

Link Analysis

Stony Brook University
CSE545, Fall 2016

The Web , circa 1998

The screenshot shows the AltaVista homepage with the following elements:

- Logo:** ALTA VISTA Technology
- Slogan:** View Multimedia From Our Vantage Point
- Advertisement:** A red banner for "AUTOPOSTAL" (USA CANADA) offering "Car Buying & Car Insurance Pain Relief" with a "LOW COST!" callout.
- Search:** A search bar with the text "Search the Web and Display the Results in Standard Form" and a "Submit" button.
- Links:** "Click here for advertising information - reach millions every month!", "Search with Digital's Alta Vista [Advanced Search] [Add URL]", "Contests" (Make Me Laugh...), and "Creative Web" (Create a Site...).

The screenshot shows the Yahoo! homepage with the following elements:

- Logos:** Yahoo! logo and various service icons like Calendar, Messenger, Chat, etc.
- Search:** A search bar with "Search" and "advanced search" buttons.
- Navigation:** Links for "Y! Shopping", "Depts: Books, CDs, Computers, DVDs", "Stores: Gap, Clinique, Coach and more", "Shop Auctions", "Aster", "Classifieds", "Shopping", "Travel", "Yellow Pages", "Maps", "Media", "Finance", "Quotes", "News", "Sports", "Weather", "Connect", "Careers", "Chat", "Clubs", "SiteClubs", "Greetings", "Mail", "Members", "Messenger", "Mobile", "Personal", "People Search", "Photos", "Personal", "Add-Books", "Business", "Calendar", "My Yahoo!", "FastDirect", "Fun", "Games", "Kids", "Movies", "Music", "Radio", "TV", "more..."
- Yahoo! Auctions:** A section for bidding, buying, or selling anything, with categories like Antiques, Cameras, Cars, Comic Books, etc.
- Arts & Humanities:** Links to literature, photography, business, and economy.
- Business & Economy:** Links to B2B, finance, shopping, and jobs.
- Computers & Internet:** Links to internet, WWW, software, and games.
- Education:** Links to college and university resources.
- Entertainment:** Links to coalitions, movies, music, and more.
- Government:** Links to elections, history, law, and taxes.
- Health:** Links to medicine, diseases, drugs, fitness, and more.
- News & Media:** Links to full coverage, newspapers, and TV.
- Recreation & Sports:** Links to sports, travel, and outdoor activities.
- Reference:** Links to libraries, dictionaries, and quotations.
- Regional:** Links to countries, regions, and US states.
- Science:** Links to animals, astronomy, engineering, and more.
- Social Science:** Links to archaeology, economics, and languages.
- Society & Culture:** Links to people, environment, and religion.
- In the News:** A list of recent news items.
- Marketplace:** Links to eBay, shops, and London.
- Broadcast Events:** Links to Open ET, PDA, Western Open, and more.
- Inside Yahoo!:** Links to Y! Games, backgammon, checkers, hearts, chess, pool, etc.

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, FL, LA, NYC, SE Bay, Wash. DC, more...

More Yahoo!'s: Austria, Brazil, India, Canada, Health, Living, Outdoors, Pets, Real Estate, Yahoo! Japan!, 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, Lottery, Politics, Sports, Technology, Weather, Publishing: Business, Clubs, Experts, Quotes, Photos, Home Pages, Message Boards, Small Business: Biz Marketplace, Business 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

ALTA VISTA
Technology
View Multimedia From Our Vantage Point

AUTOPOSTER
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

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

Contests **Creative Web**
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

Excite Direct
Turbo Search! Download Excite Direct

Take an Excite Seeing Tour

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

INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION. **NETSCAPE NOW!**



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

Yahoo! Messenger
new! create your own webcams

Yahoo! Mail
free from anywhere

[advanced search](#)

Y! Shopping Depts: [Books](#) [CDs](#) [Computers](#) [DVDs](#) [Stores](#) [Gap](#) [Clingie](#) [Coach](#) and more

Shop Auctions: [Aston](#) [Classifieds](#) [Shopping](#) [Travel](#) [Yellow Pages](#) [Maps](#) [Media](#) [Finance](#) [Quotes](#) [News](#) [Sports](#) [Weather](#)
Connect: [Careers](#) [Chat](#) [Clubs](#) [GeoCities](#) [Greetings](#) [Mail](#) [Members](#) [Messenger](#) [Mobile](#) [Personal](#) [People Search](#) [Photos](#)
Personal: [Add Book](#) [Business](#) [Calendar](#) [My Yahoo!](#) [FastDirect](#) [Fun Games](#) [Kids](#) [Movies](#) [Music](#) [Radio](#) [TV](#) [more...](#)

Yahoo! Auctions Bid, buy, or sell anything!
Categories: [Antiques](#) [Computers](#) [Electronics](#) [Golf Clubs](#) [Hobbies](#) [Jewelry](#) [Musical](#) [Novelty](#) [Real Estate](#) [Stamps](#) [Toys](#) [Video](#) [VHS](#) [Watches](#) [Wedding](#) [Wholesale](#) [Yard Sale](#) [Zoo](#) [more...](#)

In the News
[U.S. rescues 15M spp. plane fish](#)
[Suzanne Casati admits to sexual relationship with missing sister](#)
[Attorney Barry Levin found dead](#)
[Date Eamonn is now Pagan 490](#)
[Wimbledon - Tour de France](#) [more...](#)

Arts & Humanities **News & Media**
[Literature](#) [Photography](#) [Full Coverage](#) [Newspapers](#) [TV](#)

Business & Economy **Recreation & Sports**
[B2B](#) [Finance](#) [Shopping](#) [Jobs](#) [Sports](#) [Travel](#) [Automotive](#) [Outdoors](#)

Computers & Internet **Reference**
[Internet](#) [WWW](#) [Software](#) [Games](#) [Libraries](#) [Dictionaries](#) [Quotations](#)

Education **Regional**
[College and University](#) [K-12](#) [Countries](#) [Regions](#) [US States](#)

Entertainment **Science**
[Comics](#) [Movies](#) [Music](#) [Television](#) [Animals](#) [Astronomy](#) [Engineering](#)

Government **Social Science**
[Elections](#) [Military](#) [Law](#) [Taxes](#) [Archaeology](#) [Economics](#) [Languages](#)

Health **Society & Culture**
[Medicine](#) [Diseases](#) [Drugs](#) [Fitness](#) [People](#) [Environment](#) [Religion](#)

powered by **COMPAQ**

Local Yahoo!
[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](#) [Chile](#) [Mexico](#) [Spanish](#)
[U.S. Cities](#) [Atlanta](#) [Boston](#) [Chicago](#) [Dallas TX](#) [LA](#) [NYC](#) [SE Bay](#) [Wash DC](#) [more...](#)

More Yahoo!
[Outdoor](#) [Aston](#) [Bazz Index](#) [Careers](#) [Health](#) [Living](#) [Outdoors](#) [Pets](#) [Real Estate](#) [Yahoo! Japan!](#)
[Entertainment](#) [Astrology](#) [Broadcast](#) [Events](#) [Games](#) [Movies](#) [Music](#) [Radio](#) [Tickets](#) [TV](#) [more](#)
[Finance](#) [Banking](#) [Bill Pay](#) [Insurance](#) [Leads](#) [Loans](#) [Finance/Investment](#) [more](#)
[Local](#) [Classifieds](#) [Events](#) [Locations](#) [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!](#) [vtr](#) [Pages](#) [PDA's](#) [Web-enabled Phones](#) [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

AUTOSYTEL
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

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

[\[Help\]](#) [\[Advanced Search\]](#)

Excite Direct [Download](#) [Excite Direct](#)

Take an [ExciteSeeing Tour](#)

Excite on TV

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

INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION. [NETSCAPE](#) [Now!](#)



YAHOO! [What's Hot](#) [MSN](#) [Excite](#) [Hot](#)

Yahoo! Messenger [new! create your own webcams](#) [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](#) [Aster](#) [Classifieds](#) [Shopping](#) [Travel](#) [Yahoo! Pgs](#) [Maps](#) [Media](#) [Finance](#) [Quotes](#) [News](#) [Sports](#) [Weather](#) [Connect](#) [Careers](#) [Chat](#) [Chats](#) [SiteCritic](#) [Greetings](#) [Mail](#) [Members](#) [Messenger](#) [Mobile](#) [Personal](#) [People Search](#) [Photos](#) [Personal](#) [Add-Books](#) [Browsers](#) [Calendar](#) [My Yahoo!](#) [FaxDirect](#) [Fun](#) [Games](#) [Kids](#) [Movies](#) [Music](#) [Radio](#) [TV](#) [more...](#)

Yahoo! Auctions Bid, buy, or sell anything!
Categories: [Antiques](#) [Cameras](#) [Cars](#) [Class Books](#) [Computers](#) [Electronics](#) [Sports Cards](#) [Stamps](#) [Baseball Cards](#) [McGraw-Hill](#) [Audi](#) [Inter. Bonds](#) [Sosa](#) [Golfey Jr.](#) [Lithes](#) [Xmas](#) [Date Exchange!](#) [Golf Clubs](#) [Books](#) [Yesterday!](#) [Wed](#) [Longboards](#) [Pokemon](#) [Ice.com](#)

In the News
[U.S. rescues 15M spp. plans fail](#)
[Suzanne Casati admits to sexual relationship with missing sister](#)
[Attorney Barry Levin found dead](#)
[Date Exchange! is now Pagan 499](#)
[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
[4pm ET, 3pm Western](#) [Open](#)
[blink-182](#) - Artist of the month [more...](#)

Inside Yahoo!
[Y! Games](#) - background, e-books, hearts, chess, pinacle
[Y! Movies](#) - [Scary Movie 2](#), [King of the Dragon Kings](#) and [Dogs](#)
[new!](#) Play five [Fantasy Baseball](#) - midseason version
[Y! Photos](#) - post your party pics

Arts & Humanities [Literature](#) [Photography](#)
News & Media [Full Coverage](#) [Newspapers](#) [TV](#)
Business & Economy [B2B](#) [Finance](#) [Shopping](#) [Jobs](#)
Recreation & Sports [Sports](#) [Travel](#) [Aster](#) [Outdoors](#)
Computers & Internet [Internet](#) [WWW](#) [Software](#) [Games](#)
Reference [Libraries](#) [Dictionaries](#) [Quotations](#)
Education [College and University](#) [K-12](#)
Regional [Countries](#) [Regions](#) [US States](#)
Entertainment [Coalitions](#) [Movies](#) [Music](#) [Music](#)
Science [Animals](#) [Astronomy](#) [Engineering](#)
Government [Elections](#) [Military](#) [Law](#) [Taxes](#)
Social Science [Archaeology](#) [Economics](#) [Languages](#)
Health [Medicine](#) [Diseases](#) [Drugs](#) [Fitness](#)
Society & Culture [People](#) [Environment](#) [Religion](#)

[powered by COMPAQ](#)

Local Yahoo!
[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](#) [Chile](#) [Mexico](#) [Spanish](#)
[US Cities](#) [Atlanta](#) [Boston](#) [Chicago](#) [Dallas TX](#) [LA](#) [NYC](#) [SE Bay](#) [Wash DC](#) [more...](#)

More Yahoo!
[Outdoor](#) [Aster](#) [Bazaars](#) [Cancers](#) [Health](#) [Living](#) [Outdoors](#) [Pets](#) [Real Estate](#) [Yahoo! Games!](#)
[Entertainment](#) [Astrology](#) [Broadcast](#) [Events](#) [Games](#) [Movies](#) [Music](#) [Radio](#) [Tickets](#) [TV](#) [more](#)
[Finance](#) [Banking](#) [Bill Pgs](#) [Insurance](#) [Leads](#) [Loans](#) [Finance/Visions](#) [more](#)
[Local](#) [Classifieds](#) [Events](#) [Locations](#) [Maps](#) [Restaurants](#) [Yellow Pages](#) [more](#)
[News](#) [Top Stories](#) [Business](#) [Entertainment](#) [Lifestyle](#) [Politics](#) [Sports](#) [Technology](#) [Weather](#)
[Publishing](#) [Business](#) [Chats](#) [Experts](#) [Quotes](#) [Photos](#) [Home Pages](#) [Message Boards](#)
[Small Business](#) [Biz Marketplace](#) [Domain Registration](#) [Small Biz Center](#) [Store Building](#) [Web Hosting](#)
[Access Yahoo!](#) [vtr](#) [Pages](#) [PDA's](#) [Web-enabled Phones](#) [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 the use of the structure-preserving PageRank algorithm to analyze and produce much more meaningful text and hyperlink data.

The PageRank Citation Ranking: Bringing Order to the Web

January 29, 1998

...

Abstract

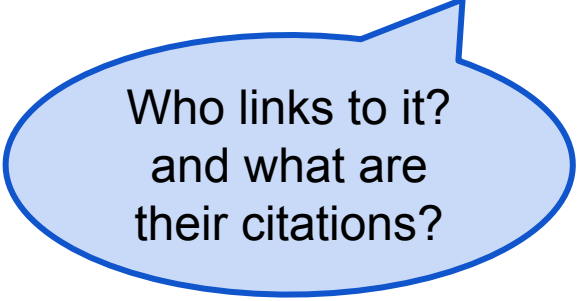
The importance of a Web page is an inherently subjective matter, which depends on the reader's interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes the PageRank algorithm, which computes a numerical importance for each page in the Web, and discusses how it is used by Google to rank search results.

PageRank

Key Idea: Consider the **citations** of the website in addition to keywords.

PageRank

Key Idea: Consider the **citations** of the website in addition to keywords.



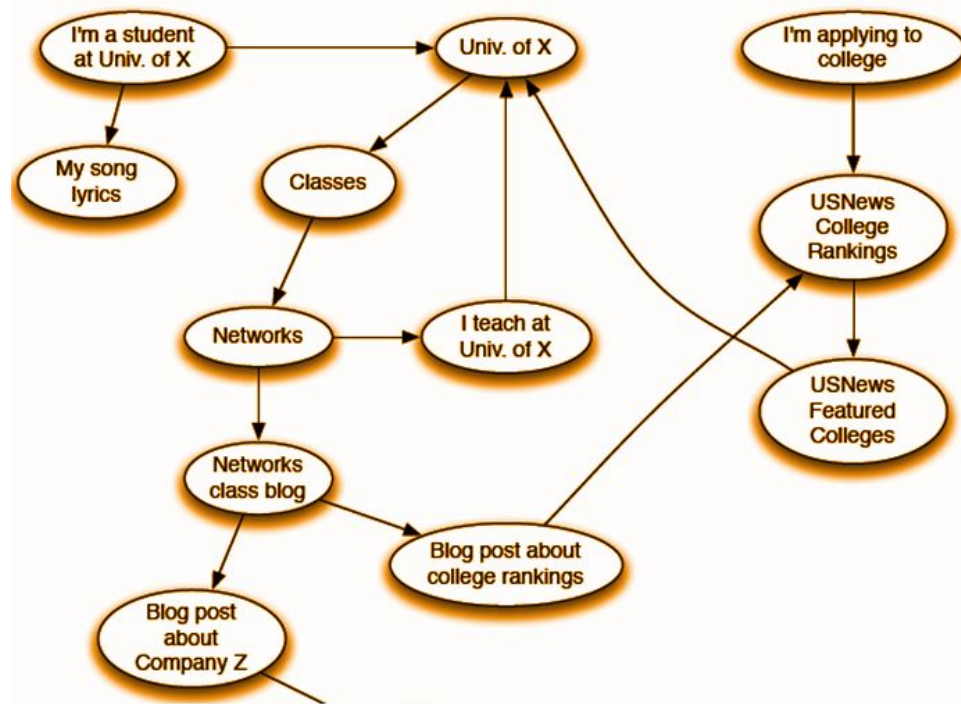
Who links to it?
and what are
their citations?

PageRank

Key Idea: Consider the **citations** of the website in addition to keywords.

Who links to it?
and what are
their citations?

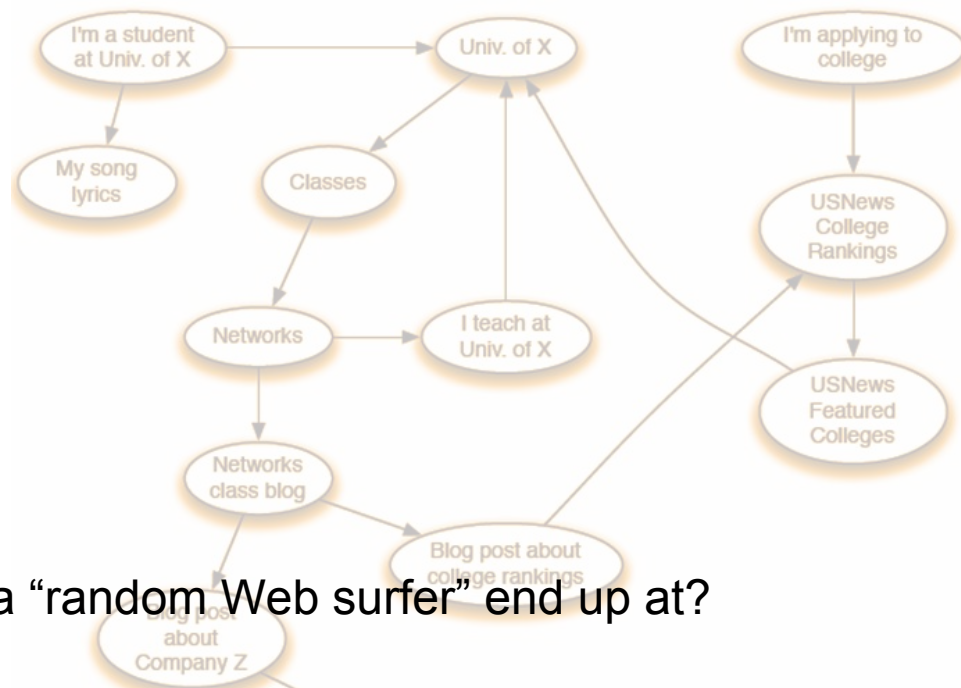
The Web as a directed graph:



PageRank

Key Idea: Consider the **citations** of the website in addition to keywords.

Who links to it?
and what are
their citations?



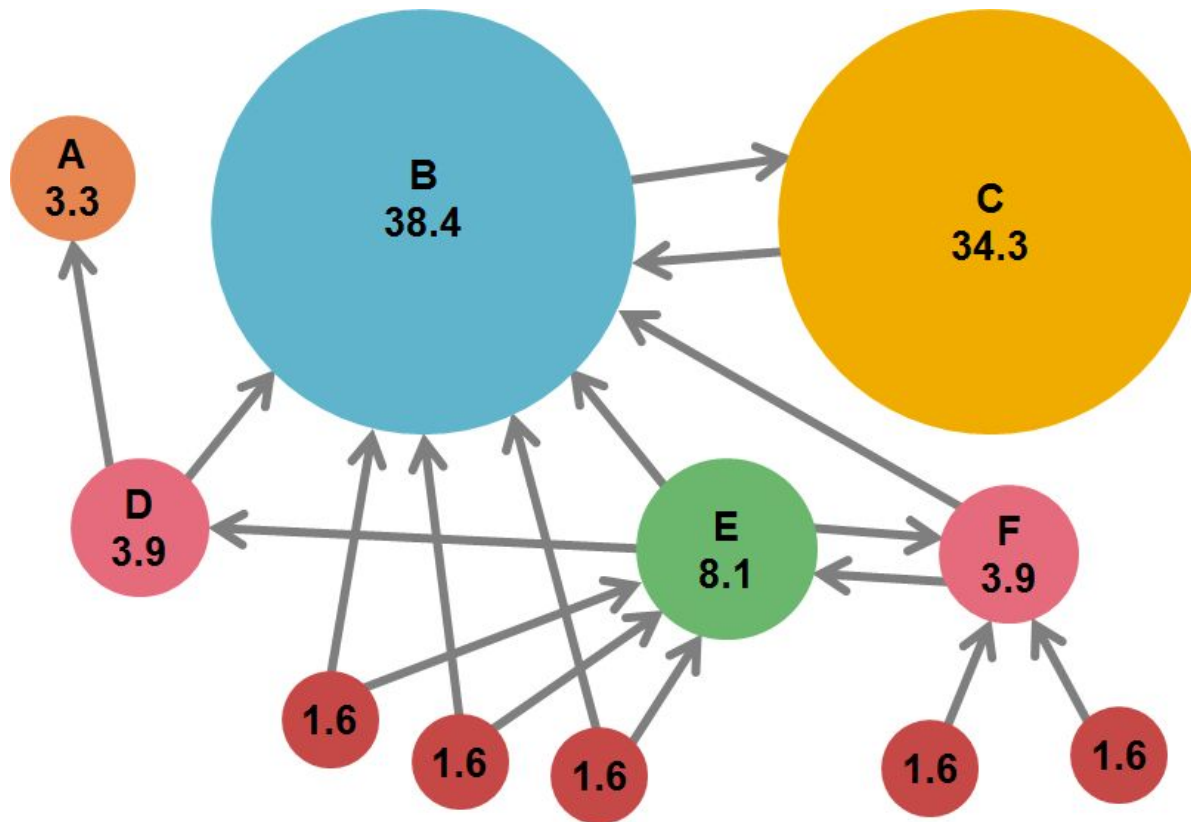
Innovation 1: What pages would a “random Web surfer” end up at?

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

PageRank

Key Idea: Consider the

Flow Model:
in-links as votes



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

Key Idea: Consider the citations of the website in addition to keywords.

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

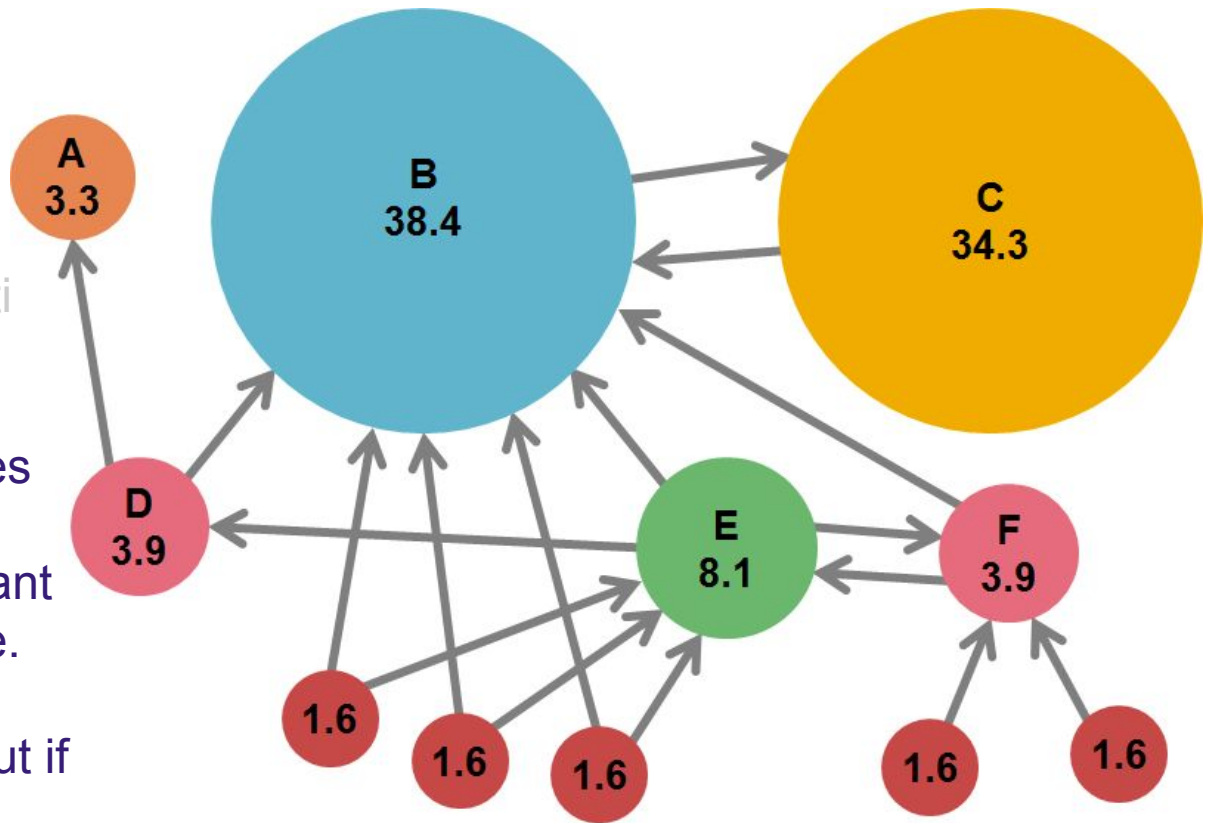
Key Idea: Consider the citati

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.



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

Key Idea: Consider the citations of the website in addition to keywords.

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.

How to compute?

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

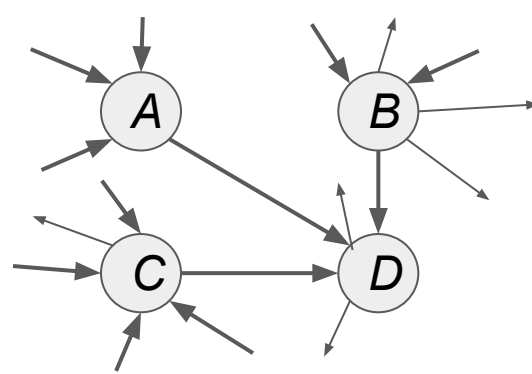
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



How to compute?

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

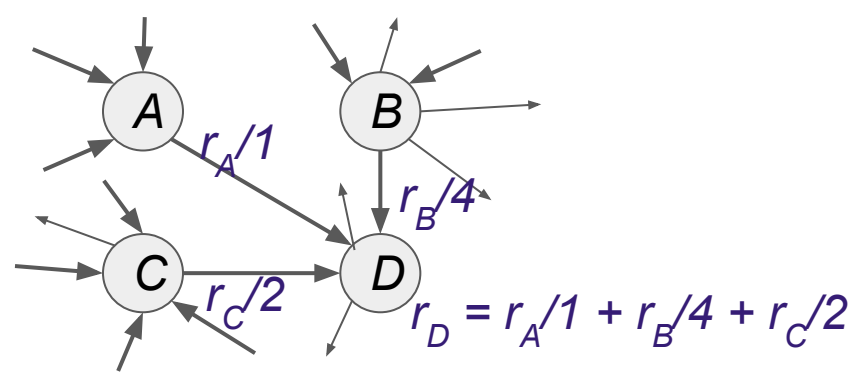
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



How to compute?

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

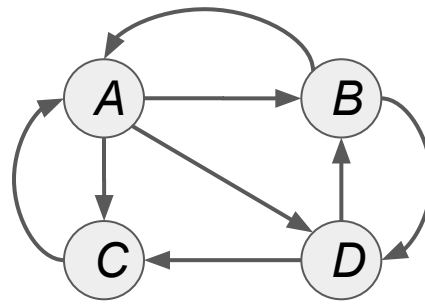
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



How to compute?

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

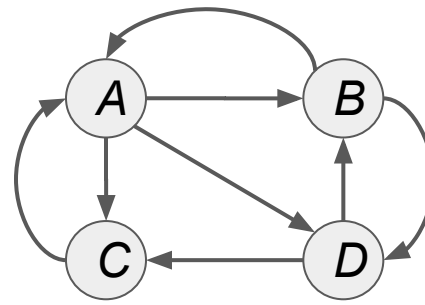
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



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

A system of equations?

How to compute?

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

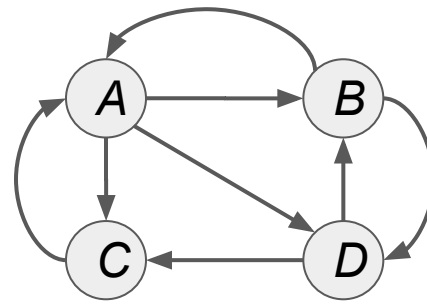
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



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

A system of equations?

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

How to compute?

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

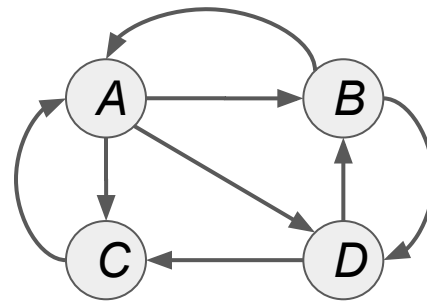
$$vote_j = \frac{r_j}{n_j} \quad (n_j \text{ is } |\text{out-links}|)$$

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

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



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

A system of equations?

Provides intuition, but impractical to solve at scale.

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

How to compute?

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

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

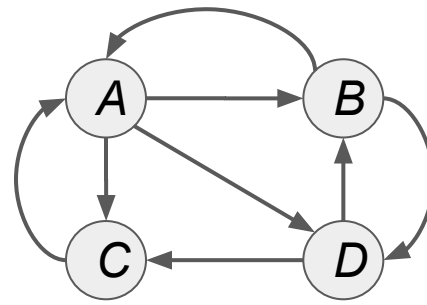
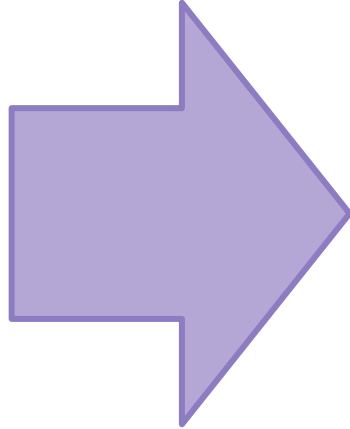
$$r_j = \sum_{i \in \text{inLinks}(j)} vote_i$$

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

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



<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

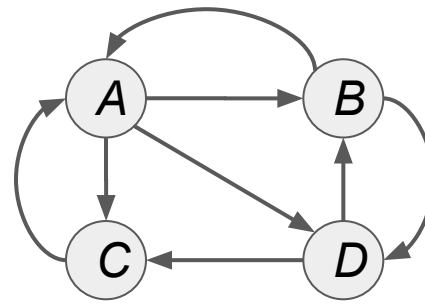
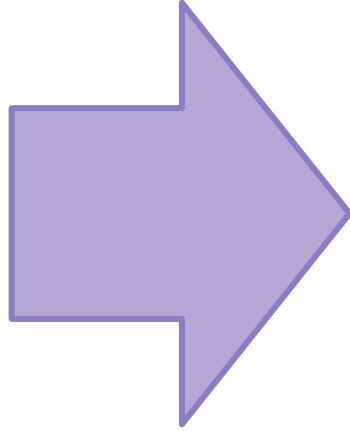
$$1 = r_A + r_B + r_C + r_D$$

“Transition Matrix”, M

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

PageRank

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



<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

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

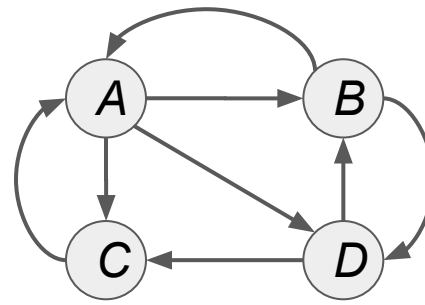
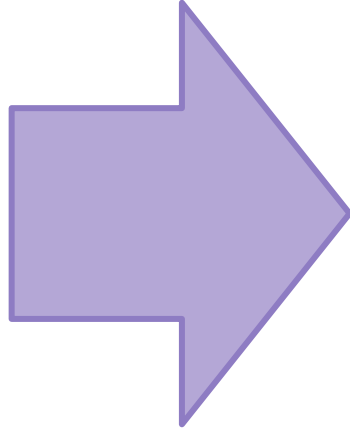
“Transition Matrix”, M

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

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

PageRank

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



<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

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

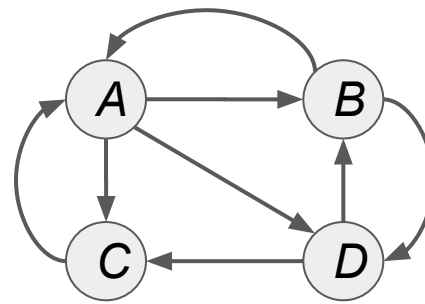
“Transition Matrix”, M

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

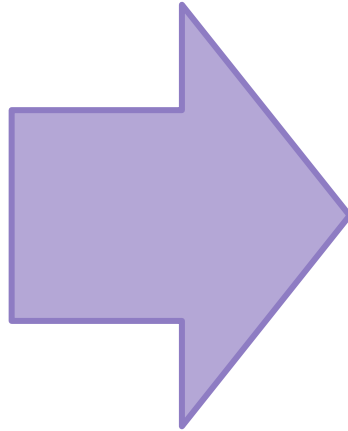
To start: $N=4$ nodes, so $r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$

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

PageRank



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



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

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

“Transition Matrix”, M

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

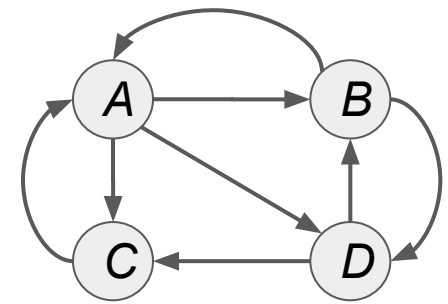
after second iteration: $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

PageRank

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
```



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

To start: $N=4$ nodes, so $r = [1/4, 1/4, 1/4, 1/4]$

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

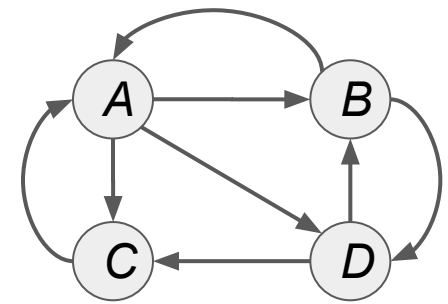
after second iteration: $M(M \cdot r) = M^2 \cdot r = [15/48, 11/48, \dots]$

PageRank

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 =  $r[t]$ 
```

$\text{err_norm}(v1, v2) = |v1 - v2|$ #L1 norm



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

To start: $N=4$ nodes, so $r = [1/4, 1/4, 1/4, 1/4,]$

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

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

PageRank

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
```

As err_norm gets smaller
we are moving toward:

$$r = M \cdot r$$

We are actually just
finding the
eigenvector of M .

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

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

PageRank

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
```

As err_norm gets smaller
we are moving toward:

$$r = M \cdot r$$

We are actually just
finding the
eigenvector of M .

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

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

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

PageRank

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...

As err_norm gets smaller
we are moving toward:

$$r = M \cdot r$$

We are actually just
finding the
eigenvector of M .

x is an
eigenvector of λ if:

$$A \cdot x = \lambda \cdot x$$

$$A = 1$$

since columns of M sum to 1.
thus, $1r = Mr$

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

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

PageRank

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...

As err_norm gets smaller
we are moving toward:

$$r = M \cdot r$$

We are actually just
finding the
eigenvector of M .

x is an
eigenvector of λ if:

$$A \cdot x = \lambda \cdot x$$

$$A = 1$$

since columns of M sum to 1.
thus, $1r = Mr$

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

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

PageRank

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
```

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.

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

PageRank

aka 1st order Markov Process

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

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.

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

PageRank

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.

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.

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

PageRank

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.

(technically: it needs to be:

stochastic, irreducible, and aperiodic)

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**.

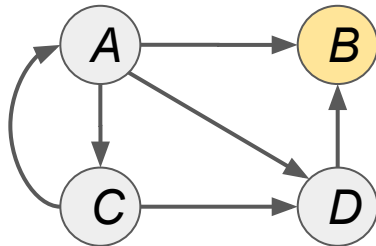
Thus, r is a stationary
distribution. Probability of
being at given node.

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

PageRank

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.



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

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.

r would eventually
converge to
[0, 0, ...]

PageRank

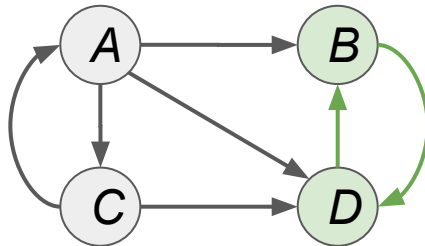
aka 1st order Markov Process

- 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.

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.



PageRank

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.

(technically: it needs to be:

stochastic, *irreducible*, and *aperiodic*)

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.

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

columns sum to 1

same node doesn't repeat at a regular interval

non-zero chance of going to any another node

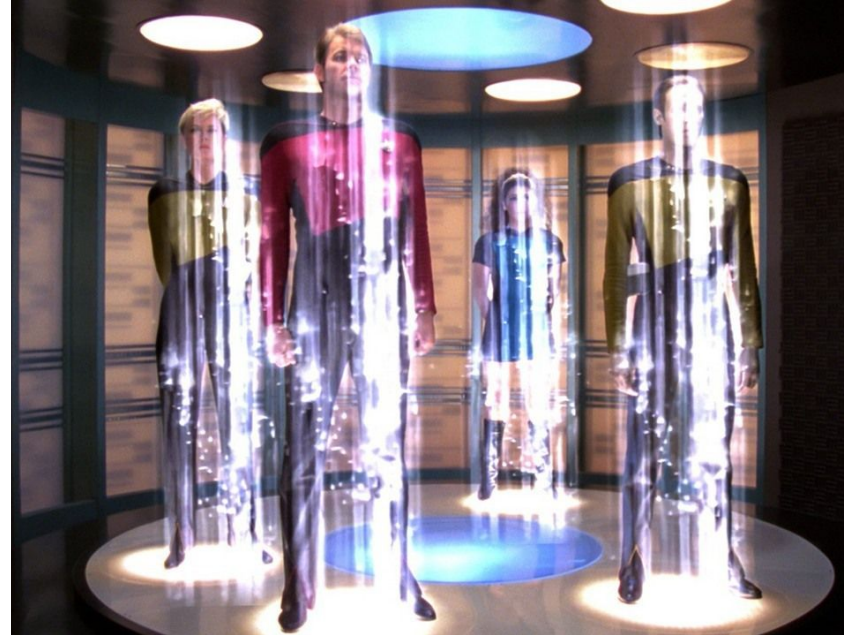
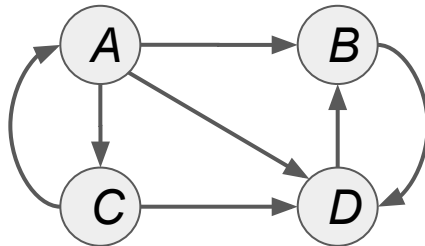
PageRank

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, $1-\beta$)



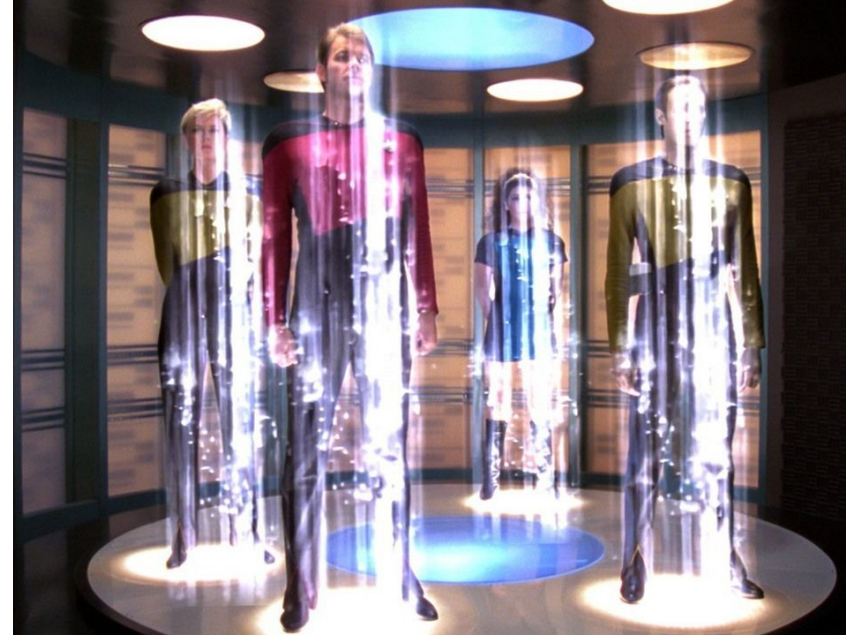
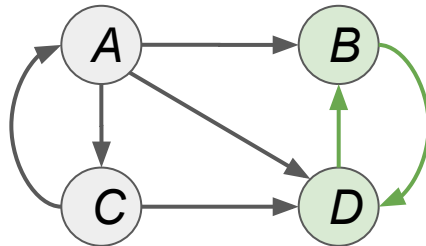
PageRank

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, $1-\beta$)



PageRank

No “dead-ends”
No “spider traps”

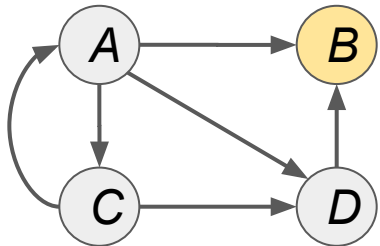
The “Google” PageRank Formulation

Add teleportation chance at all nodes.

i.e. at each step, two choices

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

Add teleportation from dead end with probability 1



<i>to \ from</i>	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

PageRank

No “dead-ends”
No “spider traps”

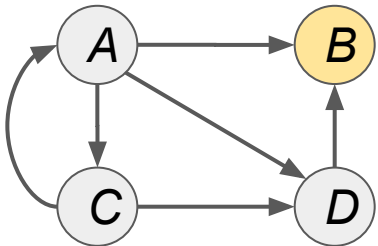
The “Google” PageRank Formulation

Add teleportation chance at all nodes.

i.e. at each step, two choices

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

Add teleportation from dead end with probability 1



<i>to \ from</i>	A	B	C	D
A	0	1/4	1	0
B	1/3	1/4	0	1
C	1/3	1/4	0	0
D	1/3	1/4	0	0

PageRank

No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation

Add teleportation chance at all nodes.

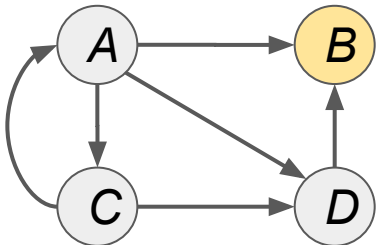
i.e. at each step, two choices

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

Add teleportation from dead end with probability 1

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

(Brin and Page, 1998)



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

PageRank

No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation

Add teleportation chance at all nodes.

i.e. at each step, two choices

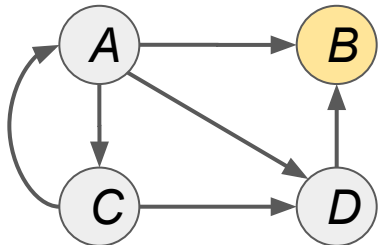
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

Add teleportation from dead end with probability 1

Flow Model

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

(Brin and Page, 1998)



Matrix Model

$$M' = \beta M + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$$

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

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

assume
 $\beta = \frac{4}{5}$

PageRank

No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation

Add teleportation chance at all nodes.

i.e. at each step, two choices

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

Add teleportation from dead end with probability 1

Matrix Model

$$M' = \beta M + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$$

To apply:

run power iterations over M'
instead of M .

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

assume
 $\beta = \frac{4}{5}$