

Link Analysis

Stony Brook University
CSE545, Spring 2019

The Web , circa 1998

ALTA VISTA
Technology
View Multimedia From Our Vantage Point

AUTOSITE
USA CANADA
Car Buying & Car Insurance
Pain Relief
Buy and insure new cars & trucks online

Click here for advertising information - reach millions every month!

Search the Web and Display the Results in Standard Form

Search with Digital's Alta Vista [Advanced Search](#) [Add URL](#)

Contests
Make Me Laugh...

Creative Web
Create a Site...

Match keywords, language (information retrieval)

Explore directory

excite

search reviews city.net live! tours
people finder maps yellow pages news

Excite Search: twice the power of the competition.

What:

Where: World Wide Web

INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION.

Excite Reviews: site reviews by the web's best editorial team.

Take an Excite Seeing Tour

Excite on TV

WRED
MIXED NEWS HOT MIXED WRED MAGAZINE

HOT
The WRED Search Center

YAHOO!

Canada Messenger Chat Room
Know when friends are online! Click to download Yahoo! Messenger
Yahoo! Mail free from anywhere

[advanced search](#)

Y! Shopping Depts: Books, CDs, Computers, DVDs Stores: Gap, Clinique, Coach and more

Shop Auctions Autos Classifieds Shopping Travel Yellow Pages Maps Media Finance Quotes News Sports Weather Connect Careers Chat Clubs Sex/Games Greetings Mail Members Messenger Mobile Personal People Search Photos Personal Adk Books Software Calendar My Yahoo! FaxDirect Fun Games Kids Movies Music Radio TV more...

Yahoo! Auctions Bid, buy, or sell anything!
Categories: Antiques Computers Games Date Equipment Cameras Electronics Golf Clubs Stamps Yesterday's Cars Sports Cards Video Longaberger Classic Books Stamps Tokeman Ice.com Baseball Cards McGraw-Hill Inter Bonds Soda Giffey Jr. Items

In the News
U.S. rescues 15M spp. plans fish
Suzanne Casati admits to sexual relationship with missing sister
Attorney Barry Levin found dead
Date Eamhand Jr. www.Paper400
Wimbledon - Tour de France more...

Arts & Humanities Literature, Photography... Full Coverage, Newspapers, TV...

Business & Economy B2B, Finance, Shopping, Jobs...

Computers & Internet Internet, WWW, Software, Games...

Education College and University, K-12...

Entertainment Cool Links, Movies, Music, Music...

Government Elections, Military, Law, Taxes...

Health Medicine, Diseases, Drugs, Fitness...

News & Media Full Coverage, Newspapers, TV...

Recreation & Sports Sports, Travel, Autos, Outdoors...

Reference Libraries, Dictionaries, Quotations...

Regional Countries, Regions, US States...

Science Animals, Astronomy, Engineering...

Social Science Archaeology, Economics, Languages...

Society & Culture People, Environment, Religion...

Marketplace
new! eBay, shops, London
Epinet - sponsored by Pepsi
Y! Store - become part of Yahoo! Shopping
Y! Careers - find a job, post your resume
Mobile phones, service plans and accessories

Broadcast Events
Open ET, PGA, Western Open
blink-182 - Artist of the month more...

Inside Yahoo!
Y! Games - backgammon, checkers, hearts, chess, pool/billiard
Y! Movies - Scarie Movie 2, King of the Dragon, Cars and Dogs
new! Play five Fantasy Baseball - mid-season version
Y! Photos - post your party pics

powered by **COMPAQ**

Local Yahoo!'s
Europe - Denmark - France - Germany - Italy - Norway - Spain - Sweden - UK & Ireland
Asia Pacific - Asia - Australia & NZ - China - HK - India - Japan - Korea - Singapore - Taiwan
Americas - Argentina - Brazil - Canada - Chinese - Mexico - Spanish
U.S. Cities - Atlanta - Boston - Chicago - Dallas/FW - LA - NYC - SE Bay - Wash DC - more...

More Yahoo!'s
Outdoor - Autos - Bazaars - Careers - Health - Living - Outdoors - Pets - Real Estate - Technology
Entertainment - Astrology - Breakfast - Events - Games - Movies - Music - Radio - Tickets - TV - more
Finance - Banking - Bill Pay - Insurance - Loans - Taxes - Finance/Investment - more
Local - Classifieds - Events - Listings - Maps - Restaurants - Yellow Pages - more
News - Top Stories - Business - Entertainment - Lifestyle - Politics - Sports - Technology - Weather
Publishing - Business - Clubs - Experts - Quotes - Photos - Home Pages - Message Boards
Small Business - Biz Marketplace - Domain Registration - Small Biz Center - Store Building - Web Hosting
Access Yahoo! via: Pages, PDAs, Web-enabled Phones and Voice (1-800-MY-Yahoo)

Make Yahoo! your home page

How to Suggest a Site - Company Info - Copyright Policy - Terms of Service - Contributors - Jobs - Advertising

Copyright © 2001 Yahoo! Inc. All rights reserved.
[Privacy Policy](#)

The Web , circa 1998



Easy to game with
"term spam"

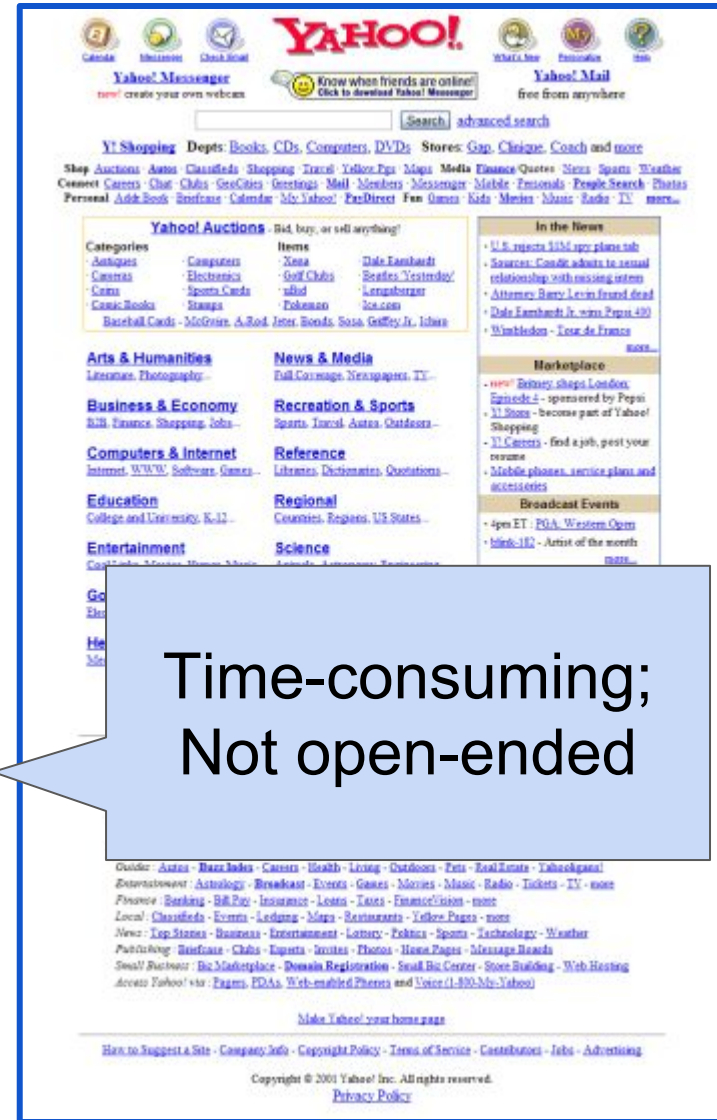
Match keywords, language (*information retrieval*)



Explore directory



Time-consuming;
Not open-ended



Enter PageRank

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

*Computer Science Department,
Stanford University, Stanford, CA 94305, USA*
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure and produce much text and hyperlink c

The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

...

Abstract

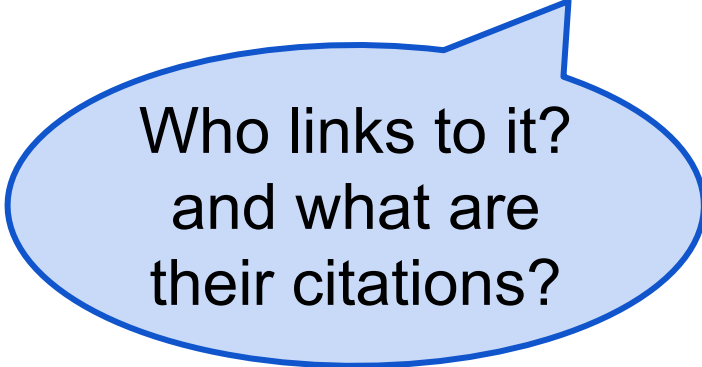
The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively

PageRank

Key Idea: Consider the **citations** of the website.

PageRank

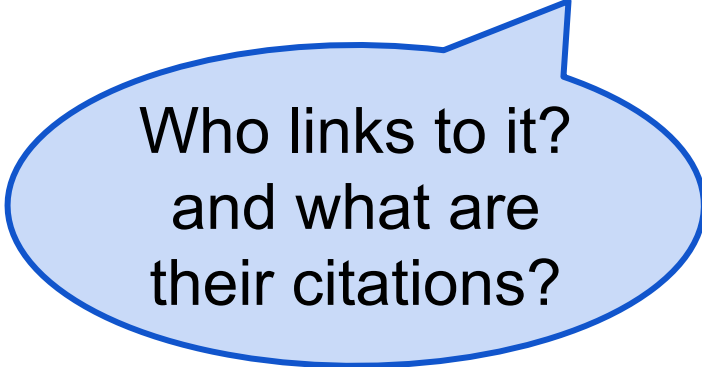
Key Idea: Consider the **citations** of the website.



Who links to it?
and what are
their citations?

PageRank

Key Idea: Consider the **citations** of the website.



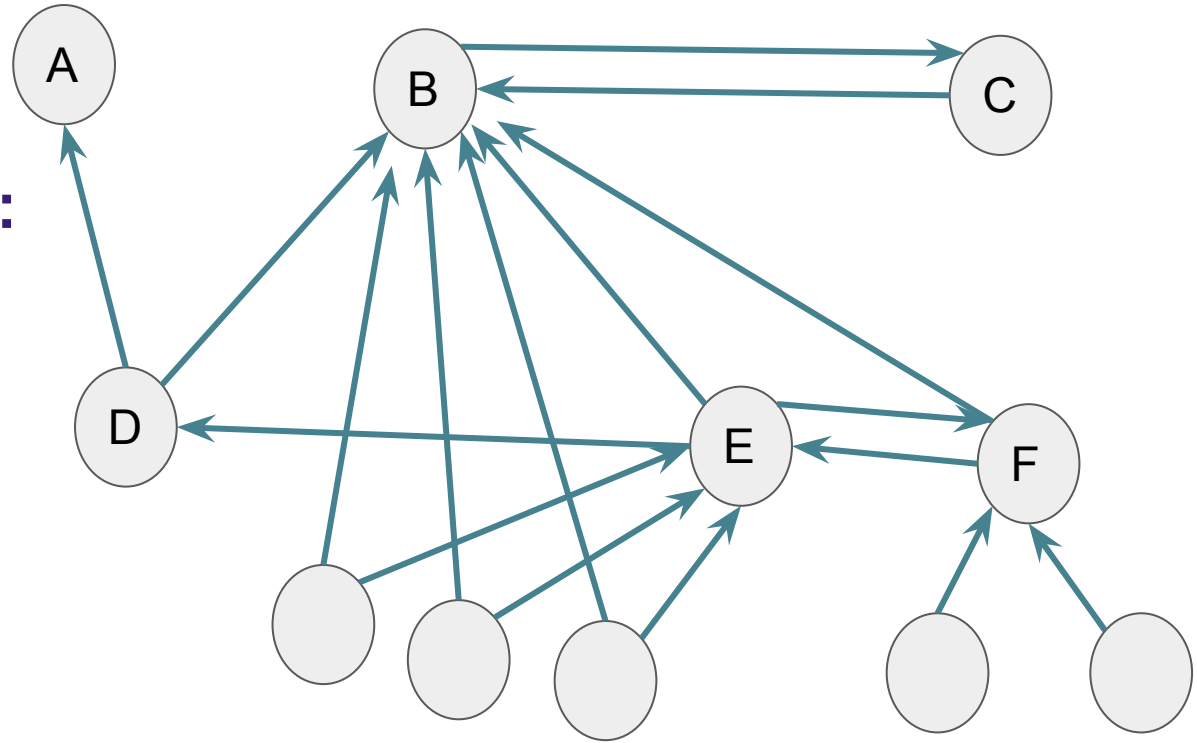
Who links to it?
and what are
their citations?

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?

PageRank

View 1: Flow Model:
in-links as votes

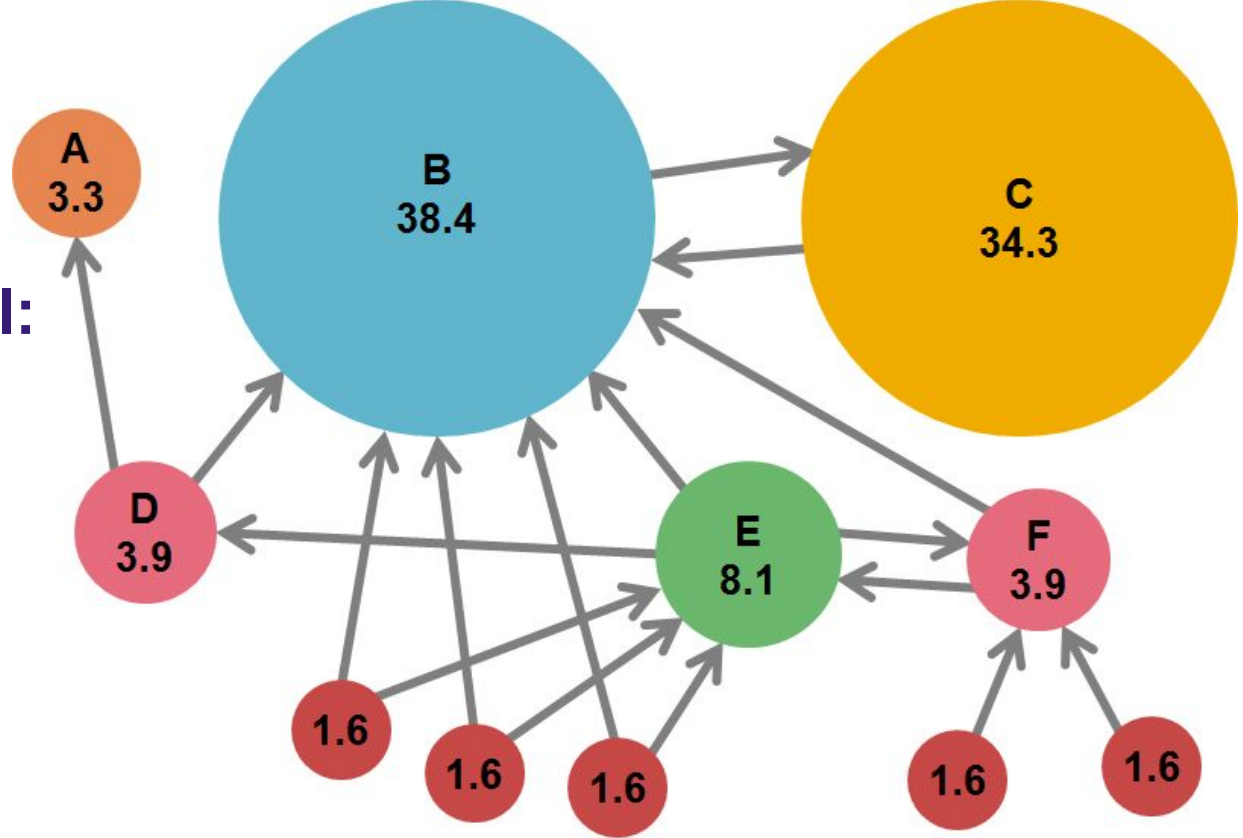


Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?

PageRank

View 1: Flow Model:
in-links as votes



J. Leskovec, A. Rajaraman, J. Ullman: Mining of Massive Datasets, <http://www.mmds.org>

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?

PageRank

View 1: Flow Model:

in-links (citations) as votes

but, citations from important pages should count more.

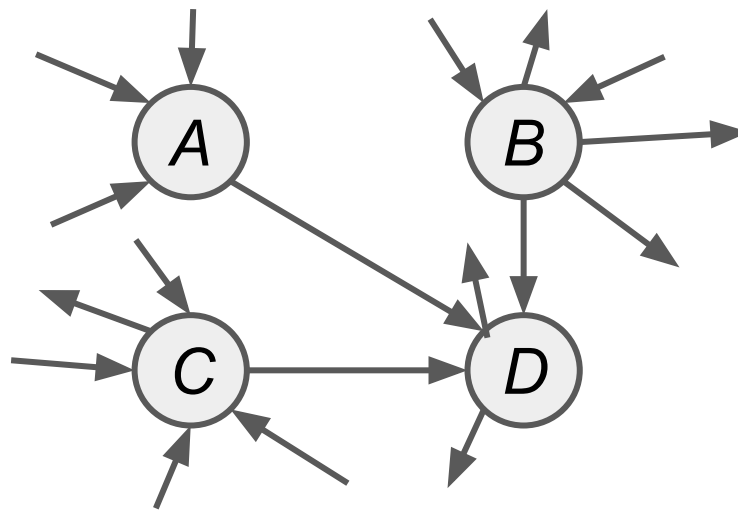
=> Use recursion to figure out if each page is important.

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?

PageRank

View 1: Flow Model:



How to compute?

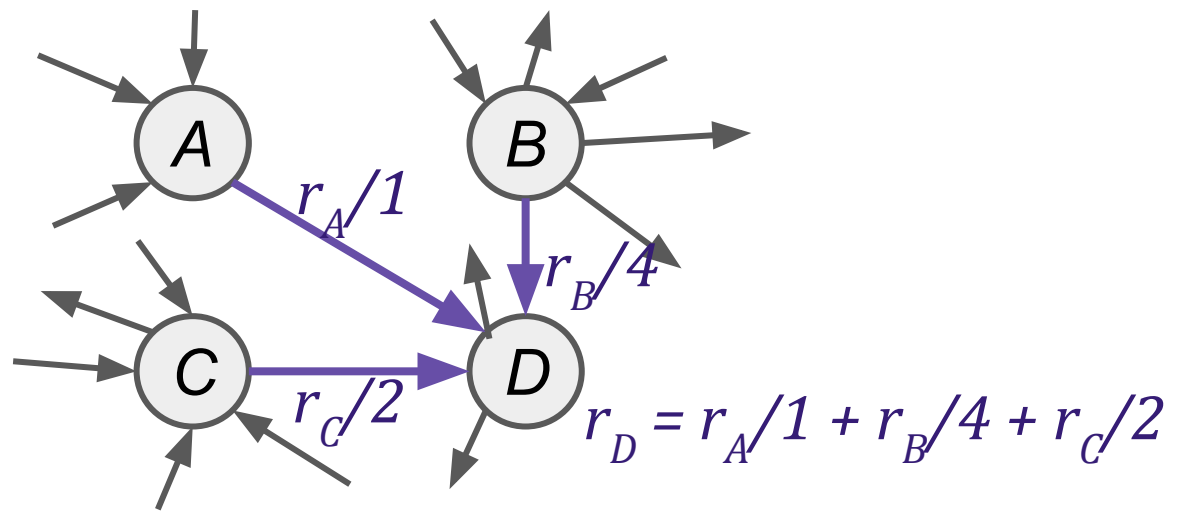
Each page (j) has an importance (i.e. rank, r_j)

$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

PageRank

View 1: Flow Model:



How to compute?

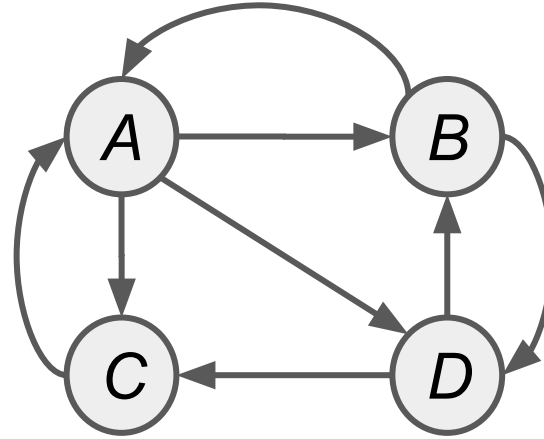
Each page (j) has an importance (i.e. rank, r_j)

$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

PageRank

View 1: Flow Model:



How to compute?

Each page (j) has an importance (i.e. rank, r_j)

$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

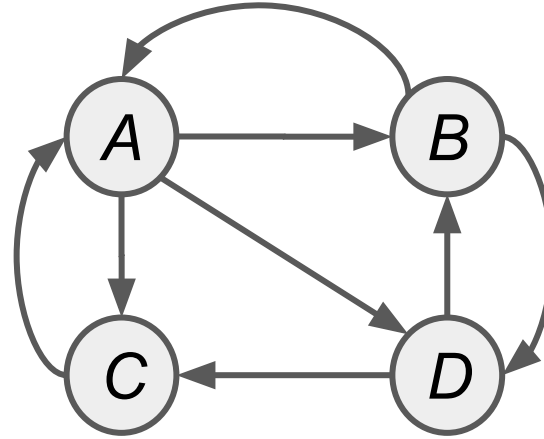
$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

PageRank

View 1: Flow Model:

A System of Equations:

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$



How to compute?

Each page (j) has an importance (i.e. rank, r_j)

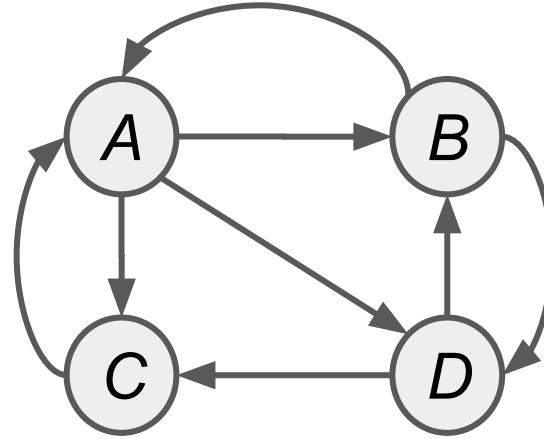
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

PageRank

View 1: Flow Model:

A System of Equations:



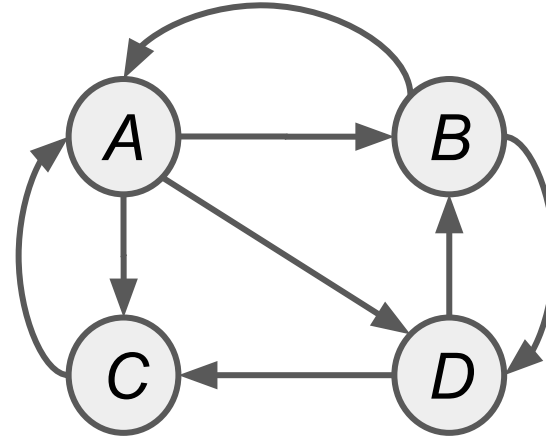
$$\begin{aligned}r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\r_D &= \frac{r_A}{3} + \frac{r_B}{2}\end{aligned}$$

How to compute?

Each page (j) has an importance (i.e. rank, r_j)

$$\begin{aligned}vote_j &= \frac{r_j}{n_j} && (n_j \text{ is } |\text{out-links}|) \\r_j &= \sum_{i \in \text{inLinks}(j)} vote_i\end{aligned}$$

PageRank



View 1: Flow Model: Solve

$$1 = r_A + r_B + r_C + r_D$$

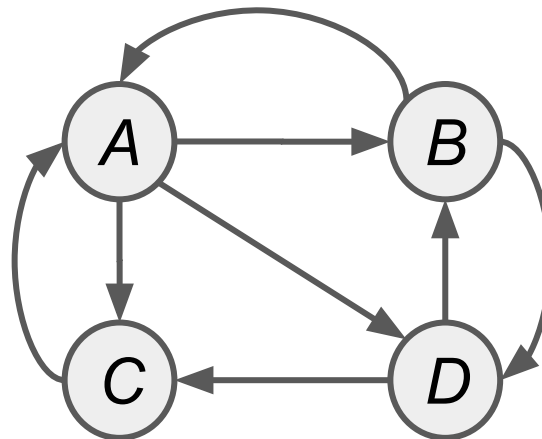
$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$

How to compute?

Each page (j) has an importance (i.e. rank, r_j)

$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$
$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

PageRank



$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$

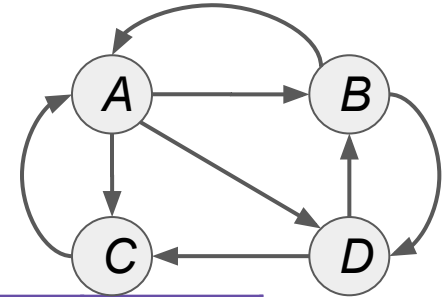
<i>to \ from</i>	A	B	C	D
A	0	1/2	1	0
B	1/3	0	0	1/2
C	1/3	0	0	1/2
D	1/3	1/2	0	0

Transition Matrix, M

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

View 2: Matrix Formulation

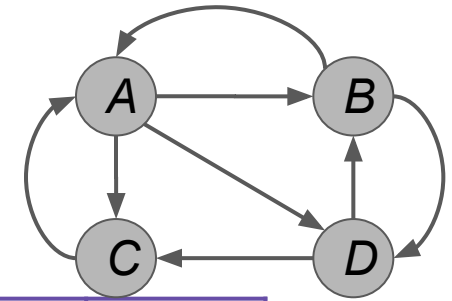
$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$

$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

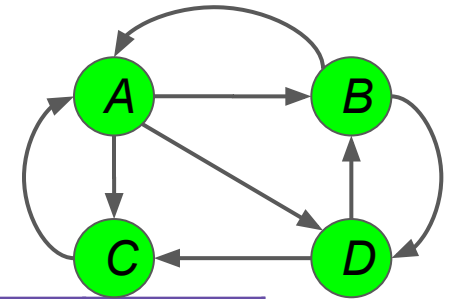
Innovation: What pages would a “random Web surfer” end up at?

To Start, all are equally likely at $\frac{1}{4}$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

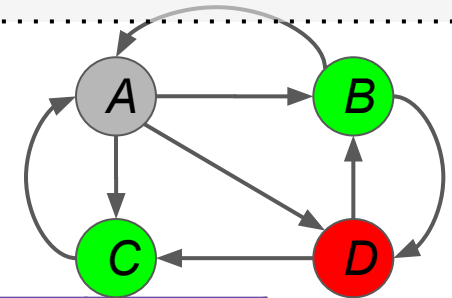
Innovation: What pages would a “random Web surfer” end up at?

To Start, all are equally likely at $\frac{1}{4}$: ends up at D

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

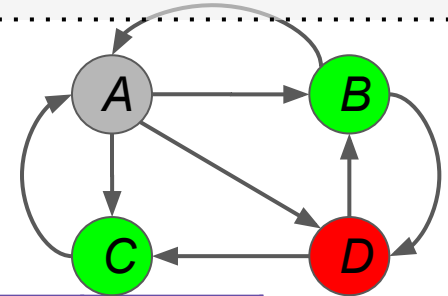
To Start, all are equally likely at $\frac{1}{4}$: ends up at D

C and B are then equally likely: $\rightarrow D \rightarrow B = \frac{1}{4} * \frac{1}{2}$; $\rightarrow D \rightarrow C = \frac{1}{4} * \frac{1}{2}$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

To Start, all are equally likely at $\frac{1}{4}$: ends up at D

C and B are then equally likely: $\rightarrow D \rightarrow B = \frac{1}{4} * \frac{1}{2}$; $\rightarrow D \rightarrow C = \frac{1}{4} * \frac{1}{2}$

Ends up at C: then A is only option: $\rightarrow D \rightarrow C \rightarrow A = \frac{1}{4} * \frac{1}{2} * 1$

View 2: Matrix Formulation

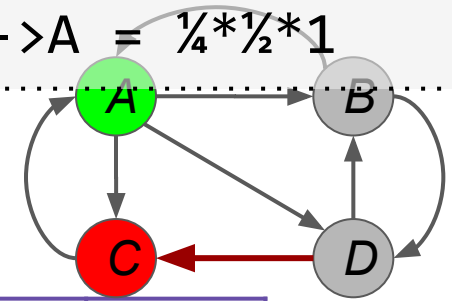
$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$

$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

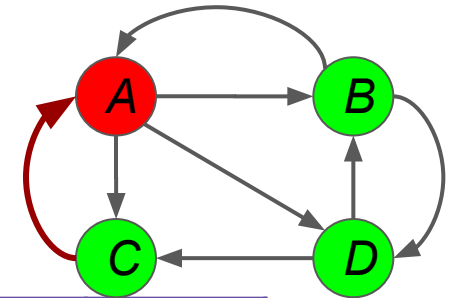
Innovation: What pages would a “random Web surfer” end up at?

...

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

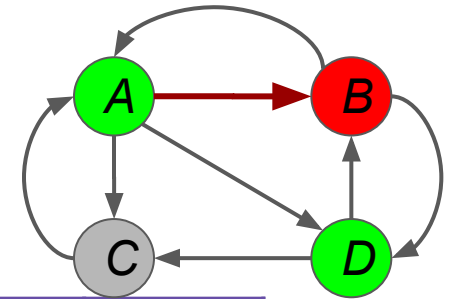
Innovation: What pages would a “random Web surfer” end up at?

...

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

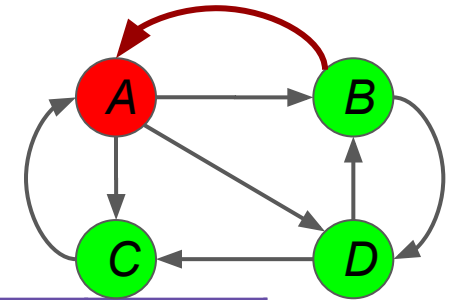
Innovation: What pages would a “random Web surfer” end up at?

...

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

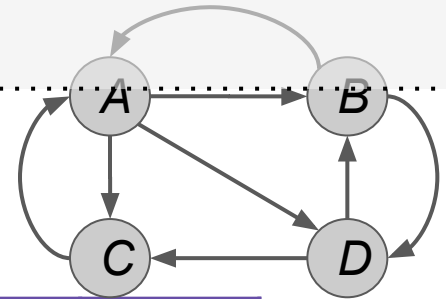
Innovation: What pages would a “random Web surfer” end up at?

To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4},]$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

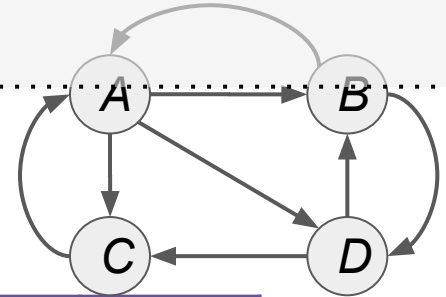
To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

after 1st iteration: $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4},]$

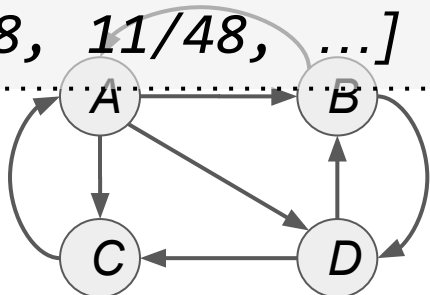
after 1st iteration: $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

after 2nd iteration: $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$
$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$



<i>to \ from</i>	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

Transition Matrix, M

Innovation: What pages would a “random Web surfer” end up at?

To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

after 1st iteration: $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

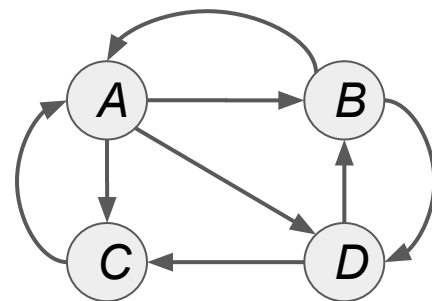
after 2nd iteration: $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

Power iteration algorithm

```
initialize:  $r[0] = [1/N, \dots, 1/N]$ ,  
            $r[-1] = [0, \dots, 0]$ 
```

```
while (err_norm( $r[t]$ ,  $r[t-1]$ ) > min_err):
```

```
err_norm( $v1$ ,  $v2$ ) =  $|v1 - v2|$  #L1 norm
```



to \ from	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

“Transition Matrix”, M

Innovation: What pages would a “random Web surfer” end up at?

To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

after 1st iteration: $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

after 2nd iteration: $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

Power iteration algorithm

```
initialize:  $r[0] = [1/N, \dots, 1/N]$ ,  
            $r[-1] = [0, \dots, 0]$ 
```

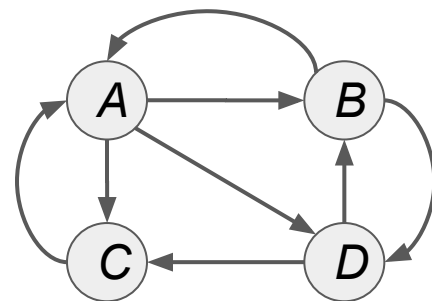
```
while (err_norm( $r[t]$ ,  $r[t-1]$ ) > min_err):
```

```
     $r[t+1] = M \cdot r[t]$ 
```

```
     $t += 1$ 
```

```
solution =  $r[t]$ 
```

```
err_norm( $v1$ ,  $v2$ ) =  $|v1 - v2|$  #L1 norm
```



to \ from	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

“Transition Matrix”, M

As err_norm gets smaller we are moving toward: $r = M \cdot r$

View 3: Eigenvectors:

Power iteration algorithm

```
initialize:   $r[0] = [1/N, \dots, 1/N],$   
             $r[-1] = [0, \dots, 0]$   
while ( $\text{err\_norm}(r[t], r[t-1]) > \text{min\_err}$ ):  
     $r[t+1] = M \cdot r[t]$   
     $t += 1$   
solution =  $r[t]$   
  
 $\text{err\_norm}(v1, v2) = |v1 - v2|$  #L1 norm
```

As `err_norm` gets smaller we are moving toward: $r = M \cdot r$

View 3: Eigenvectors:

We are actually just finding the *eigenvector* of M .

Power iteration algorithm

```
initialize:  r[0] = [1/N, ..., 1/N]
             r[-1]=[0,...,0]
while (err_norm(r[t],r[t-1])>min_err):
    r[t+1] = M·r[t]
    t+=1
solution = r[t]

err_norm(v1, v2) = |v1 - v2| #L1 norm
```

finds the...

x is an
eigenvector of A if:
 $A \cdot x = \lambda \cdot x$

As `err_norm` gets smaller we are moving toward: $r = M \cdot r$

View 3: Eigenvectors:

We are actually just finding the *eigenvector* of M .

Power iteration algorithm

```
initialize:  r[0] = [1/N, ..., 1/N]
             r[-1]=[0,...,0]
while (err_norm(r[t],r[t-1])>min_err):
    r[t+1] = M·r[t]
    t+=1
solution = r[t]

err_norm(v1, v2) = sum(|v1 - v2|)
                  #L1 norm
```

finds the... →

x is an
eigenvector of A if:
 $A \cdot x = \lambda \cdot x$

$\lambda = 1$ (eigenvalue for 1st principal eigenvector)
since columns of M sum to 1.
Thus, if r is x , then $Mr=1r$

View 4: Markov Process

Where is surfer at time $t+1$? $p(t+1) = M \cdot p(t)$

Suppose: $p(t+1) = p(t)$, then $p(t)$ is a *stationary distribution* of a **random walk**.

Thus, r is a stationary distribution. Probability of being at given node.

View 4: Markov Process

Where is surfer at time $t+1$? $p(t+1) = M \cdot p(t)$

Suppose: $p(t+1) = p(t)$, then $p(t)$ is a *stationary distribution* of a **random walk**.

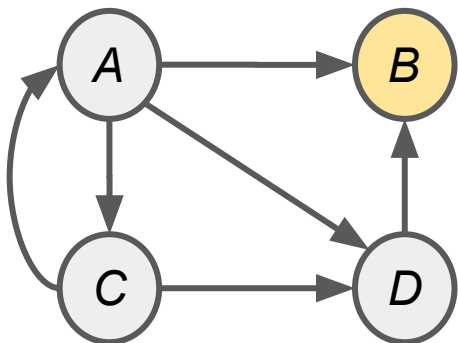
Thus, r is a stationary distribution. Probability of being at given node.

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
 - Stationary distributions have a unique distribution if:
 - No “*dead-ends*”: a node can’t propagate its rank
 - No “*spider traps*”: set of nodes with no way out.

Also known as being *stochastic*, *irreducible*, and *aperiodic*.

View 4: Markov Process - Problems for vanilla PI



to \ from	A	B	C	D
A	0	0	1	0
B	1/3	0	0	1
C	1/3	0	0	0
D	1/3	0	0	0

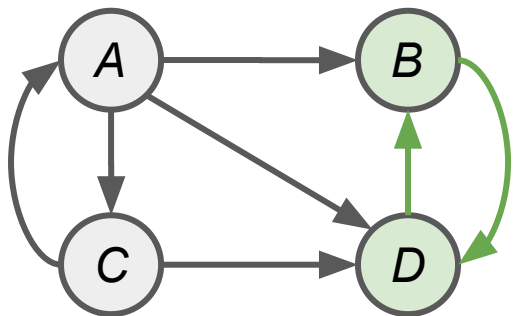
What would r converge to?

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
 - Stationary distributions have a unique distribution if:
 - No “**dead-ends**”: a node can’t propagate its rank
 - No “**spider traps**”: set of nodes with no way out.

Also known as being *stochastic*, *irreducible*, and *aperiodic*.

View 4: Markov Process - Problems for vanilla PI



to \ from	A	B	C	D
A	0	0	1	0
B	1/3	0	0	1
C	1/3	0	0	0
D	1/3	1	0	0

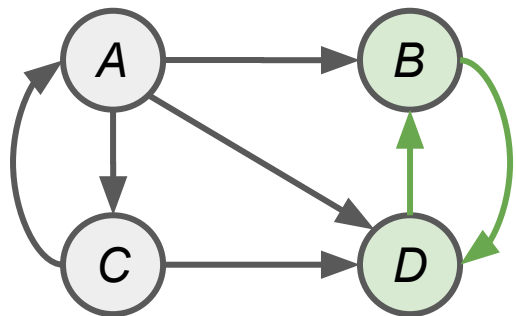
What would r converge to?

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
 - Stationary distributions have a unique distribution if:
 - No “*dead-ends*”: a node can’t propagate its rank
 - No “*spider traps*”: set of nodes with no way out.

Also known as being *stochastic*, *irreducible*, and *aperiodic*.

View 4: Markov Process - Problems for vanilla PI



to \ from	A	B	C	D
A	0	0	1	0
B	1/3	0	0	1
C	1/3	0	0	0
D	1/3	1	0	0

What would r converge to?

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
 - Stationary distributions have a unique distribution if:

same node doesn't repeat at regular intervals

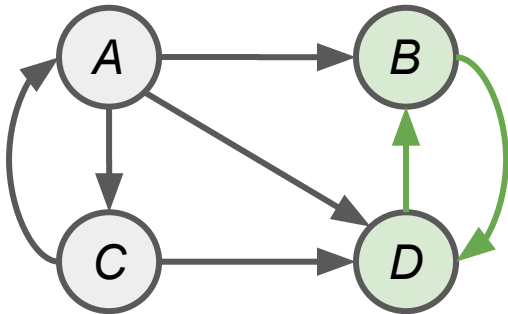
columns sum to 1 non-zero chance of going to any other node

Also known as being *stochastic*, *irreducible*, and *aperiodic*.

Goals:

No “dead-ends”

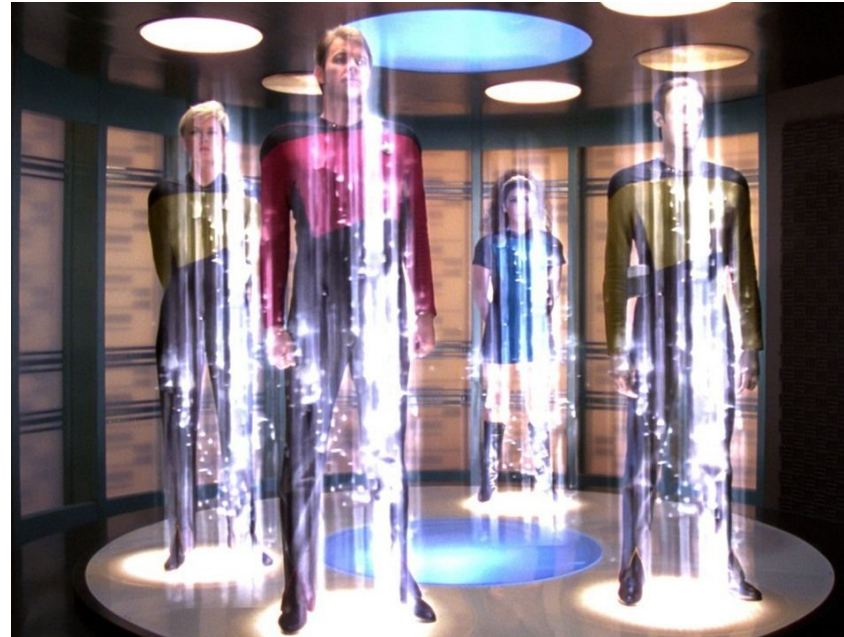
No “spider traps”



The “Google” PageRank Formulation

Add teleportation: At each step, two choices

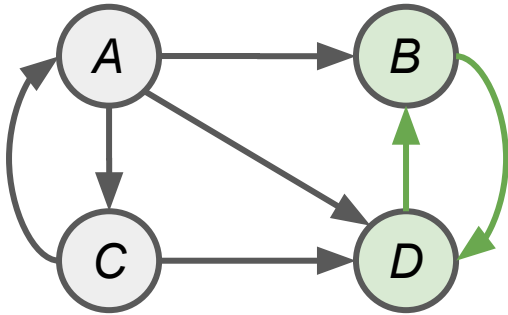
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)



Goals:

No “dead-ends”

No “spider traps”



The “Google” PageRank Formulation

Add teleportation: At each step, two choices

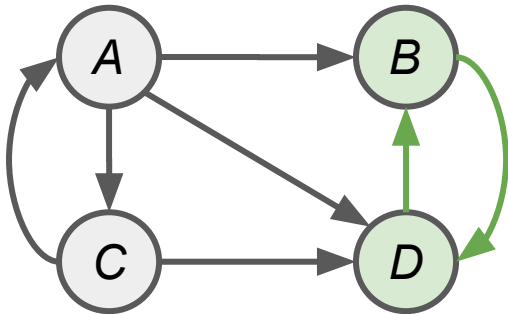
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

<i>to \ from</i>	A	B	C	D
A	0	0	1	0
B	$\frac{1}{3}$	0	0	1
C	$\frac{1}{3}$	0	0	0
D	$\frac{1}{3}$	1	0	0

Goals:

No “dead-ends”

No “spider traps”



The “Google” PageRank Formulation

Add teleportation: At each step, two choices

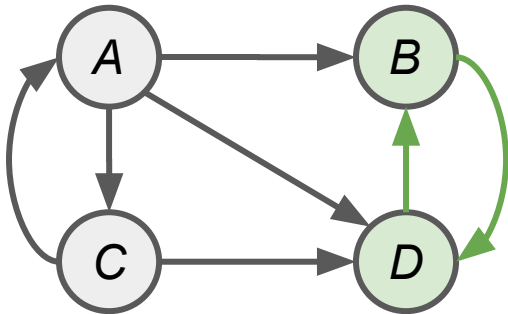
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

<i>to \ from</i>	A	B	C	D
A	0	$0 + .15 * \frac{1}{4}$	1	$0 + .15 * \frac{1}{4}$
B	$\frac{1}{3}$	$0 + .15 * \frac{1}{4}$	0	$.85 * 1 + .15 * \frac{1}{4}$
C	$\frac{1}{3}$	$0 + .15 * \frac{1}{4}$	0	$0 + .15 * \frac{1}{4}$
D	$\frac{1}{3}$	$.85 * 1 + .15 * \frac{1}{4}$	0	$0 + .15 * \frac{1}{4}$

Goals:

No “dead-ends”

No “spider traps”



The “Google” PageRank Formulation

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

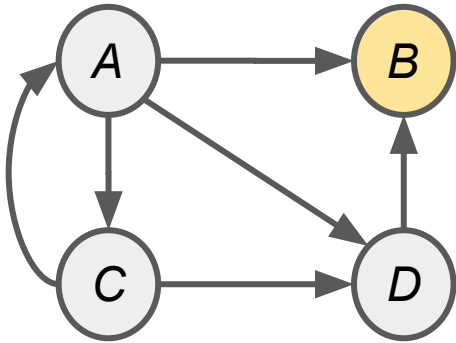
No “dead-ends”

No “spider traps”

The “Google” PageRank Formulation

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)



<i>to \ from</i>	A	B	C	D
A	0	0	1	0
B	$\frac{1}{3}$	0	0	1
C	$\frac{1}{3}$	0	0	0
D	$\frac{1}{3}$	0	0	0

Goals:

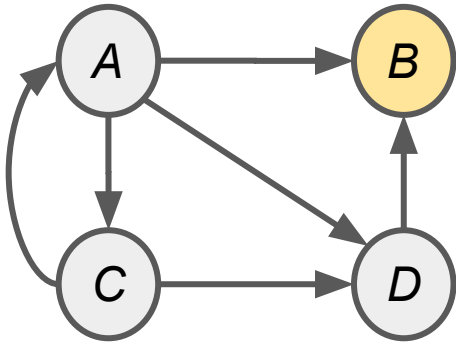
No “dead-ends”

No “spider traps”

The “Google” PageRank Formulation

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)



<i>to \ from</i>	A	B	C	D
A	0	$\frac{1}{4}$	1	0
B	$\frac{1}{3}$	$\frac{1}{4}$	0	1
C	$\frac{1}{3}$	$\frac{1}{4}$	0	0
D	$\frac{1}{3}$	$\frac{1}{4}$	0	0

Goals:

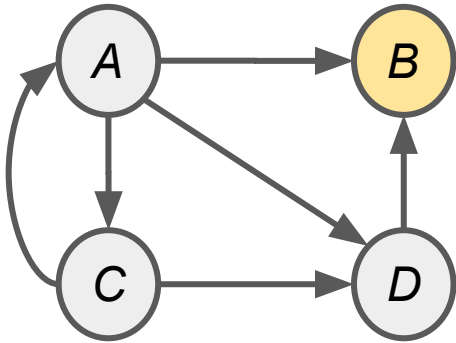
No “dead-ends”

No “spider traps”

The “Google” PageRank Formulation

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)



<i>to \ from</i>	A	B	C	D
A	0	$.85 \cdot \frac{1}{4} + .15 \cdot \frac{1}{4}$	1	0
B	$\frac{1}{3}$	$.85 \cdot \frac{1}{4} + .15 \cdot \frac{1}{4}$	0	1
C	$\frac{1}{3}$	$.85 \cdot \frac{1}{4} + .15 \cdot \frac{1}{4}$	0	0
D	$\frac{1}{3}$	$.85 \cdot \frac{1}{4} + .15 \cdot \frac{1}{4}$	0	0

Goals:

No “dead-ends”

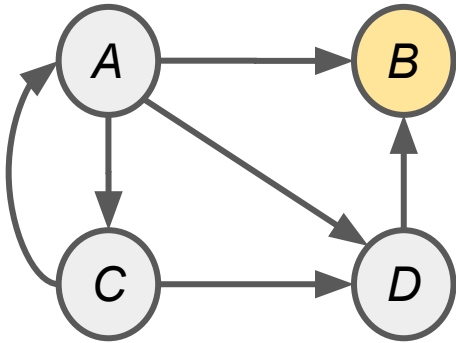
No “spider traps”

The “Google” PageRank Formulation

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

(Teleport from a dead-end has probability 1)

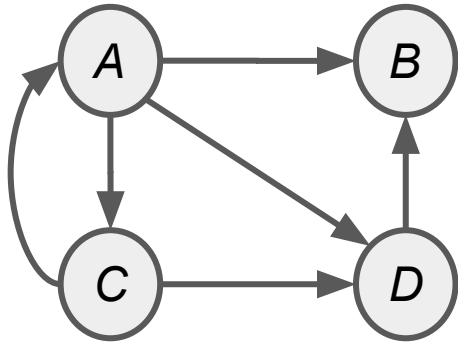


<i>to \ from</i>	A	B	C	D
A	$0 + .15 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{4}$	$.85 \cdot 1 + .15 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$
B	$.85 \cdot \frac{1}{3} + .15 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$	$.85 \cdot 1 + .15 \cdot \frac{1}{4}$
C	$.85 \cdot \frac{1}{3} + .15 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$
D	$.85 \cdot \frac{1}{3} + .15 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$

Goals:

No “dead-ends”

No “spider traps”



Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

No “dead-ends”

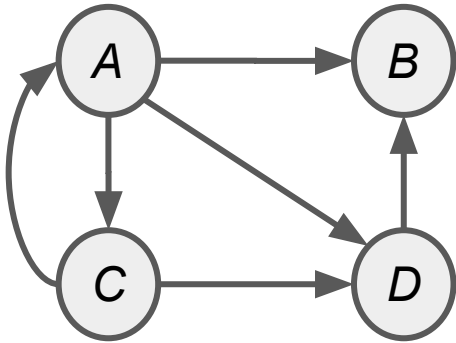
No “spider traps”

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation,
as Matrix Model: $M' = \beta M + (1 - \beta) \begin{bmatrix} 1 \\ \frac{1}{N} \end{bmatrix}_{N \times N}$



<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

No “dead-ends”

No “spider traps”

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation,
as Matrix Model: $M' = \beta M + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$

<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$.85 * \frac{1}{4} + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$.85 * \frac{1}{4} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$.85 * \frac{1}{4} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$.85 * \frac{1}{4} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

No “dead-ends”
No “spider traps”

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation,
as Matrix Model: $M' = \beta M + (1 - \beta) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{N \times N}$

To apply:
run power
iterations over M'
instead of M .

<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * 1 + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

No “dead-ends”
No “spider traps”

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation,
as Matrix Model: $M' = \beta M + (1 - \beta) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \frac{1}{N}$

Steps:

1. Compute M
2. Add $1/N$ to all dead-ends.
3. Convert M to M'
4. Run Power Iterations.

<i>to \ from</i>	A	B	C	D
A	$0 + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
B	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$
C	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

Goals:

- No “dead-ends”
- No “spider traps”

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation, as Matrix Model: $M' = \beta M + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$

Steps:

1. Compute M
2. Add $1/N$ to all dead-ends.
3. Convert M to M'
4. Run Power Iterations.

to \ from	A	B	C	D
A				$+ .15 * \frac{1}{4}$
B				$5 * 1 + .15 * \frac{1}{4}$
C				$+ .15 * \frac{1}{4}$
D	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$

In Practice, Just store βM as sparse matrix and distribute r according to above.

In other words, you only need to store M (as a sparse matrix) and r (as a vector), but never store M' . Use this function within the inner loop of power iterations to achieve the same result as if using M' .

Teleportation, as Flow Model:

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

(Brin and Page, 1998)

Teleportation, as Matrix Model: $M' = \beta M + (1 - \beta) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \frac{1}{N}$

Steps:

1. Compute M
2. Add $1/N$ to all dead-ends.
3. Convert M to M'
4. Run Power Iterations.

to \ from	A	B	C	D
A	$.15 \cdot \frac{1}{4}$			
B	$5 \cdot 1 + .15 \cdot \frac{1}{4}$			
C		$.15 \cdot \frac{1}{4}$		
D	$.85 \cdot \frac{1}{3} + .15 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$	$0 + .15 \cdot \frac{1}{4}$

In Practice, Just store βM as sparse matrix and distribute r according to above.

Summary

- Flow View: Link Voting
- Matrix View: Linear Algebra
 - Eigenvectors View
- Markov Process View
- How to remove:
 - Dead Ends
 - Spider Traps

In practice, sparse matrix, implement teleportation functionally rather than update M'