


# Link Analysis



**CSE545 - Spring 2020**  
Stony Brook University

H. Andrew Schwartz

# Big Data Analytics, The Class

**Goal: Generalizations**  
*A model or summarization of the data.*

*Data Frameworks*

Hadoop File System ✓  
Streaming ✓  
MapReduce ✓  
Spark ✓  
Tensorflow ✓

*Algorithms and Analyses*

Similarity Search ✓  
Link Analysis  
Hypothesis Testing  
Recommendation Systems  
Deep Learning

# The Web, circa 1998

ALTA VISTA  
Technology  
View Multimedia From Our Vantage Point

**AUTOPOSTAL** 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](#)

Make Me Laugh...  
 Create a Site...



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 ExciteSeeing Tour  
Excite on TV



YAHOO!  
Yahoo! Messenger  
Know when friends are online!  
Click to download Yahoo! Messenger

Yahoo! Mail  
free from anywhere

Search 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 Theater  
Connect Careers Chat Clubs Sex/Games Greetings Mail Members Messenger Mobile Personal People Search Photos  
Personal Adult Books Business Calendar My Yahoo! FaxDirect Fax Games Kids Movies Music Radio TV more...

**Yahoo! Auctions** Bid, buy, or sell anything!  
Categories: Antiques - Computers - Games - Date Forward/ - Casinos - Electronics - Golf Clubs - Grades Yesterday/ - Cars - Sports Cards - Mail - Longboards - Comic Books - Stamps - Pokémon - Ice.com - Baseball Cards - McGraw-Hill - Inter Bonds - Sosa Jeffrey Jr. - Items

**Items**  
Xena - Dale Earnhardt/ - Golf Clubs - Grades Yesterday/ - Mail - Longboards - Pokémon - Ice.com - Baseball Cards - McGraw-Hill - Inter Bonds - Sosa Jeffrey Jr. - Items

**In the News**  
U.S. rescues 15M spp. plans fish  
Senator: Calif. admits to sexual relationship with missing intern  
Attorney: Barry Levin found dead  
Dale Earnhardt Jr. wins Pocono 400  
Wimbledon - Tour de France more...

**Marketplace**  
new! eBay shops London  
Epicred - 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 - Scarface Movie 2, King of the Dragon, Cars and Dogs  
new! Play free Fantasy Baseball - midseason version  
Y! Photos - post your party pics

**Arts & Humanities**  
Literature, Photography

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

**Computers & Internet**  
Internet, WWW, Software, Games

**Education**  
College and University, K-12

**Entertainment**  
Cartoons, Comics, Horror, Music

**Government**  
Elections, Military, Law, Taxes

**Health**  
Medicine, Diseases, Drugs, Fitness

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

**Recreation & Sports**  
Sports, Travel, Action, Outdoors

**Reference**  
Libraries, Dictionaries, Quotations

**Regional**  
Countries, Regions, US States

**Science**  
Animals, Astronomy, Engineering

**Social Science**  
Archaeology, Economics, Languages

**Society & Culture**  
People, Environment, Religion

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  
US Cities - Atlanta - Boston - Chicago - Dallas TX - LA - NYC - SE Bay - Wash DC - more...

More Yahoo!'s  
Outdoor - Autos - Book Index - Careers - Health - Living - Outdoors - Pets - Real Estate - Telecommunications  
Entertainment - Astronomy - Baseball - 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, PDA's, 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

ALTA VISTA  
Technology  
View Multimedia From Our Vantage Point

**AUTOPOSTAL**  
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](#)

Make Me Laugh... 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.

YAHOO!

Yahoo! Messenger Know when friends are online! Click to download Yahoo! Messenger

Yahoo! Mail free from anywhere

Search 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 Theater 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	Items
Antiques	Books
Cameras	Computers
Cars	Electronics
Cass.Books	Games
CDs	Home
Computers	Internet
Electronics	Jobs
Games	Legal
Home	Medical
Internet	Real Estate
Jobs	Software
Legal	Travel
Medical	Video
Real Estate	Web Services
Software	World
Travel	Y! Store
Video	Y! Web
Web Services	Y! Web Services
World	Y! Web Services

Baseball Cards - McGraw, A.Rod Inter Bonds, Soda Goffey, Jr., Lites

**In the News**

- U.S. rescues 15M spp. plans feb
- Source: ConAgra admits to sexual relationship with missing sister
- Attorney Rary: Levin found dead
- Date Embassy in vnm Pcpri 4/9
- Wimbledon - Tour de France more...

**Marketplace**

- new! eBay: shops London
- Epinec: 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 - midseason 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 TX - LA - NYC - SE Bay - Wash DC - more...

More Yahoo!'s  
Outdoor - Autos - Bazaars - Careers - Health - Living - Outdoors - Pets - Real Estate - Telephones  
Entertainment - Astrology - Breakfast - Events - Games - Movies - Music - Radio - Tickets - TV - more  
Finance - Banking - Bill Pay - Insurance - Loans - Taxes - Finance/Invest - 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"

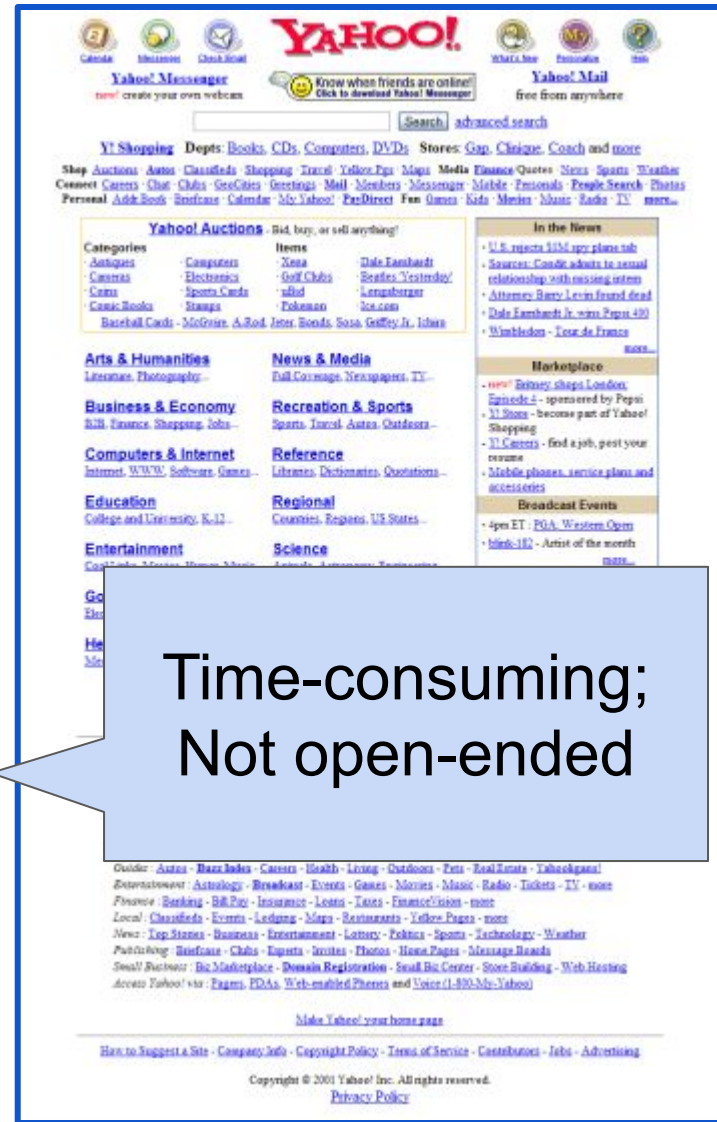
Match keywords, language (*information retrieval*)



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 present in hypertext. Google is designed to crawl and index the Web efficiently and produce much text and hyperlink

...

## The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

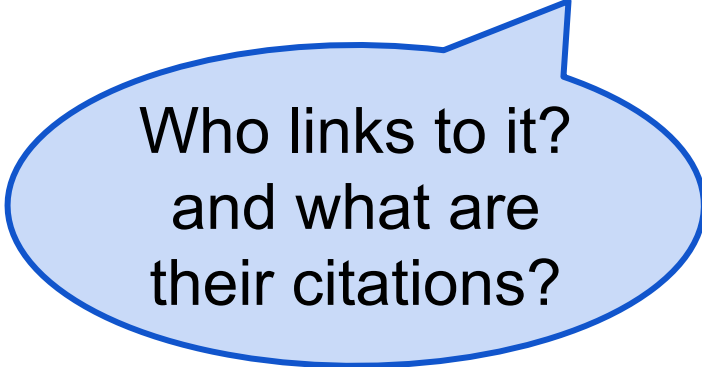
Abstract

# PageRank

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

# PageRank

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

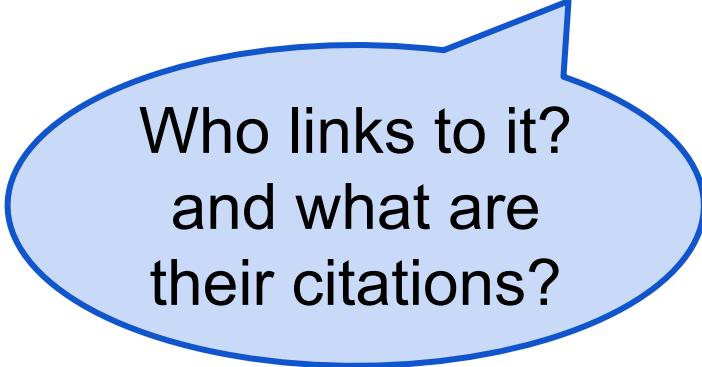


Who links to it?  
and what are  
their citations?



# PageRank

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



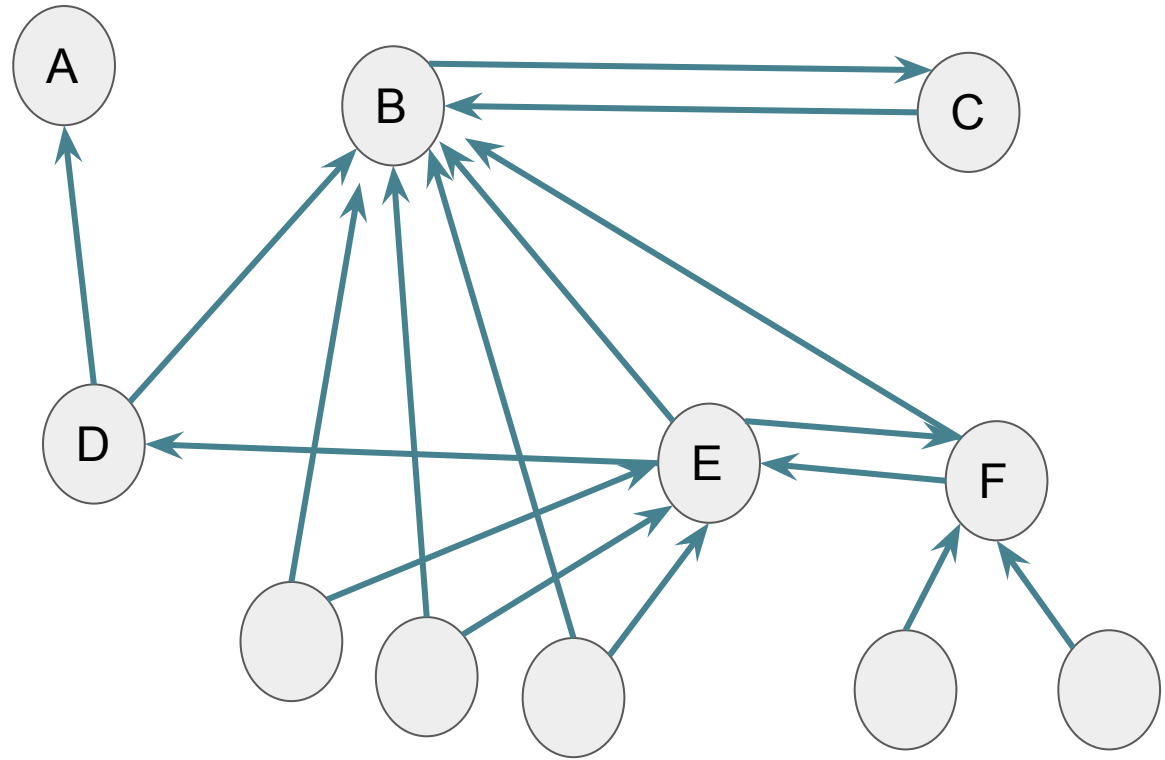
Who links to it?  
and what are  
their citations?

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

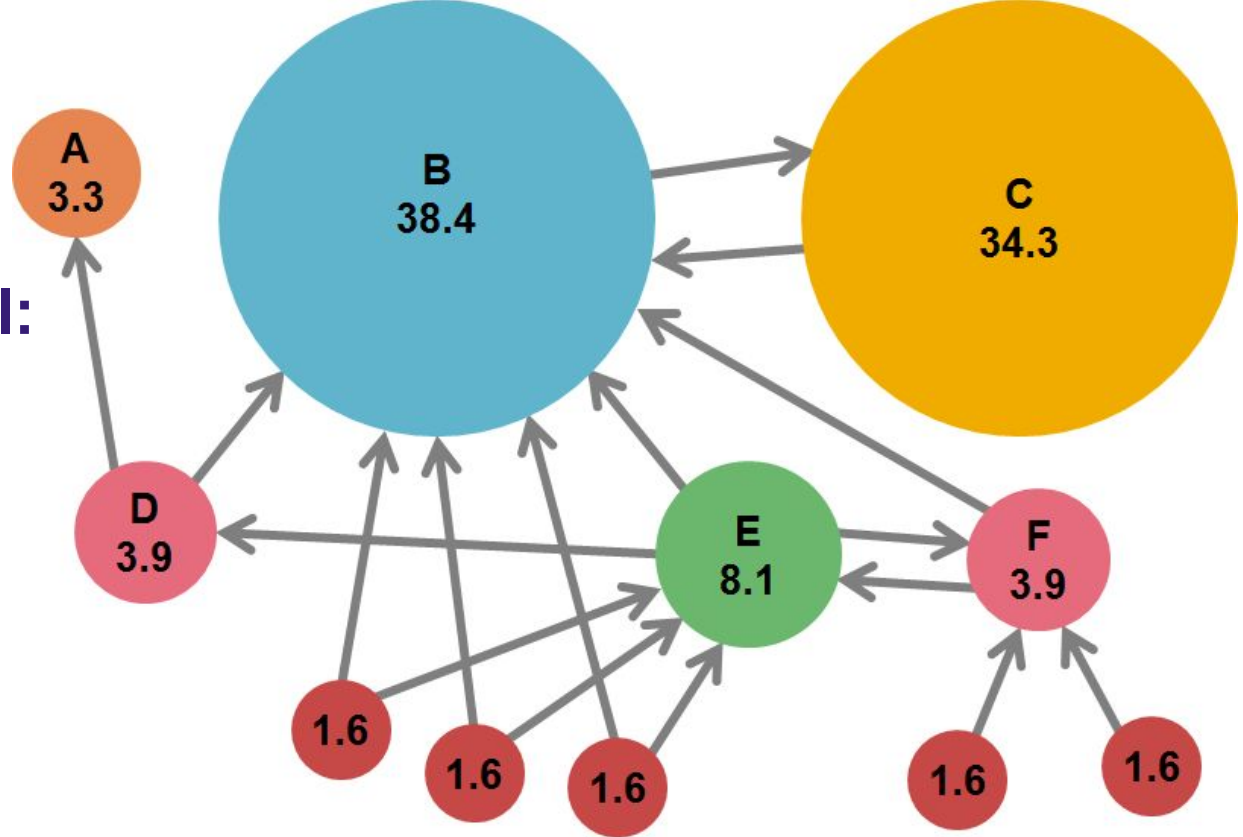


**Innovation 1: What pages would a “random Web surfer” end up at?**

**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

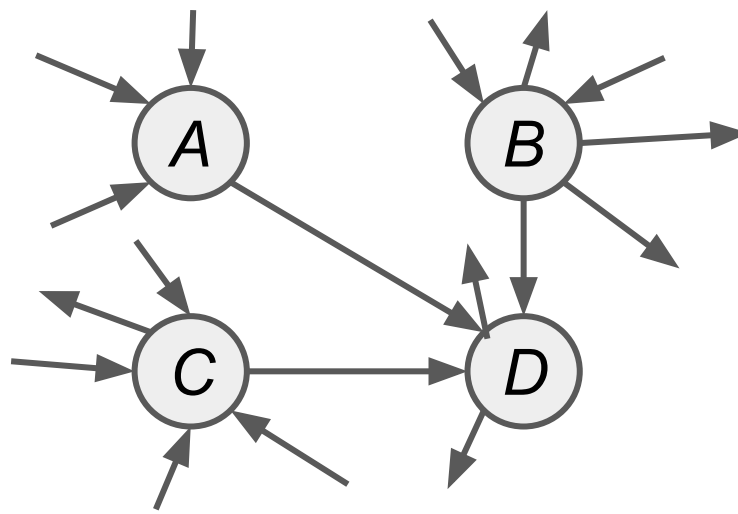
but, citations from important pages should count more.

=> 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?

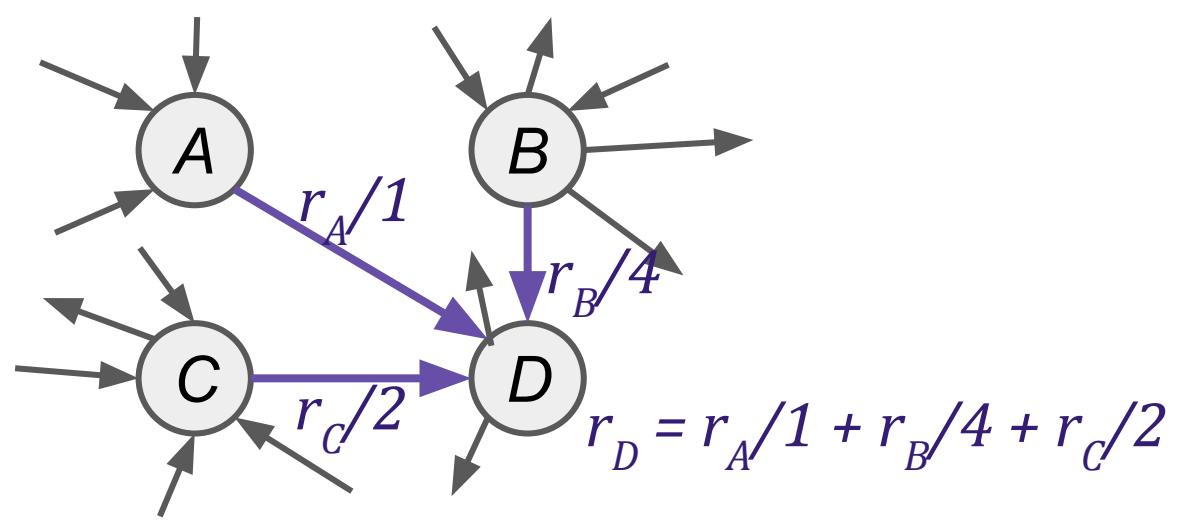
Each page ( $j$ ) has an importance (i.e. rank,  $r_j$ )

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

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

# PageRank

## View 1: Flow Model:



## How to compute?

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

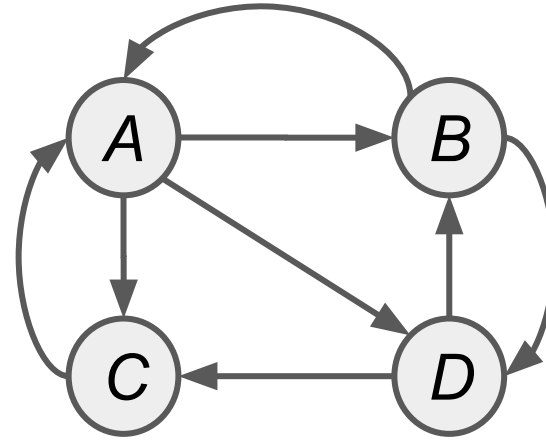
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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



# PageRank

## View 1: Flow Model:



## How to compute?

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

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

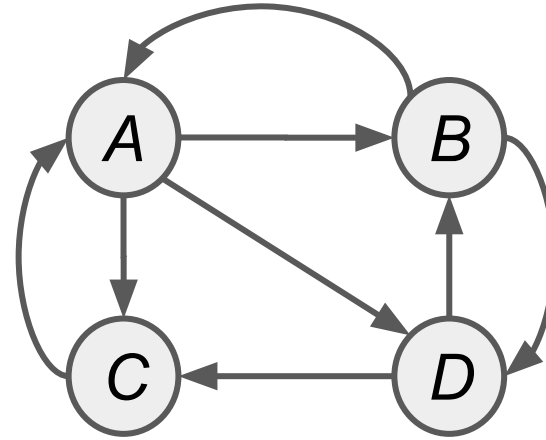
$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

# 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,  $r_j$ )

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

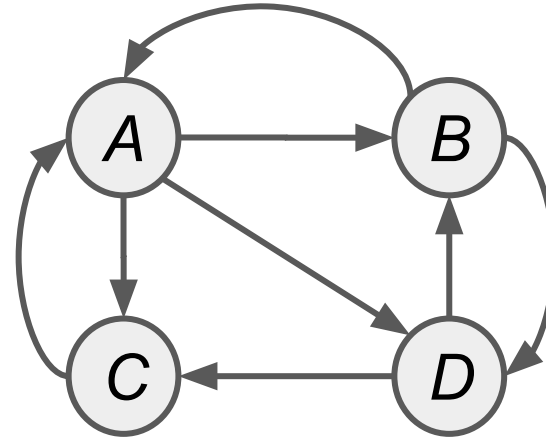
$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

# PageRank

## View 1: Flow Model:

A System of Equations:

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

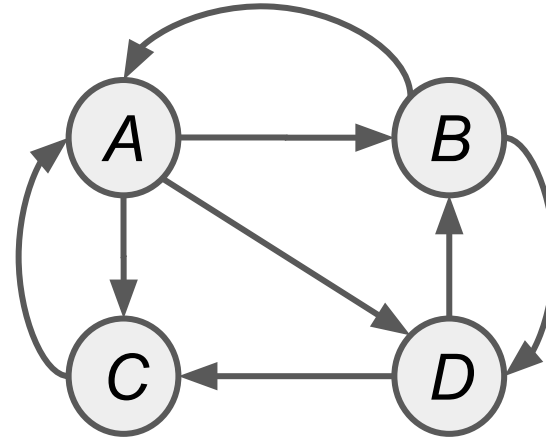


## How to compute?

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

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

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

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

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

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

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

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

How to compute?

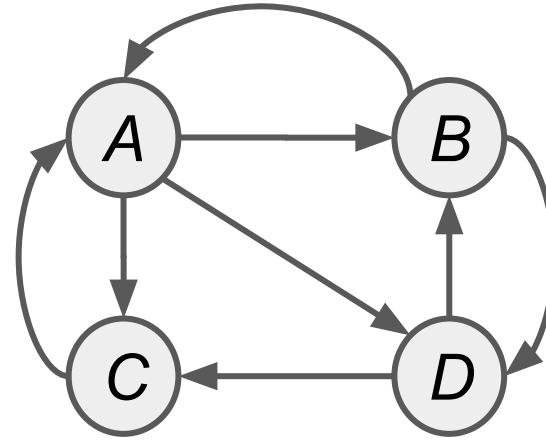
Each page ( $j$ ) has an importance (i.e. rank,  $r_j$ )

$$vote_j = \frac{r_j}{n_j}$$

( $n_j$  is |out-links|)

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

# 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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

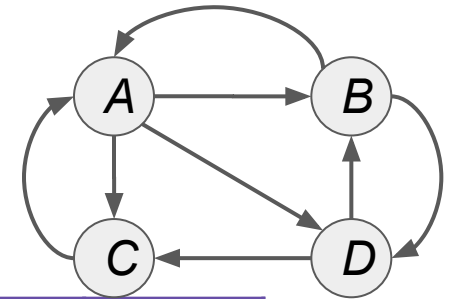
Transition Matrix, M

# PageRank

## 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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M



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

## View 2: Matrix Formulation

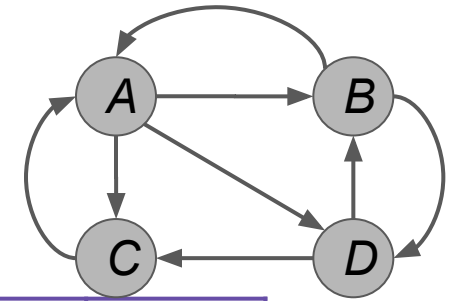
$$1 = r_A + r_B + r_C + r_D$$

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

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

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

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



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

Transition Matrix, M

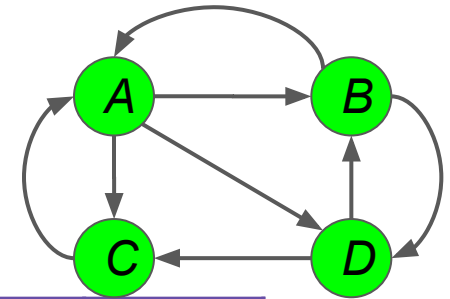
**Innovation: What pages would a “random Web surfer” end up at?**

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

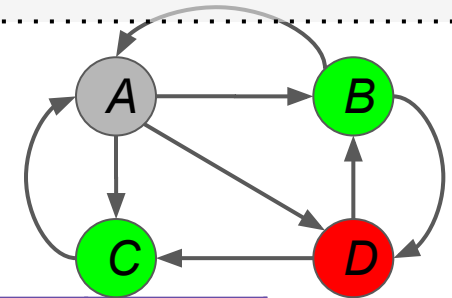
**Innovation: What pages would a “random Web surfer” end up at?**

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

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

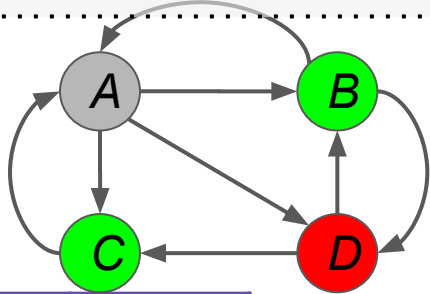
To Start, all are equally likely at  $\frac{1}{4}$ : ends up at D

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	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

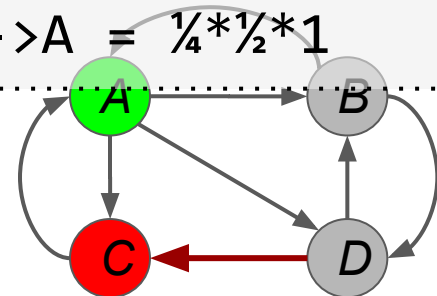
C and B are then equally likely:  $\rightarrow D \rightarrow B = \frac{1}{4} * \frac{1}{2}$ ;  $\rightarrow D \rightarrow C = \frac{1}{4} * \frac{1}{2}$

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

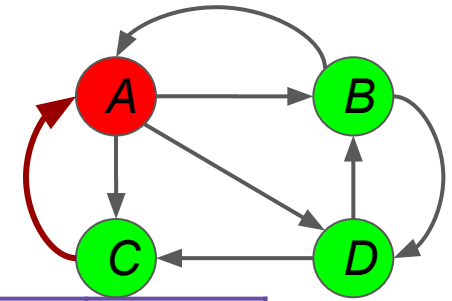
Innovation: What pages would a “random Web surfer” end up at?

...

## 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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M



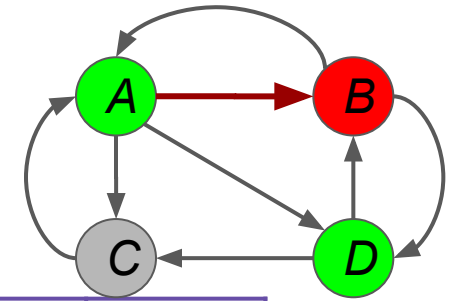
Innovation: What pages would a “random Web surfer” end up at?

...

## 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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

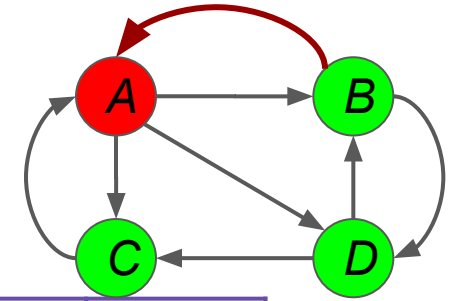
Innovation: What pages would a “random Web surfer” end up at?

...

## 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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

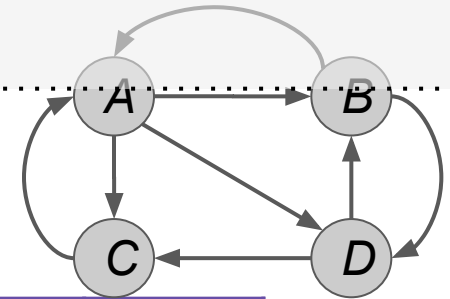
**Innovation: What pages would a “random Web surfer” end up at?**

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	1/3	1/2	0	0

Transition Matrix, M

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

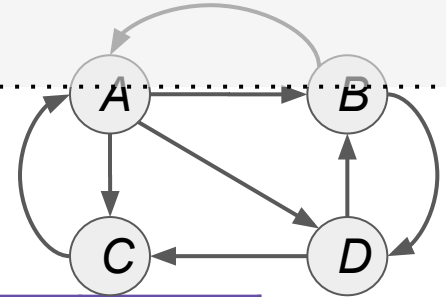
To start:  $N=4$  nodes, so  $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	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  $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4},]$

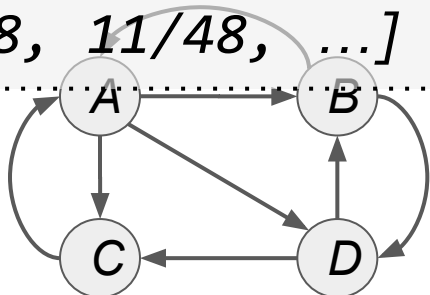
after 1st iteration:  $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	1/2	1	0
<b>B</b>	1/3	0	0	1/2
<b>C</b>	1/3	0	0	1/2
<b>D</b>	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:  $M \cdot r = [3/8, 5/24, 5/24, 5/24]$

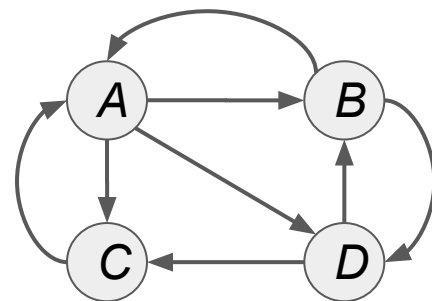
after 2nd iteration:  $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

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

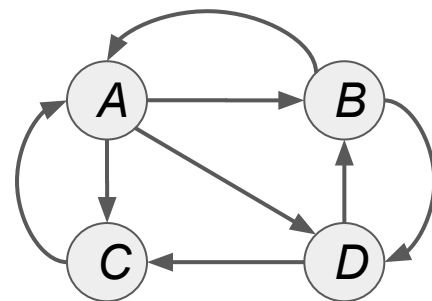
```
while (err_norm( $r[t], r[t-1]$ ) > min_err):
```

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

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

```
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  $\text{err\_norm}$  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:  $r[0] = [1/N, \dots, 1/N]$   
 $r[-1] = [0, \dots, 0]$   
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

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**.

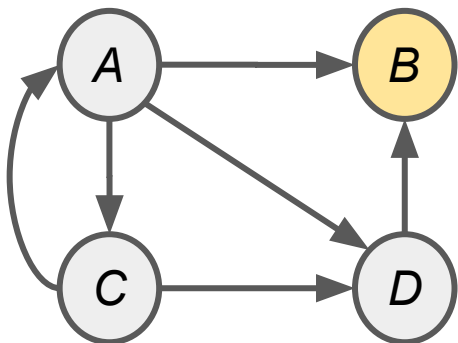
Thus,  $r$  is a stationary distribution. Probability of being at given node.

### 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

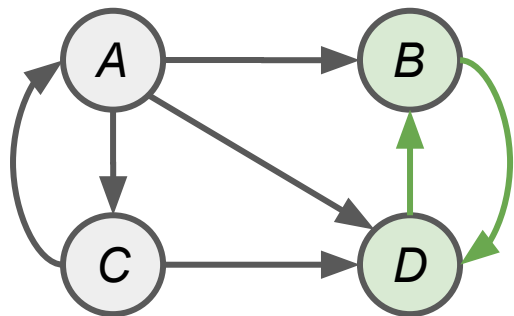
What would  $r$  converge to?

### 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

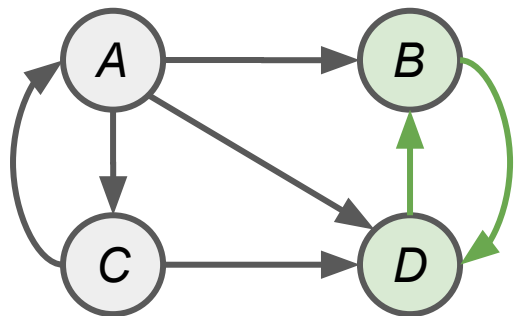
What would  $r$  converge to?

### 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

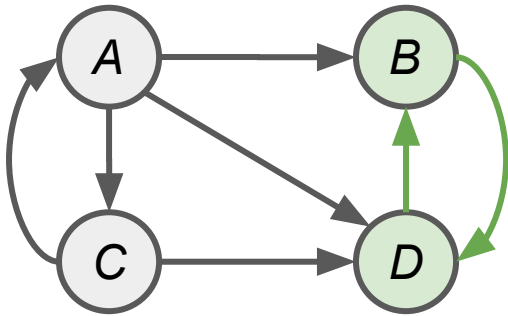
columns sum to 1    non-zero chance of going to any other node

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

## Goals:

No “dead-ends”

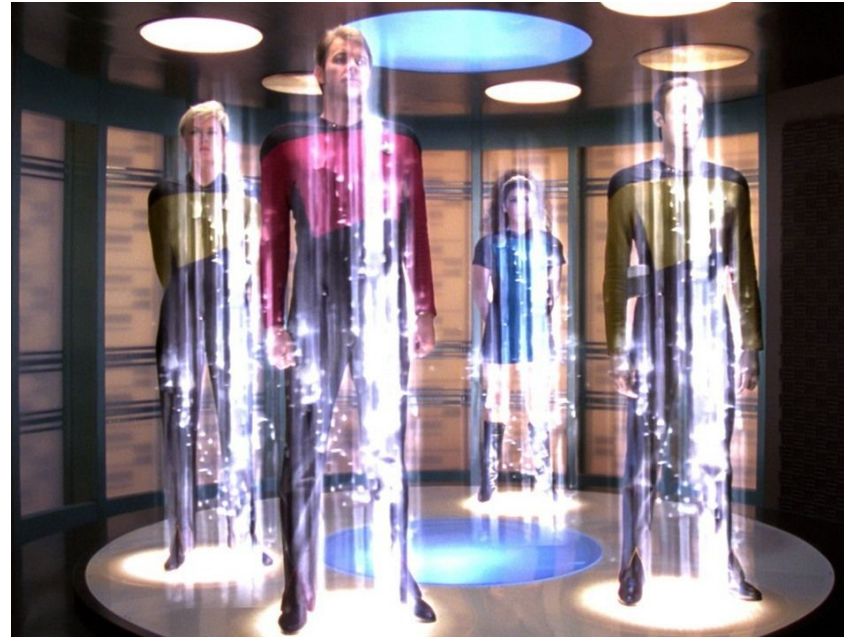
No “spider traps”



## 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,  $1-\beta$ )

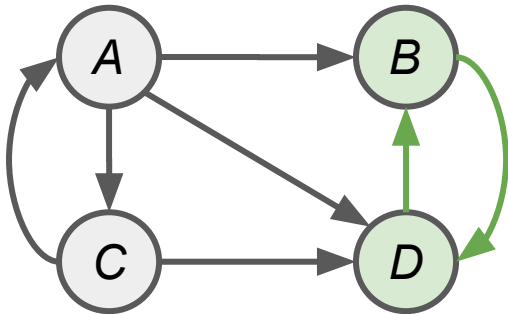




## Goals:

No “dead-ends”

No “spider traps”



## The “Google” PageRank Formulation

Add teleportation: At each step, two choices

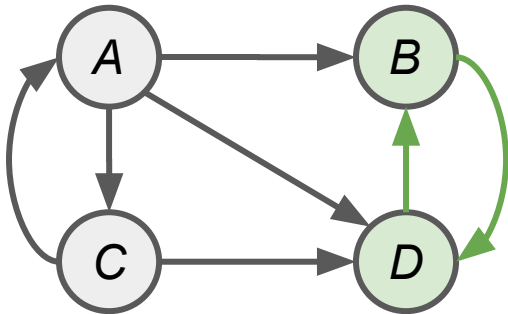
1. Follow a random link (probability,  $\beta = \sim .85$ )
2. Teleport to a random node (probability,  $1-\beta$ )

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

## Goals:

No “dead-ends”

No “spider traps”



## The “Google” PageRank Formulation

Add teleportation: At each step, two choices

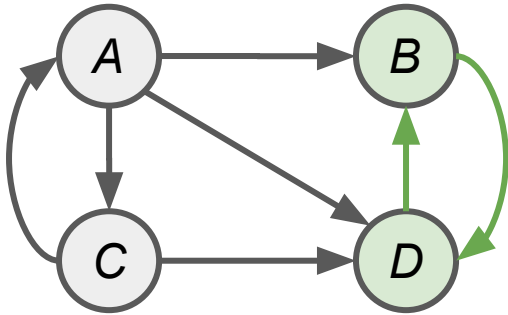
1. Follow a random link (probability,  $\beta = \sim .85$ )
2. Teleport to a random node (probability,  $1-\beta$ )

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

## Goals:

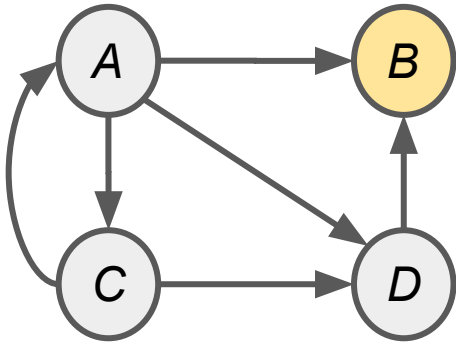
No “dead-ends”

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>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	0	0	1	0
<b>B</b>	$\frac{1}{3}$	0	0	1
<b>C</b>	$\frac{1}{3}$	0	0	0
<b>D</b>	$\frac{1}{3}$	0	0	0

## Goals:

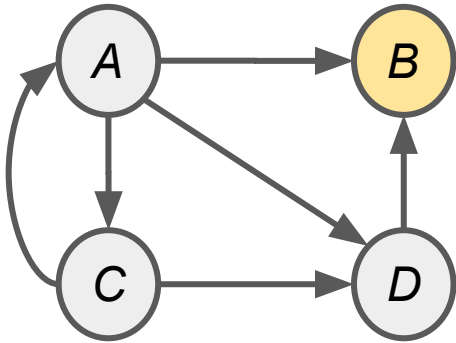
No “dead-ends”

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

## Goals:

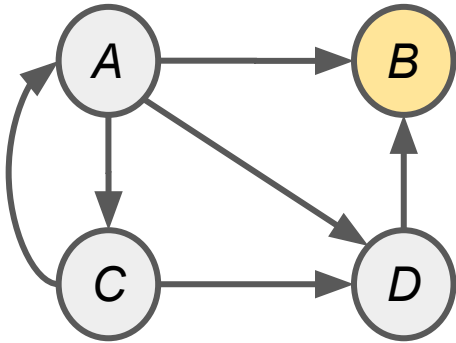
No “dead-ends”

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

## Goals:

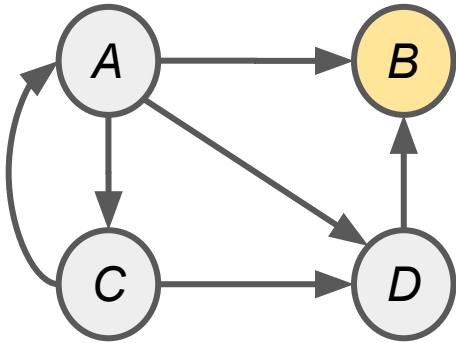
No “dead-ends”

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)

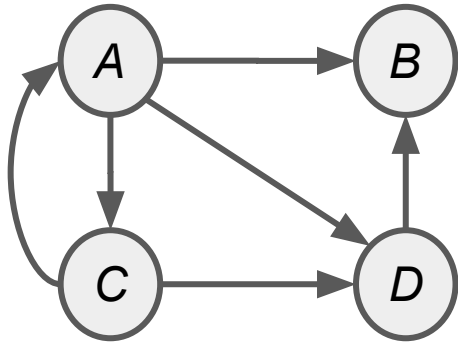


<i>to \ from</i>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	$0 + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$
<b>B</b>	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$	$1 * \frac{1}{4}$	$0 + .15 * \frac{1}{4}$	$.85 * \frac{1}{3} + .15 * \frac{1}{4}$
<b>C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

## 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)

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



## Goals:

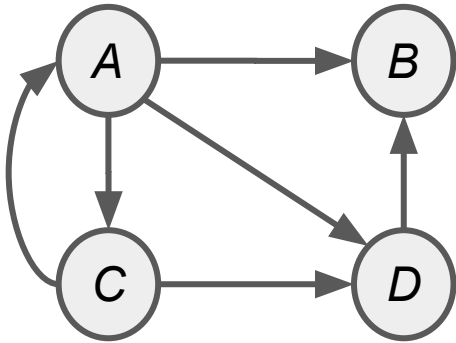
No “dead-ends”  
No “spider traps”

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



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

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

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

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

## 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.

*But, M' is now a dense matrix!*

*E.g. 1.7B webpages as nodes.*

*1.7B x 1.7B = 2.9 x 10<sup>18</sup>!*

to				<b>D</b>
				0+.15*1/4
				.85*1+.15*1/4
<b>C</b>	.85*1/3+.15*1/4	1*1/4	0+.15*1/4	0+.15*1/4
<b>D</b>	.85*1/3+.15*1/4	1*1/4	0+.15*1/4	0+.15*1/4

# PageRank, in Practice

## Steps:

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

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

*But,  $M'$  is now a dense matrix!*

*E.g. 1.7B webpages as nodes.*

$$1.7B \times 1.7B = 2.9 \times 10^{18}!$$

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

# PageRank, in Practice

...  $M$  is sparse...

## Steps:

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

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

But,  $M'$  is now a dense matrix!

E.g.  $1.7B$  webpages as nodes.

$$1.7B \times 1.7B = 2.9 \times 10^{18}!$$

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

# PageRank, in Practice

...  $M$  is sparse... Can we just work with  $M$ ?

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

## Steps:

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

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

But,  $M'$  is now a dense matrix!

E.g.  $1.7B$  webpages as nodes.  
 $1.7B \times 1.7B = 2.9 \times 10^{18}$ !



# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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.

```
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]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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.

```
initialize: r[0] = [1/N, ..., 1/N],  
           r[-1] = [0, ..., 0]
```

```
M = addToDeadEnds(1/N, M)
```

```
while (err_norm(r[t], r[t-1]) > min_err):  
    r[t+1] = M * r[t]  
    t += 1  
solution = r[t]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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.

```
initialize: r[0] = [1/N, ..., 1/N],  
           r[-1] = [0, ..., 0]  
M = addToDeadEnds(1/N, M)  
M' = beta*M + (1-beta)*[1/N]NxN  
while (err_norm(r[t], r[t-1]) > min_err):  
    r[t+1] = M' * r[t]  
    t += 1  
solution = r[t]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

```
initialize:  r[0] = [1/N, ..., 1/N],  
             r[-1]=[0,...,0]  
M = addToDeadEnds(1/N, M)  
M' = beta*M + (1-beta)*[1/N]NxN  
while (err_norm(r[t],r[t-1])>min_err):  
    r[t+1] = M'·r[t]  
    t+=1  
solution = r[t]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

*Yes! Work with the  
calculation of M'  
instead of simply M.*

```
initialize:  r[0] = [1/N, ..., 1/N],  
            r[-1]=[0,...,0]  
M = addToDeadEnds(1/N, M)  
M' = beta*M + (1-beta)*[1/N]NxN  
while (err_norm(r[t],r[t-1])>min_err):  
    r[t+1] = M'·r[t]  
    t+=1  
solution = r[t]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

*Yes! Work with the  
calculation of M'  
instead of simply M.*

```
initialize: r[0] = [1/N, ..., 1/N],  
           r[-1]=[0,...,0]  
M = addToDeadEnds(1/N, M)  
M' = beta*M + (1-beta)*[1/N]NxN  
while (err_norm(r[t],r[t-1])>min_err):  
    r[t+1] = (beta*M + (1-beta)*[1/N]NxN)·r[t]  
    t+=1  
solution = r[t]
```

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

```
initialize: r[0] = [1/N, ..., 1/N],  
            r[-1] = [0, ..., 0]
```

```
M = addToDeadEnds(1/N, M)
```

```
while (err_norm(r[t], r[t-1]) > min_err):  
    r[t+1] = (beta*M + (1-beta)*[1/N]_{N x N}) * r[t]  
    t += 1
```

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

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

*The second half of  
the M' equation is  
just a constant*

```
initialize: r[0] = [1/N, ..., 1/N],  
           r[-1] = [0, ..., 0]  
M = addToDeadEnds(1/N, M)  
tele = (1-beta) * (1/N)  
While (err_norm(r[t], r[t-1]) > min_err):  
    r[t+1] = (beta * M + (1-beta) * [1/N]_{N x N}) * r[t]  
    t += 1  
solution = r[t]
```



# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

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

# PageRank, in Practice

*... M is sparse... Can we just work with M?*

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

*If M larger than it  
needs to be because  
of the dead-ends?*

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

# PageRank, in Practice

...  $M$  is sparse... Can we just work with  $M$ ?

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

Exercise:

Get rid of this step. How to adjust algorithm?

Hint: at least 2 options:

1. Track dead ends
2. Consider  $r$  should sum to 1.

```
initialize:  r[0] = [1/N, ..., 1/N],  
             r[-1]=[0,...,0]
```

```
→ M = addToDeadEnds(1/N, M)
```

```
tele = (1-beta)* (1/N)
```

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

```
    r[t+1] = (beta*M .+ tele).r[t]
```

```
    t+=1
```

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

# PageRank: 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  $M'$

# 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 present in hypertext. Google is designed to crawl and index the Web efficiently and produce much text and hyperlink

...

## The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

Abstract

# Search, 20+ years later

## 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

Many innovations since  
examples:

### Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and present the information in the most relevant form. We describe the system architecture, the algorithms for crawling, indexing and searching that underlie the current system, and the plans for future directions. We describe the main design principles that we used in creating the system, which are: fast search, quality of search results, and personalization.

- Content Specific, “Personalized PageRank”
- Search Engine Optimization (SEO) countermeasures
- Location/user-specific Search

...

January 29, 1998

Abstract

# Search, 20+ years later

## 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

Many innovations since  
examples:

### Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and present the results in a form that is useful to users. It has a number of novel features that are based on the analysis of the structure of hypertext and the analysis of usage statistics.

- Content Specific, “Personalized PageRank”
- Search Engine Optimization (SEO) countermeasures
- Location/user-specific Search

but still core of approach: PageRank