

67-300 SEARCH ENGINES

---

# LINK ANALYSIS

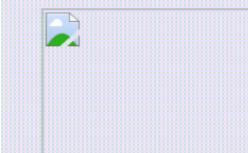
LECTURER: JOAO PALOTTI ([JPALOTTI@ANDREW.CMU.EDU](mailto:JPALOTTI@ANDREW.CMU.EDU))  
12TH APRIL 2017

### LECTURE'S GOAL

- ▶ Finish assessments on Juno ([juno.ifs.tuwien.ac.at:9999](http://juno.ifs.tuwien.ac.at:9999))
- ▶ Short history on Web Retrieval and advertisement on the Web
- ▶ Link Analysis & PageRank
- ▶ Incorporating link analysis in retrieval

# AT THE BEGINNING THERE WAS CHAOS...

- ▶ Two very different approaches to find information:
  - ▶ Full-text index search engines
  - ▶ Taxonomies populated engines
- ▶ Some early search engines mixed both approaches



Researching stocks?

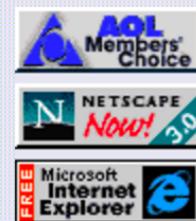
Buying a car?

Planning a wedding?

[Check out](#)

[ExciteSeeing Tours.](#)

[Bill Mitchell:](#)  
[Satire that clicks!](#)



[Make your website  
searchable, FREE!](#)



## [Excite City.Net](#)

Your guide to the world.

Search City.Net

[search](#)

[maps](#) • [concierge](#) • [top cities](#)

## [Excite Live!](#)

Your news, your way.

- [Latest news](#)
- [Stock quotes](#)
- [Sports scores](#)
- [TV listings](#)
- [Local weather](#)
- [Horoscopes](#)
- [Movie reviews](#)
- [Site reviews](#)
- [The Web's Smartest News Agent](#)



©1996 Excite Inc.

[Feedback](#)

[About Excite](#)

[Jobs at Excite](#)

[Link to Excite](#)

[Advertising](#)

[Disclaimer](#)

**EXCITE.COM IN 1996**  
**THANKS [WEB.ARCHIVE.ORG](#)**



**AUTO.BY-TEL**  
USA CANADA  
Buy and insure new cars & trucks online

**Car Buying & Car Insurance  
Pain Relief**



[Click here for advertising information - reach millions every month!](#)

Search  and Display the Results

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



[Download free demo versions of AltaVista Technology software](#)





[ New Search | Big Yellow | World News | Fast Facts | The iZone ]

## Search for information about:

in the World Wide Web Search Now

Infoseek Guide is best viewed with:



Want personalized news? [Get Personal now!](#)

### Basic Search Tips:

- Click in the box above and type a few words that describe what you want to find. For example, typing **growing orchids indoors** will find sites about caring for orchids.
- If you are looking for a person or place, type the name, starting with capital letters. For example, typing **Florence Italy** will find sites about this famous city.
- These detailed [search tips](#) describe how to use the features of Infoseek Guide to find what you are looking for.
- For the broadest results, you can search the entire **World Wide Web**.
- To restrict your search to hand-picked and categorized sites, choose **Infoseek Select Sites**.
- Or just search for a category within Infoseek Select by choosing **Categories of Sites**.
- To search through Internet discussion forums (similar to bulletin boards), choose **Usenet Newsgroups**.
- To search for someone's e-mail address, choose **E-mail Addresses**.
- To search through news stories within the past month, choose **Reuters News**.
- To search through answers to Frequently Asked Questions, choose **Web FAQs**.

### Explore these popular Infoseek Select topics:

- ▶ [Arts & Entertainment](#)
- ▶ [Business & Finance](#)
- ▶ [Computers & Internet](#)
- ▶ [Education](#)
- ▶ [Government & Politics](#)
- ▶ [Health & Medicine](#)
- ▶ [Living](#)
- ▶ [News](#)
- ▶ [Reference](#)
- ▶ [Science & Technology](#)
- ▶ [Sports](#)
- ▶ [Travel](#)

Try [Infoseek Personal](#), your personalized news service

[Customize](#) | [Add Site](#) | [Help](#) | [Feedback](#)  
[Download iSeek](#) | [About Infoseek](#)



[Banking online. Is it as easy as 1,2,3? Click here.]

# Welcome to Google

GOOGLE IN 1998

Google Search Engine Prototype

Might-work-some-of-the-time-prototype that is much more up to date.

# Welcome to Google

GOOGLE IN 1998

Google Search Engine Prototype

Might-work-some-of-the-time-prototype that is much more up to date.



GOOGLE IN JAN 1999

Search the web using Google!

Google Search

Special Searches  
[Stanford Search](#)  
[Linux Search](#)

[Help!](#)  
[About Google!](#)  
[Company Info](#)  
[Google! Logos](#)

Get Google!  
updates monthly:  
  
 [Archive](#)

# Welcome to Google

GOOGLE IN 1998

Google Search Engine Prototype

Might-work-some-of-the-time-prototype that is much more up to date.



GOOGLE IN JAN 1999

Special Searches  
[Stanford Search](#)  
[Linux Search](#)

[Help!](#)  
[About Google!](#)  
[Company Info](#)  
[Google! Logos](#)

Get Google!  
updates monthly:  
your e-mail  
[Subscribe](#) [Archive](#)



GOOGLE IN DEC 1999

Search the web using Google

[Google Search](#) [I'm feeling lucky](#)

Google Wins PC Magazine's Technical Excellence Award  
for Innovation in Web Application Development

# YAHOO IN 1996



- [Arts and Humanities](#) - [Architecture](#), [Photography](#), [Literature](#)...
- [Business and Economy \[Xtra!\]](#) - [Companies](#), [Investments](#), [Classifieds](#)...
- [Computers and Internet \[Xtra!\]](#) - [Internet](#), [WWW](#), [Software](#), [Multimedia](#)...
- [Education](#) - [Universities](#), [K-12](#), [College Entrance](#)...
- [Entertainment \[Xtra!\]](#) - [Cool Links](#), [Movies](#), [Music](#), [Humor](#)...
- [Government](#) - [Politics \[Xtra!\]](#), [Agencies](#), [Law](#), [Military](#)...
- [Health \[Xtra!\]](#) - [Medicine](#), [Drugs](#), [Diseases](#), [Fitness](#)...
- [News and Media \[Xtra!\]](#) - [Current Events](#), [Magazines](#), [TV](#), [Newspapers](#)...
- [Recreation and Sports \[Xtra!\]](#) - [Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors](#)...
- [Reference](#) - [Libraries](#), [Dictionaries](#), [Phone Numbers](#)...
- [Regional](#) - [Countries](#), [Regions](#), [U.S. States](#)...
- [Science](#) - [CS](#), [Biology](#), [Astronomy](#), [Engineering](#)...
- [Social Science](#) - [Anthropology](#), [Sociology](#), [Economics](#)...
- [Society and Culture](#) - [People](#), [Environment](#), [Religion](#)...

[Yahooligans!](#) - [Yahoo! Shop](#) - [Yahoo! Internet Life](#)

[Daily Picks](#) - [Random Yahoo! Link](#) - [Weekly Picks](#)

[National Yahoos](#) - [Canada](#) - [France](#) - [Germany](#) - [Japan](#) - [U.K. and Ireland](#)

[Yahoo! Metros](#) - [New York](#) - [San Francisco Bay Area](#) - [Los Angeles](#)

[How to Include Your Site](#) - [Company Information](#)

# DMOZ IN 2000

**dmoz** open directory project

[about dmoz](#) | [add URL](#) | [feedback](#) | [link](#) | [editor login](#)

Search [advanced](#)

## Arts

[Movies](#), [Television](#), [Music](#)...

## Games

[Video Games](#), [RPGs](#), [Gambling](#)...

## Kids and Teens

[Arts](#), [School Time](#), [Teen Life](#)...

## Reference

[Maps](#), [Education](#), [Libraries](#)...

## Shopping

[Autos](#), [Clothing](#), [Gifts](#)...

## World

[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Korean](#), [Nederlands](#), [Polska](#), [Svenska](#)...

## Business

[Jobs](#), [Industries](#), [Investing](#)...

## Health

[Fitness](#), [Medicine](#), [Alternative](#)...

## News

[Media](#), [Newspapers](#), [Weather](#)...

## Regional

[US](#), [Canada](#), [UK](#), [Europe](#)...

## Society

[People](#), [Religion](#), [Issues](#)...

## Computers

[Internet](#), [Software](#), [Hardware](#)...

## Home

[Family](#), [Consumers](#), [Cooking](#)...

## Recreation

[Travel](#), [Food](#), [Outdoors](#), [Humor](#)...

## Science

[Biology](#), [Psychology](#), [Physics](#)...

## Sports

[Baseball](#), [Soccer](#), [Basketball](#)...

[Become an Editor](#)

Help build the largest human-edited directory of the web



Copyright © 1998-2000 Netscape

2,209,355 sites - 32,063 editors - 323,996 categories

# DMOZ IN 2000

**dmoz** open directory project

[about dmoz](#) | [add URL](#) | [feedback](#) | [link](#) | [editor login](#)

 Search [advanced](#)

## Arts

[Movies](#), [Television](#), [Music](#)...

## Games

[Video Games](#), [RPGs](#), [Gambling](#)...

## Kids and Teens

[Arts](#), [School Time](#), [Teen Life](#)...

## Reference

[Maps](#), [Education](#), [Libraries](#)...

## Shopping

[Autos](#), [Clothing](#), [Gifts](#)...

## World

[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Korean](#), [Nederlands](#), [Polska](#), [Svenska](#)...

## Business

[Jobs](#), [Industries](#), [Investing](#)...

## Health

[Fitness](#), [Medicine](#), [Alternative](#)...

## News

[Media](#), [Newspapers](#), [Weather](#)...

## Regional

[US](#), [Canada](#), [UK](#), [Europe](#)...

## Society

[People](#), [Religion](#), [Issues](#)...

## Computers

[Internet](#), [Software](#), [Hardware](#)...

## Home

[Family](#), [Consumers](#), [Cooking](#)...

## Recreation

[Travel](#), [Food](#), [Outdoors](#), [Humor](#)...

## Science

[Biology](#), [Psychology](#), [Physics](#)...

## Sports

[Baseball](#), [Soccer](#), [Basketball](#)...

[Become an Editor](#) Help build the largest human-edited directory of the web



Copyright © 1998-2000 Netscape

2,209,355 sites - 32,063 editors - 323,996 categories

# DMOZ IN 2000

**dmoz** open directory project

[about dmoz](#) | [add URL](#) | [feedback](#) | [link](#) | [editor login](#)

 Search [advanced](#)

## Arts

[Movies, Television, Music...](#)

## Games

[Video Games, RPGs, Gambling...](#)

## Kids and Teens

[Arts, School Time, Teen Life...](#)

## Reference

[Maps, Education, Libraries...](#)

## Shopping

[Autos, Clothing, Gifts...](#)

## World

[Deutsch, Español, Français, Italiano, Japanese, Korean, Nederl](#)

## Business

[Jobs, Industries, Investing](#)

## Health

[Fitness, Medicine, Alterna](#)

## News

[Media, Newspapers, Weat](#)

## Regional

[US, Canada, UK, Europe..](#)

## Society

[People, Religion, Issues...](#)



[Become an Editor](#) Help build the largest human-edited directo

Copyright © 1998-2000 Netscape

2,209,355 sites - 32,063 editors - 32

**As of Mar 17, 2017, dmoz.org is no longer available.**

**The editors have set up a static mirror [here](#).**

If you are interested in staying in touch with the DMOZ community,  
please visit [www.resource-zone.com](http://www.resource-zone.com).

Thank you all, especially the editors, for your interest and dedication to this project.

# WHICH REPRESENTATION WON THIS BATTLE?

- ▶ On one hand...
  - ▶ Taxonomies are an organized way to find information, but...

# WHICH REPRESENTATION WON THIS BATTLE?

- ▶ On one hand...
  - ▶ Taxonomies are an organized way to find information, but...
  - ▶ Maintenance is hard
  - ▶ Do not scale well
  - ▶ Many clicks to find stuff

# WHICH REPRESENTATION WON THIS BATTLE?

- ▶ On one hand...
  - ▶ Taxonomies are an organized way to find information, but...
    - ▶ Maintenance is hard
    - ▶ Do not scale well
    - ▶ Many clicks to find stuff
- ▶ On the other hand:
  - ▶ Full-text search are exposed to all sort of spammers...

# FIRST GENERATION OF SPAMS

- ▶ Query: “doha real estate”
- ▶ How can I boost my website to sell houses in doha?



That is a way to hide text....

# FIRST GENERATION OF SPAMS

- ▶ Query: “doha real estate”
- ▶ How can I boost my website to sell houses in doha?
- ▶ Adding/Hiding keywords (such as a lot of repetitions of doha real estate) in the same color as the page background. **Invisible to humans, not to machines.**

# ADVERSARIAL SEARCH

- ▶ Research area dedicated to combat spam...

# ADVERSARIAL SEARCH

- ▶ Research area dedicated to combat spam...
- ▶ Search Engine Optimization (SEO):
  - ▶ Not necessarily spam (fine line between spam and legitimate SEO)
  - ▶ Hired to restructure your content in a way that makes it easy to index
  - ▶ Influential bloggers hired to link to your site
  - ▶ Suggestions on addition of interesting and original content

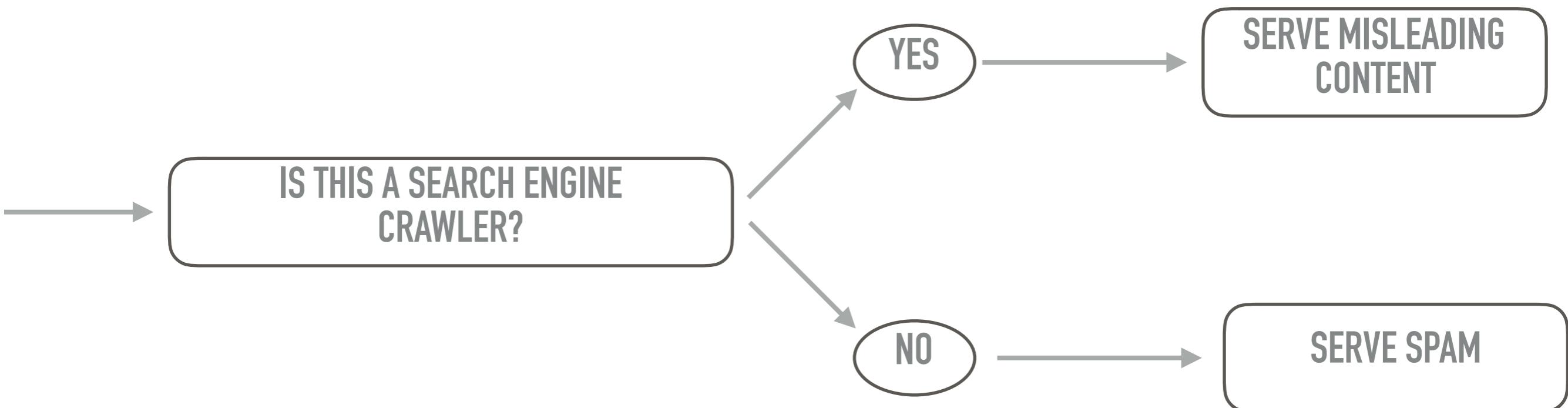
# ADVERSARIAL SEARCH

- ▶ Research area dedicated to combat spam...
- ▶ Search Engine Optimization (SEO):
  - ▶ Not necessarily spam (fine line between spam and legitimate SEO)
  - ▶ Hired to restructure your content in a way that makes it easy to index
  - ▶ Influential bloggers hired to link to your site
  - ▶ Suggestions on addition of interesting and original content

**TALKING ABOUT SPAM...**

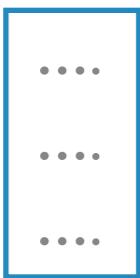
# ADVERSARIAL SEARCH

- ▶ Cloaking:



# ADVERSARIAL SEARCH

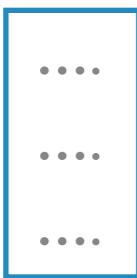
- ▶ Duplicates:



Cool original page

# ADVERSARIAL SEARCH

- ▶ Duplicates:



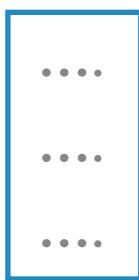
Cool original page



Another cool  
original page

# ADVERSARIAL SEARCH

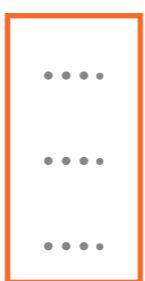
- ▶ Duplicates:



Cool original page



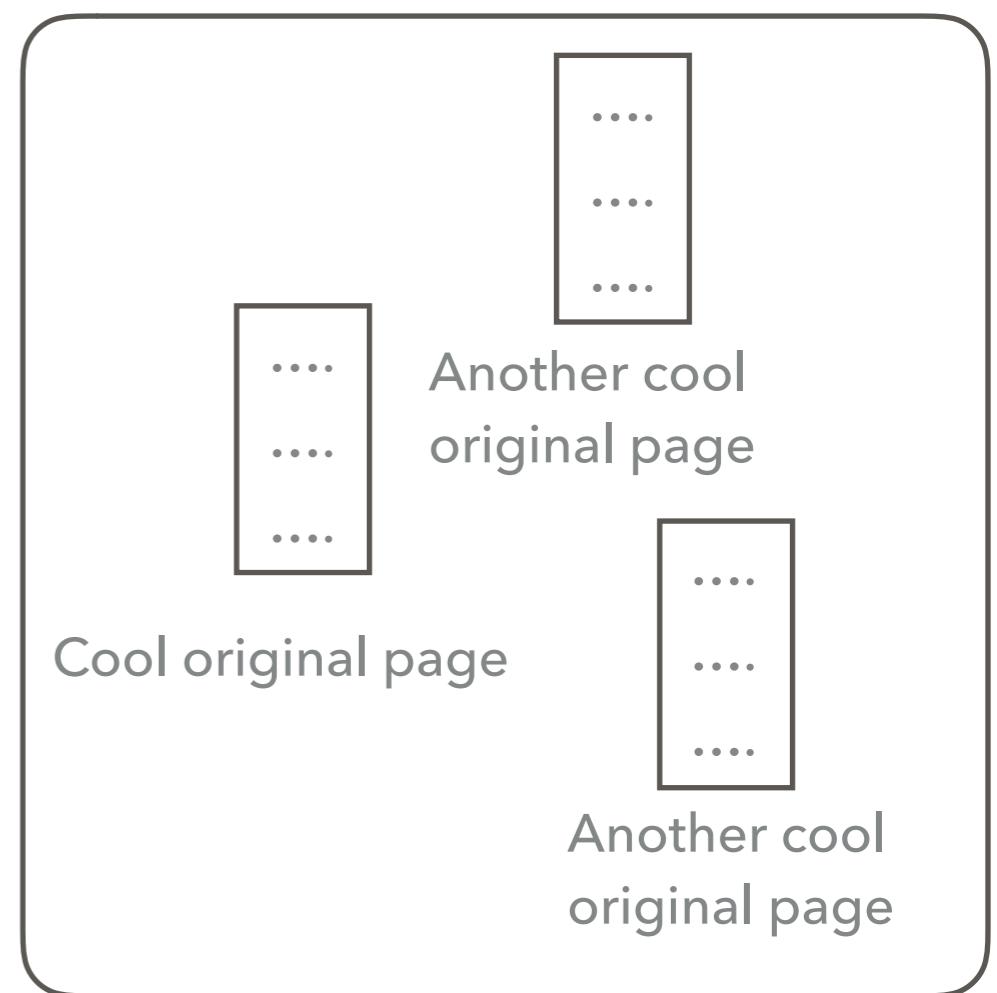
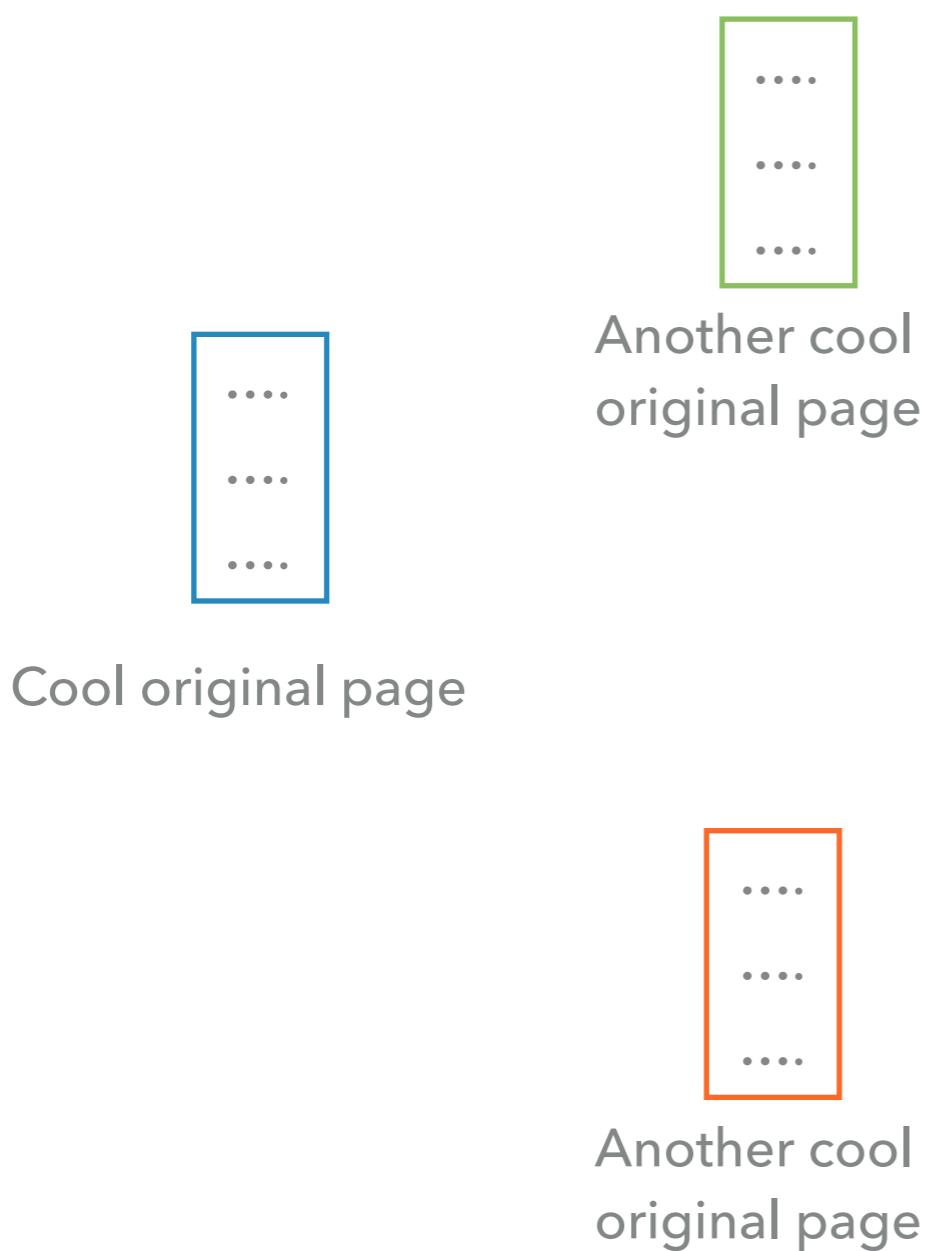
Another cool  
original page



Another cool  
original page

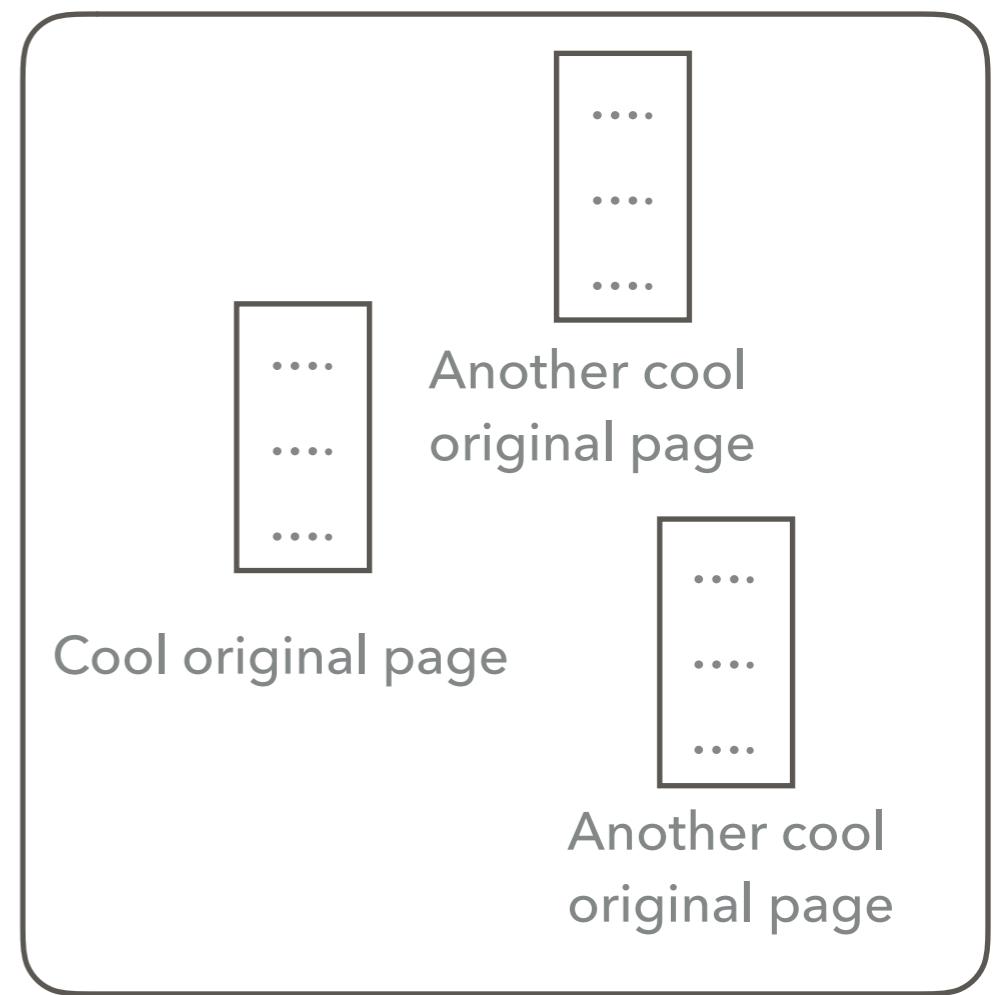
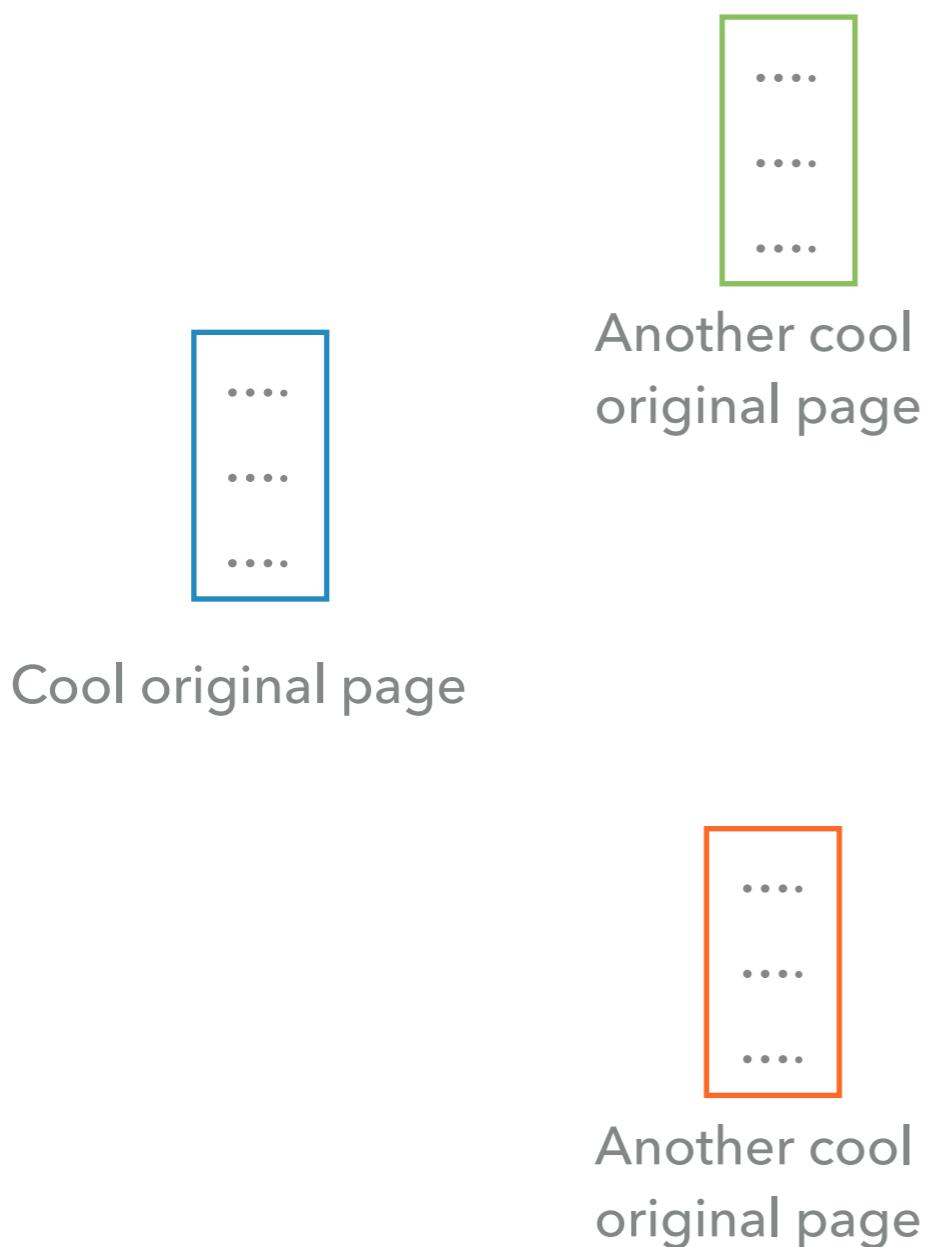
# ADVERSARIAL SEARCH

## ▶ Duplicates:



# ADVERSARIAL SEARCH

## ▶ Duplicates:



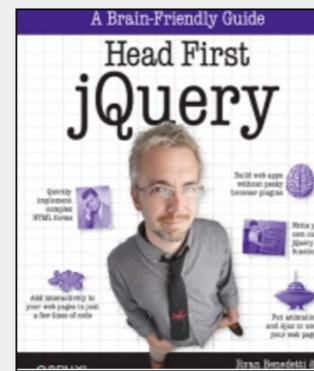
**EXAMPLE: HOW TO WEBSITES**

# ADVERSARIAL SEARCH

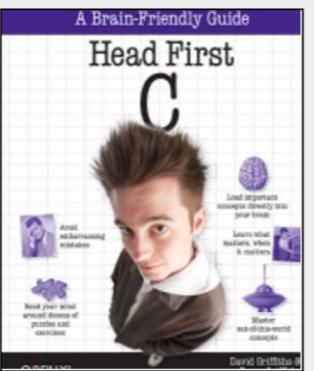
- ▶ Doorway / lander pages
- ▶ Have you search for “**free**” stuff on the Web?



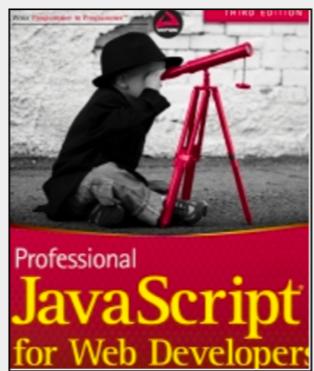
## Top Download eBooks



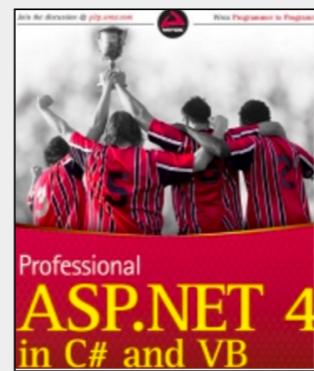
[Head First jQuery](#)



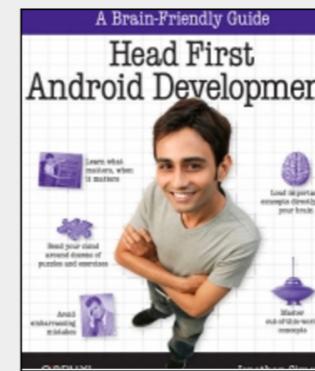
[Head First C](#)



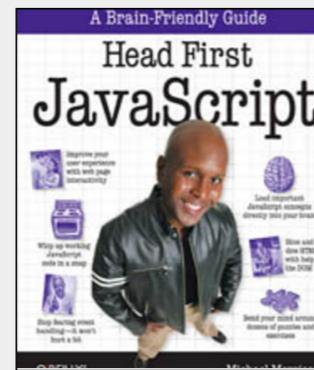
[Professional JavaScript for Web Developers, 3rd Edition](#)



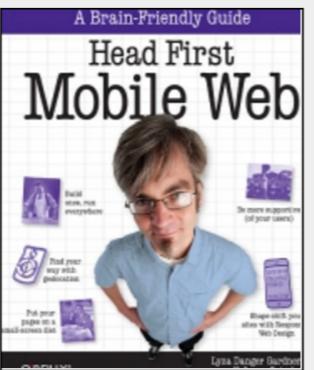
[Professional ASP.NET 4 in C# and VB](#)



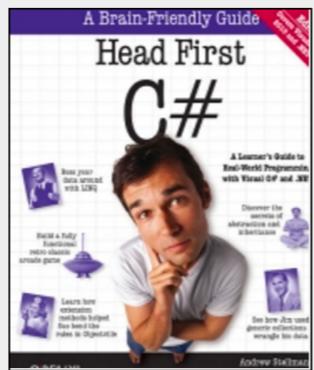
[Head First Android Development](#)



[Head First JavaScript](#)



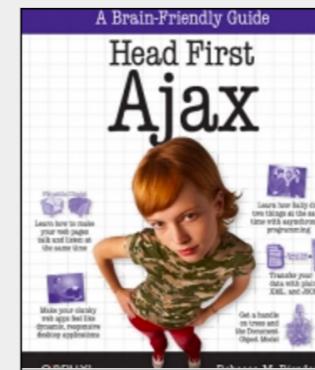
[Head First Mobile Web](#)



[Head First C#, 2nd Edition](#)

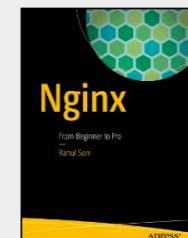


[Head First Servlets and JSP, 2nd Edition](#)



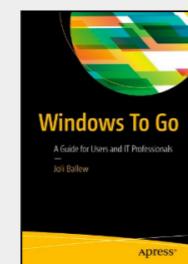
[Head First Ajax](#)

## Last Upload eBooks



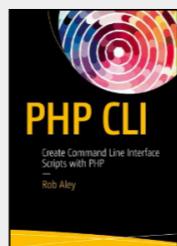
[Nginx](#)

Teaches you to start up Nginx and quickly take your expertise to a level where you can comfortably work with various aspects of the web server and make informed design decisions for your web farm. Nginx powers more than 40% of the top 1000 websites a...



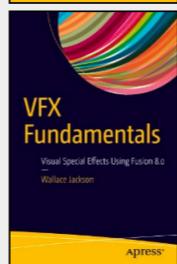
[Windows To Go](#)

Discover how to use, manage, deploy, and secure Windows To Go, the ultimate mobile edition of Windows 10, designed to boot directly from a USB stick. This book shows you how to use your Windows To Go drive to work at home and on the move and access I...



[PHP CLI](#)

This concise book shows you how to create PHP Command Line Interface (CLI) scripts, including user interaction and scripts to automate and assist your workflow. Learn to quickly create useful and effective Command Line software and scripts using the ...



[VFX Fundamentals](#)

This book is a comprehensive title covering concepts central to Visual Special Effects using the free Black Magic Design Fusion 8.0 software package. The book also provides foundational background information regarding digital image compositing, digi...

[Free Ebook Downloads](#)

[Download Digital Books](#)

[Download Books For Free](#)

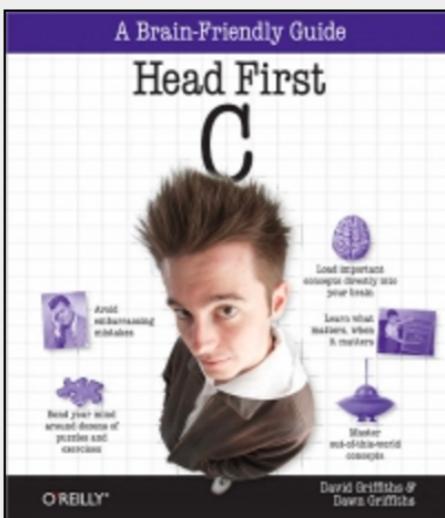
[Read Free Books Online](#)

[Free IT Ebooks](#)

[Free Computer Ebooks](#)



## Head First C



### Book Description

Ever wished you could learn C from a book? Head First C provides a complete learning experience for C and structured imperative programming. With a unique method that goes beyond syntax and how-to manuals, this guide not only teaches you the language, it helps you understand how to be a great programmer. You'll learn key areas such as language basics, pointers and pointer arithmetic, and dynamic memory management. Advanced topics include multi-threading and [network programming](#) - topics typically covered on a college-level course.

We think your time is too valuable to waste struggling with new concepts. Using the latest research in cognitive science and learning theory to craft a multi-sensory learning experience, Head First C uses a visually rich format designed for the way your brain works, not a text-heavy approach that puts you to sleep.

### Book Details

Publisher:	<a href="#">O'Reilly Media</a>
By:	<a href="#">David Griffiths, Dawn Griffiths</a>
ISBN:	<b>978-1-4493-9991-7</b>
Year:	<b>2012</b>
Pages:	<b>640</b>
Language:	<b>English</b>
File size:	<b>56.3 MB</b>
File format:	<b>PDF</b>



### Subscribe via email

enter your email address

Subscribe

[Free Ebook Downloads](#)

[Read Free Books Online](#)

[Download Digital Books](#)

[Free IT Ebooks](#)

[Create A Website](#)



### Paper Book

Buy: [Head First C](#)

[Download EBooks](#)

[Download Books For Free](#)

[Free It Ebooks](#)

[Free EBook PDF](#)



# ADVERSARIAL SEARCH

- ▶ Doorway / lander pages
  - ▶ Have you search for “**free**” stuff on the Web?
  - ▶ Optimized pages for a single search... “free books”
  - ▶ Only purpose: get people to click on the ads and make money this way

# SOLUTION?

- ▶ Search engines have guidelines you should follow:
  - ▶ Ignore these guidelines and have your website removed from their index...
- ▶ Alternative: you pay to be high ranked!

# ADVERTISEMENTS

ALSO FROM 1996...

The screenshot shows a search results page for "lactose intolerance". At the top, there's a yellow header bar with the "GO TO" logo, a search bar containing "lactose intolerance", and a "CLICK HERE" button for Capital One. To the right, there's a banner for "0% intro + 9.9% fixed APR" with "60-sec responses". Below the header, it says "You searched for lactose intolerance." and "Search Tips". On the left, there's a sidebar with categories like Computing, Education & Caesar, Entertainment, Finance, Health, Homelife, Reference, Shopping, and Travel. A yellow callout box in the center says "Access 75% of all users now! Premium Listings reach 75% of all Internet users. [Sign up](#) for Premium Listings today!".

1. [No More Lactose Intolerance!](#)  
Try Sammi's best whey and rice. For anyone with lactose intolerance this is it! 75% less lactose, with Acidophilus and bifidus bacteria. Buy factory direct with free shipping.  
[www.bestlifeint.com](http://www.bestlifeint.com) (Cost to advertiser: [\\$0.30](#))

2. [Lactose Intolerance: Whey Drinks](#)  
Lactose intolerance: Swiss Whey Powder. Delicious, natural, refreshing and easy to mix. Low in lactose. Great for beverages, cooking. 5 year shelf life. Money back guarantee. Mix to taste.  
[www.galaxymall.com](http://www.galaxymall.com) (Cost to advertiser: [\\$0.07](#))

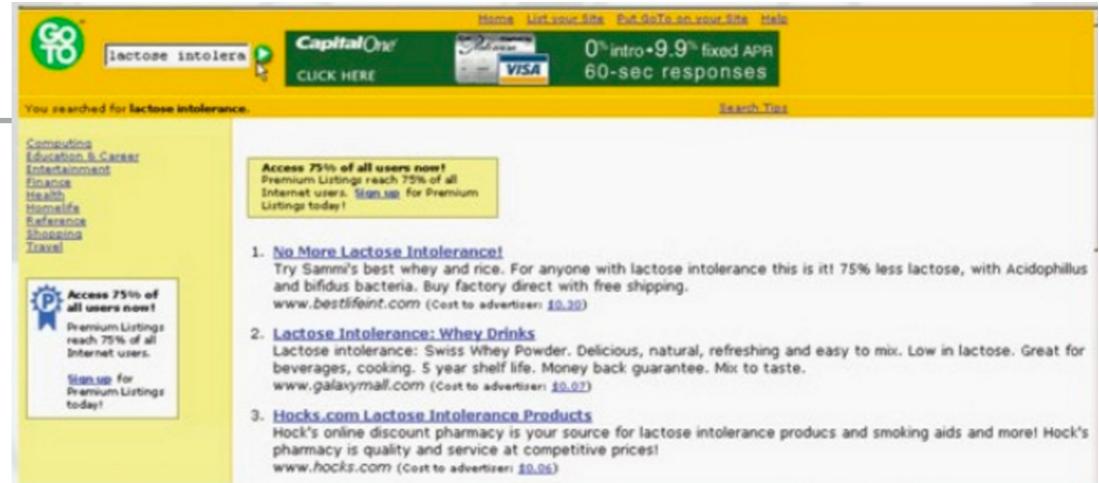
3. [Hock's.com Lactose Intolerance Products](#)  
Hock's online discount pharmacy is your source for lactose intolerance products and smoking aids and more! Hock's pharmacy is quality and service at competitive prices!  
[www.hocks.com](http://www.hocks.com) (cost to advertiser: [\\$0.06](#))

# ADVERTISEMENTS: COST PER CLICK



- ▶ Highest bid for this search was \$0.30
- ▶ The website owner has to pay \$0.30 per time a user clicks on this link (Cost Per Click = CPC)
- ▶ Pages were ranked by bid value – No TF-IDF, no VSM, no LM, no BM25 anymore...

# ADVERTISEMENTS: COST PER CLICK



- ▶ **Good:** It was honest – Goto did not pretend there was any relevance ranking
- ▶ **Bad:** Open to abuse – query: “treatment for cancer” Ad: “how to write your last will”
  - ▶ No relationship with query relevance
  - ▶ **Ugly:** No separation of ads/docs...

## LECTURE 10 - LINK ANALYSIS

# ADVERTISEMENTS — VERSION 2.0

- ▶ Also done by auction
- ▶ Second price auction
- ▶ Prices are not known
- ▶ Measure: CTR
  - ▶ clickthrough rate == per impression
- ▶ Ranking factors: the query + location
  - + time of the day + quality + loading speed

Google hotels doha

All Maps News Images More Settings Tools

About 33,000,000 results (1.12 seconds)

**105 Hotels in Doha, Qatar - Best Price Guarantee - booking.com**  
Ad [www.booking.com/Doha/Hotels](http://www.booking.com/Doha/Hotels) ▾  
Book your Hotel in Doha online. No reservation costs. Great rates.  
Get Instant Confirmation · We speak your language · 24/7 Customer Service · No Booking Fees  
Types: Hotels, Apartments, Villas, Hostels, Resorts, B&Bs

£0 - £45 Hotels up to £45.00/day Book Now

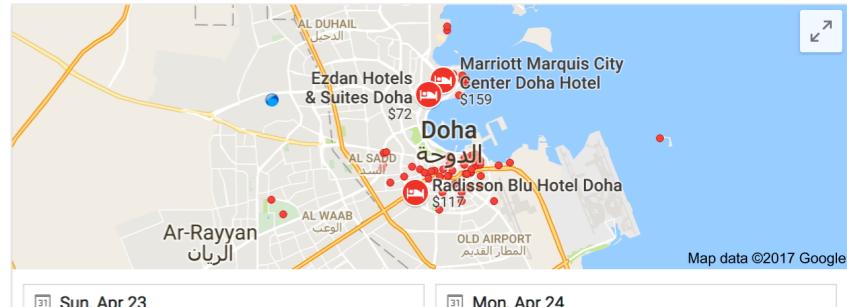
£45 - £90 Hotels up to £90.00/day Book Now

£90 - £130 Hotels up to £130.00/day Book Now

**115 Hotels in Doha - Book Great Hotels on Expedia - expedia.com**  
Ad [www.expedia.com/Doha\\_Qatar/Hotels](http://www.expedia.com/Doha_Qatar/Hotels) ▾  
Find and Compare Exclusive Deals on Hotels in Doha and Save Big!  
Packages: Save up to 20% · New Expedia + rewards · Daily Deals up to 40% Off  
Amenities: Free Wifi, Free Parking, Free Breakfast, Free Airport Transport, Spa Services, Baby Sitting Ser...

**Hilton Hotels in Doha - Book Your Doha Hotel Today - hilton.com**  
Ad [www.hilton.com/Hotels/Doha](http://www.hilton.com/Hotels/Doha) ▾  
Book on the Official Website for Best Price Guarantee and Free Wi-Fi.

**Find the Best Hotels in Doha - Never Pay Full Price on Hotels**  
Ad [ar.trivago.com/Best-Hotels/Doha](http://ar.trivago.com/Best-Hotels/Doha) ▾  
The Best Hotels in Doha. 200+ Booking Sites Compared In A Few Clicks! - trivago™

 Map data ©2017 Google

Sun, Apr 23 Mon, Apr 24 Sort by ▾

Hotel	Price	Rating	Distance	Offer	Image
Ezdan Hotels & Suites Doha	\$72	3.5 ★★★★☆ (451) · 4-star hotel	8.2 km	DEAL 17% off	
Marriott Marquis City Center Doha Hotel	\$159	4.3 ★★★★★ (306) · 5-star hotel	9.0 km		
Radisson Blu Hotel Doha	\$117	4.2 ★★★★☆ (486) · 4-star hotel	8.9 km		

More hotels About pricing

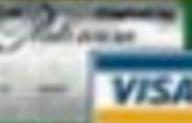
The 10 Best Doha Hotels: Places to Stay in Doha - TripAdvisor

<https://www.tripadvisor.co.uk/Hotels-g294009-Doha-Hotels.html> ▾  
Book the Best Doha Hotels on TripAdvisor. See 21521 traveller reviews, 26682 photos, and great deals.

# ADVERTISEMENTS - SECOND PRICE APPROACH

You searched for lactose intolerance.

GO TO [lactose intolerance](#)

CapitalOne<sup>®</sup>  CLICK HERE  0% intro+9.9% fixed APR  
60-sec responses

Computing  
Education & Caesar  
Entertainment  
Finance  
Health  
HomeLife  
Reference  
Shopping  
Travel

**P** Access 75% of all users now!  
Premium Listings reach 75% of all Internet users. [Sign up](#) for Premium Listings today!

- [No More Lactose Intolerance!](#)**  
Try Sammi's best whey and rice. For anyone with lactose intolerance this is it! 75% less lactose, with Acidophilus and bifidus bacteria. Buy factory direct with free shipping.  
[www.bestlifeint.com](http://www.bestlifeint.com) (Cost to advertiser: **\$0.30**)
- [Lactose Intolerance: Whey Drinks](#)**  
Lactose intolerance: Swiss Whey Powder. Delicious, natural, refreshing and easy to mix. Low in lactose. Great for beverages, cooking. 5 year shelf life. Money back guarantee. Mix to taste.  
[www.galaxymall.com](http://www.galaxymall.com) (Cost to advertiser: **\$0.07**)
- [Hock's.com Lactose Intolerance Products](#)**  
Hock's online discount pharmacy is your source for lactose intolerance products and smoking aids and more! Hock's pharmacy is quality and service at competitive prices!  
[www.hocks.com](http://www.hocks.com) (cost to advertiser: **\$0.06**)

### WIN-WIN-WIN

- ▶ Search engine company gets revenue every time somebody clicks on an ad (>90% Google's revenue)
- ▶ Users clicks on an ad only if they are interested
- ▶ Advertisers find new customers in a cost-effective way

### WIN-WIN-WIN

- ▶ Search engine company gets revenue every time somebody clicks on an ad (>90% Google's revenue)
- ▶ Users clicks on an ad only if they are interested
- ▶ Advertisers find new customers in a cost-effective way

#### Spammers' strategies to make money:

- Buy a keyword on Google and resell it for a higher price...
- Bots to click on ads of competitors...

### WIN-WIN-WIN

- ▶ Search engine company gets revenue every time somebody clicks on an ad (>90% Google's revenue)
- ▶ Users clicks on an ad only if they are interested
- ▶ Advertisers find new customers in a cost-effective way

#### Spammers' strategies to make money:

- Buy a keyword on Google and resell it for a higher price...
- Bots to click on ads of competitors...

There is a lot of research on how to stop spammers, including developing methods based on link analysis

# LECTURE 10 - LINK ANALYSIS

## BEFORE IT...

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) Read Edit View history Search Wikipedia 

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia Coordinates:  40.443322°N 79.943583°W

**Carnegie Mellon University** (Carnegie Mellon or CMU; /kɑːrni'geɪ mələn/ or /kɑːr'nɛgi' mələn/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national *U.S. News & World Report* rankings.<sup>[8]</sup> It is home to the world's first degree-granting Robotics and Drama programs,<sup>[9]</sup> as well as one of the first Computer Science departments.<sup>[10]</sup> The university spent \$242 million on research in 2015.<sup>[11]</sup>

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 19 Nobel Prize Laureates, 19 Members of the American Academy of Arts & Sciences,<sup>[12]</sup> 72 Members of the National Academies, 114 Emmy Award Winners, 43 Tony Award laureates, 7 Academy Award Winners, 12 Turing Award winners, 4 Rhodes Scholars, and one Schwarzman Scholar.<sup>[13]</sup>

**Contents** [hide]

- 1 Institutional formation
- 2 Campus
  - 2.1 Campus architecture and design
  - 2.2 Present
- 3 Admissions and enrollment
- 4 Rankings and reputation
- 5 International activities
- 6 In popular culture
- 7 Schools and divisions
  - 7.1 Libraries
  - 7.2 Collaboration with the University of Pittsburgh
- 8 Discoveries and innovation
- 9 Research
- 10 Alumni and faculty
- 11 Student life
  - 11.1 Traditions
  - 11.2 Housing
  - 11.3 Fraternities and sororities
- 12 Athletics
  - 12.1 Football
  - 12.2 Track and cross country
  - 12.3 Volleyball
  - 12.4 Cricket
- 13 See also
- 14 Notes and references
- 15 External links

Institutional formation [edit]

**Former names** Carnegie Technical Schools (1900–1912), Carnegie Institute of Technology (1912–1967), Carnegie-Mellon University (1968–1988)<sup>[1]</sup>, Carnegie Mellon University (1988–Present)

**Motto** "My heart is in the work"  
(Andrew Carnegie)

**Type** Private university

**Established** 1900 by Andrew Carnegie  
1967 (merger with Mellon Institute)

**Endowment** \$1.709 billion (2016)<sup>[2]</sup>

**President** Subra Suresh

**Provost** Farnam Jahanian<sup>[3]</sup>

**Academic staff** 1,423<sup>[4]</sup>

**Undergraduates** 6,362

**Postgraduates** 7,141

**Other students** 145

**Location** Pittsburgh, Pennsylvania, United States

**Campus** Urban, 140 acres (57 ha)<sup>[5]</sup>

**Colors** Cardinal, Black, Grey and White    

**Athletics** NCAA Division III UAA, ACHA, IRA  
17 varsity teams<sup>[6]</sup>

**Nickname** Tartans

**Mascot** Scotty the Scottish Terrier<sup>[7]</sup>

**Website** [www.cmu.edu](http://www.cmu.edu)

**Carnegie Mellon University**

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export Create a book Download as PDF Printable version

In other projects Wikimedia Commons

Languages  العربية Azərbaycanca বাংলা Български Català Čeština Dansk Deutsch Español Esperanto Euskara فارسی Français 한국어 Bahasa Indonesia Íslenska Italiano

# LECTURE 10 - LINK ANALYSIS

## BEFORE IT...

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Read [Edit](#) [View history](#) [Search Wikipedia](#) 

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia

Coordinates:  40.443322°N 79.943583°W

**Carnegie Mellon University**

**Former names** Carnegie Technical Schools (1900–1912), Carnegie Institute of Technology (1912–1967), Carnegie-Mellon University (1968–1988)<sup>[1]</sup>, Carnegie Mellon University (1988–Present)

**Motto** "My heart is in the work" (Andrew Carnegie)

**Type** Private university

**Established** 1900 by Andrew Carnegie  
1967 (merger with Mellon Institute)

**Endowment** \$1.709 billion (2016)<sup>[2]</sup>

**President** Subra Suresh

**Provost** Farnam Jahanian<sup>[3]</sup>

**Academic staff** 1,423<sup>[4]</sup>

**Undergraduates** 6,362

**Postgraduates** 7,141

**Other students** 145

**Location** Pittsburgh, Pennsylvania, United States

**Campus** Urban, 140 acres (57 ha)<sup>[5]</sup>

**Colors** Cardinal, Black, Grey and White    

**Athletics** NCAA Division III UAA, ACHA, IRA  
17 varsity teams<sup>[6]</sup>

**Nickname** Tartans

**Mascot** Scotty the Scottish Terrier<sup>[7]</sup>

**Website** [www.cmu.edu](http://www.cmu.edu)

**Carnegie Mellon University**

**Article** [Talk](#)

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national *U.S. News & World Report* rankings.<sup>[8]</sup> It is home to the world's first degree-granting Robotics and Drama programs,<sup>[9]</sup> as well as one of the first Computer Science departments.<sup>[10]</sup> The university spent \$242 million on research in 2015.<sup>[11]</sup>

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 19 Nobel Prize Laureates, 19 Members of the American Academy of Arts & Sciences,<sup>[12]</sup> 72 Members of the National Academies, 114 Emmy Award Winners, 43 Tony Award laureates, 7 Academy Award Winners, 12 Turing Award winners, 4 Rhodes Scholars, and one Schwarzman Scholar.<sup>[13]</sup>

**Contents** [hide]

- 1 Institutional formation
- 2 Campus
  - 2.1 Campus architecture and design
  - 2.2 Present
- 3 Admissions and enrollment
- 4 Rankings and reputation
- 5 International activities
- 6 In popular culture
- 7 Schools and divisions
  - 7.1 Libraries
  - 7.2 Collaboration with the University of Pittsburgh
- 8 Discoveries and innovation
- 9 Research
- 10 Alumni and faculty
- 11 Student life
  - 11.1 Traditions
  - 11.2 Housing
  - 11.3 Fraternities and sororities
- 12 Athletics
  - 12.1 Football
  - 12.2 Track and cross country
  - 12.3 Volleyball
  - 12.4 Cricket
- 13 See also
- 14 Notes and references
- 15 External links

Institutional formation [edit]

WIKIPEDIA The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Wikidata item  
Cite this page

Print/export  
Create a book  
Download as PDF  
Printable version

In other projects  
Wikimedia Commons

Languages   
[العربية](#)  
[Azərbaycanca](#)  
[ગુજરાતી](#)  
[Български](#)  
[Català](#)  
[Čeština](#)  
[Dansk](#)  
[Deutsch](#)  
[Español](#)  
[Esperanto](#)  
[Euskara](#)  
[فارسی](#)  
[Français](#)  
[한국어](#)  
[Bahasa Indonesia](#)  
[Íslenska](#)  
[Italiano](#)

# LECTURE 10 - LINK ANALYSIS

## BEFORE IT...

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Coordinates: 40.443322°N 79.943583°W

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia

**Carnegie Mellon University** (Carnegie Mellon or CMU; /kɑr̃niɡi ˈmɛlən/ or /kar niɡi ˈmɛlən/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national *U.S. News & World Report* rankings.<sup>[8]</sup> It is home to the world's first degree-granting Robotics and Drama programs,<sup>[9]</sup> as well as one of the first Computer Science departments.<sup>[10]</sup> The university spent \$242 million on research in 2015.<sup>[11]</sup>

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 19 Nobel Prize Laureates, 19 Members of the American Academy of Arts & Sciences,<sup>[12]</sup> 72 Members of the National Academies, 114 Emmy Award Winners, 43 Tony Award laureates, 7 Academy Award Winners, 12 Turing Award winners, 4 Rhodes Scholars, and one Schwarzman Scholar.<sup>[13]</sup>

**Contents** [hide]

- 1 Institutional formation
- 2 Campus
  - 2.1 Campus architecture and design
  - 2.2 Present
- 3 Admissions and enrollment
- 4 Rankings and reputation
- 5 International activities
- 6 In popular culture
- 7 Schools and divisions
  - 7.1 Libraries
  - 7.2 Collaboration with the University of Pittsburgh
- 8 Discoveries and innovation
- 9 Research
- 10 Alumni and faculty
- 11 Student life
  - 11.1 Traditions
  - 11.2 Housing
  - 11.3 Fraternities and sororities
- 12 Athletics
  - 12.1 Football
  - 12.2 Track and cross country
  - 12.3 Volleyball
  - 12.4 Cricket
- 13 See also
- 14 Notes and references
- 15 External links

**Carnegie Mellon University**

**Former names** Carnegie Technical Schools (1900–1912), Carnegie Institute of Technology (1912–1967), Carnegie-Mellon University (1968–1988)<sup>[1]</sup>, Carnegie Mellon University (1988–Present)

**Motto** "My heart is in the work"  
(Andrew Carnegie)

**Type** Private university

**Established** 1900 by Andrew Carnegie  
1967 (merger with Mellon Institute)

**Endowment** \$1.709 billion (2016)<sup>[2]</sup>

**President** Subra Suresh

**Provost** Farnam Jahanian<sup>[3]</sup>

**Academic staff** 1,423<sup>[4]</sup>

**Undergraduates** 6,362

**Postgraduates** 7,141

**Other students** 145

**Location** Pittsburgh, Pennsylvania, United States

**Campus** Urban, 140 acres (57 ha)<sup>[5]</sup>

**Colors** Cardinal, Black, Grey and White 

**Athletics** NCAA Division III UAA, ACHA, IRA  
17 varsity teams<sup>[6]</sup>

**Nickname** Tartans

**Mascot** Scotty the Scottish Terrier<sup>[7]</sup>

**Website** www.cmu.edu<sup>[8]</sup>

**Carnegie Mellon University**

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export Create a book Download as PDF Printable version

In other projects Wikimedia Commons

Languages العربية Azərbaycanca বাংলা Български Català Čeština Dansk Deutsch Español Esperanto Euskara فارسی Français 한국어 Bahasa Indonesia Íslenska Italiano

# LECTURE 10 - LINK ANALYSIS

## BEFORE IT...

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Coordinates: 40.443322°N 79.943583°W

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia

**Carnegie Mellon University** (Carnegie Mellon or CMU; /kɑr̃niɡi ˈmɛlən/ or /kar niɡi ˈmɛlən/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national *U.S. News & World Report* rankings.<sup>[8]</sup> It is home to the world's first degree-granting Robotics and Drama programs,<sup>[9]</sup> as well as one of the first Computer Science departments.<sup>[10]</sup> The university spent \$242 million on research in 2015.<sup>[11]</sup>

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 19 Nobel Prize Laureates, 19 Members of the American Academy of Arts & Sciences,<sup>[12]</sup> 72 Members of the National Academies, 114 Emmy Award Winners, 43 Tony Award laureates, 7 Academy Award Winners, 12 Turing Award winners, 4 Rhodes Scholars, and one Schwarzman Scholar.<sup>[13]</sup>

**Contents** [hide]

- 1 Institutional formation
- 2 Campus
  - 2.1 Campus architecture and design
  - 2.2 Present
- 3 Admissions and enrollment
- 4 Rankings and reputation
- 5 International activities
- 6 In popular culture
- 7 Schools and divisions
  - 7.1 Libraries
  - 7.2 Collaboration with the University of Pittsburgh
- 8 Discoveries and innovation
- 9 Research
- 10 Alumni and faculty
- 11 Student life
  - 11.1 Traditions
  - 11.2 Housing
  - 11.3 Fraternities and sororities
- 12 Athletics
  - 12.1 Football
  - 12.2 Track and cross country
  - 12.3 Volleyball
  - 12.4 Cricket
- 13 See also
- 14 Notes and references
- 15 External links

**Carnegie Mellon University**

**Former names** Carnegie Technical Schools (1900–1912), Carnegie Institute of Technology (1912–1967), Carnegie-Mellon University (1968–1988)<sup>[1]</sup>, Carnegie Mellon University (1988–Present)

**Motto** "My heart is in the work"  
(Andrew Carnegie)

**Type** Private university

**Established** 1900 by Andrew Carnegie  
1967 (merger with Mellon Institute)

**Endowment** \$1.709 billion (2016)<sup>[2]</sup>

**President** Subra Suresh

**Provost** Farnam Jahanian<sup>[3]</sup>

**Academic staff** 1,423<sup>[4]</sup>

**Undergraduates** 6,362

**Postgraduates** 7,141

**Other students** 145

**Location** Pittsburgh, Pennsylvania, United States

**Campus** Urban, 140 acres (57 ha)<sup>[5]</sup>

**Colors** Cardinal, Black, Grey and White 

**Athletics** NCAA Division III UAA, ACHA, IRA  
17 varsity teams<sup>[6]</sup>

**Nickname** Tartans

**Mascot** Scotty the Scottish Terrier<sup>[7]</sup>

**Website** www.cmu.edu<sup>[8]</sup>

**Carnegie Mellon University**

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export Create a book Download as PDF Printable version

In other projects Wikimedia Commons

Languages العربية Azərbaycanca বাংলা Български Català Čeština Dansk Deutsch Español Esperanto Euskara فارسی Français 한국어 Bahasa Indonesia Íslenska Italiano

Institutional formation [edit]

# LECTURE 10 - LINK ANALYSIS

## BEFORE IT...

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Coordinates: 40.443322°N 79.943583°W

**Carnegie Mellon University**

From Wikipedia, the free encyclopedia

**Carnegie Mellon University** (Carnegie Mellon or CMU; /kɑr̃niɡi ˈmɛlən/ or /kar niɡi ˈmɛlən/) is a private research university in Pittsburgh, Pennsylvania.

Founded in 1900 by Andrew Carnegie as the Carnegie Technical Schools, the university became the Carnegie Institute of Technology in 1912 and began granting four-year degrees. In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University.

The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national *U.S. News & World Report* rankings.<sup>[8]</sup> It is home to the world's first degree-granting Robotics and Drama programs,<sup>[9]</sup> as well as one of the first Computer Science departments.<sup>[10]</sup> The university spent \$242 million on research in 2015.<sup>[11]</sup>

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and over 5,000 faculty and staff. Past and present faculty and alumni include 19 Nobel Prize Laureates, 19 Members of the American Academy of Arts & Sciences,<sup>[12]</sup> 72 Members of the National Academies, 114 Emmy Award Winners, 43 Tony Award laureates, 7 Academy Award Winners, 12 Turing Award winners, 4 Rhodes Scholars, and one Schwarzman Scholar.<sup>[13]</sup>

**Contents** [hide]

- 1 Institutional formation
- 2 Campus
  - 2.1 Campus architecture and design
  - 2.2 Present
- 3 Admissions and enrollment
- 4 Rankings and reputation
- 5 International activities
- 6 In popular culture
- 7 Schools and divisions
  - 7.1 Libraries
  - 7.2 Collaboration with the University of Pittsburgh
- 8 Discoveries and innovation
- 9 Research
- 10 Alumni and faculty
- 11 Student life
  - 11.1 Traditions
  - 11.2 Housing
  - 11.3 Fraternities and sororities
- 12 Athletics
  - 12.1 Football
  - 12.2 Track and cross country
  - 12.3 Volleyball
  - 12.4 Cricket
- 13 See also
- 14 Notes and references
- 15 External links

**Institutional formation** [edit]

**Carnegie Mellon University**

Former names Carnegie Technical Schools (1900–1912), Carnegie Institute of Technology (1912–1967), Carnegie-Mellon University (1968–1988)<sup>[1]</sup>, Carnegie Mellon University (1988–Present)

Motto "My heart is in the work"  
(Andrew Carnegie)

Type Private university

Established 1900 by Andrew Carnegie  
1967 (merger with Mellon Institute)

Endowment \$1.709 billion (2016)<sup>[2]</sup>

President Subra Suresh

Provost Farnam Jahanian<sup>[3]</sup>

Academic staff 1,423<sup>[4]</sup>

Undergraduates 6,362

Postgraduates 7,141

Other students 145

Location Pittsburgh, Pennsylvania, United States

Campus Urban, 140 acres (57 ha)<sup>[5]</sup>

Colors Cardinal, Black, Grey and White █ █ █ █

Athletics NCAA Division III UAA, ACHA, IRA  
17 varsity teams<sup>[6]</sup>

Nickname Tartans

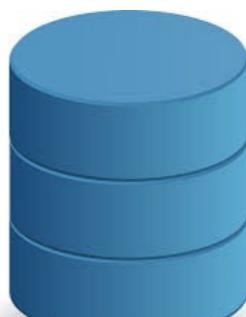
Mascot Scotty the Scottish Terrier<sup>[7]</sup>

Website [www.cmu.edu](http://www.cmu.edu)

# LECTURE 10 - LINK ANALYSIS

## REAL LIFE DOCUMENT ANALYSIS

The screenshot shows a Wikipedia article page for Carnegie Mellon University. The main content is highlighted with a green border. The sidebar contains a detailed table of contents with 15 numbered sections. The main text discusses the university's history, research, and various programs.



**INDEX1: DOCUMENT TITLES**



**INDEX2: DOCUMENT ABSTRACT**



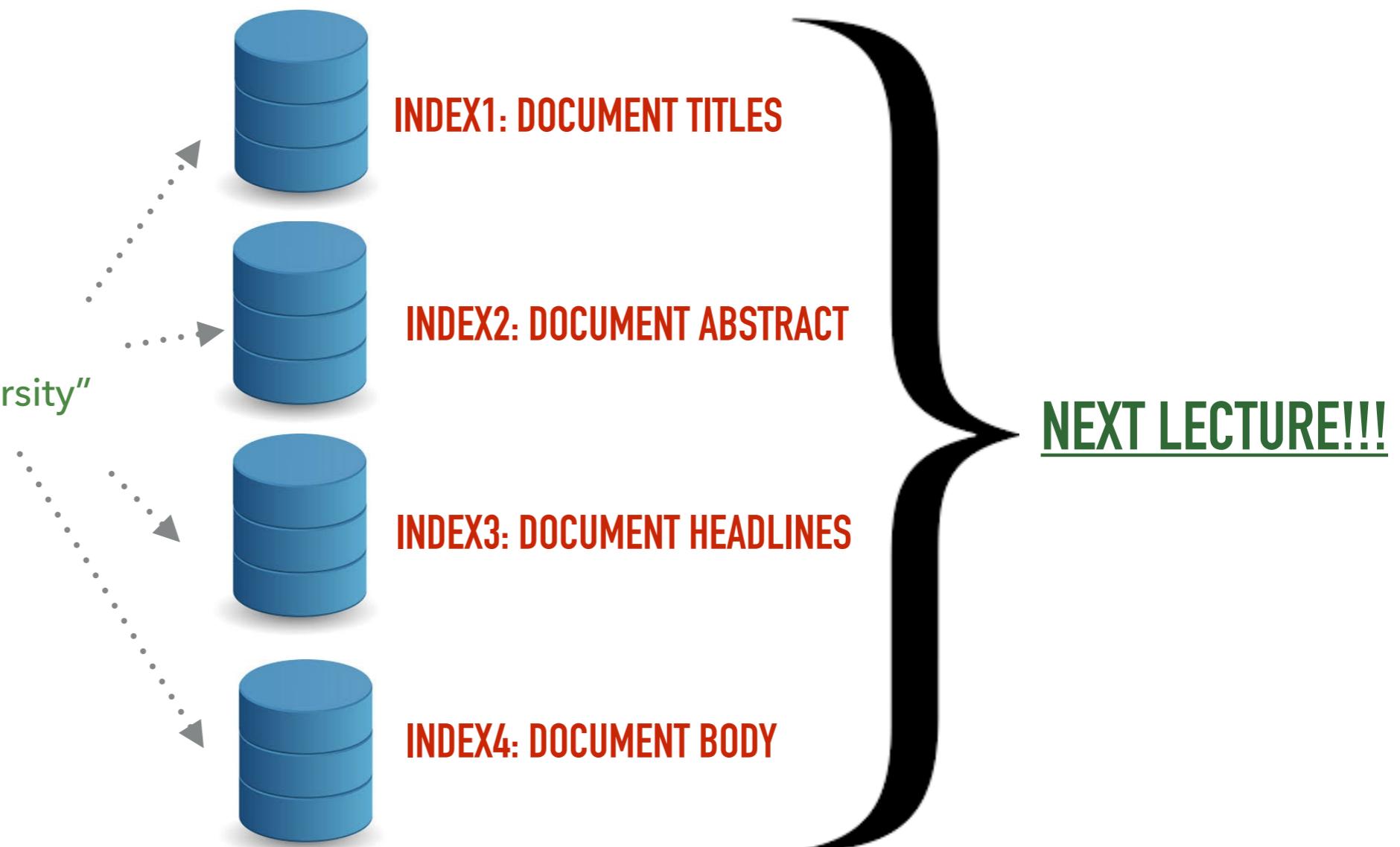
**INDEX3: DOCUMENT HEADLINES**



**INDEX4: DOCUMENT BODY**

# YOUR QUERY BEHIND THE SCENES

Query: "computer science university"



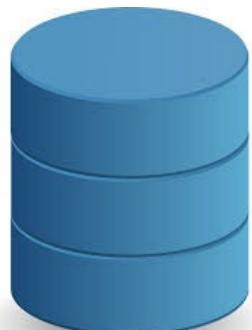
# LINK ANALYSIS

- ▶ First use in search engines:
  - ▶ <a href="<http://www.acm.org/jacm>"> Journal of the ACM </a>
  - ▶ <a href="<http://www.ibm.com>"> Good site to buy computers and related stuff </a>
  - ▶ We all fear the <a href="<http://www.ibm.com>">Big Blue </a>
  - ▶ I just found this amazing <a href="<http://www.yahoo.com>">Internet Portal</a>

**CAN WE EXPLORE THE ANCHOR TEXT TO RETRIEVE BETTER?**

# LINK ANALYSIS

- ▶ First use in search engines:
  - ▶ <a href="<http://www.acm.org/jacm>"> Journal of the ACM </a>
  - ▶ <a href="<http://www.ibm.com>"> Good site to buy computers and related stuff </a>
  - ▶ We all fear the <a href="<http://www.ibm.com>">Big Blue </a>
  - ▶ I just found this amazing <a href="<http://www.yahoo.com>">Internet Portal</a>



**INDEX5: ANCHOR TEXT**

“computer” ..... ➤ “ibm.com”

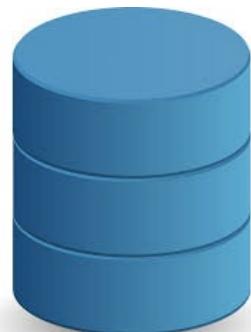
“big” ..... ➤ “ibm.com”

“blue” ..... ➤ “ibm.com”

# LINK ANALYSIS

## ► Side effects:

- IBM can be found searching for its nickname “Big Blue”, even though there is no single mention of “big blue” in its website
- Anchor text is often a better description of a page’s content than the page itself



**INDEX5: ANCHOR TEXT**

“computer” ..... ➤ “ibm.com”

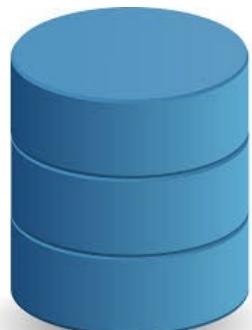
“big” ..... ➤ “ibm.com”

“blue” ..... ➤ “ibm.com”

# LINK ANALYSIS

## ► Side effects:

- ▶ IBM can be found searching for its nickname “Big Blue”, even though there is no single mention of “big blue” in its website
- ▶ **Anchor text is often a better description of a page’s content than the page itself**
- ▶ Allows orchestrated campaigns against specific sites:
  - ▶ Google Bomb – Queries like “evil empire”, “who is a failure”



**INDEX5: ANCHOR TEXT**

“computer” ..... ➤ “ibm.com”

“big” ..... ➤ “ibm.com”

“blue” ..... ➤ “ibm.com”

## LECTURE 10 - LINK ANALYSIS

# LINK ANALYSIS - PARENTHESIS

big blue

Web Images News

73,200,000 RESULTS Date Language Region

**The Big Blue (1988) - IMDb**  
[www.imdb.com/title/tt0095250](http://www.imdb.com/title/tt0095250) ★★★★ 7.6/10 · 39K ratings · Adventure/Drama/Sport · R  
Title: The Big Blue (1988) 7.6 /10. Want to share IMDb's rating on your own site? Use the HTML below. You must be a registered ... Blue is the Title;

**Big Blue View - Official Site**  
[www.bigblueview.com](http://www.bigblueview.com)  
Your best source for quality New York Giants news, rumors, analysis, stats and scores from the fan perspective.

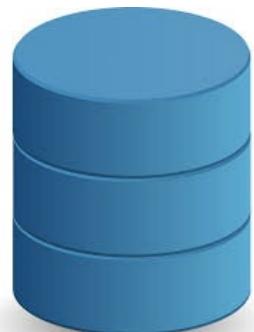
**Bigblue Dive Lights**  
[www.bigbluedivelights.com/home.asp](http://www.bigbluedivelights.com/home.asp)  
Please Select Your Region ... Please Select Your Region + -

**Images of big blue**  
[bing.com/images](http://bing.com/images)  


See more images of big blue

**BigBlueInteractive - New York Giants News and Discussion**  
[www.bigblueinteractive.com](http://www.bigblueinteractive.com)  
For 22 years, BBI has been the premier fan site for news, analysis, rumors, and discussion of the New York Giants, including NFL Draft, free agency, & more.

**IBM - Official Site**  
[www.ibm.com](http://www.ibm.com)  
For more than a century IBM has been dedicated to every client's success and to creating innovations that matter for the world



**INDEX5: ANCHOR TEXT**

"computer" ..... ➤ "ibm.com"

"big" ..... ➤ "ibm.com"

"blue" ..... ➤ "ibm.com"

## LECTURE 10 - LINK ANALYSIS

# LINK ANALYSIS - PARENTHESIS

 big blue 

**Web** Images News

73,200,000 RESULTS Date ▾ Language ▾ Region ▾

**The Big Blue (1988) - IMDb**  
[www.imdb.com/title/tt0095250](http://www.imdb.com/title/tt0095250)   
★★★★★ 7.6/10 · 39K ratings · Adventure/Drama/Sport · R  
Title: The Big Blue (1988) 7.6 /10. Want to share IMDb's rating on your own site? Use the HTML below. You must be a registered ... Blue is the Title;

**Big Blue View - Official Site**  
[www.bigblueview.com](http://www.bigblueview.com)   
Your best source for quality New York Giants news, rumors, analysis, stats and scores from the fan perspective.

**Bigblue Dive Lights**  
[www.bigbluedivelights.com/home.asp](http://www.bigbluedivelights.com/home.asp)  
Please Select Your Region ... Please Select Your Region + -

**Images of big blue**  
[bing.com/images](http://bing.com/images)  
  
See more images of big blue

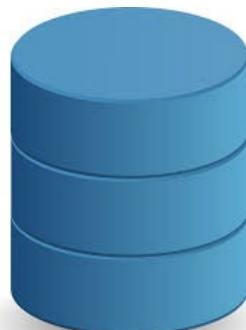
**BigBlueInteractive - New York Giants News and Discussion**  
[www.bigblueinteractive.com](http://www.bigblueinteractive.com)   
For 22 years, BBI has been the premier fan site for news, analysis, rumors, and discussion of the New York Giants, including NFL Draft, free agency, & more.

**IBM - Official Site**  
[www.ibm.com](http://www.ibm.com)   
For more than a century IBM has been dedicated to every client's success and to creating innovations that matter for the world

 evil empire 

**Web** Images News

Your country or region requires a strict Bing SafeSearch setting, which filters out results that might return adult content. To learn more about SafeSearch requirements in [your country or region](#), see [How Bing delivers search results](#).



**INDEX5: ANCHOR TEXT**

"computer" ..... ➤ "ibm.com"

"big" ..... ➤ "ibm.com"

"blue" ..... ➤ "ibm.com"

# LINK ANALYSIS - PARENTHESIS

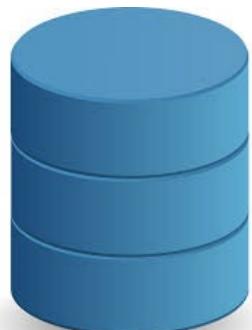
## Bush – Tops For “Who Is A Failure” On Google

Danny Sullivan on February 21, 2008 at 6:07 pm

The screenshot shows the Google search interface with the query 'who is a failure' entered into the search bar. Below the search bar, a blue header bar indicates the search is for 'Web'. The main results area shows a single search result for 'President of the United States - George W. Bush'. The result includes a snippet of text from the Encarta Encyclopedia article, a link to the original source ('www.whitehouse.gov/president/'), and options for 'Cached' and 'Similar pages'.

### ANOTHER GOOGLE BOMB

For years, US President George W. Bush was tops on Google for searches on [miserable failure](#) until [last year's Googlebomb "fix"](#) solved his problem on Google. But Bush is back, sort of.

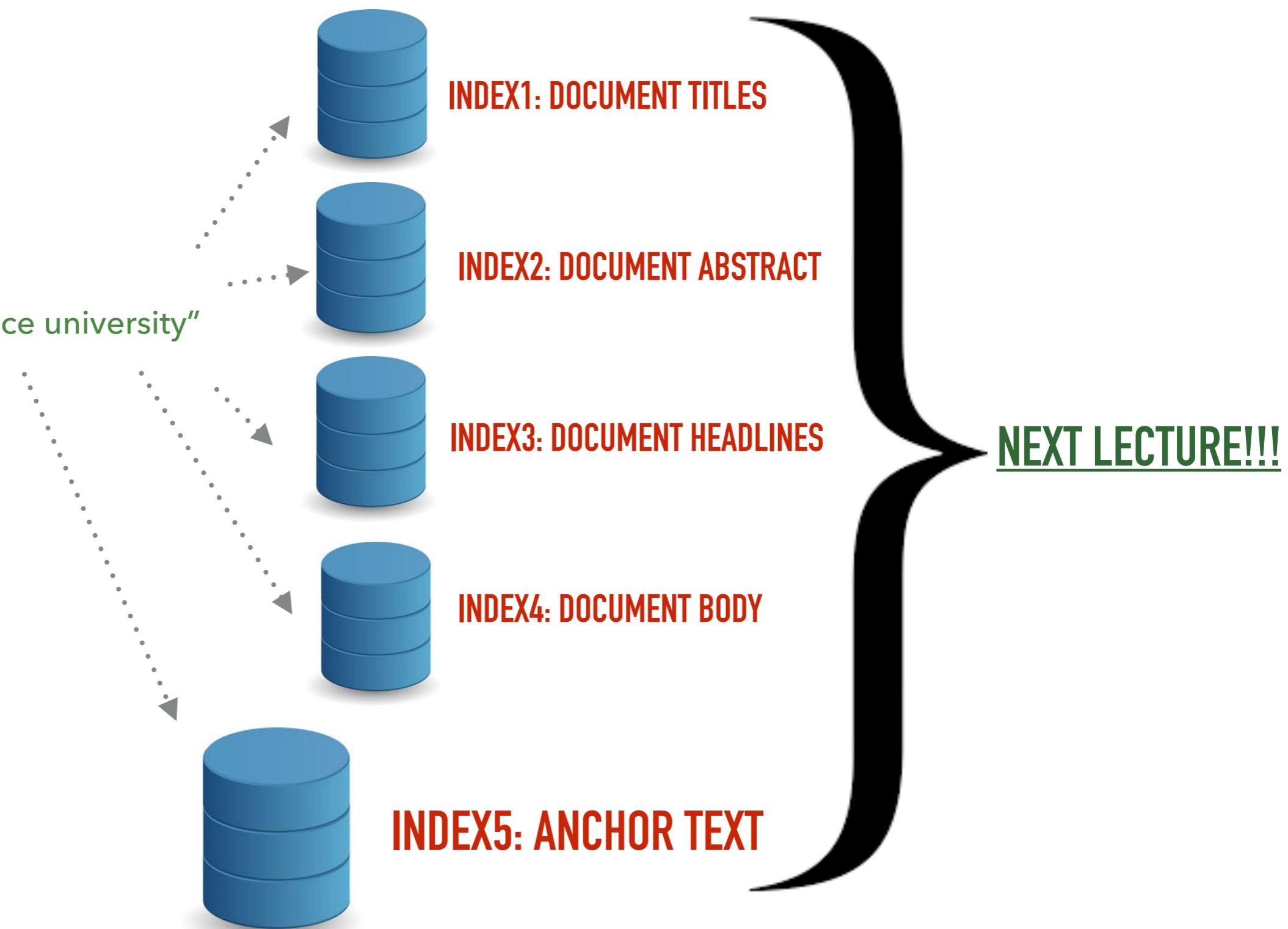


**INDEX5: ANCHOR TEXT**

“who”	.....► “whitehouse.gov/president”
“is”	.....► “whitehouse.gov/president”
“failure”	.....► “whitehouse.gov/president”

# YOUR QUERY BEHIND THE SCENES

Query: "computer science university"



# LINK ANALYSIS

- ▶ Observations:
  - ▶ Surrounding anchor text is often usable in the same manner:
    - There is a good discussion of traffic rules in Qatar  
`<a>here</a>`.
  - ▶ IDF plays an important rule again to decrease importance of common text as “here”, “click”, “link”
  - ▶ Current search engines assign a **substantial weighting** to anchor text terms

**APART FROM THAT, LINKS CAN BE USED AS A PROXY FOR QUALITY...**

# LINK ANALYSIS

- ▶ There is a good discussion of traffic rules in Qatar [here](#).



# LINK ANALYSIS

- ▶ There is a good discussion of traffic rules in Qatar [here](#).



- ▶ Parenthesis:
  - ▶ Power Laws again: Very few websites are highly cited!
  - ▶ Number of pages with in-degree  $i$  proportional to  $1 / i^a$
  - ▶  $a \sim= 2.1$

# LINK ANALYSIS

- ▶ There is a good discussion of traffic rules in Qatar [here](#).



- ▶ Inspiration used here comes from citation analysis in the **scientific literature**. Important numbers for researchers...
- ▶ Co-citation similarity: similarity of two articles comparing the overlap of citations

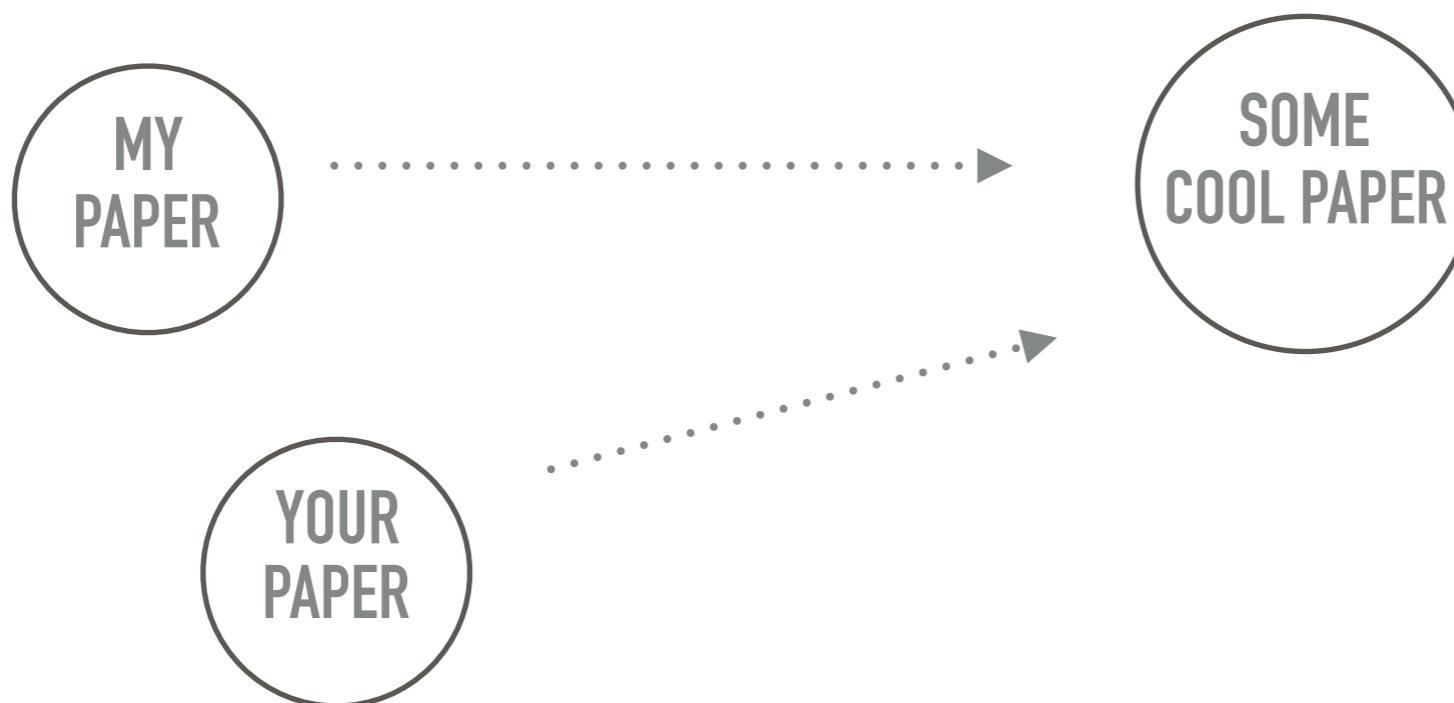
# LINK ANALYSIS

- ▶ Can we use the citation graph to calculate quality?



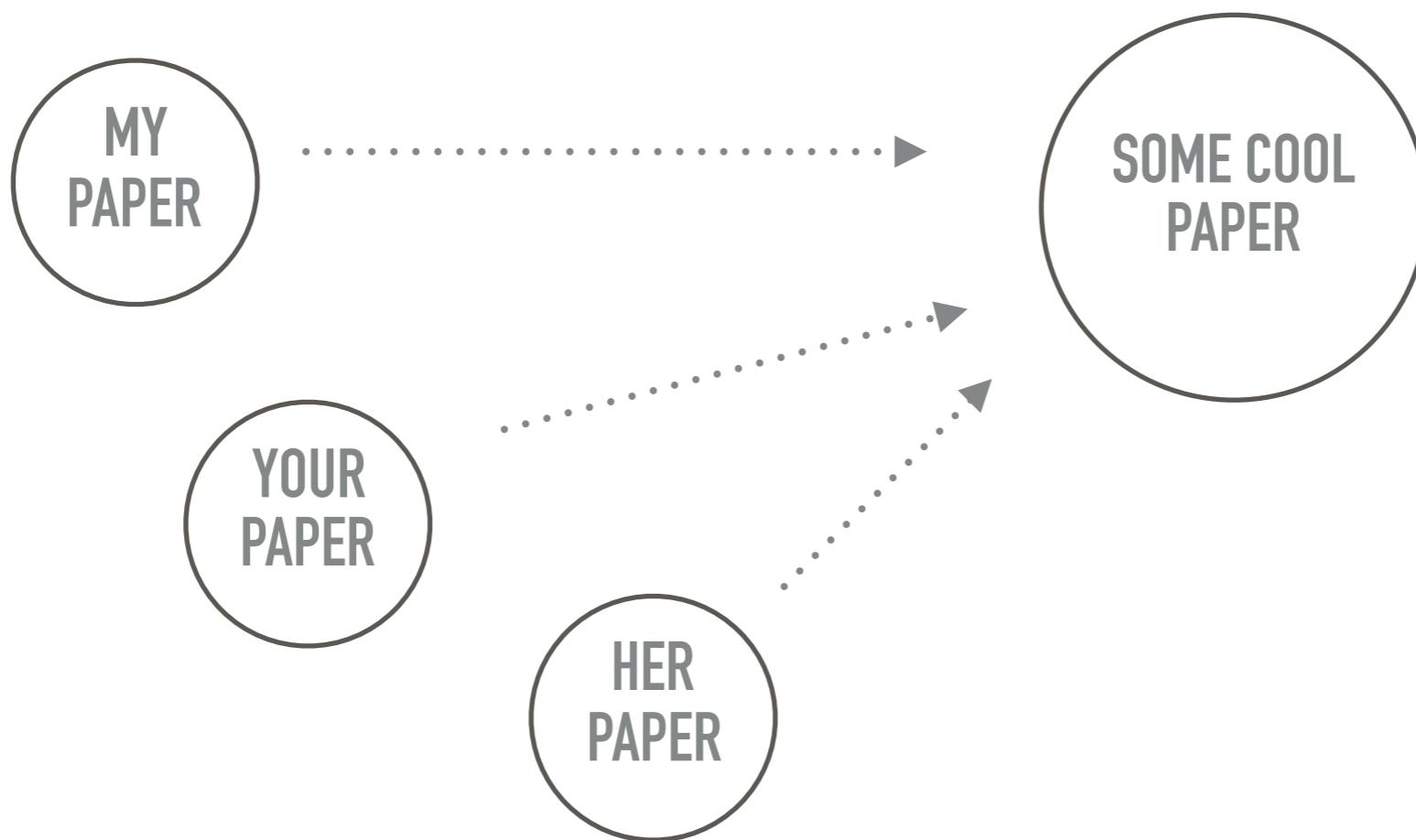
# LINK ANALYSIS

- ▶ Can we use the citation graph to calculate quality?



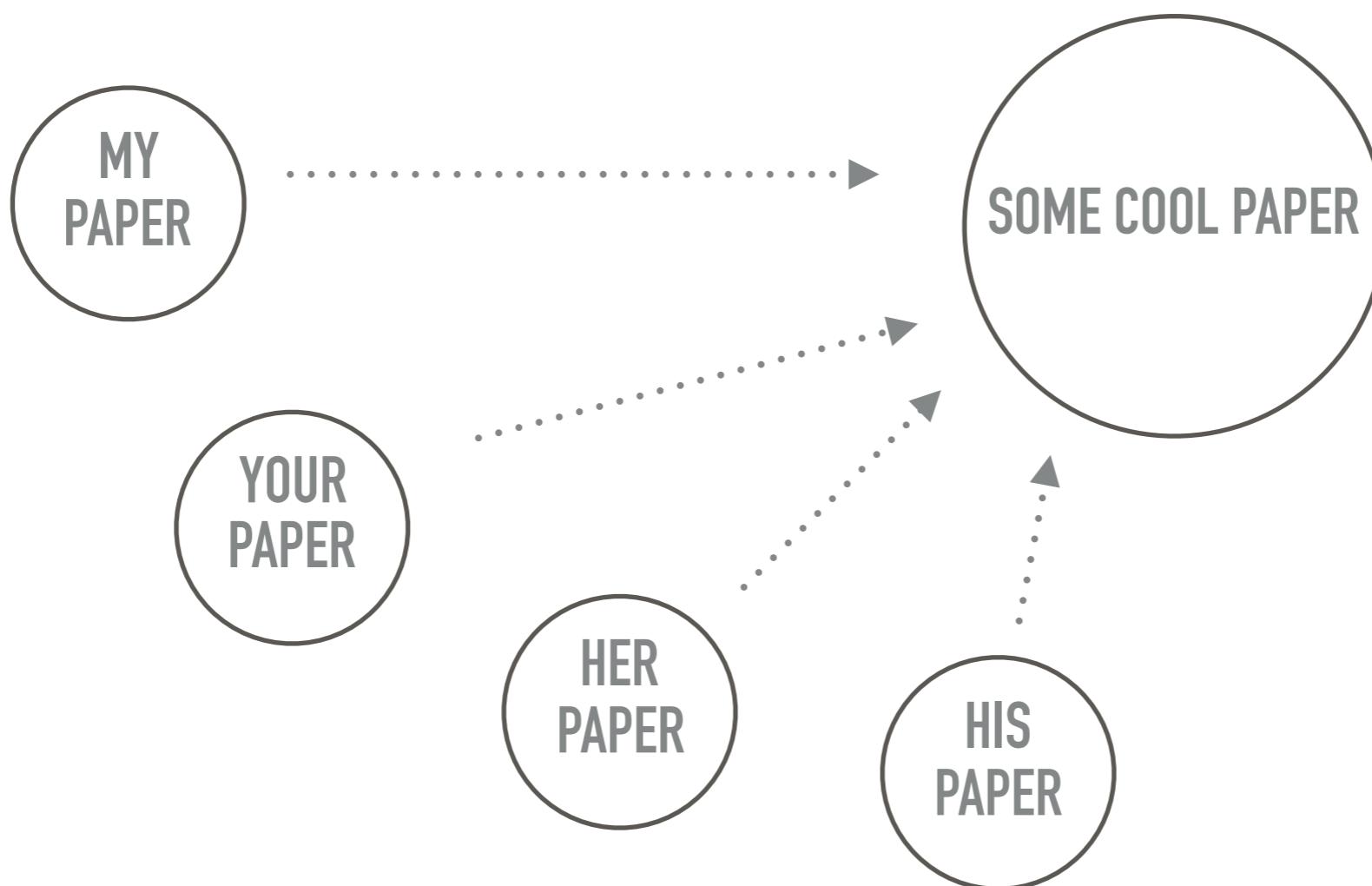
# LINK ANALYSIS

- ▶ Can we use the citation graph to calculate quality?



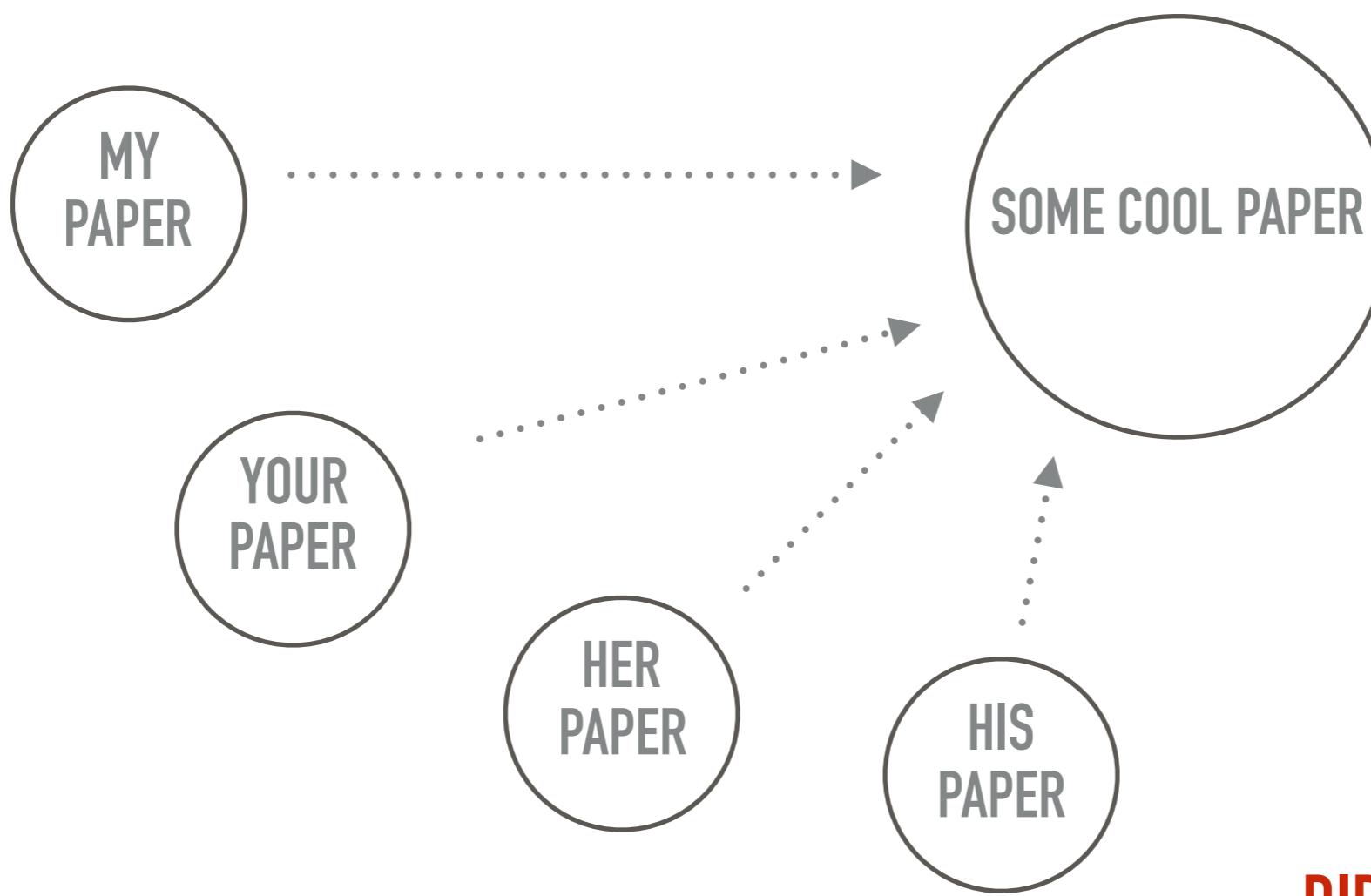
# LINK ANALYSIS

- ▶ Can we use the citation graph to calculate quality?



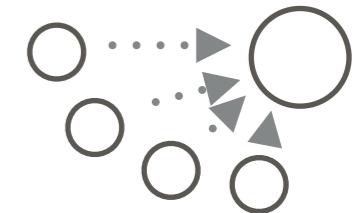
## LINK ANALYSIS

- ▶ Can we use the citation graph to calculate quality?

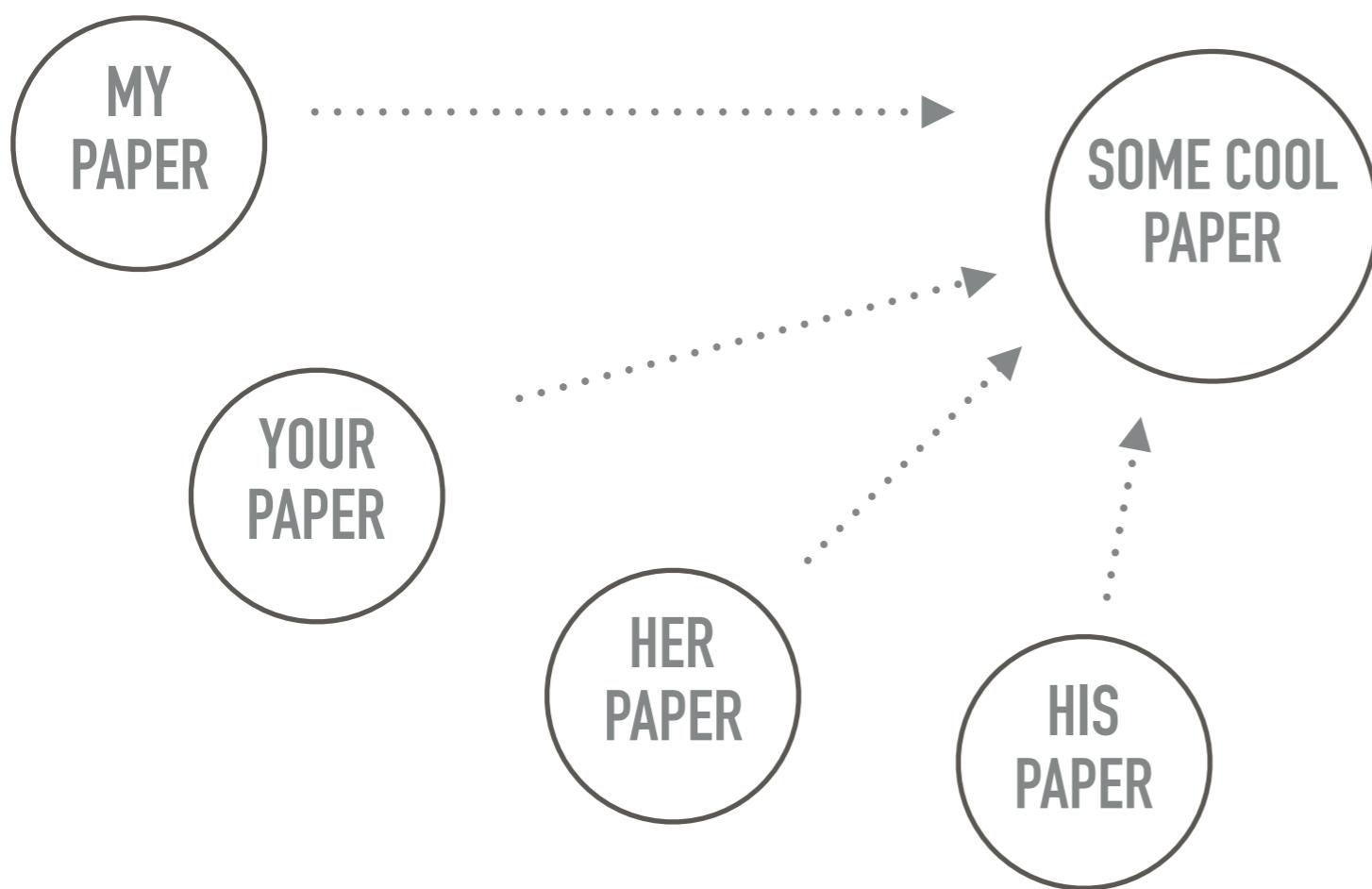


**DOES IT WORK?  
DID WE GET RID OF SPAM?**

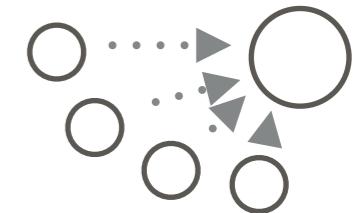
# LINK ANALYSIS - PAGE RANK



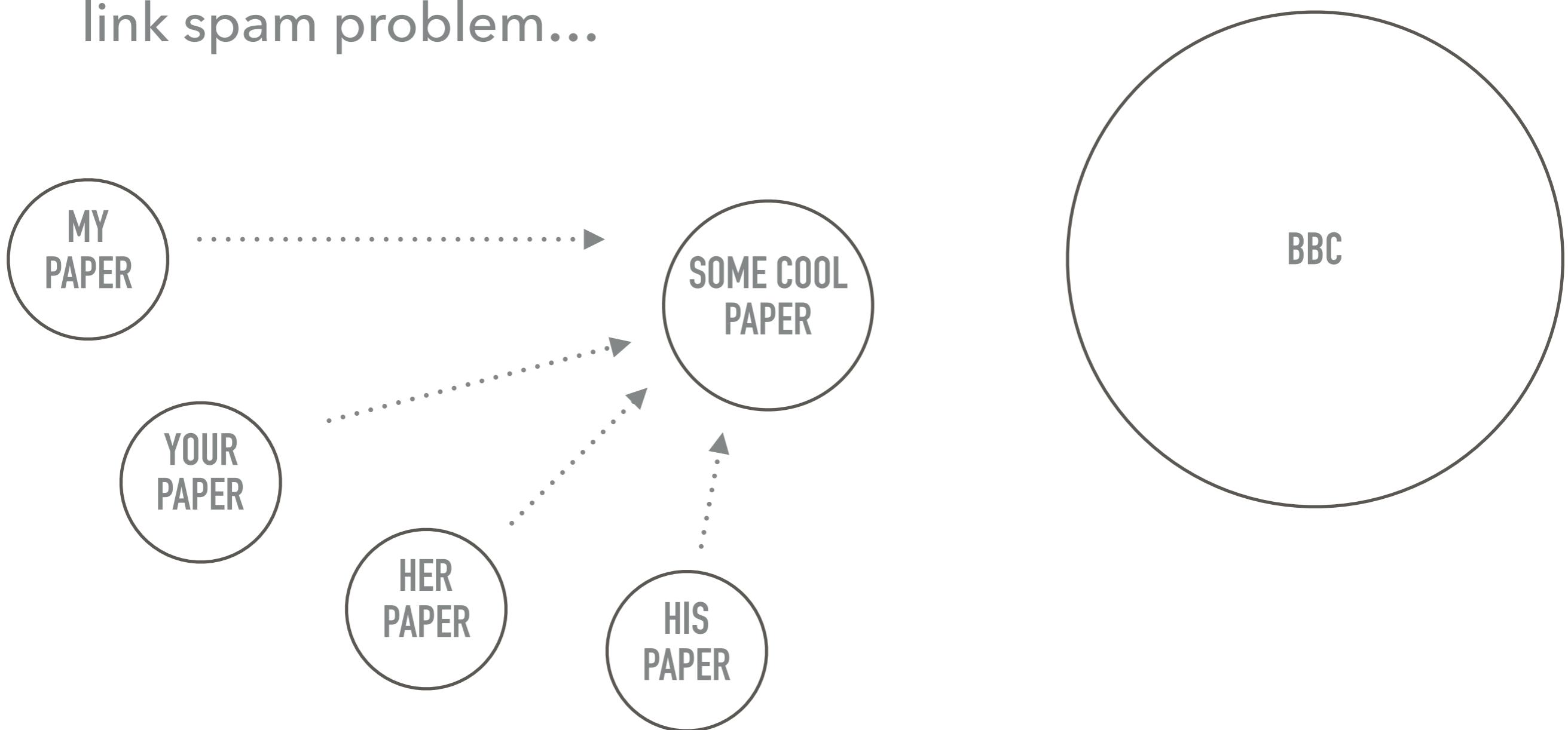
- ▶ A weighted version of citation frequency can cope with link spam problem...



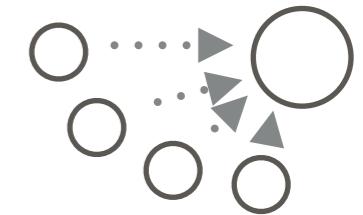
# LINK ANALYSIS - PAGE RANK



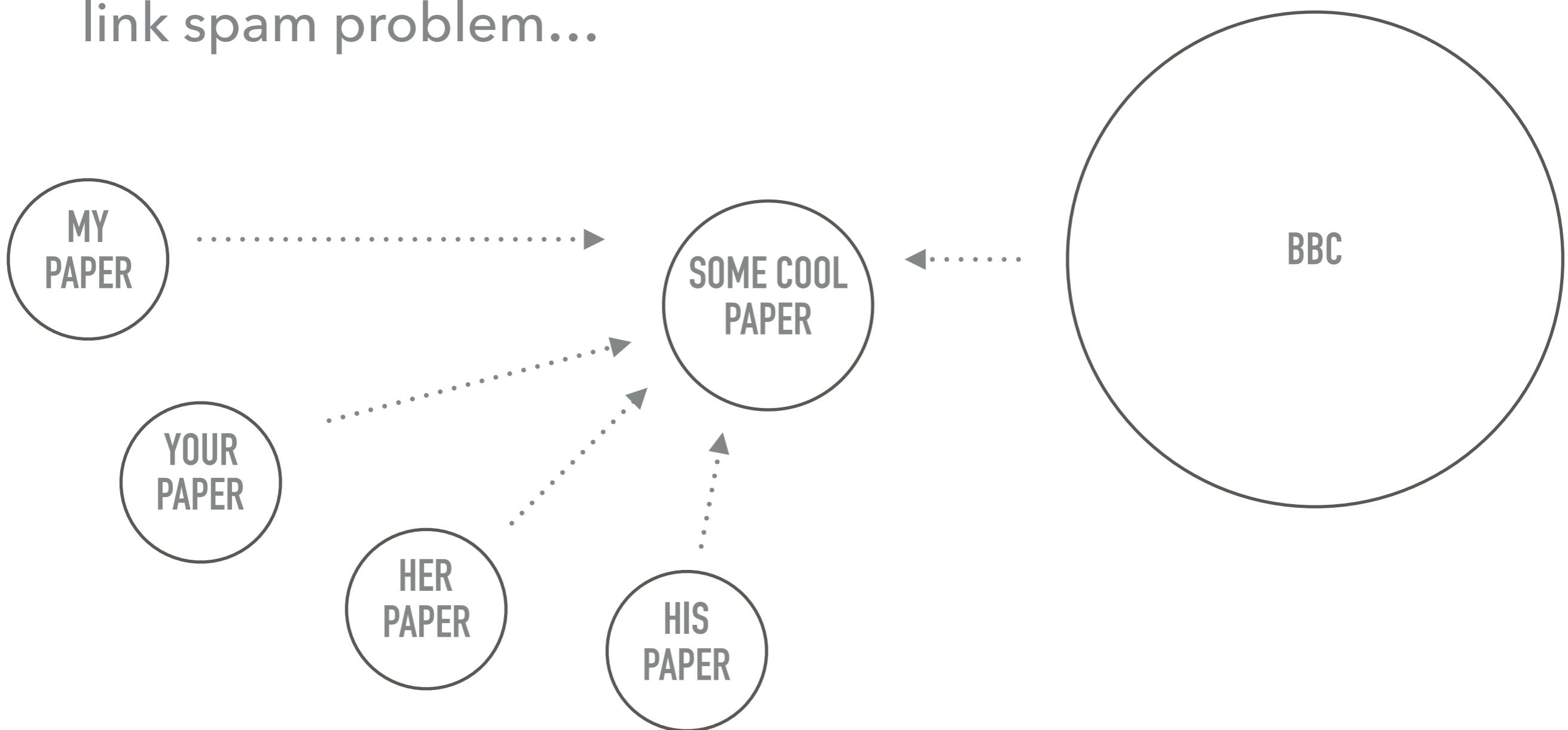
- ▶ A weighted version of citation frequency can cope with link spam problem...



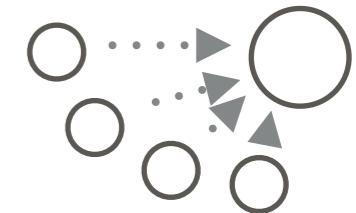
# LINK ANALYSIS - PAGE RANK



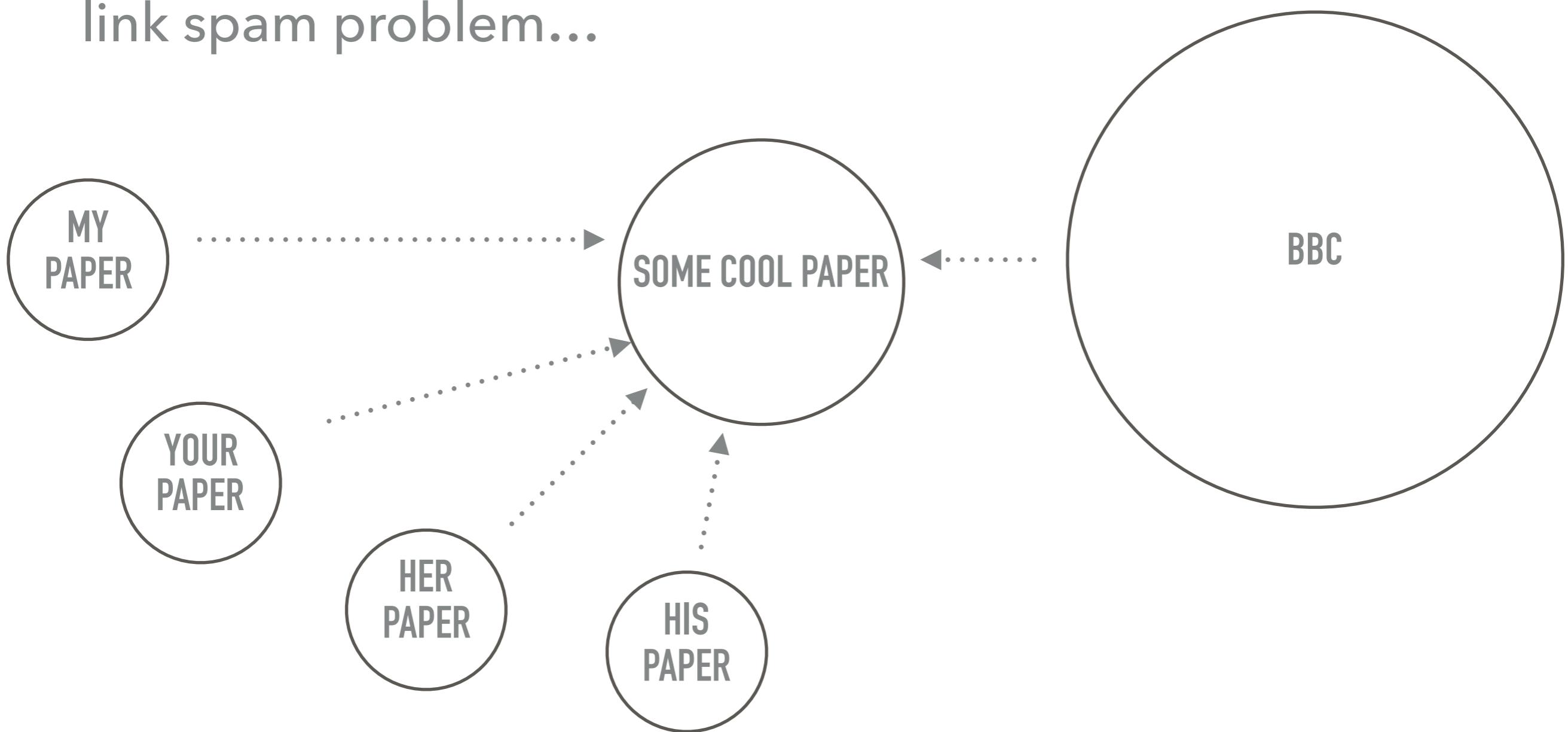
- ▶ A weighted version of citation frequency can cope with link spam problem...



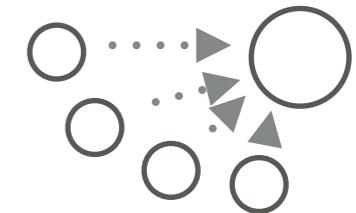
# LINK ANALYSIS - PAGE RANK



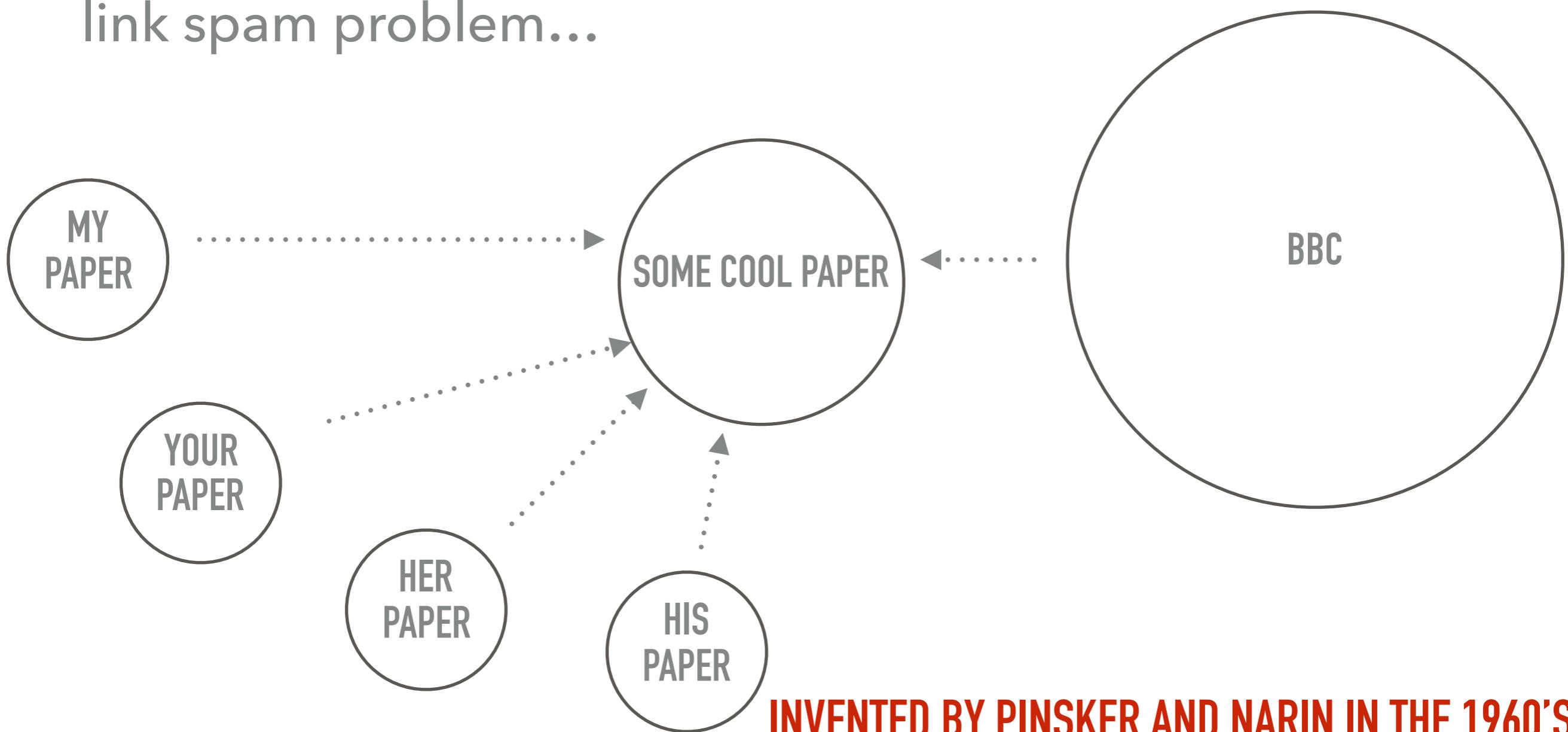
- ▶ A weighted version of citation frequency can cope with link spam problem...



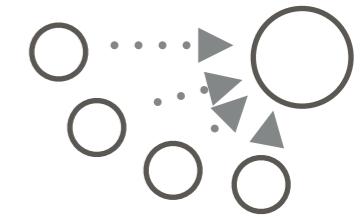
# LINK ANALYSIS - PAGE RANK



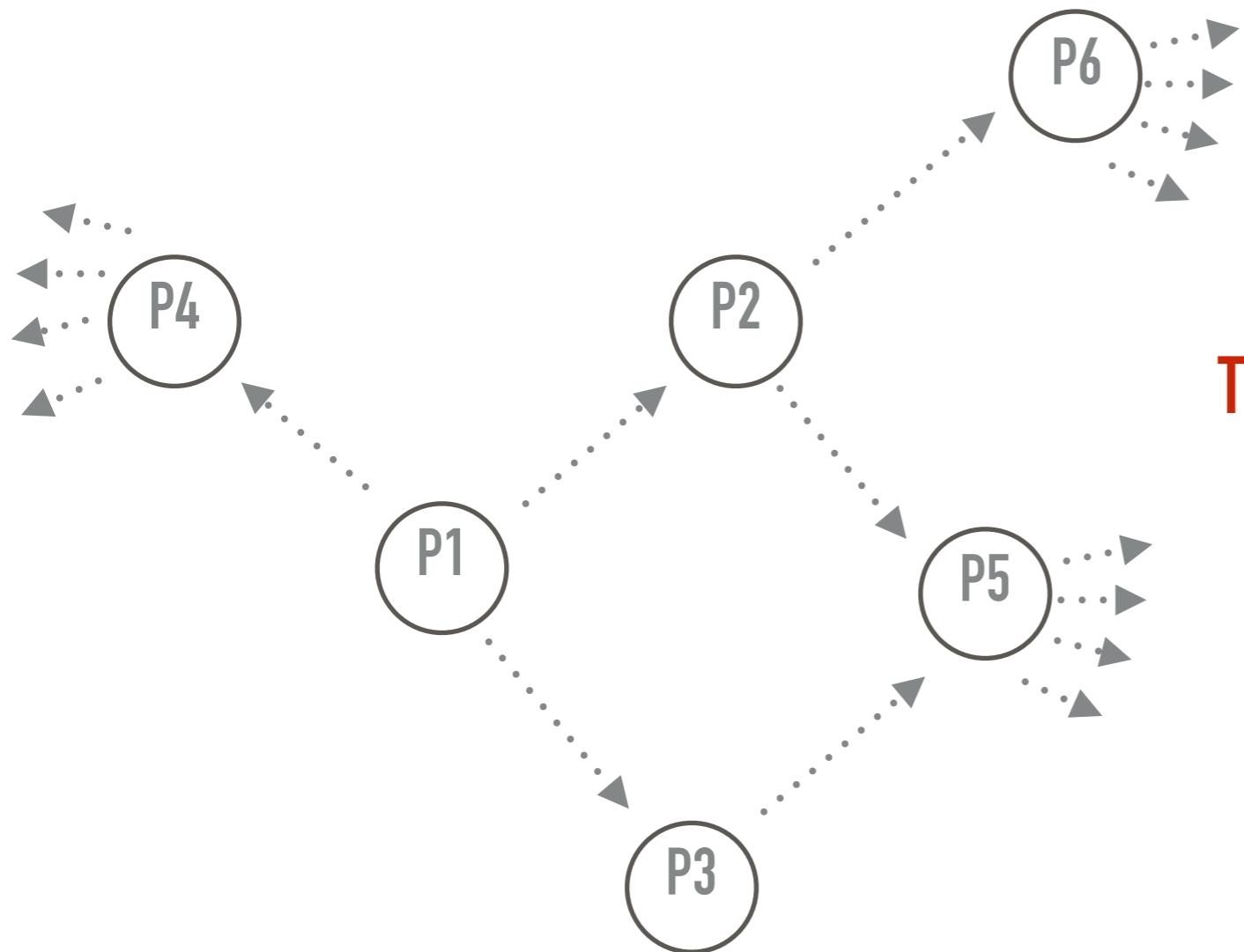
- ▶ A weighted version of citation frequency can cope with link spam problem...



# LINK ANALYSIS - PAGE RANK

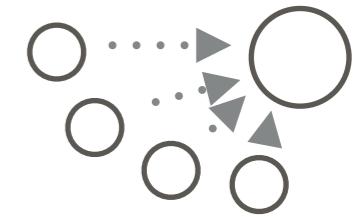


- ▶ The Web Surfer random walking on the Web

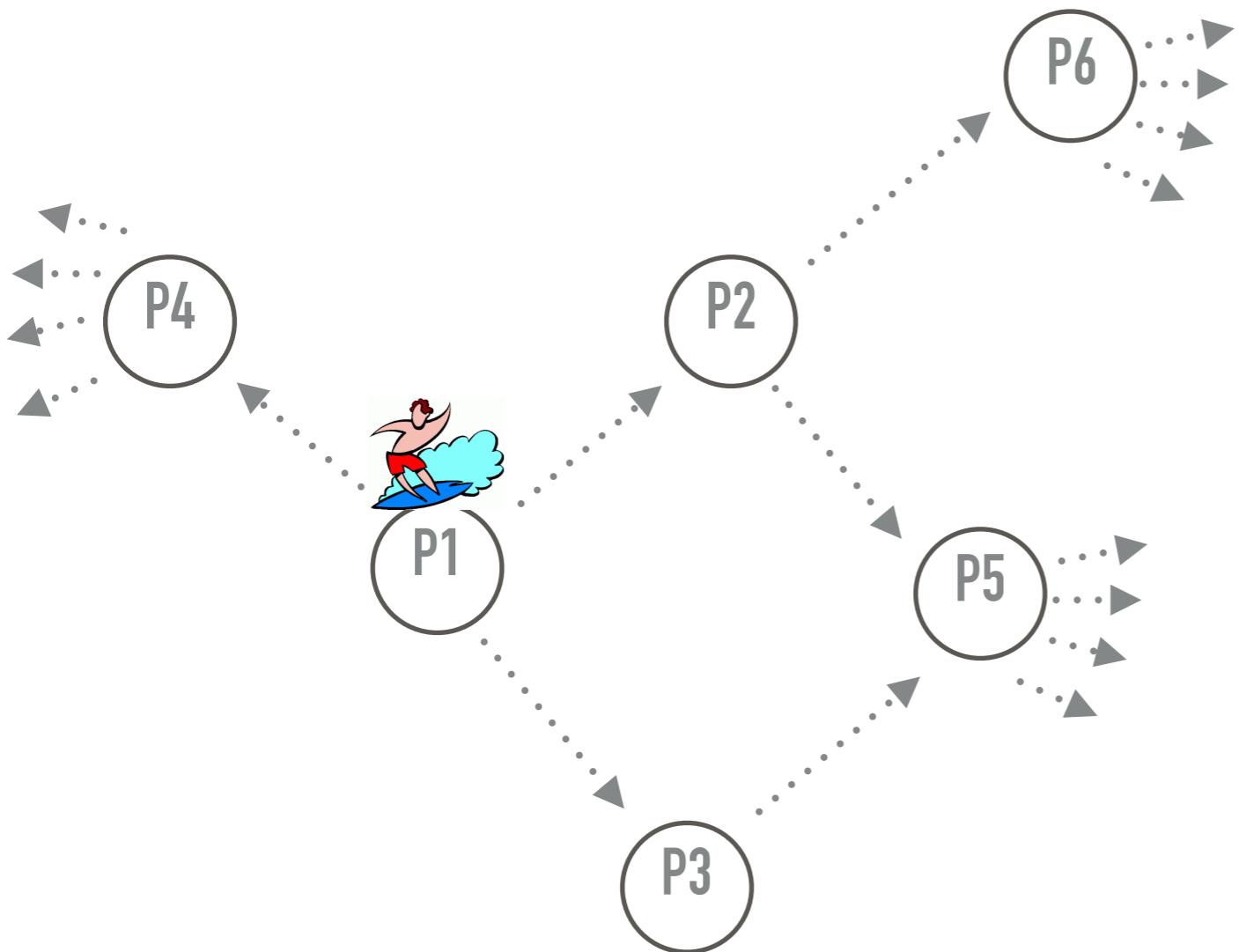


**THAT IS A TINY PIECE OF THE WEB GRAPH**

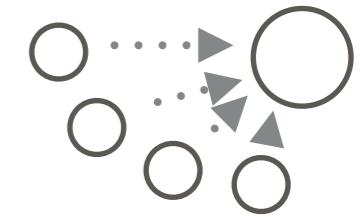
# LINK ANALYSIS - PAGE RANK



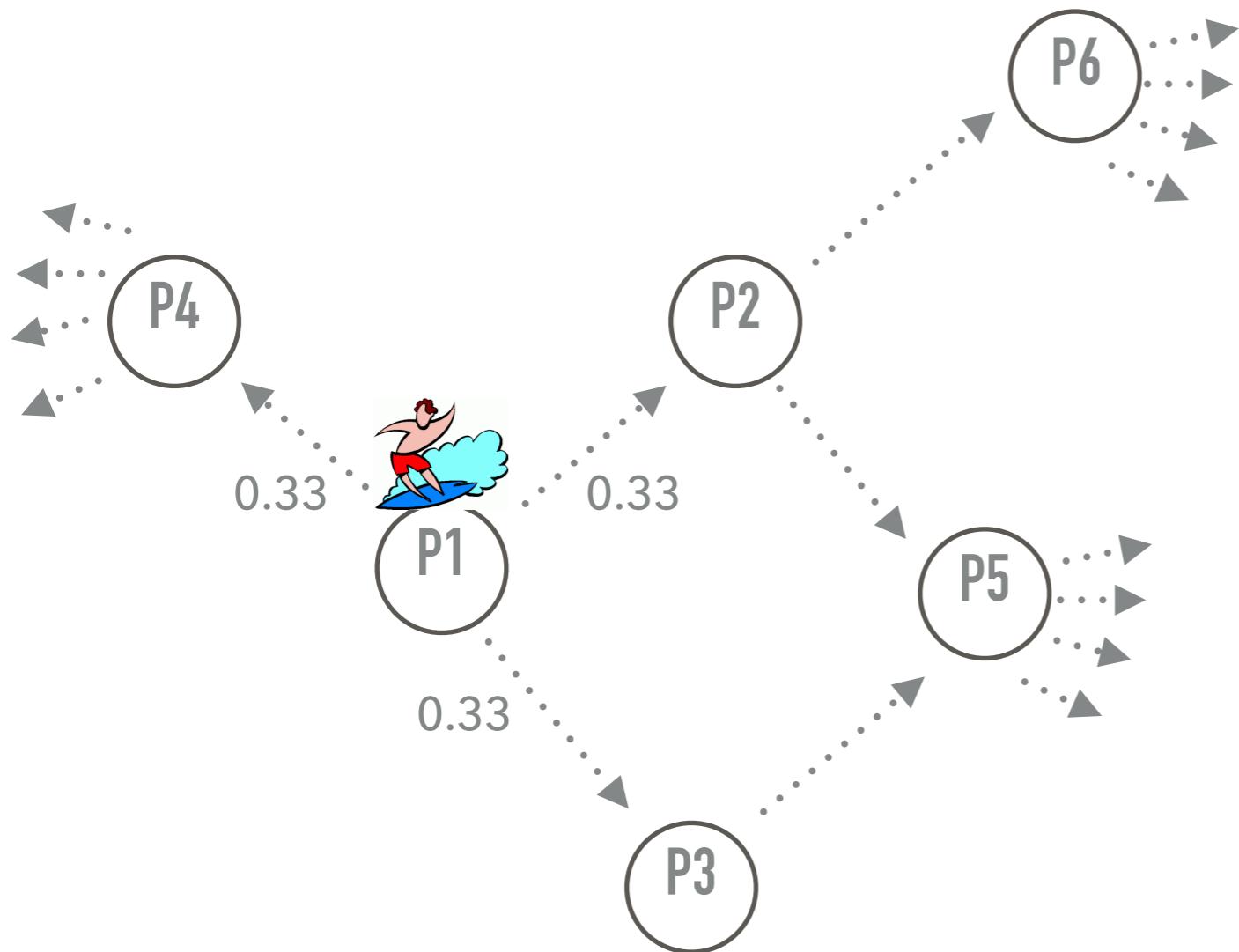
- ▶ The Web Surfer random walking on the Web



# LINK ANALYSIS - PAGE RANK

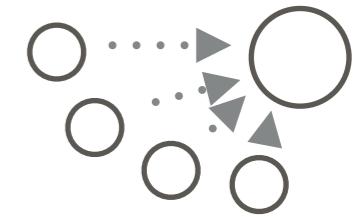


- ▶ The Web Surfer random walking on the Web

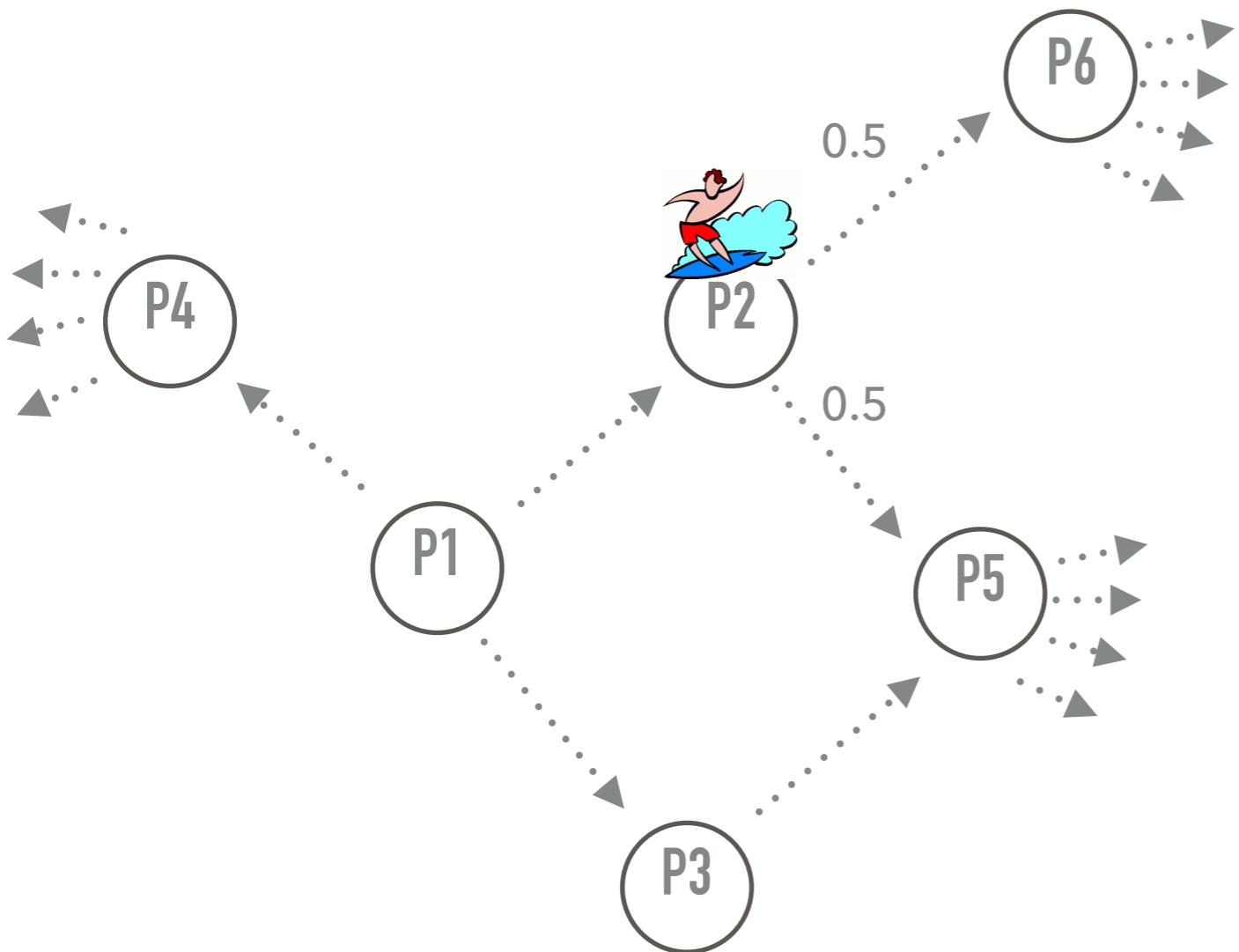


$$\forall i, \sum_{j=1}^N P_{ij} = 1.$$

# LINK ANALYSIS - PAGE RANK

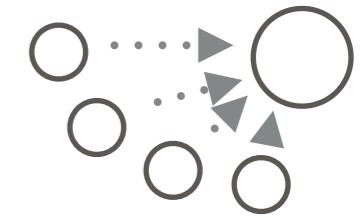


- ▶ The Web Surfer random walking on the Web

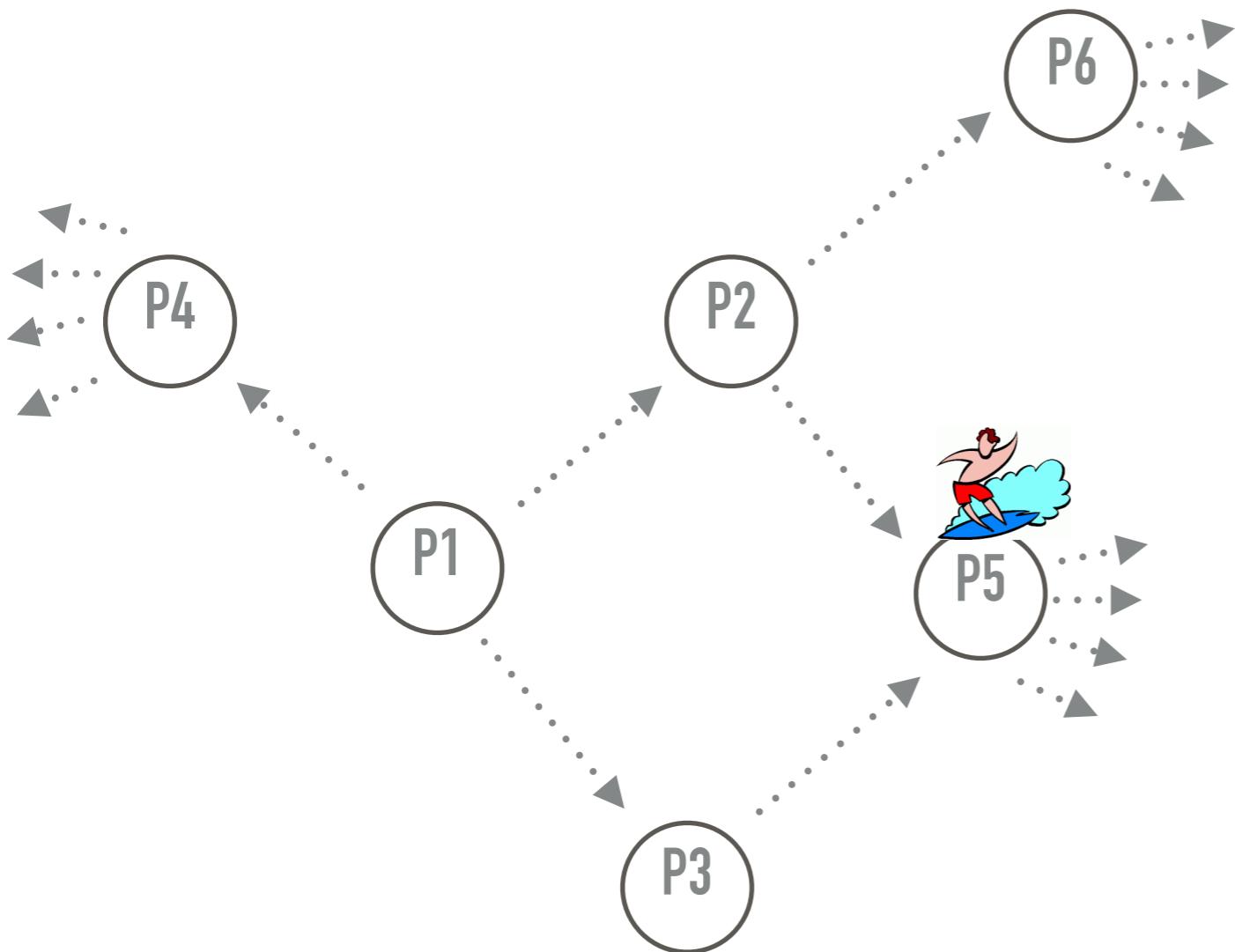


$$\forall i, \sum_{j=1}^N P_{ij} = 1.$$

# LINK ANALYSIS - PAGE RANK



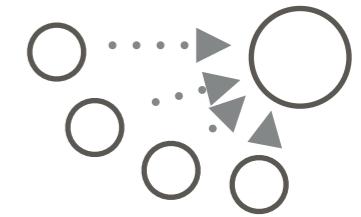
- ▶ The Web Surfer random walking on the Web



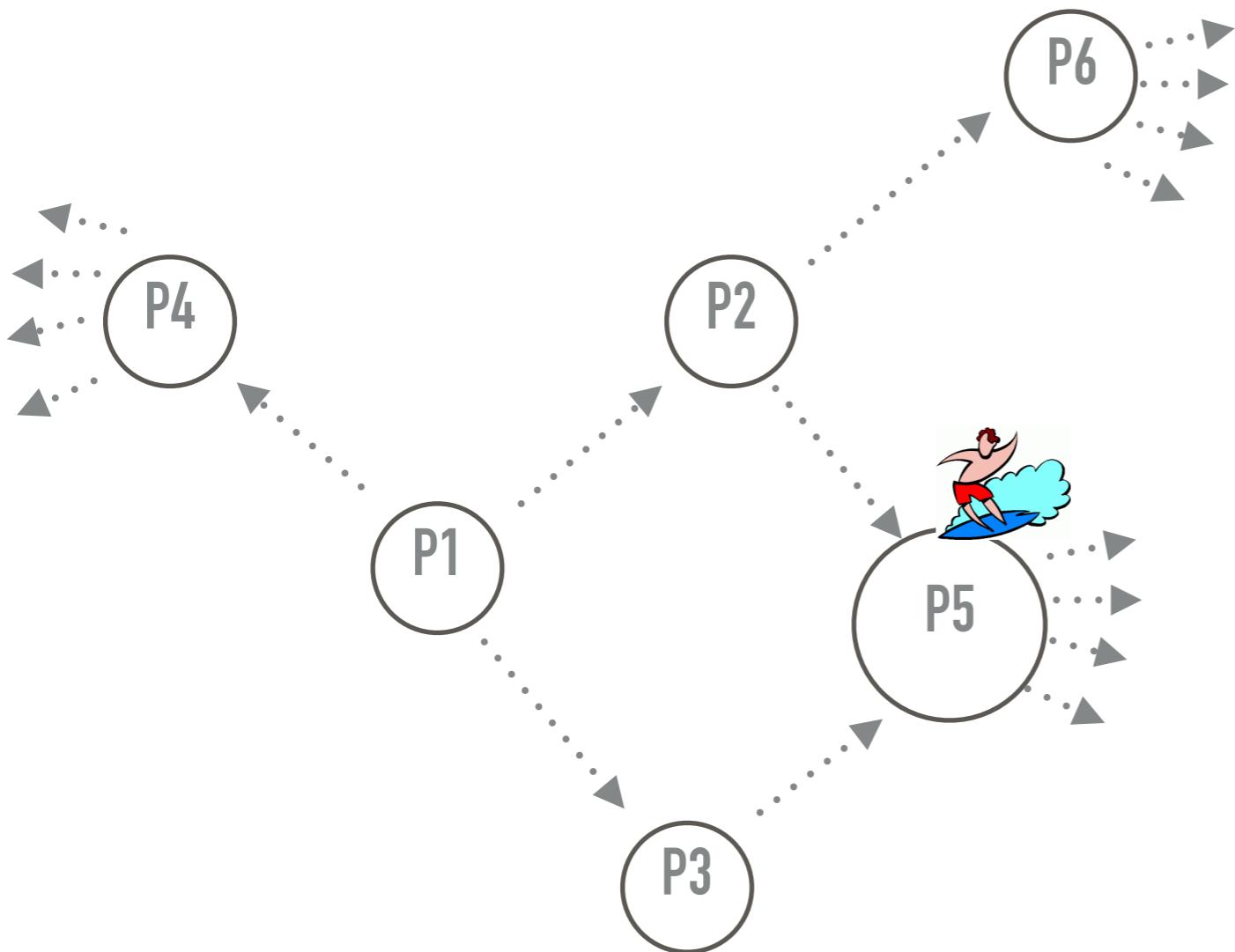
## WHAT IS PAGERANK?

The probability of this random surfer  
being at page P in the long run

# LINK ANALYSIS - PAGE RANK



- ▶ The Web Surfer random walking on the Web

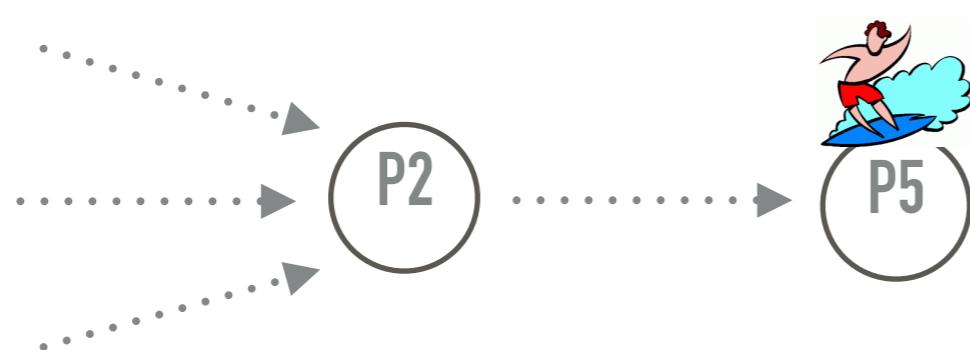


**WHAT IS PAGERANK?**

The probability of this random surfer  
being at page P in the long run

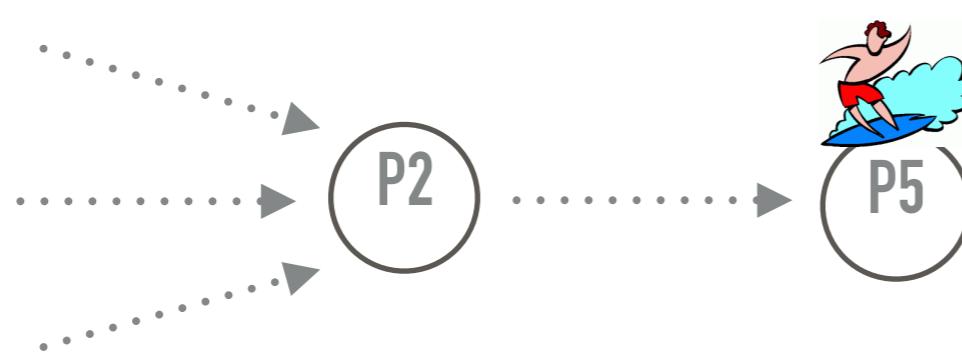
# PAGE RANK

- ▶ Considerations:
  - ▶ A page that has many in-links has a higher probability of being visited
  - ▶ Pages linked by popular pages have higher probability of being visited
  - ▶ What if a page is a dead end?



# PAGE RANK

- ▶ Considerations:
  - ▶ A page that has many in-links has a higher probability of being visited
  - ▶ Pages linked by popular pages have higher probability of being visited
  - ▶ What if a page is a dead end?



**TELEPORTING!**

# PAGE RANK

- ▶ At a **dead end**, surfer jumps to a random page with probability  $1/N$
- ▶ At a **non-dead end**:
  - ▶ with probability **10%**, jumps to a random web page
  - ▶ with probability **90%**, go out on a random regular link
- ▶ Teleportation rate (alpha =  $\alpha$ ): **10% in this case**

**DOES TELEPORTATION MAKE SENSE?**

# RESEMBLES THE WAY WE SEARCH THE INTERNET

1. We type a URL address in our Browser
2. We browser the content of the page and follow a link that we are interested
3. When we the current page is not interesting anymore, we type another URL address and start browsing again
4. Repeat steps 2 and 3 until we are tired or satisfied

### DUE TO TELEPORTATION

- ▶ Using teleportation, we have an Ergodic Markov Chain
- ▶ Irreducibility: there is a path from any page to any other page
- ▶ Aperiodicity: pages cannot be partitioned such that the random walker visits the partitions sequentially

### DUE TO TELEPORTATION

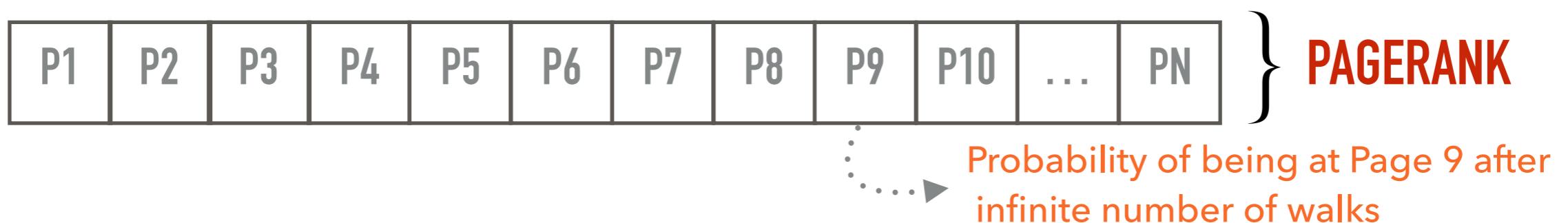
- ▶ Using teleportation, we have an Ergodic Markov Chain
- ▶ Irreducibility: there is a path from any page to any other page
- ▶ Aperiodicity: pages cannot be partitioned such that the random walker visits the partitions sequentially
- ▶ In an Ergodic Markov Chain the probabilities of being in any state reaches a **steady-state**:



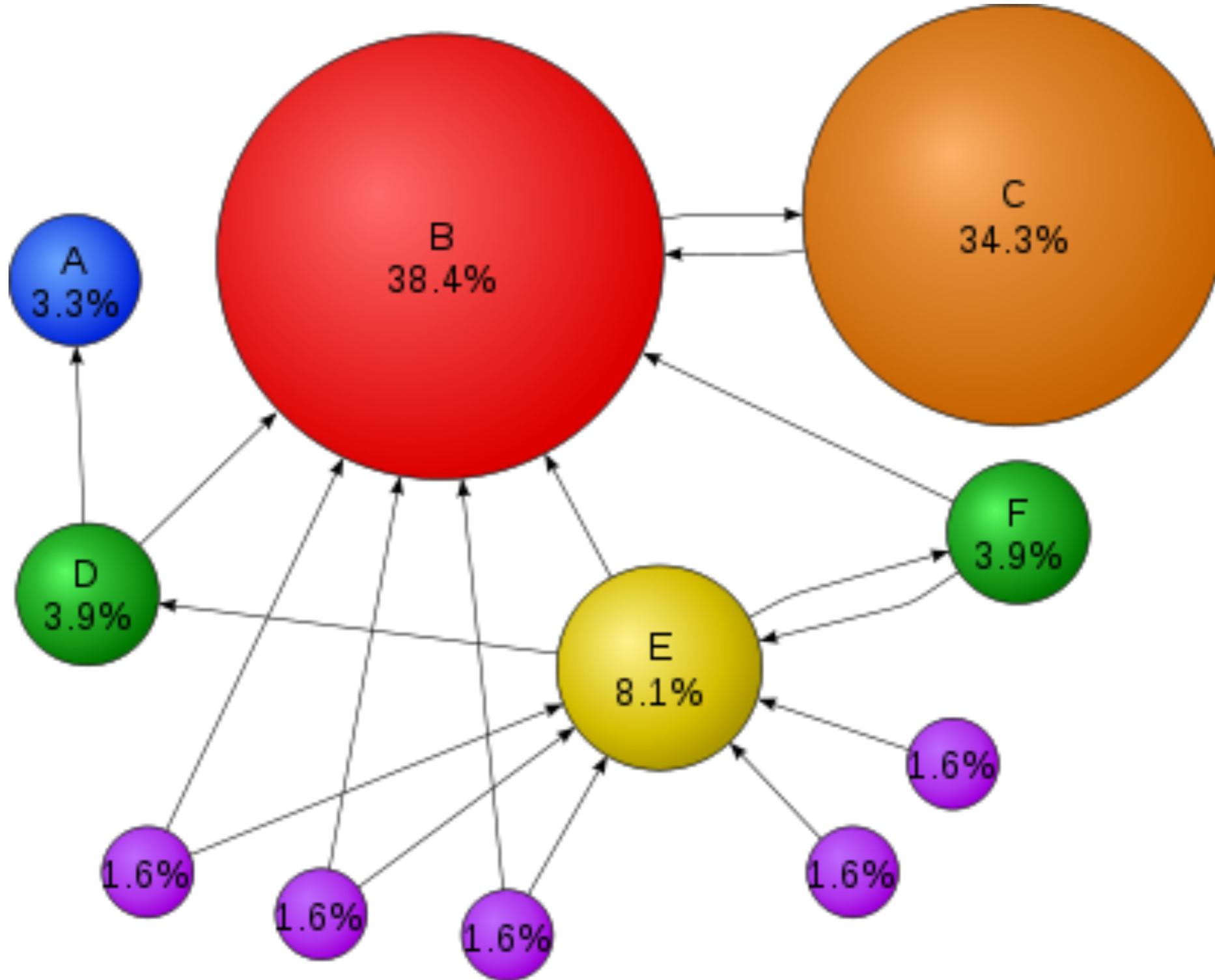
Probability of being at Page 9 after infinite number of walks

### DUE TO TELEPORTATION

- ▶ Using teleportation, we have an Ergodic Markov Chain
- ▶ Irreducibility: there is a path from any page to any other page
- ▶ Aperiodicity: pages cannot be partitioned such that the random walker visits the partitions sequentially
- ▶ In an Ergodic Markov Chain the probabilities of being in any state reaches a **steady-state**:

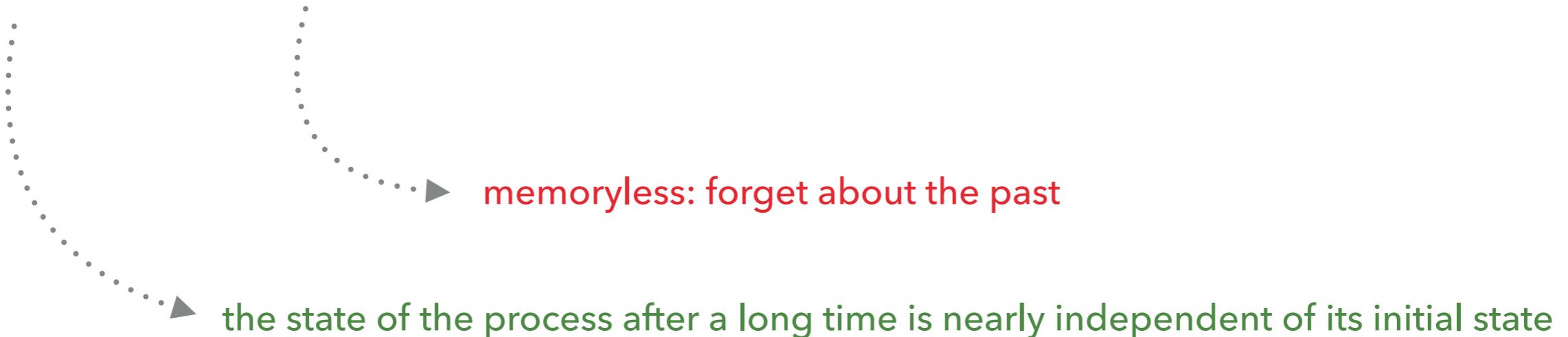


### PAGE RANK EXAMPLE



### PAUSE

- ▶ We know how to represent the Web graph
- ▶ We know what we need to represent the Web graph as a  
**Ergodic Markov Chain**

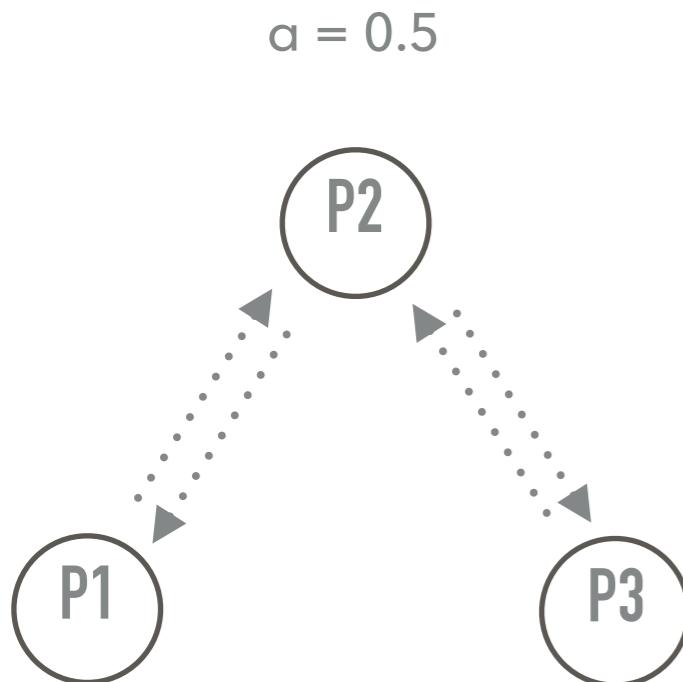


- ▶ Missing: We do not know how to compute PageRank yet!

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



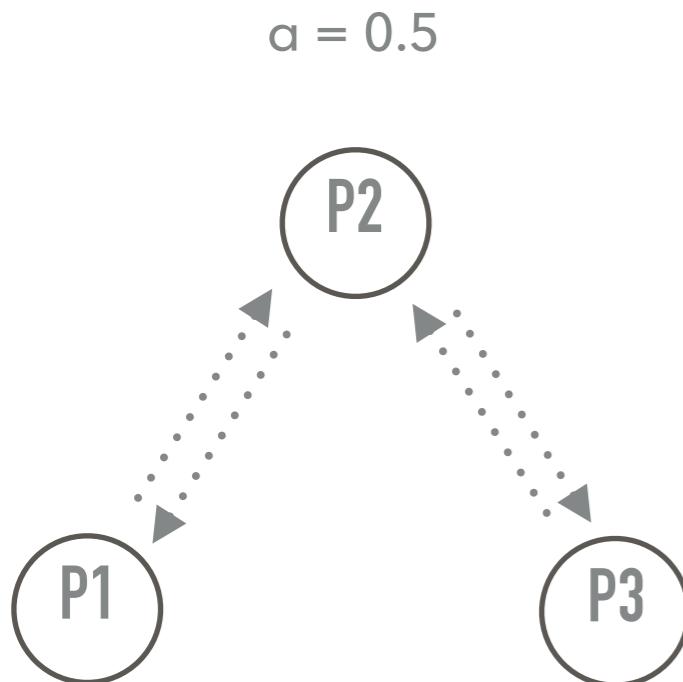
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by  $1/N$ . For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by  $(1-a)$
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



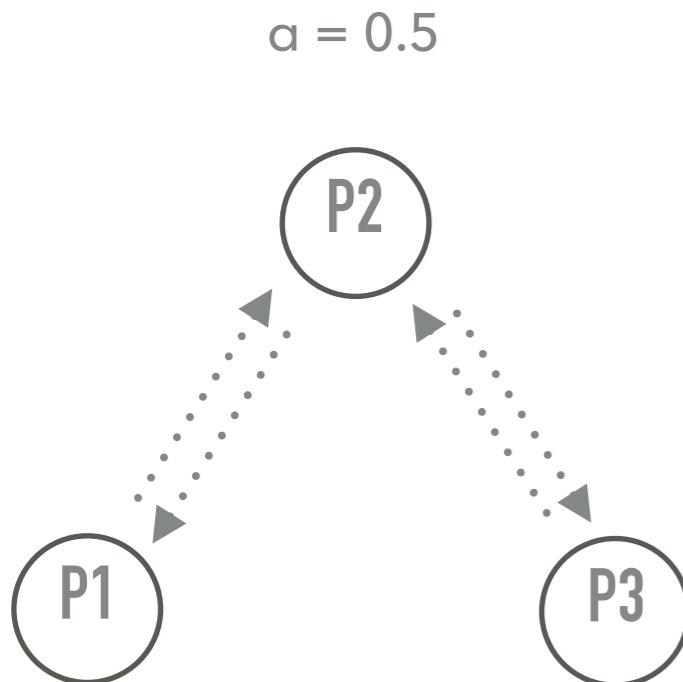
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by  $1/N$ . For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by  $(1-a)$
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



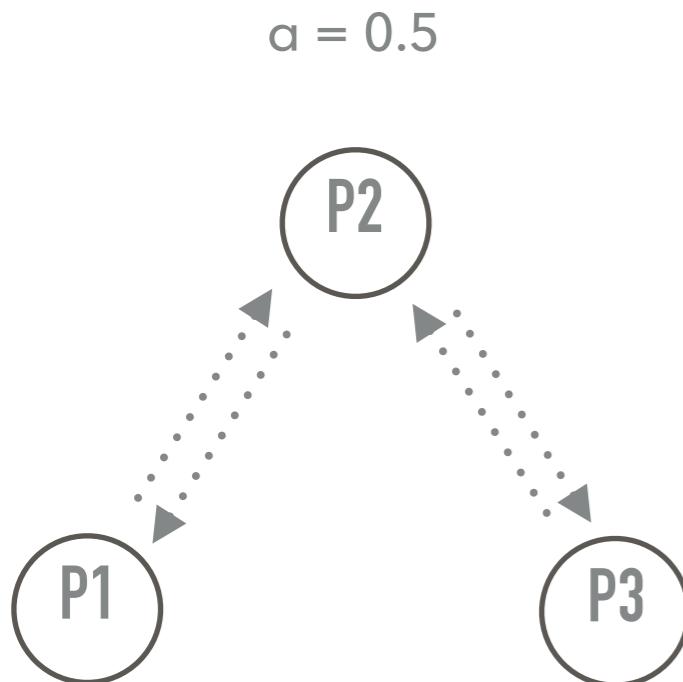
$$\begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by  $1/N$ . For all other rows proceed as follows.
2. **Divide each 1 in A by the number of 1's in its row.**
3. Multiply the resulting matrix by  $(1-a)$
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



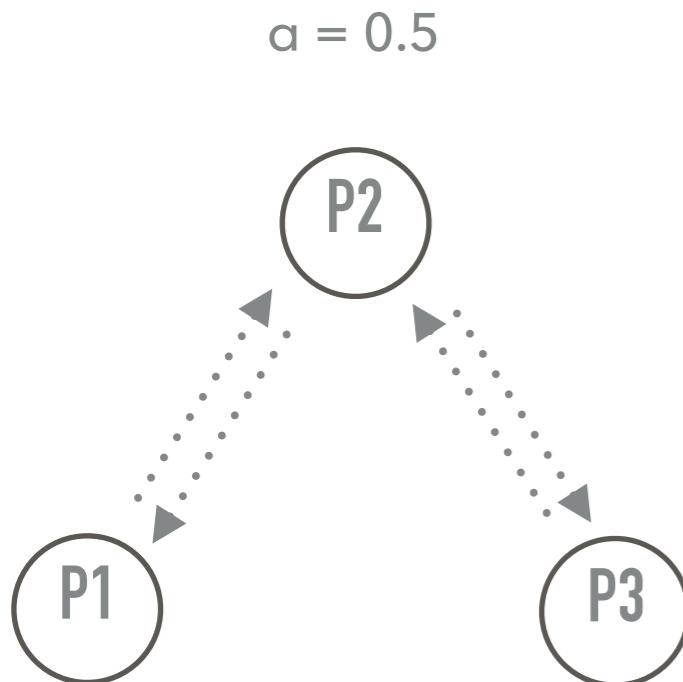
$$\begin{bmatrix} 0 & 0.5 & 0 \\ 0.25 & 0 & 0.25 \\ 0 & 0.5 & 0 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by  $1/N$ . For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. **Multiply the resulting matrix by  $(1-a)$**
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



$$P = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix}$$

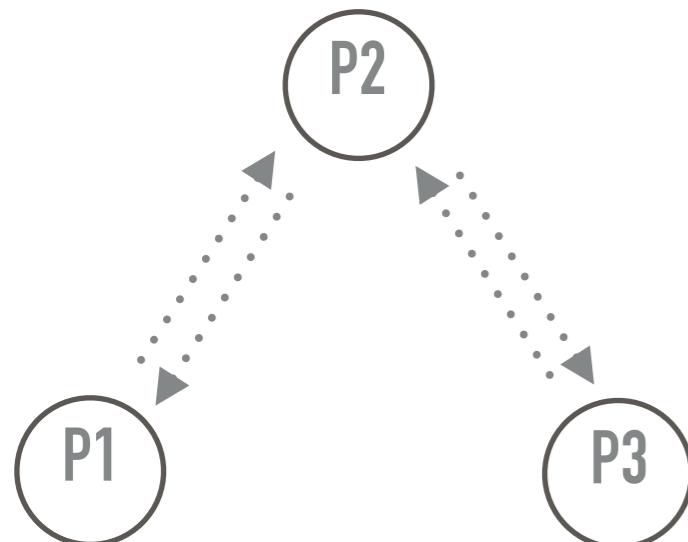
1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE

$\alpha = 0.5$



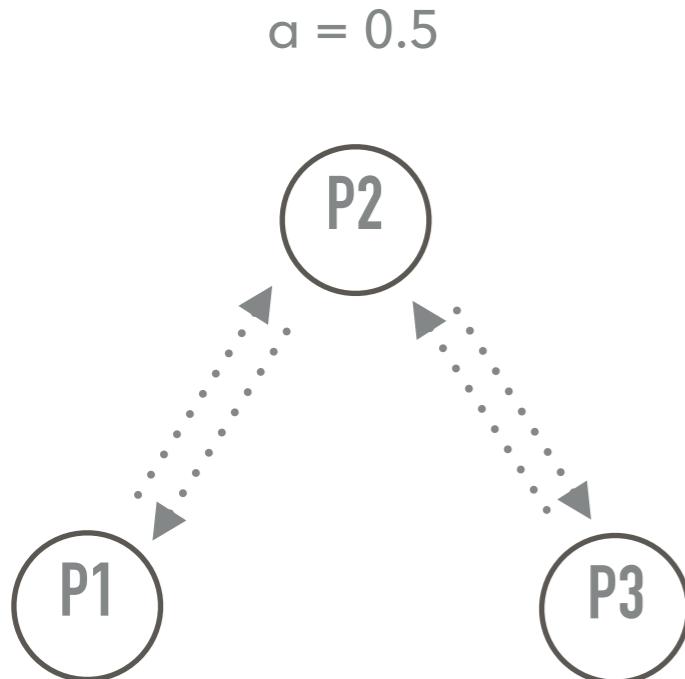
$$P = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add  $a/N$  to every entry of the resulting matrix, to obtain P.

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



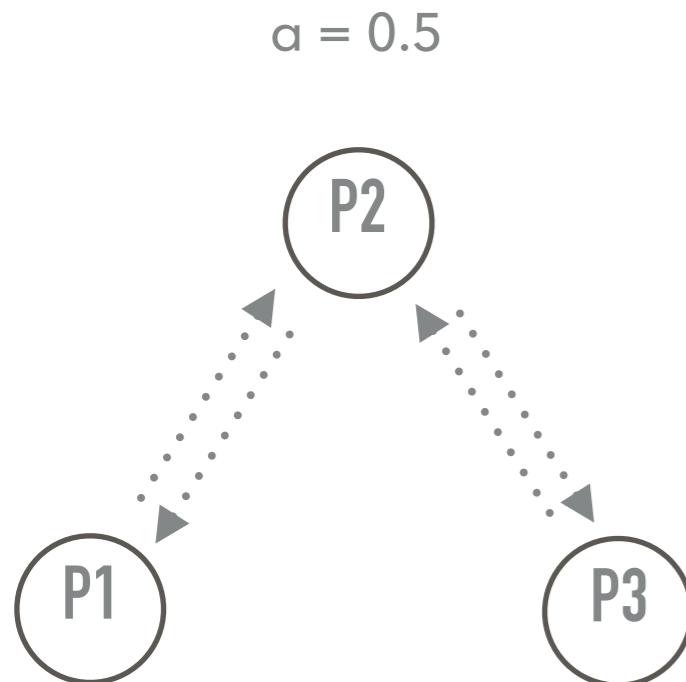
$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$$\text{PR} = \begin{bmatrix} X & Y & Z \end{bmatrix} \quad \text{PAGERANK VALUES}$$

## LECTURE 10 - LINK ANALYSIS

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

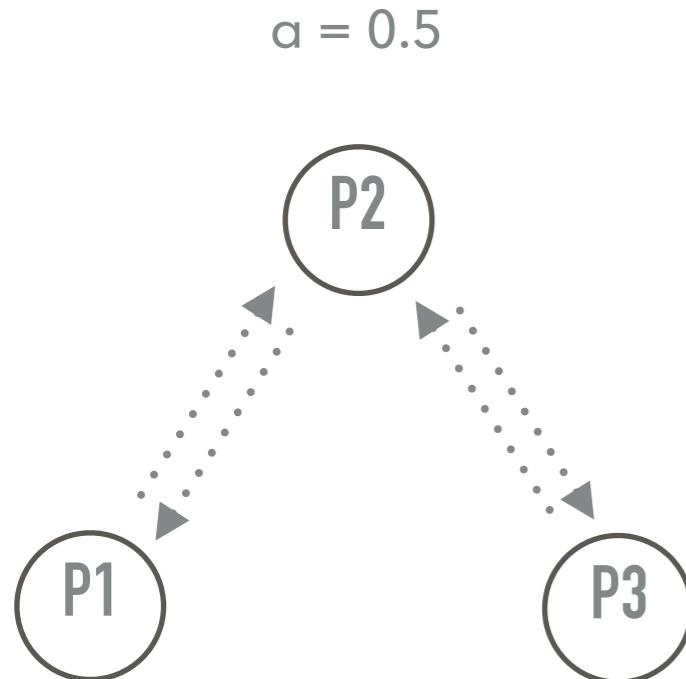
$$PR = \begin{bmatrix} X & Y & Z \end{bmatrix}$$

### PROPERTY OF A MARVOK CHAIN

$$PR_t(d) = PR_{t-1}(d) * M$$

## LECTURE 10 - LINK ANALYSIS

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

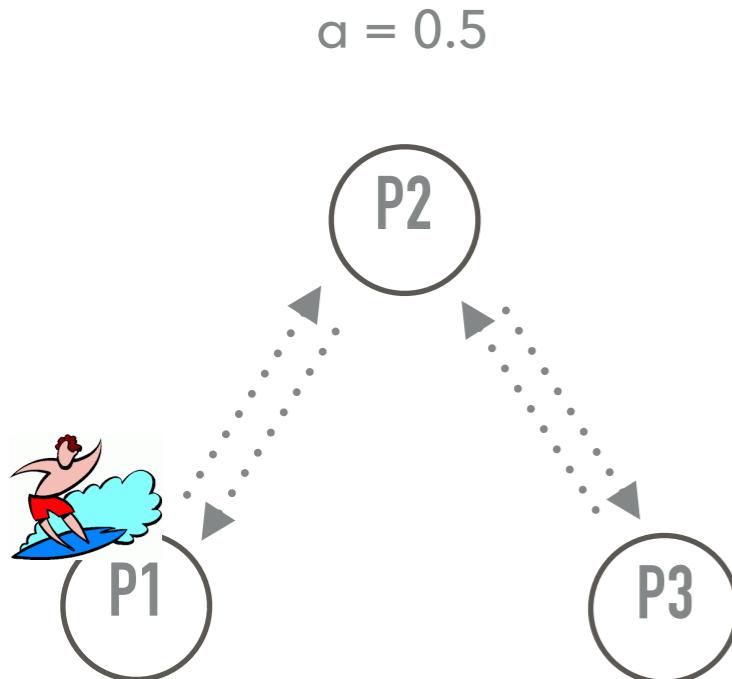
1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

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

$$PR_t(d) = PR_{t-1}(d) * M$$

## LECTURE 10 - LINK ANALYSIS

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

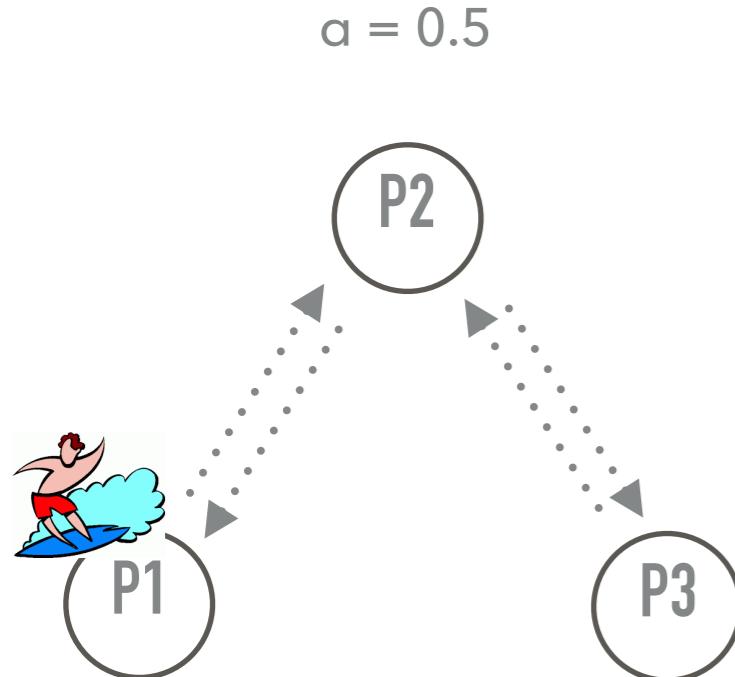
$$PR = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \text{ Arbitrarily chosen. Does not matter!}$$

$$PR_t(d) = PR_{t-1}(d) * M$$

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$$PR = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \text{ Arbitrarily chosen. Does not matter!}$$

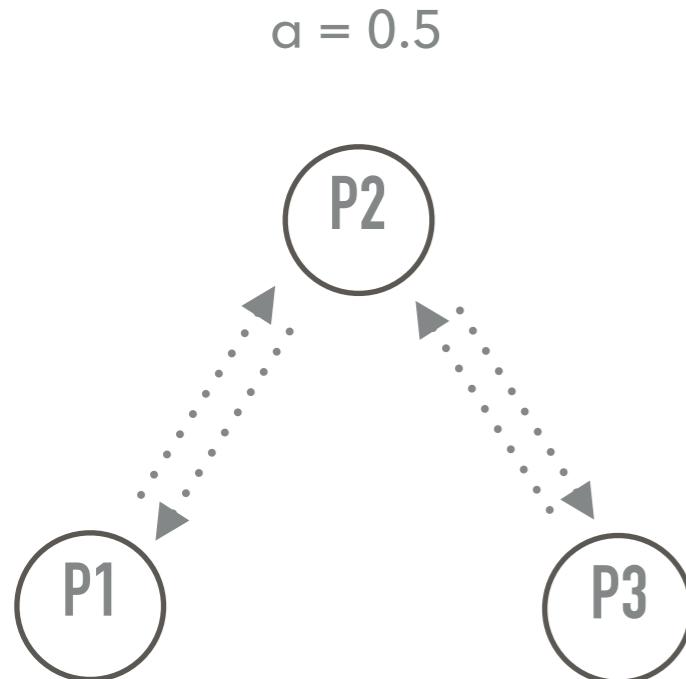
$$PR_t(d) = PR_{t-1}(d) * M$$

$$PR(1) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \end{bmatrix}$$

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$$PR = \begin{bmatrix} 1/6 & 2/3 & 1/6 \end{bmatrix}$$

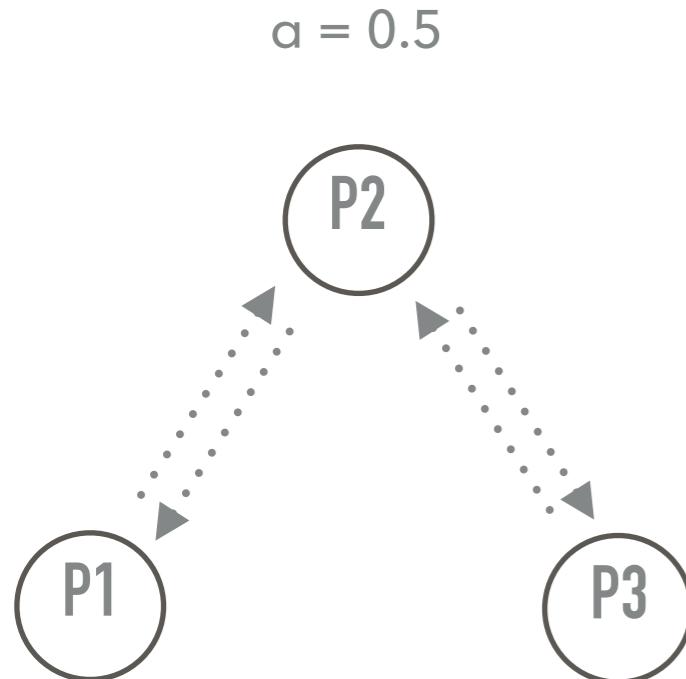
$$PR_t(d) = PR_{t-1}(d) * M$$

$$PR(1) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \end{bmatrix}$$

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$$PR = \begin{bmatrix} 1/6 & 2/3 & 1/6 \end{bmatrix}$$

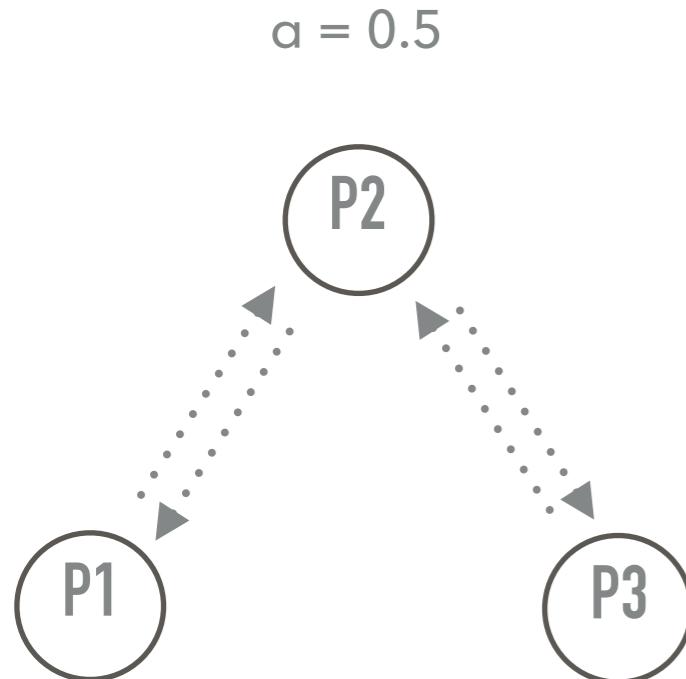
$$PR_t(d) = PR_{t-1}(d) * M$$

$$PR(2) = \begin{bmatrix} 1/6 & 2/3 & 1/6 \end{bmatrix} \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

## LECTURE 10 - LINK ANALYSIS

---

### EXAMPLE



$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$$PR = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

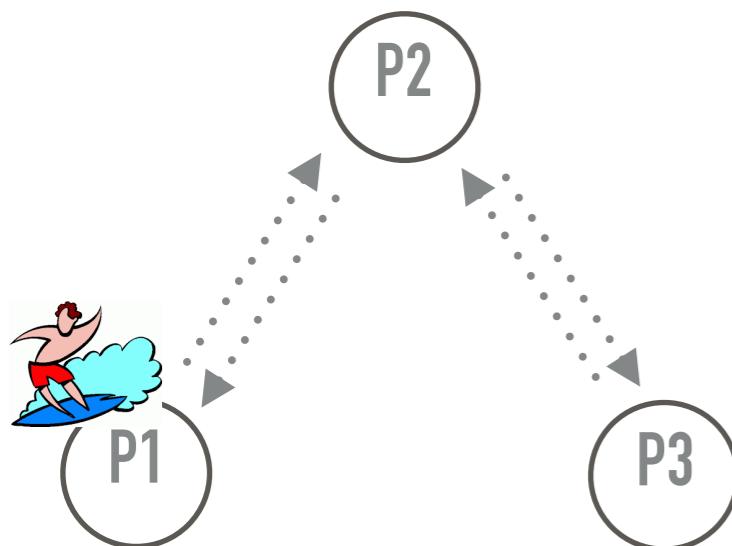
$$PR_t(d) = PR_{t-1}(d) * M$$

$$PR(2) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} = \begin{bmatrix} \dots \end{bmatrix}$$

## LECTURE 10 - LINK ANALYSIS

### EXAMPLE

$$a = 0.5$$



$$\text{PR} = \begin{bmatrix} 5/18 & 4/9 & 5/18 \end{bmatrix}$$

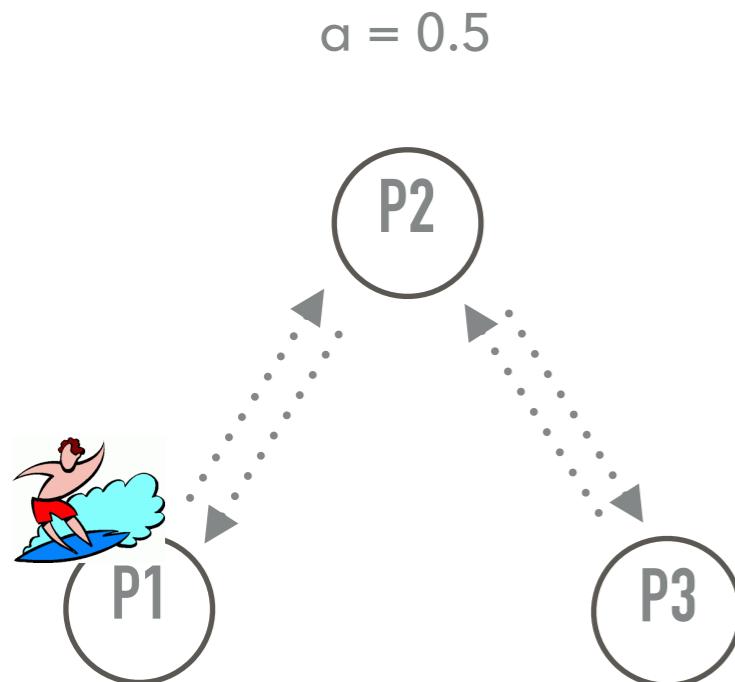
$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$\vec{x}_0$	1	0	0
$\vec{x}_1$	1/6	2/3	1/6
$\vec{x}_2$	1/3	1/3	1/3
$\vec{x}_3$	1/4	1/2	1/4
$\vec{x}_4$	7/24	5/12	7/24
...	...	...	...
$\vec{x}$	5/18	4/9	5/18

## LECTURE 10 - LINK ANALYSIS

### EXAMPLE



$$PR = \begin{bmatrix} 5/18 & 4/9 & 5/18 \end{bmatrix}$$

$$M = \begin{bmatrix} 0.16 & 0.56 & 0.16 \\ 0.41 & 0.16 & 0.41 \\ 0.16 & 0.56 & 0.41 \end{bmatrix} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

1. If a row of A has no 1's, then replace each element by 1/N. For all other rows proceed as follows.
2. Divide each 1 in A by the number of 1's in its row.
3. Multiply the resulting matrix by (1-a)
4. Add a/N to every entry of the resulting matrix, to obtain M.

$\vec{x}_0$	1	0	0
$\vec{x}_1$	1/6	2/3	1/6
$\vec{x}_2$	1/3	1/3	1/3
$\vec{x}_3$	1/4	1/2	1/4
$\vec{x}_4$	7/24	5/12	7/24
...	...	...	...
$\vec{x}$	5/18	4/9	5/18

$$PR = PR_0 * (M)^t$$

### SO WHAT?

- ▶ Given that we have a quality value for a page, what can we do with it?

### SO WHAT?

- Given that we have a quality value for a page, what can we do with it?

$$P(d|q) = \frac{P(q|d)P(d)}{P(q)}$$

It can be a constant...  
or it can be a proxy for  
document popularity,  
document credibility,  
document readability...

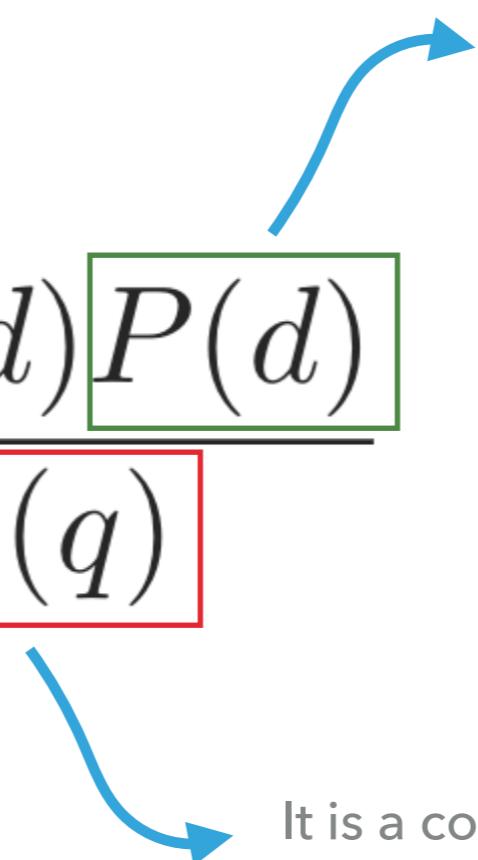
It is a constant for every  
document in the collection.

### SO WHAT?

- Given that we have a quality value for a page, what can we do with it?

It can be a constant...  
or it can be a proxy for  
document popularity,  
document credibility,  
document readability...

$$P(d|q) = \frac{P(q|d)P(d)}{P(q)}$$



It is a constant for every  
document in the collection.

**ANOTHER USE OF THIS VALUE WILL BE PRESENTED NEXT LECTURE!**

# TODAY'S LECTURE IN THE STANFORD IR BOOK

- ▶ Chapter 19: Web search basics
  - ▶ Chapter 19.1: Background and history
  - ▶ Chapter 19.2: Web characteristics
  - ▶ Chapter 19.3: Advertising as the economic model
- ▶ Chapter 21: Link Analysis
  - ▶ Chapter 21.1: The Web as a graph
  - ▶ Chapter 21.2: Page Rank