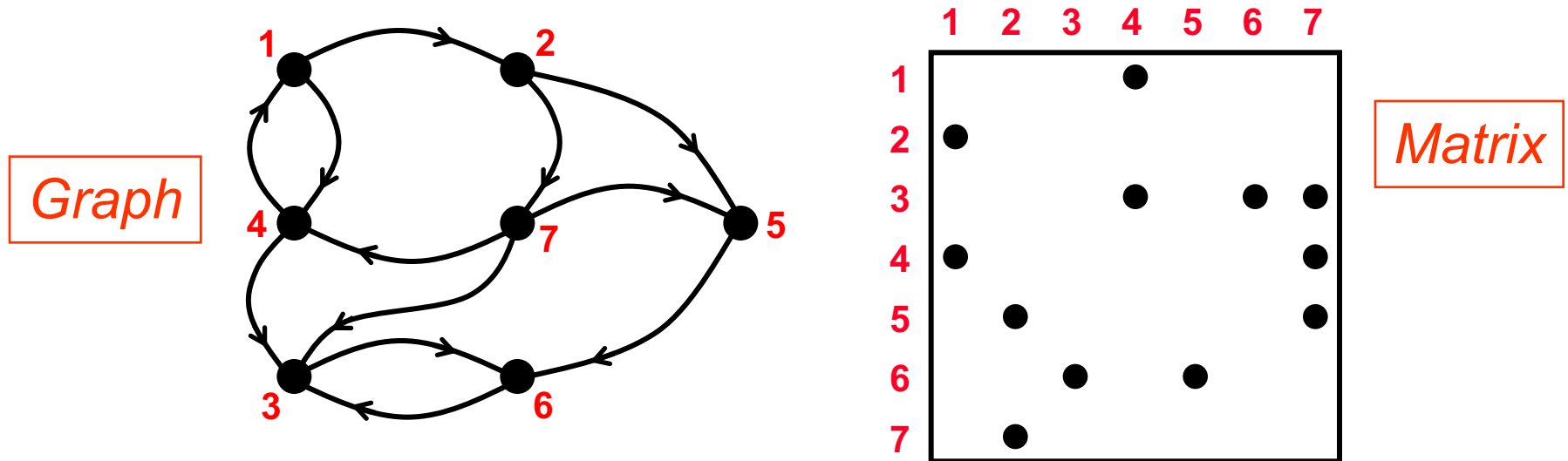
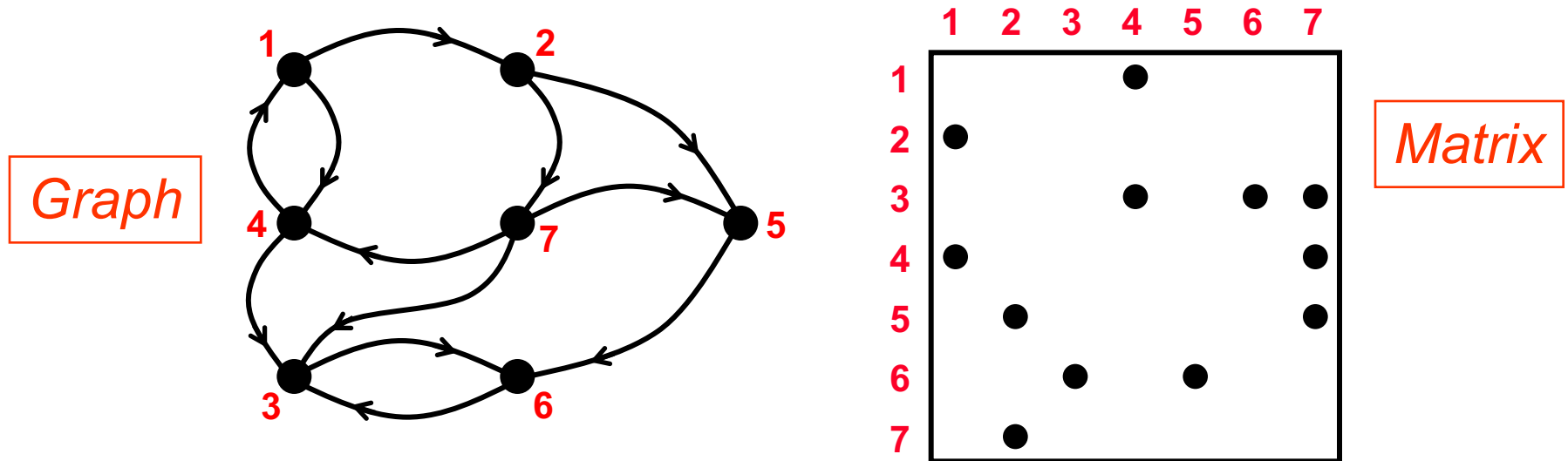


Analyzing the Web with graphs and matrices



- 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

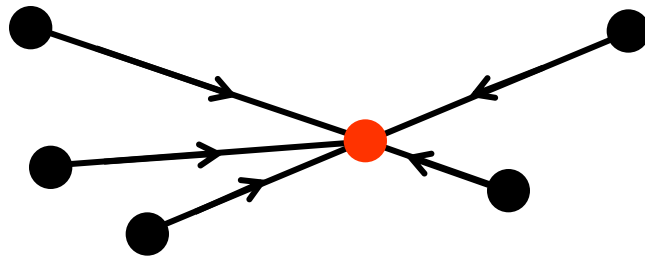
Analyzing the Web with graphs and matrices



- 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 **130,000,000,000,000** rows & columns (2016)

Google and the Random Surfer

How does Google figure out which web pages are most important?

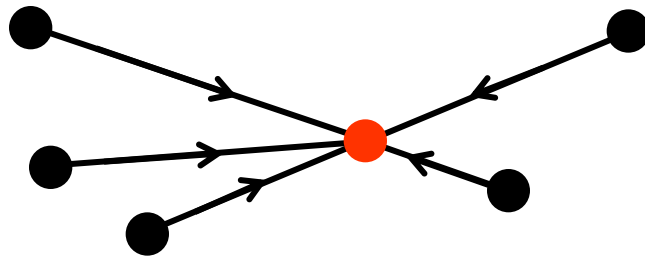


- An important page is one that lots of pages point to?

No ...

Google and the Random Surfer

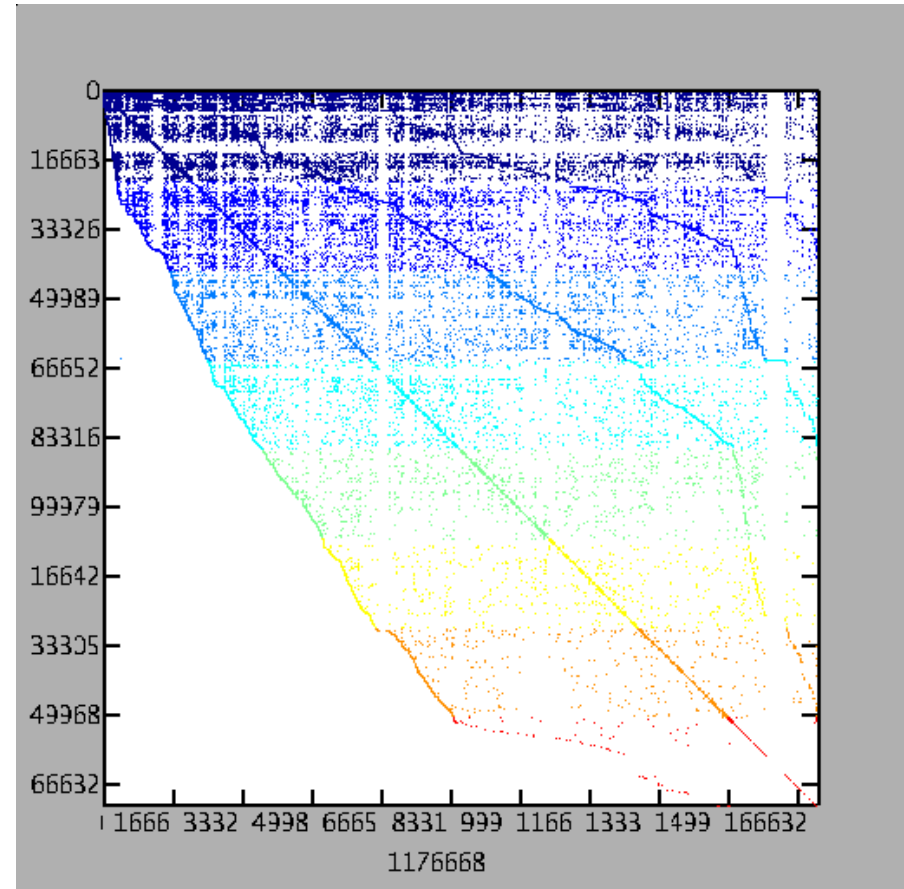
How does Google figure out which web pages are most important?



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

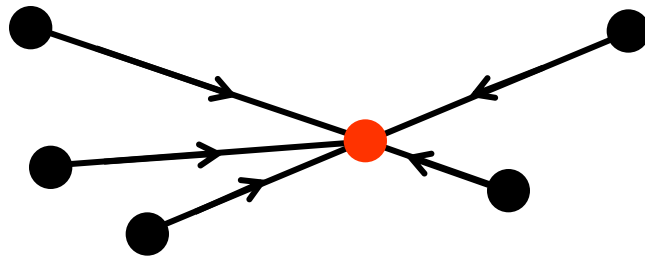
A Page Rank Matrix

- 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)



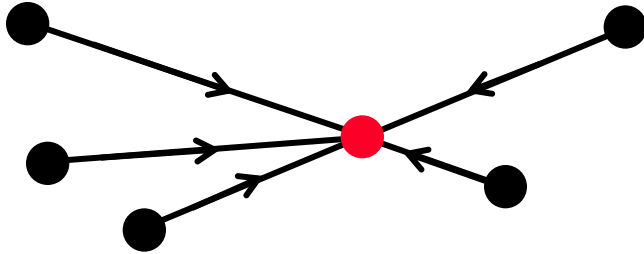
Google and the Random Surfer

How does Google figure out which web pages are most important?



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

Random Surfer Rule



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

- With probability p ($= .85$), follow a random link from this page.
- With probability $m = 1 - p$ ($= .15$), choose a page at random.
- (If there are no links out of this page, pretend it links to all other pages.)