

Link Analysis (Part I)

Professor Wei-Min Shen
University of Southern California

With slide contributions from
J. Leskovec, Anand Rajaraman, Jeffrey D. Ullman

OUTLINE

Problem and Motivation

- How to rank a web page?

Three Approaches

1. Page Rank
2. Topic-Specific (personalized) Page Rank
3. Web Spam Detection

Web as a Graph

- Web as a directed graph:

- Nodes: Webpages
- Edges: Hyperlinks

I teach a class on Networks.

CS224W:
Classes are in the Gates building

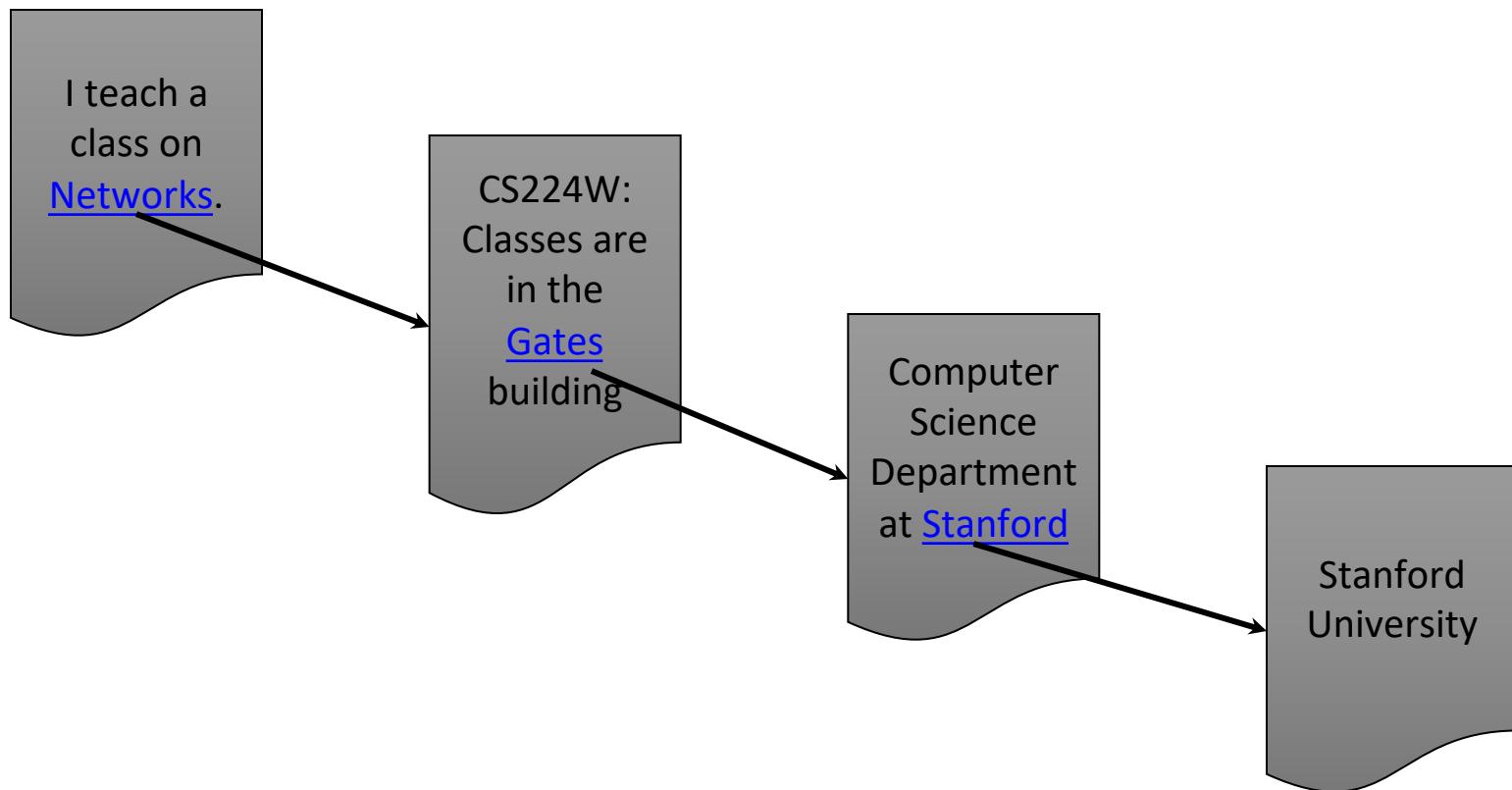
Computer Science Department at Stanford

Stanford University

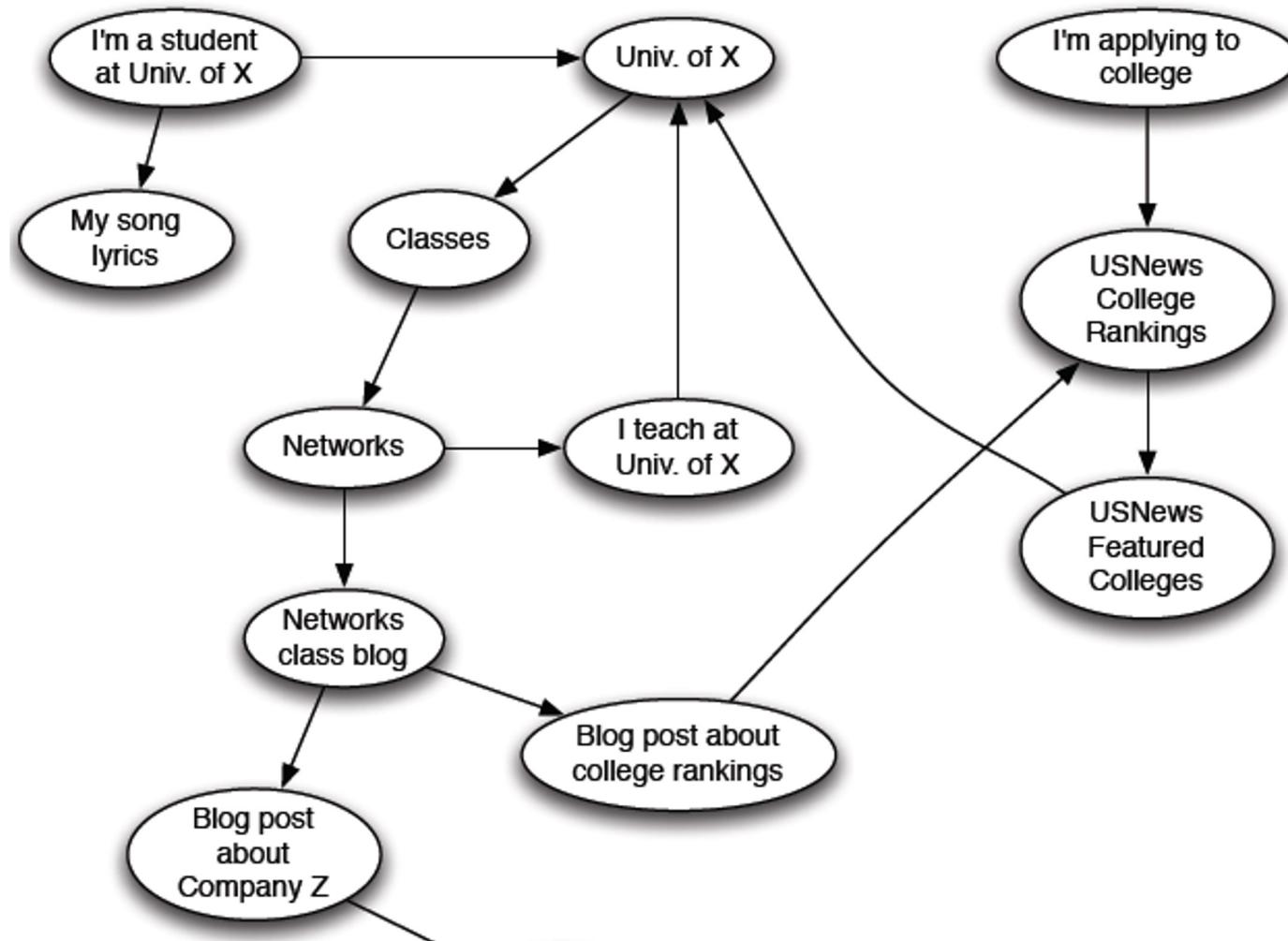
Web as a Graph

- Web as a directed graph:

- Nodes: Webpages
- Edges: Hyperlinks



Web as a Directed Graph



Broad Question

• How to organize the Web?

• First try: Human curated Web directories

- Yahoo, DMOZ, LookSmart

The screenshot shows the homepage of DMOZ (Open Directory Project). At the top, there's a green header bar with the DMOZ logo, a search bar, and social media links. Below the header is a grid of 12 categories: Arts, Business, Computers, Health, Home, Reference, Games, Recreation, Science, News, Shopping, and Sports. Each category has a small icon and a brief description. At the bottom, there's a footer with statistics: 91,730 Editors, 1,031,547 Categories, 3,887,669 Sites, and 90 Languages.

dmoz About Become an Editor Suggest a Site Help Login

Welcome to DMOZ!
It's the Web, Organized.

Learn more #OrganizeTheWeb

+ Search DMOZ

Arts
Movies, Television, Music...

Business
Jobs, Real Estate, Investing...

Computers
Internet, Software, Hardware...

Health
Fitness, Medicine, Alternative...

Home
Family, Consumers, Cooking...

Reference
Maps, Education, Libraries...

Games
Video Games, RPGs, Gambling...

Recreation
Travel, Food, Outdoors, Humor...

Science
Biology, Psychology, Physics...

News
Media, Newspapers, Weather...

Shopping
Clothing, Food, Gifts...

Society
People, Religion, Issues...

Kids & Teens Directory
Arts, School Time, Teen Life...

Regional
US, Canada, UK, Europe...

Sports
Baseball, Soccer, Basketball...

DMOZ around the World
Deutsch, Français, 日本語, Italiano, Español, Русский, Nederlands, Polski, Türkçe, Dansk, 简体中文, ...

91,730 Editors 1,031,547 Categories 3,887,669 Sites 90 Languages

The screenshot shows the Yahoo homepage from September 2016. It features a large search bar at the top with the text "Remove the Shopping Gap to Play". Below the search bar are several promotional banners for "NEW COOL RANDOM", "Yahoo! Shopping", and "Web Launch". To the right of the search bar are "Options" and "Text-Only Yahoo - Combinations". On the left side, there's a sidebar with a list of categories: Arts, Business and Economy, Computers and Internet, Education, Entertainment, Government, Health, News, Recreation, Reference, Regional, Science, Social Science, and Society and Culture. At the bottom of the sidebar, it says "#OrganizeTheWeb".

NEW COOL RANDOM

Yahoo! Shopping Remove the Shopping Gap to Play Web Launch Options

Now Open: Yahoo! Shopping

● Arts
Humanities, Photography, Archaeology, ...

● Business and Economy [Dmoz]
Directory, Investments, Classifieds, Taxes, ...

● Computers and Internet [Dmoz]
Hardware, WWW, Software, Mathematics, ...

● Education
Universities, K-12, Colleges, ...

● Entertainment [Dmoz]
TV, Movies, Music, Magazines, ...

● Government
Politics [Dmoz], Agencies, Law, Military, ...

● Health
Medicine, Drugs, Diseases, Fitness, ...

● News [Dmoz]
World [Dmoz], Daily, Current Events, ...

● Recreation
Sports [Dmoz], Games, Travel, Avatars, ...

● Reference
Libraries, Dictionaries, Fact Books, ...

● Regional
Countries, Regions, U.S. States, ...

● Science
Cell Biology, Astronomy, Engineering, ...

● Social Science
Anthropology, Sociology, Economics, ...

● Society and Culture
People, Environment, Religion, ...

#OrganizeTheWeb

Text-Only Yahoo - Combinations

Broad Question (Cont'd)

- **How to organize the Web?**

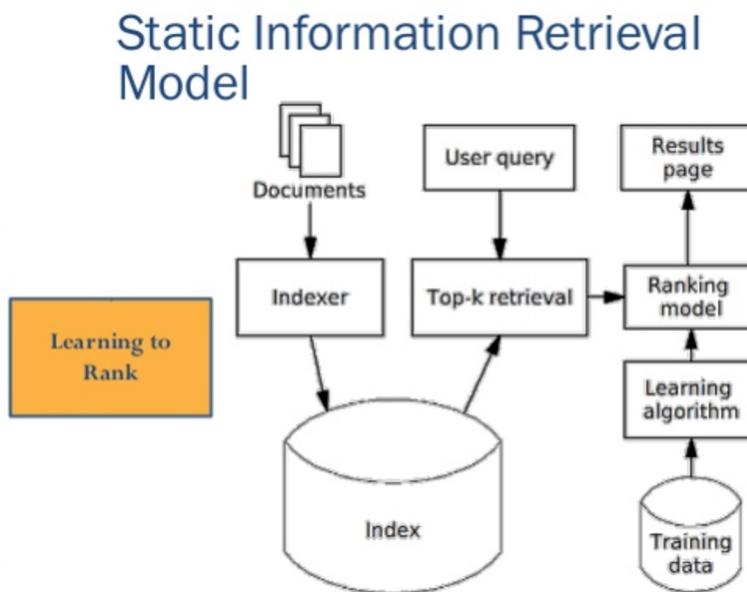
- Second try: **Web Search**

- **Information Retrieval** investigates:

Find relevant docs in a small
and trusted set

- Newspaper articles,
patents, etc.

But: Web is **huge**, full of untrusted
documents, random things, web spam,
etc.



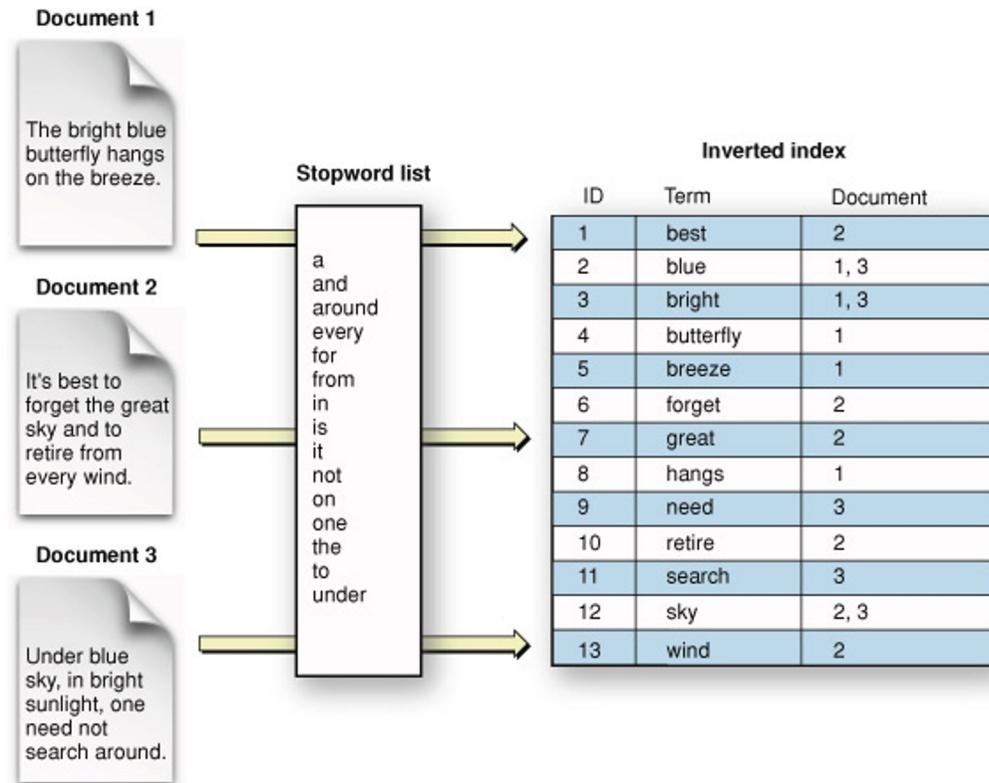
Early Web Search

- Keywords extracted from web pages
 - E.g., title, content
 - Used to build **inverted index**
- Queries are matched with web pages
 - Via lookup in the **inverted index**
 - Pages ranked by occurrences of query keywords

```
"a": {2}  
"banana": {2}  
"is": {0, 1, 2}  
"it": {0, 1, 2}  
"what": {0, 1}
```

Inverted Index

- Problem: susceptible to **term spam**



https://developer.apple.com/library/mac/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html

Term Spam

- Disguise a page as something it is not about
 - E.g., adding thousands of keyword “movies”
 - Actual content may be some advertisement
 - Fool search engine to return it for query “movies”
- May even fade spam words into background
- Spam pages may be based on top-ranked pages

Web Search: Two Challenges

Two challenges of web search:

- (1) Web contains many sources of information

Who to “trust”?

- Trick: Trustworthy pages may point to each other!

- (2) What is the “best” answer to query “newspaper”?

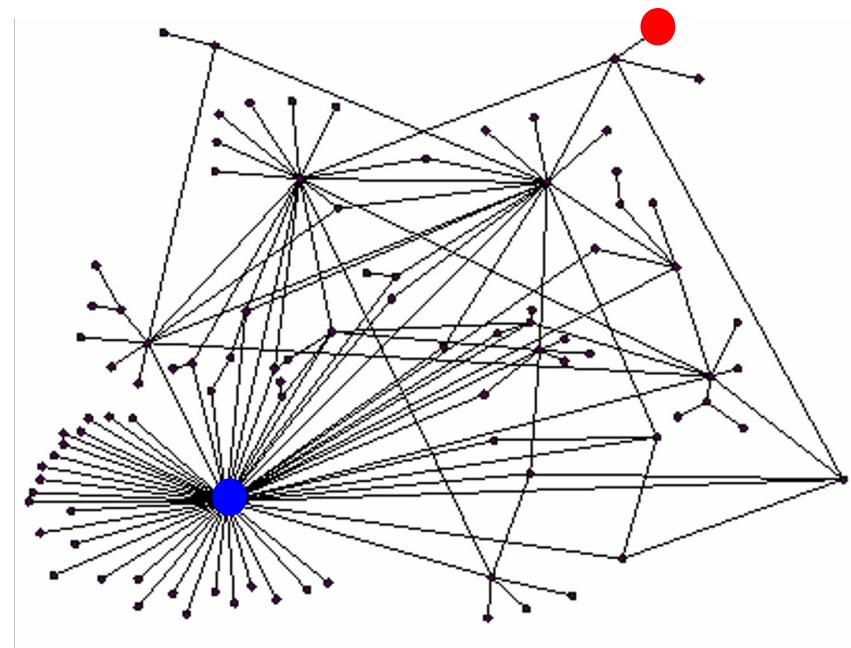
- No single right answer

- Trick: Pages that actually know about newspapers might all be pointing to many newspapers

Ranking Nodes on the Graph

- All web pages are not equally “important”
blog.bob.com vs. www.usc.edu

- There is large diversity in the web-graph node connectivity
- Let's rank the pages by the link structure!



Link Analysis Algorithms

- We will cover the following **Link Analysis approaches** for computing **importance** of nodes in a graph:
 - Page Rank
 - Topic-Specific (Personalized) Page Rank
 - Web Spam Detection Algorithms

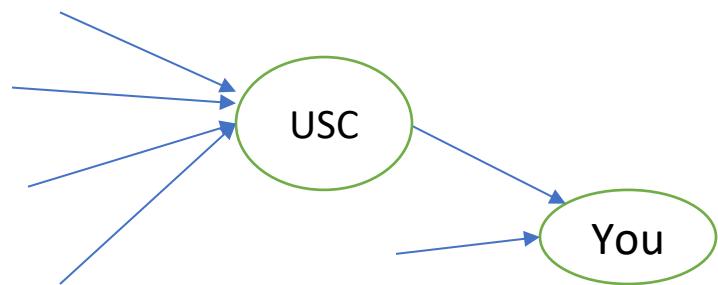
PageRank: The “Flow” Formulation

PageRank: Combating Term Spam

- Key idea: rank pages by linkage too

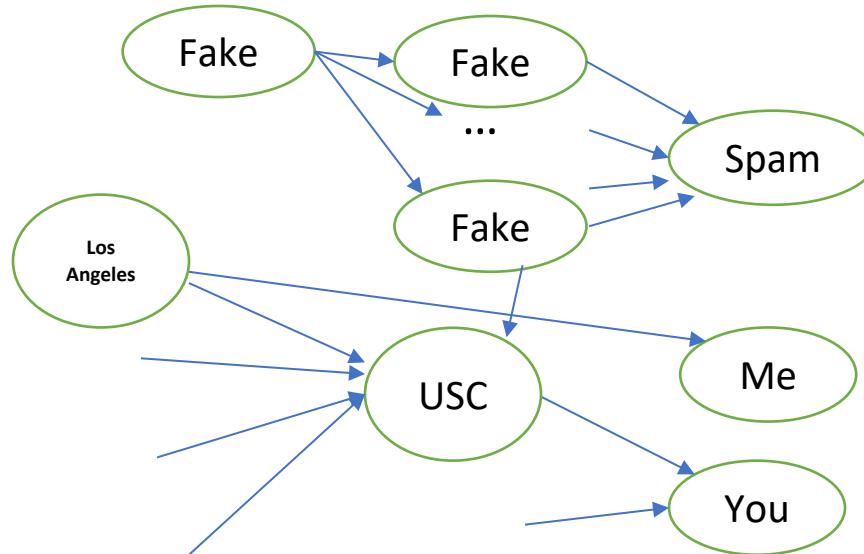
- How **many** pages point to a page
- How **important** these pages are
=> PageRank

- USC.edu can be important
 - because many pages point to it
- Your home page can be important
 - If it is pointed to by USC 😊



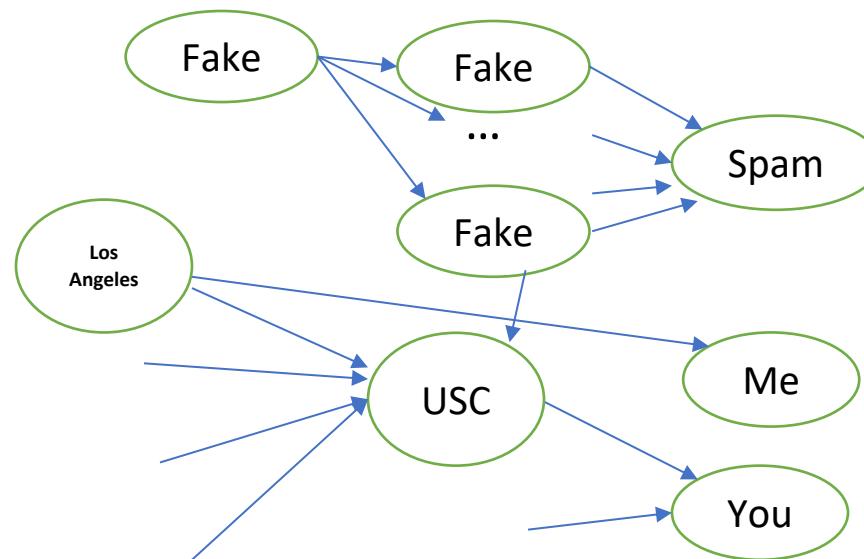
Random Surfer Model

- Random surfer of web
 - starts from any page
 - follows its outgoing links randomly
- Page is important if it attracts a large # of surfers



PageRank

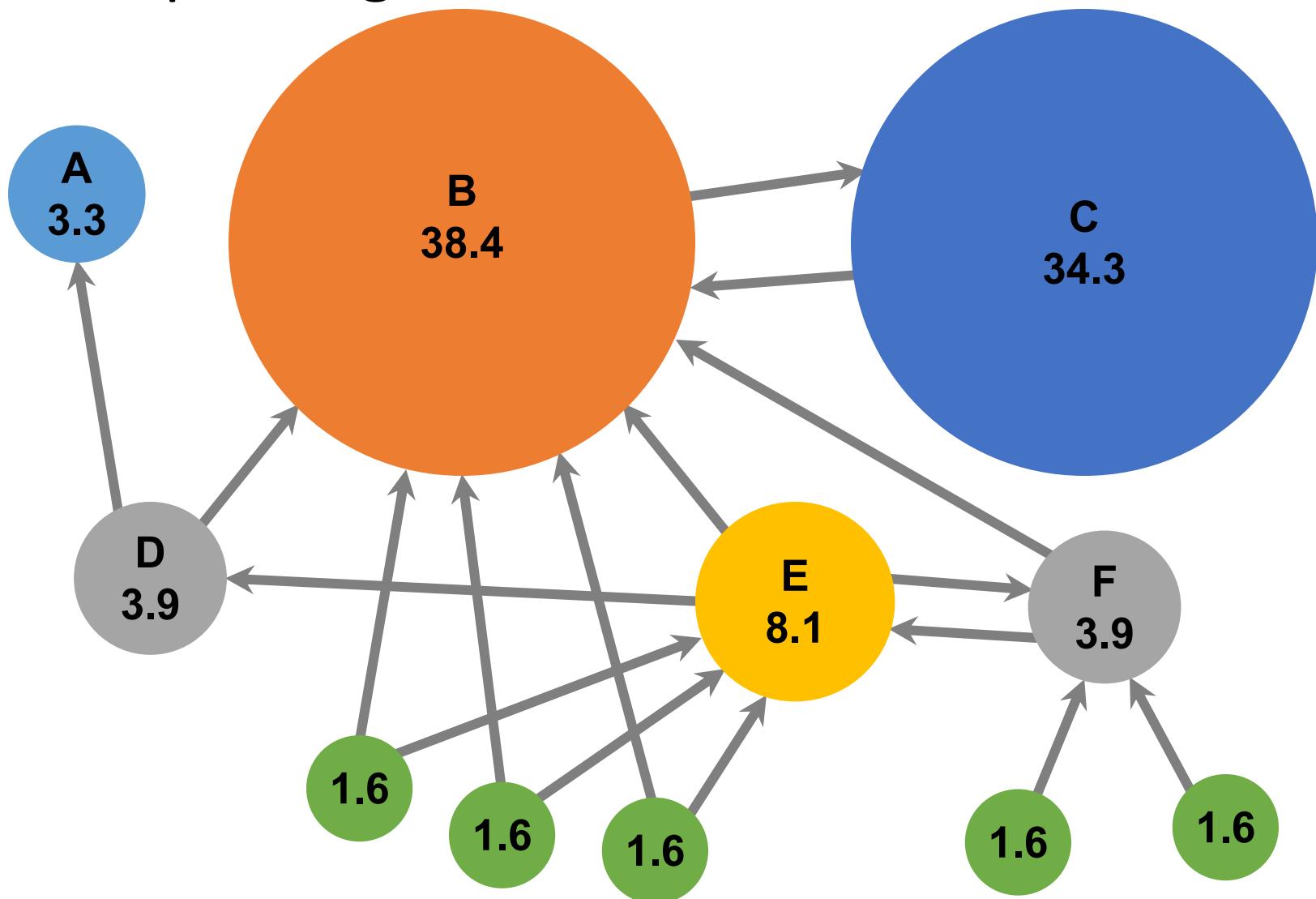
- What is the Probability that a random surfer lands on the page?



Intuition

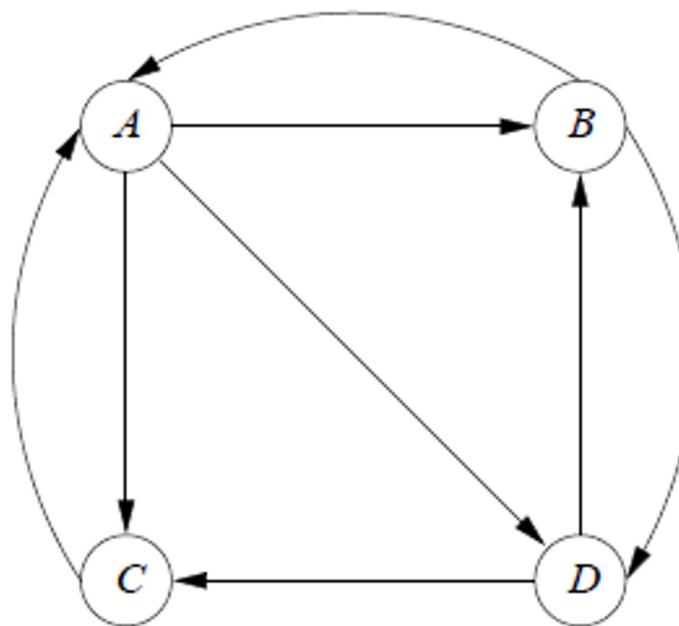
- If a page is important, then
 - many other pages may directly/indirectly link to it
 - random surfer can easily find it
- Spam pages are **less connected**
 - So less chance to attract random surfer
- Random surfer model more robust than manual approach
 - A **collective voting scheme**

Example: PageRank Scores



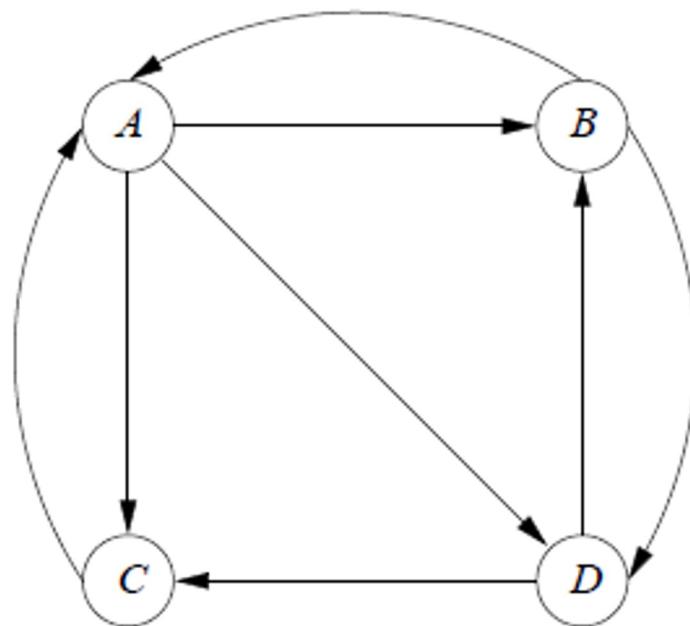
Assumption: A Strongly Connected Web Graph

- Nodes = pages
 - Edges = hyperlinks between pages
 - every node is reachable from every other node
-



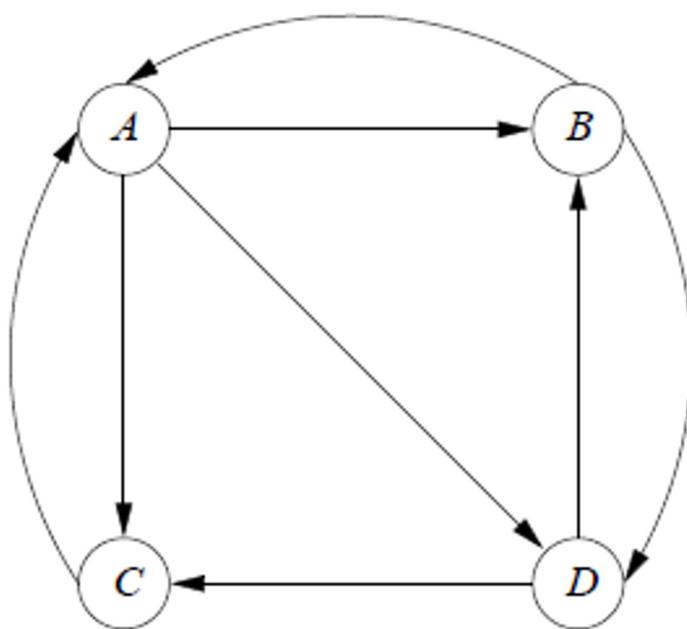
Model: Random Surfer on the Graph

- Can start at any node
 - If at A, can next go to B, C, or D, each with 1/3 prob.
 - If at B, can go to A and D, each with 1/2 prob.
 - So on...



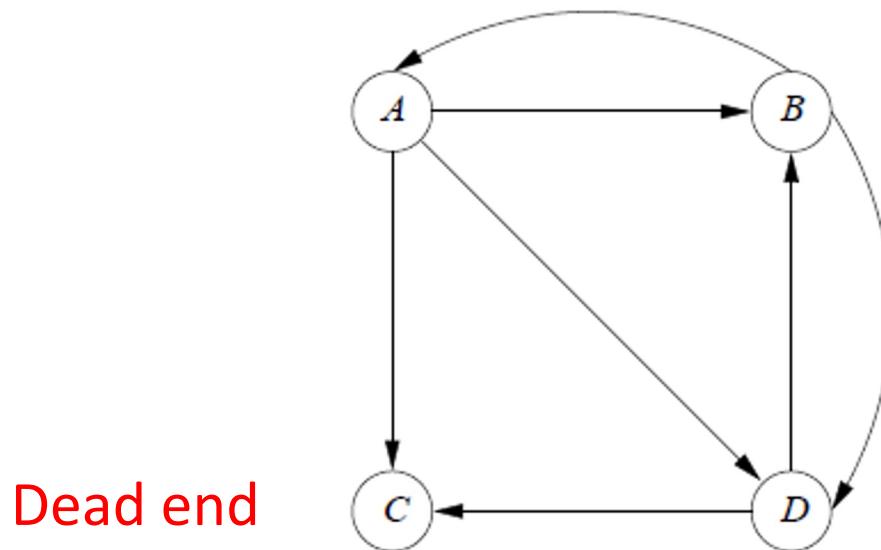
Random Surfer Property: Memoryless

- Where to go from node X is not affected by how the surfer got to X



Extreme Case: Dead End

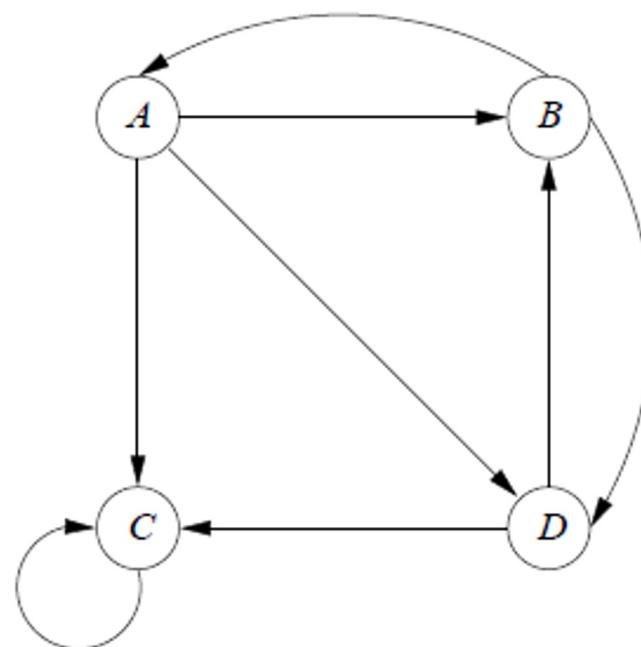
- Dead end: a page with no edges out
 - Absorb PageRanks
 - PageRank $\Rightarrow 0$ for any page that can reach the dead end (including the dead end itself)



Extreme Case: Spider Trap

- Group of pages with no edges going out of group
 - Absorb all PageRanks (rank of C =>1, others =>0)
 - Surfer can never leave, once trapped
 - Can have > 1 such trap nodes

Spider trap



PageRank: Formulation Details

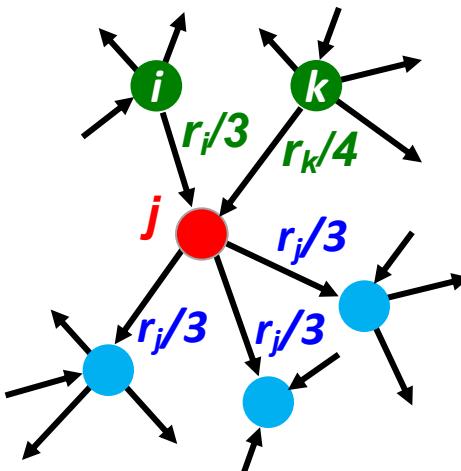
PageRank: Links as Votes

- **Idea: Links as votes**
 - **Page is more important if it has more links**
 - In-coming links? Out-going links?
- **Think of in-links as votes:**
 - www.stanford.edu has 23,400 in-links
 - www.inf553.cs.usc.edu has 1 in-link
- **Are all in-links equal?**
 - **Links from important pages count more**
 - A Recursive question!

Simple Recursive Formulation

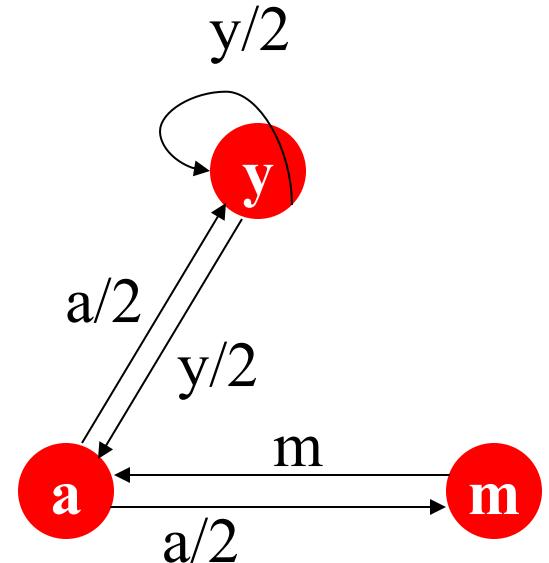
- Each link's vote is proportional to the **importance** of its source page
- If page j with importance r_j has n out-links, each link gets r_j/n votes
- Page j 's own importance is the sum of the votes on its in-links

$$r_j = r_i/3 + r_k/4$$



PageRank: The “Flow” Model

- A “vote” from an important page is worth more
- A page is important if it is pointed to by other important pages
- Define a “rank” r_j for page j



$$r_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}$$

d_i = out-degree of node i

“Flow” equations:

$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

“Flow” equations:

$$\mathbf{r}_y = \mathbf{r}_y/2 + \mathbf{r}_a/2$$

$$\mathbf{r}_a = \mathbf{r}_y/2 + \mathbf{r}_m$$

$$\mathbf{r}_m = \mathbf{r}_a/2$$

Solving the flow equations

- **3 equations, 3 unknowns, no constants**

- No unique solution
- All solutions equivalent modulo scale factor

- **Additional constraint forces uniqueness**

- **Add some constant constraints:** e.g., $\mathbf{r}_y + \mathbf{r}_m + \mathbf{r}_a = 1$
- Solve for unique solutions: $\mathbf{r}_y = 2/5, \mathbf{r}_a = 2/5, \mathbf{r}_m = 1/5$

- **Gaussian elimination method (in the later slides) works for small examples, but we need a better method for large graphs**

$$\left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 5 & 35 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 2 & 8 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 0 & 0 \end{array} \right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & -2 & -3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

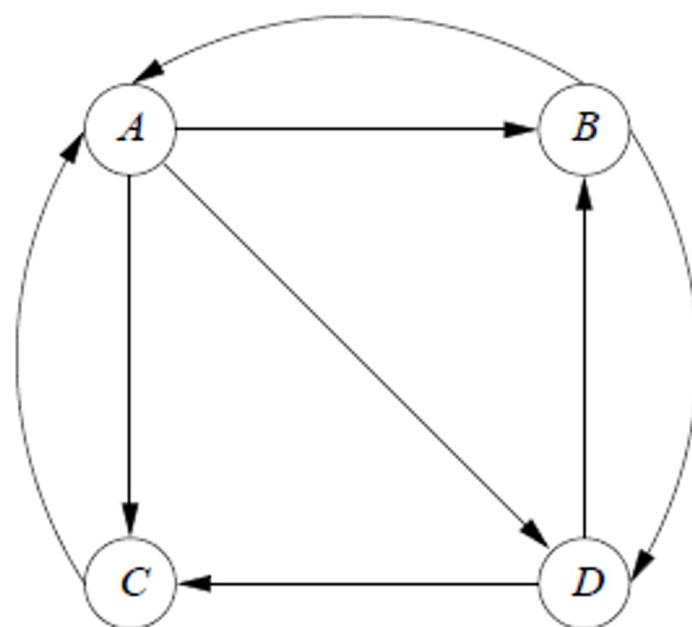
PageRank: Matrix formulation

- Stochastic Transition (or adjacency) Matrix M
- Suppose page j has n outlinks
 - If outlink $j \rightarrow i$, then $M_{ij}=1/n$
 - Else $M_{ij}=0$
- M is a column stochastic matrix
 - Each column sums to 1

Transition Matrix

- $M[i,j] = \text{prob. of going from node } j \text{ to node } i$
 - If j has k outgoing edges, prob. for each edge = $1/k$

$$M = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \end{matrix}$$



PageRank: Matrix formulation (Cont'd)

- Stochastic Transition (or adjacency) Matrix M
- Suppose page j has n outlinks
 - If outlink $j \rightarrow i$, then $M_{ij}=1/n$
 - Else $M_{ij}=0$
- M is a column stochastic matrix
 - Each column sums to 1
- Rank vector r is a vector with one entry per web page
 - r_i is the importance score of page i
- The flow equations can be written as

$$r = Mr$$

Example

- Flow equation in matrix form: $\mathbf{M}\mathbf{r} = \mathbf{r}$
- Suppose page j links to 3 pages, including i

$$\begin{matrix} & j \\ i & \left[\begin{array}{c|c} & \\ \hline & \\ & \end{array} \right] \\ 1/3 & \left[\begin{array}{c|c} & \\ \hline & \\ & \end{array} \right] \end{matrix} \quad \cdot \quad \begin{matrix} r_j \\ \vdash \end{matrix} \quad = \quad \begin{matrix} r_i \\ \vdash \end{matrix}$$
$$M \quad \cdot \quad r \quad = \quad r$$

Stationary Distribution

- Limiting probability distribution of random surfer
 - PageRanks are based on **limiting distribution**
 - **the probability distribution will converge eventually**
- Requirement for its existence
 - Graph is **strongly connected**: a node can reach any other node in the graph
=> Cannot have **dead ends, spider traps**

Eigenvectors and Eigenvalues

- An **eigenvector** of a square matrix \mathbf{A} is a non-zero vector \mathbf{v} that, when the matrix multiples \mathbf{v} , yields the same as when some scalar multiplies \mathbf{v} , the scalar multiplier often being denoted by λ
- That is:

$$\mathbf{Av} = \lambda\mathbf{v}$$

- The number λ is called the **eigenvalue** of \mathbf{A} corresponding to \mathbf{v}

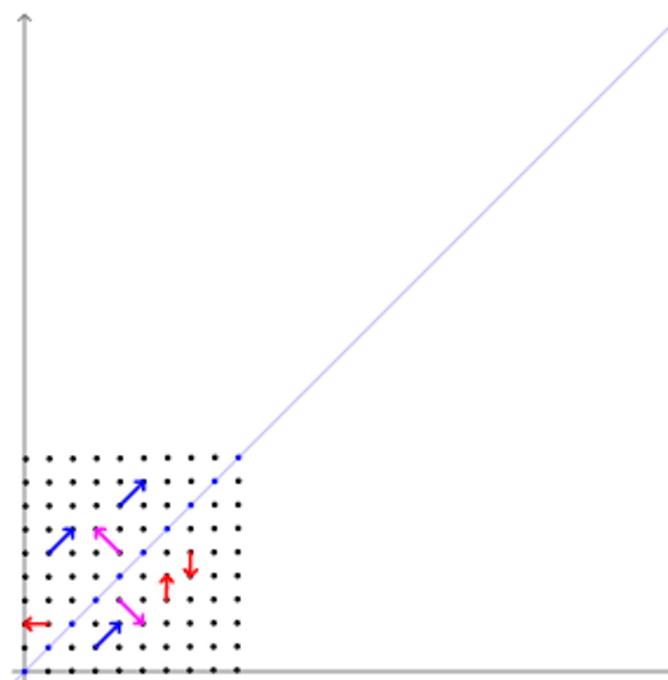
Eigenvalues and Eigenvectors Example

- The transformation matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ preserves the direction of vectors parallel to $\mathbf{v} = (1, -1)^T$ (in purple) and $\mathbf{w} = (1, 1)^T$ (in blue). The vectors in red are not parallel to either eigenvector, so, their directions are changed by the transformation.

$$A\mathbf{v} = \lambda\mathbf{v}$$

<http://setosa.io/ev/eigenvectors-and-eigenvalues/>

https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors



Eigenvector Formulation

- The flow equations can be written limiting distribution

$$\mathbf{r} = \mathbf{M} \cdot \mathbf{r}$$

- So the rank vector \mathbf{r} is an eigenvector of the stochastic web matrix \mathbf{M}

- \mathbf{r} is \mathbf{M} 's first or principal eigenvector,
with corresponding eigenvalue 1
- Largest eigenvalue of \mathbf{M} is **1** since \mathbf{M} is
column stochastic (with non-negative entries)
 - *We know \mathbf{r} is unit length and each column of \mathbf{M} sums to one*

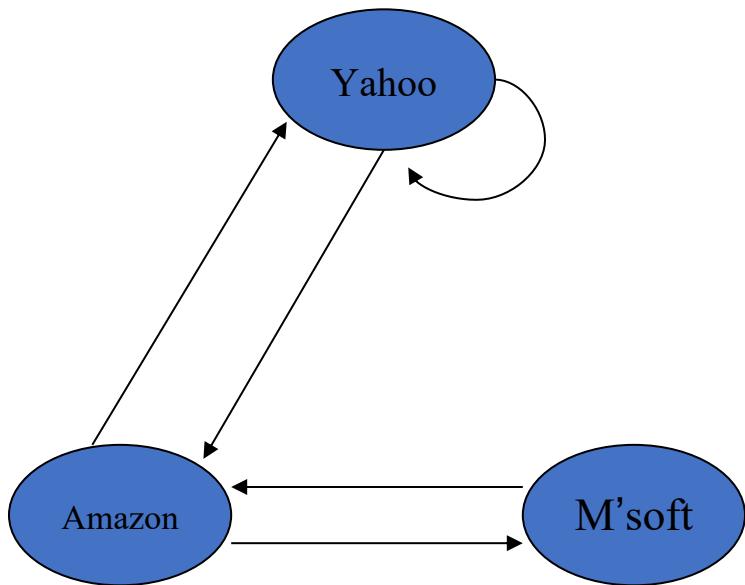
NOTE: \mathbf{x} is an eigenvector with the corresponding eigenvalue λ if:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

- We can now efficiently solve for \mathbf{r}

- 1. Power Iteration: https://en.wikipedia.org/wiki/Power_iteration
- 2. Use the principal eigenvector

Example



$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2 + r_m$$

$$r_m = r_a/2$$

	y	a	m
y	1/2	1/2	0
a	1/2	0	1
m	0	1/2	0

$$\mathbf{r} = \mathbf{M}\mathbf{r}$$

$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix}$$

Two Approaches for Solving $Mv = v$

- Power Iteration
 - Repeatedly solving $Mv^k = v^{k+1}$
- Finding the Eigenvectors
 - Solving $Mv = \lambda v$, or $(M - I)v = 0$
 - Gaussian Elimination

Random Walk Interpretation

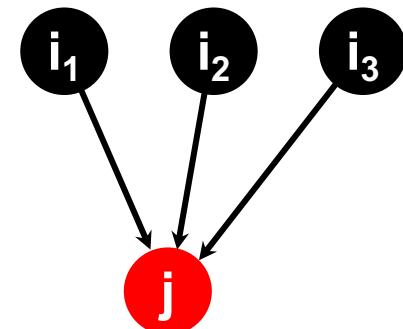
Imagine a random web surfer:

- At any time t , surfer is on some page i
- At time $t+1$, the surfer follows an out-link from i uniformly at random
- Ends up on some page j linked from i
 - Process repeats indefinitely

Let:

$p(t)$... vector whose i^{th} coordinate is the prob. that the surfer is at page i at time t

- So, $p(t)$ is a probability distribution over pages
 - -> rank vector

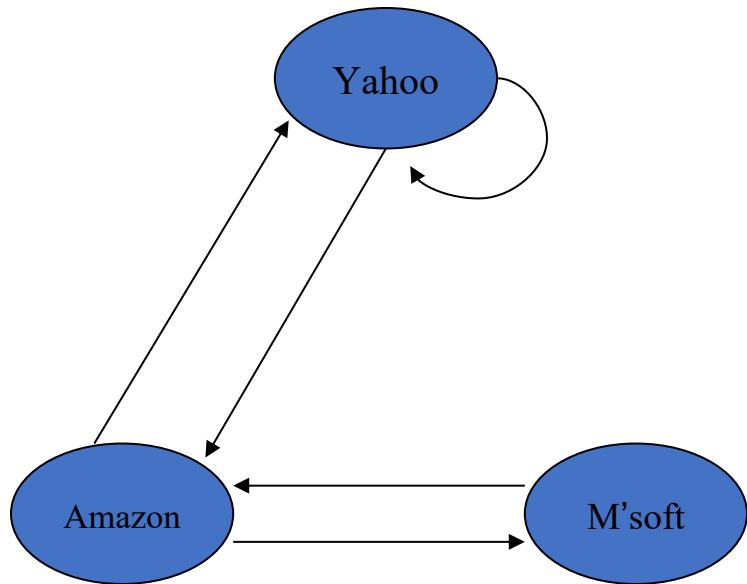


$$r_j = \sum_{i \rightarrow j} \frac{r_i}{d_{\text{out}}(i)}$$

Power Iteration Method

- Simple iterative scheme (aka **relaxation**)
- Suppose there are N web pages
- Initialize: $\mathbf{r}^0 = [1/N, \dots, 1/N]^T$
- Iterate: $\mathbf{r}^{k+1} = \mathbf{M}\mathbf{r}^k$
- Stop when $|\mathbf{r}^{k+1} - \mathbf{r}^k|_1 < \epsilon$
- $|\mathbf{x}|_1 = \sum_{1 \leq i \leq N} |x_i|$ is the L₁ norm
 - Can use any other vector norm e.g., Euclidean

Power Iteration Example 1



	y	a	m
y	1/2	1/2	0
a	1/2	0	1
m	0	1/2	0

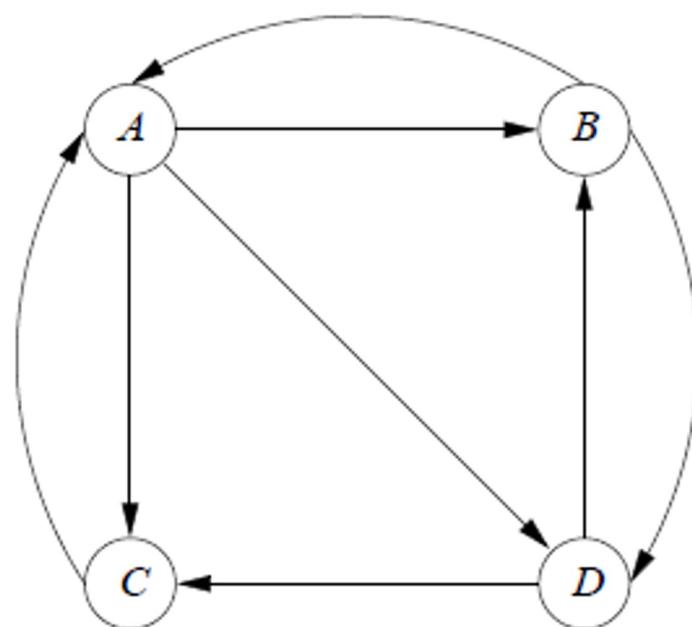
$$\mathbf{r}^{k+1} = \mathbf{M}\mathbf{r}^k$$

$$\begin{array}{lcl} r_y & = & \frac{1}{3} \quad \frac{1}{3} \quad \frac{5}{12} \quad \frac{3}{8} \quad \dots \quad \frac{2}{5} \\ r_a & = & \frac{1}{3} \quad \frac{1}{2} \quad \frac{1}{3} \quad \frac{11}{24} \quad \dots \quad \frac{2}{5} \\ r_m & = & \frac{1}{3} \quad \frac{1}{6} \quad \frac{1}{4} \quad \frac{1}{6} \quad \dots \quad \frac{1}{5} \end{array}$$

Power Iteration Example 2

- $M[i,j] = \text{prob. of going from node } j \text{ to node } i$
 - If j has k outgoing edges, prob. for each edge = $1/k$

$$M = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \end{matrix}$$



Prob. of Locations of Surfer

- Represented as a **column vector**, v
- Initially, surfer can be at any page with equal probability

$$v^0 = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Limiting Distribution

- ◆ $v^1 = Mv^0$
- ◆ $v^2 = Mv^1 (= M^2 v^0)$
- ◆ ...
- ◆ $v^i = Mv^{i-1} (= M^i v^0)$
- ◆ ...
- ◆ $v = Mv$

(from some step k on, **v** does not change any more)

Compute Next Distribution

◆ $v^1 = Mv^0$

➤ $v_i^1 = \sum_{j=1}^n M_{ij} v_j^0$

$$\begin{array}{cccc} & A & B & C & D \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \left[\begin{matrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{matrix} \right] & \left[\begin{matrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{matrix} \right] \end{array}$$

◆ E.g., $v_0^1 = 0 * \frac{1}{4} + \frac{1}{2} * \frac{1}{4} + 1 * \frac{1}{4} + 0 * \frac{1}{4} = \frac{3}{8}$

➤ i.e., prob. at A is 3/8 (or 9/24) after step 1

$$v^0 = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \left[\begin{matrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{matrix} \right] \quad v^1 = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \left[\begin{matrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{matrix} \right] \quad V^2 = ? \quad V^3 = ?$$

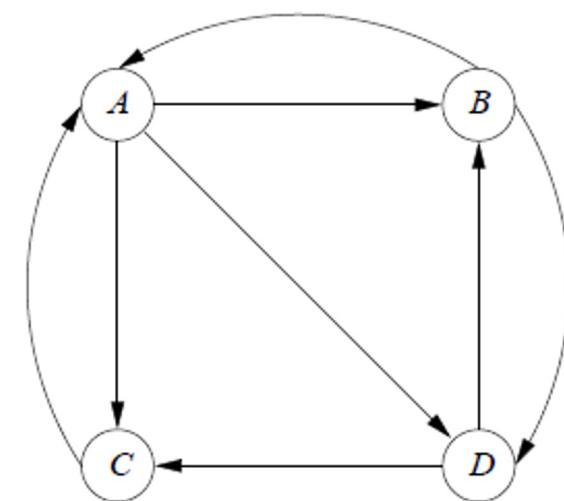
Solving $Mv = v$ by Power Iteration

- Note: $\sum_{i=1}^n v_i^k = 1$, after every step k
- Usually stop when little change btw iterations
- In practice, 50-75 iterations for Web

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \quad \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix}$$

NOTE: x is an eigenvector with the corresponding eigenvalue λ if:
 $Ax = \lambda x$

$$\begin{bmatrix} 11/32 \\ 7/32 \\ 7/32 \\ 7/32 \end{bmatrix}, \dots, \begin{bmatrix} 3/9 \\ 2/9 \\ 2/9 \\ 2/9 \end{bmatrix}$$



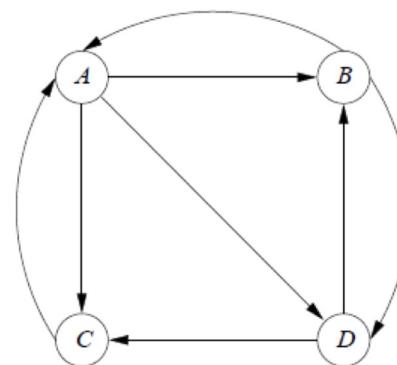
Two Approaches for Solving $Mv = v$

- Power Iteration
 - Repeatedly solving $Mv^k = v^{k+1}$
- Finding the Eigenvectors
 - Solving $Mv = \lambda v$, or $(M - I)v = 0$
 - Gaussian Elimination

Solving $Mv = v$ by Finding the Principal Eigenvector (when $\lambda=1$)

- $Mv=\lambda v$
 - λ : eigenvalue, v : eigenvector
- M is a (left) stochastic matrix
 - Each column adds up to 1
- Largest eigenvalue of a stochastic matrix is $\lambda=1$
 - Corresponding eigenvector is the principal eigenvector
 - Proof in the next few slides

$$M = \begin{bmatrix} & A & B & C & D \\ A & 0 & 1/2 & 1 & 0 \\ B & 1/3 & 0 & 0 & 1/2 \\ C & 1/3 & 0 & 0 & 1/2 \\ D & 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$



Characteristic Polynomial

Calculating the Eigenvalues

- ◆ $Mv = \lambda v$
- ◆ $(M - \lambda I)v = 0$
- ◆ v is not a null vector , so
➤ $\det(M - \lambda I) = 0$

$$\begin{vmatrix} -\lambda & 1/2 & 1 & 0 \\ 1/3 & -\lambda & 0 & 1/2 \\ 1/3 & 0 & -\lambda & 1/2 \\ 1/3 & 1/2 & 0 & -\lambda \end{vmatrix} = 0$$

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

I is the identity matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} |A| &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$

Compute the determinant using cofactors in this column (next slide)

Characteristic Polynomial

$$\begin{vmatrix} \frac{1}{3} & -\lambda & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & -\lambda \end{vmatrix} - \lambda \begin{vmatrix} -\lambda & \frac{1}{2} & 0 \\ \frac{1}{3} & -\lambda & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & -\lambda \end{vmatrix}$$

$$= \lambda \begin{vmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & -\lambda \end{vmatrix} - 1/2 \begin{vmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{vmatrix} - \lambda(-1/2) \begin{vmatrix} -\lambda & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{vmatrix} - \lambda \begin{vmatrix} -\lambda & \frac{1}{2} \\ \frac{1}{3} & -\lambda \end{vmatrix}$$

$$= \lambda \left(-\frac{1}{3} \lambda - \frac{1}{6} \right) - \lambda \left[-\frac{1}{2} \left(-\frac{1}{2} \lambda - \frac{1}{6} \right) - \lambda \left(\lambda^2 - \frac{1}{6} \right) \right]$$

$$= \lambda^4 - \frac{3}{4} \lambda^2 - \frac{\lambda}{4} = \lambda(\lambda-1)(\lambda^2 + \lambda + \frac{1}{4}) = \lambda(\lambda-1)(\lambda + \frac{1}{2})^2$$

Characteristic Polynomial

- ◆ $Mv = \lambda v$
- ◆ $(M - \lambda I)v = 0$
- ◆ $\det(M - \lambda I) = 0$

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

$$\begin{vmatrix} -\lambda & 1/2 & 1 & 0 \\ 1/3 & -\lambda & 0 & 1/2 \\ 1/3 & 0 & -\lambda & 1/2 \\ 1/3 & 1/2 & 0 & -\lambda \end{vmatrix} = 0$$

$$\lambda(\lambda-1)(\lambda + \frac{1}{2})^2 = 0 \Rightarrow \lambda_1 = 0, \lambda_2 = 1, \lambda_3 = \lambda_4 = -\frac{1}{2}$$

Solving $Mv = v$ by Finding the Principal Eigenvector (when $\lambda=1$)

- $Mv = \lambda v$
- $(M - \lambda I)v = 0$
- $\det(M - \lambda I) = 0$

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

$$\lambda(\lambda-1)(\lambda+\frac{1}{2})^2=0 \Rightarrow \lambda_1 = 0, \lambda_2 = 1, \lambda_3 = \lambda_4 = -\frac{1}{2}$$

- Problem becomes solving $(M - I)v = 0$
- Use Gaussian elimination

Solving $Mv = v$ by Gaussian Elimination

$$\bullet (M - I)v = 0$$

$$\begin{bmatrix} -1 & 1/2 & 1 & 0 \\ 1/3 & -1 & 0 & 1/2 \\ 1/3 & 0 & -1 & 1/2 \\ 1/3 & 1/2 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & -1 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ \frac{1}{3} & -1 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & -1 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & -\frac{5}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{6} & -\frac{2}{3} & \frac{1}{2} \\ 0 & \frac{2}{3} & \frac{1}{3} & -1 \end{bmatrix}$$

Solving $Mv = v$ by Gaussian Elimination

$$\bullet (M - I)v = 0$$

$$\begin{bmatrix} -1 & 1/2 & 1 & 0 \\ 1/3 & -1 & 0 & 1/2 \\ 1/3 & 0 & -1 & 1/2 \\ 1/3 & 1/2 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & -\frac{5}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{6} & -\frac{2}{3} & \frac{1}{2} \\ 0 & \frac{2}{3} & \frac{1}{3} & -1 \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & -\frac{6}{15} & -\frac{3}{5} \\ 0 & \frac{1}{6} & -\frac{2}{3} & \frac{1}{2} \\ 0 & \frac{2}{3} & \frac{1}{3} & -1 \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & -\frac{6}{15} & -\frac{3}{5} \\ 0 & 0 & -\frac{3}{5} & \frac{3}{5} \\ 0 & 0 & \frac{3}{5} & -\frac{3}{5} \end{bmatrix}$$

Solving $Mv = v$ by Gaussian Elimination

$$\bullet (M - I)v = 0$$

$$\begin{bmatrix} -1 & 1/2 & 1 & 0 \\ 1/3 & -1 & 0 & 1/2 \\ 1/3 & 0 & -1 & 1/2 \\ 1/3 & 1/2 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & -\frac{6}{15} & -\frac{3}{5} \\ 0 & 0 & -\frac{3}{5} & \frac{3}{5} \\ 0 & 0 & \frac{3}{5} & -\frac{3}{5} \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & -\frac{6}{15} & -\frac{3}{5} \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\quad} \begin{bmatrix} 1 & -\frac{1}{2} & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\xrightarrow{\quad} \begin{bmatrix} 1 & 0 & 0 & -\frac{3}{2} \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Solving $Mv = v$ by Gaussian Elimination

$$\bullet (M - I)v = 0$$

$$\begin{bmatrix} -1 & 1/2 & 1 & 0 \\ 1/3 & -1 & 0 & 1/2 \\ 1/3 & 0 & -1 & 1/2 \\ 1/3 & 1/2 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cccc} 1 & 0 & 0 & -\frac{3}{2} \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right] \quad \left. \begin{array}{l} v_1 - 3/2 v_4 = 0 \\ v_2 - v_4 = 0 \\ v_3 - v_4 = 0 \\ v_1 + v_2 + v_3 + v_4 = 1 \end{array} \right\} \quad \begin{array}{l} v_1 = 1/3 \\ v_2 = v_3 = v_4 = 2/9 \end{array}$$

The “principal” eigenvector is $[1/3, 2/9, 2/9, 2/9]$, when $\lambda = 1$

Comparison of Solutions

- $Mv = v$
 - Solution 1: power iteration
 - Solution 2: Gaussian elimination
- Power iteration
 - Complexity: $O(kn^2)$, where $k = \# \text{ of iterations}$
- Gaussian elimination
 - Complexity: $O(n^3)$
 - $n^2 + (n-1)^2 + \dots + 2^2$

Next Week

- Link Analysis (Part II)
 - Solving the flow model: $Mv = v$
 - Power Iteration and Gaussian Elimination
 - Dead-ends and spider traps (teleporting and taxation)
 - PageRank for Search Engines (MapReduce)
 - Topic-specific: (augment for general-popularity)
 - Trust-Rank: (fighting against link spams)
 - Hubs-and-Authorities: (multiple importances)