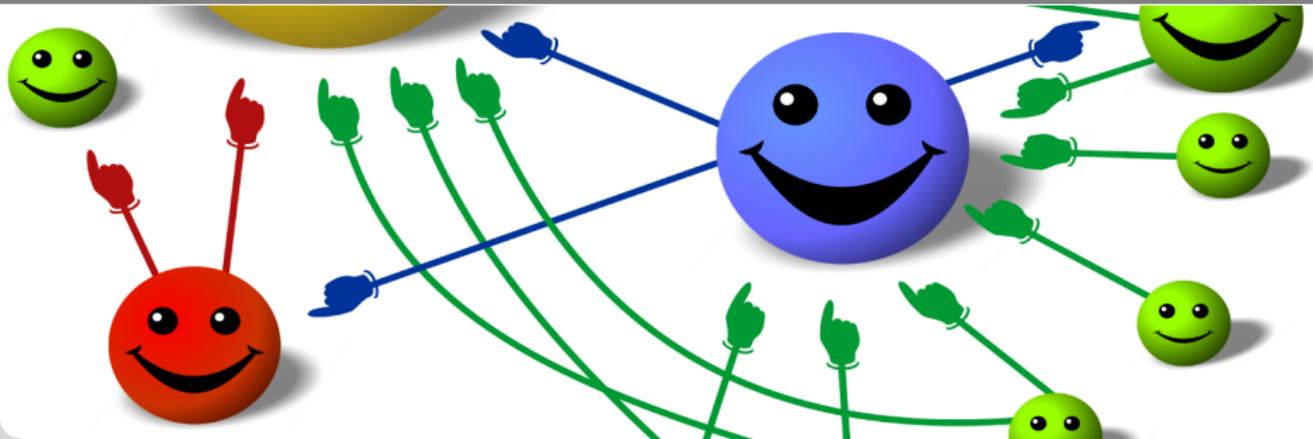


How Google Searches Work

History of searching and the PageRank algorithm

Martin Thoma, Benjamin Lipp | 7th of February, 2013

SPRACHENZENTRUM



Contents

1 Introduction

2 PageRank

3 End

The Early Days: Web Directories

 open directory project In partnership with
Aol Search.

[about dmoz](#) | [dmoz blog](#) | [suggest URL](#) | [help](#) | [link](#) | [editor login](#)

[advanced](#)

Arts Movies , Television , Music ...	Business Jobs , Real Estate , Investing ...	Computers Internet , Software , Hardware ...
Games Video Games , RPGs , Gambling ...	Health Fitness , Medicine , Alternative ...	Home Family , Consumers , Cooking ...
Kids and Teens Arts , School Time , Teen Life ...	News Media , Newspapers , Weather ...	Recreation Travel , Food , Outdoors , Humor ...
Reference Maps , Education , Libraries ...	Regional US , Canada , UK , Europe ...	Science Biology , Psychology , Physics ...
Shopping Clothing , Food , Gifts ...	Society People , Religion , Issues ...	Sports Baseball , Soccer , Basketball ...
World Català , Dansk , Deutsch , Español , Français , Italiano , 日本語 , Nederlands , Polski , Русский , Svenska ...		

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



Copyright © 2013 Netscape

5,114,083 sites - 96,877 editors - over 1,014,849 categories

The Early Days: Web Directories

 open directory project

Search

the entire directory



Top: Reference: **Education** (37,399)

-
- [Career, Specialized and Technical Training@](#) (852)
 - [Colleges and Universities](#) (50,589)
 - [Early Childhood](#) (160)
 - [Home Schooling@](#) (1,283)
 - [K through 12](#) (2,348)
 - [Special Education](#) (328)

-
- [Distance Learning](#) (660)
 - [Educational Testing](#) (20)
 - [How to Study](#) (28)

-
- | | |
|--|---|
| • Alumni Directories (13) | • Issues@ (375) |
| • Conferences (0) | • Journals (44) |
| • Directories (84) | • Methods and Theories (215) |
| • Educators (186) | • News and Media (1) |
| • Employment@ (85) | • Organizations (191) |
| • History of Education@ (20) | • Products and Services (1,176) |
| • Instructional Technology (212) | • Regional (0) |
| • International (1,117) | • Subjects (26) |

The Early Days: Web Directories

 open directory project

Search the entire directory

Top: **Reference:** **Education:** **Colleges and Universities** (30,589)

- [Africa](#) (202)
- [Asia](#) (1,810)
- [Caribbean](#) (18)
- [Central America](#) (50)
- [Europe](#) (4,258)
- [Middle East](#) (247)
- [North America](#) (22,540)
- [Oceania](#) (653)
- [South America](#) (10)

- [Academic Departments](#) (0)
- [Admissions](#) (120)
- [Campus Media](#)@ (866)
- [Career, Specialized and Technical Training](#)@ (852)
- [Christian](#)@ (61)
- [College Life](#) (122)
- [Colleges for Students with Disabilities](#)@ (43)
- [Directories](#) (96)
- [Educators](#) (57)
- [Faculty Experts Directories](#) (16)
- [Financial Aid](#) (233)
- [Graduate Education](#) (73)
- [Guides](#) (44)
- [History of College Architecture](#)@ (52)
- [Instructional Technology](#)@ (29)
- [Islamic](#)@ (6)
- [Jewish](#)@ (36)
- [Libraries](#)@ (55)
- [News and Media](#) (4)
- [Online Learning](#)@ (660)
- [Organizations](#) (8)
- [Student Organizations](#)@ (2,860)
- [Transdisciplinary](#) (15)
- [University Presses](#)@ (102)
- [Virtual Tours](#) (13)

The Early Days: Web Directories

 open directory project

[about dmoz](#) | [dmoz blog](#) | [suggest](#)

Search

the entire directory

[Top](#): **[Reference](#)**: **[Education](#)**: **[Colleges and Universities](#)**: **[Europe](#)** (4,258)

- [Academic Departments](#) (0)

- [Albania](#) (8)
- [Austria](#) (54)
- [Belarus](#) (11)
- [Belgium](#) (24)
- [Bosnia and Herzegovina](#) (8)
- [Bulgaria](#) (30)
- [Croatia](#) (18)
- [Cyprus@](#) (1)
- [Czech Republic](#) (59)
- [Denmark](#) (11)
- [Estonia](#) (23)
- [Finland](#) (23)
- [France](#) (19)
- [Germany](#) (121)
- [Greece](#) (22)
- [Hungary](#) (16)
- [Iceland](#) (4)
- [Lithuania](#) (9)
- [Luxembourg](#) (1)
- [Macedonia](#) (19)
- [Malta](#) (14)
- [Moldova](#) (1)
- [Montenegro](#) (1)
- [Netherlands](#) (30)
- [Norway](#) (160)
- [Poland](#) (175)
- [Portugal](#) (6)
- [Romania](#) (18)
- [Russia](#) (78)
- [Serbia](#) (6)
- [Slovakia](#) (12)
- [Slovenia](#) (7)
- [Spain](#) (20)
- [Sweden](#) (56)

The Early Days: Web Directories

 open directory project

[about dmoz](#) |

Search the entire directory ▾

[Top](#): **[Reference](#)**: **[Education](#)**: **[Colleges and Universities](#)**: **[Europe](#)**: **[Germany](#)** (121)

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#)

- [Baden-Württemberg](#) (15)
- [Bavaria](#) (10)
- [Berlin](#) (35)
- [Brandenburg](#) (5)
- [Bremen](#) (2)
- [Hamburg](#) (3)
- [Hesse](#) (7)
- [Lower Saxony](#) (3)
- [Mecklenburg-Western Pomerania](#) (0)
- [North Rhine-Westphalia](#) (23)
- [Rhineland-Palatinate](#) (4)
- [Saarland](#) (2)
- [Saxony](#) (1)
- [Saxony-Anhalt](#) (0)
- [Schleswig-Holstein](#) (2)
- [Thuringia](#) (5)

See also:

- [Regional: Europe: Germany](#) (3,998)
- [Regional: Europe: Germany: Education](#) (11)

This category in other languages:

[French](#) (6)

[German](#) (2,229)

The Early Days: Web Directories



[Top](#): [Reference](#): [Education](#): [Colleges and Universities](#): [Europe](#): [Germany](#): **[Baden-Württemberg](#) (15)**

- [Hochschule Aalen - University of Applied Sciences](#) (1)
- [Hochschule Esslingen - University of Applied Sciences](#) (1)
- [Hochschule Furtwangen - University of Applied Sciences](#) (0)
- [Hochschule Heilbronn - Heilbronn University](#) (1)
- [Hochschule Konstanz - University of Applied Sciences](#) (0)
- [International University in Germany](#) (1)
- [Stuttgart Institute of Management and Technology](#) (1)
- [University of Freiburg](#) (1)
- [University of Heidelberg](#) (1)
- [University of Hohenheim](#) (1)
- [University of Konstanz](#) (2)
- [University of Mannheim](#) (1)
- [University of Maryland University College - Mannheim](#)
- [University of Stuttgart](#) (1)
- [University of Ulm](#) (1)

See also:

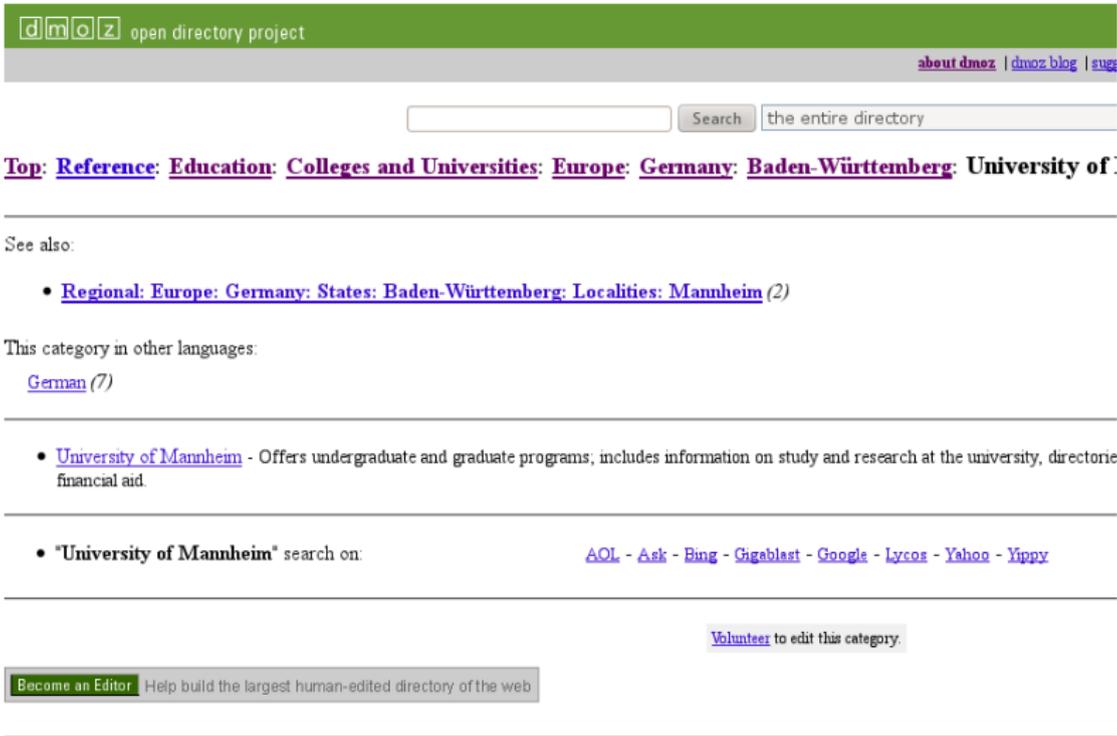
- [Regional: Europe: Germany: States: Baden-Württemberg: Education](#) (1)

This category in other languages:

[German](#) (247)

- **"Baden-Württemberg"** search on: [AOL](#) - [Ask](#) - [Bing](#) - [Cgables](#) - [Google](#) - [Lycos](#) - [Yahoo](#) - [Yippy](#)

The Early Days: Web Directories



dmoz open directory project

[about dmoz](#) | [dmoz blog](#) | [sug](#)

Search the entire directory

[Top](#): [Reference](#): [Education](#): [Colleges and Universities](#): [Europe](#): [Germany](#): [Baden-Württemberg](#): **University of**

See also:

- [Regional: Europe: Germany: States: Baden-Württemberg: Localities: Mannheim](#) (2)

This category in other languages:

[German](#) (7)

- [University of Mannheim](#) - Offers undergraduate and graduate programs, includes information on study and research at the university, directorie financial aid.

- "University of Mannheim" search on: [AOL](#) - [Ask](#) - [Bing](#) - [Gigablast](#) - [Google](#) - [Lycos](#) - [Yahoo](#) - [Yippy](#)

[Volunteer](#) to edit this category.

Become an Editor Help build the largest human-edited directory of the web

Search: **uni mannheim**

Open Directory Categories (1-5 of 100)

1. [World: Français: Régional: Amérique: Etats-Unis: Etat et politique: Ambassades et consulats: Pays représentés](#) (39)
2. [World: Deutsch: Wissen: Bibliotheken: Universitäts- und Hochschulbibliotheken: Deutschland](#) (33)
3. [World: Français: Régional: Amérique: Etats-Unis: Etat et politique: Ambassades et consulats: Représentations à l'étranger](#) (24)
4. [World: Nederlands: Maatschappij: Overheid: Europese Unie](#) (22)
5. [World: Deutsch: Gesundheit: Krankenhäuser und Kliniken: Universitätskliniken: Deutschland](#) (21)

[more...](#)

Open Directory Sites (1-20 of 15698)

1. [amnesty international Hochschulgruppe Mannheim](#) - Die amnesty-Gruppe an der Uni Mannheim setzt sich unter anderem mit Petitionen, Appellbriefen und stellt aktuelle Aktivitäten vor
-- <http://www.amnesty-uni-mannheim.de/> [World: Deutsch: Regional: Europa: Deutschland: Baden-Württemberg: Städte und Gemeinden: M. Mannheim: Gesellschaft](#) (23)
2. [LHG - Liberale Hochschulgruppe Mannheim](#) - Die liberale Studierenden an der Uni Mannheim stellen sich vor
-- <http://www.uni-mannheim.de/studorg/liberale/> [World: Deutsch: Wissen: Bildung: Hochschulen: Europa: Deutschland: Baden-Württemberg: Universität Mannheim](#) (0)
3. [Uni Mannheim](#) - Marktübersicht deutscher Anbieter für Online-Preisvergleiche
-- <http://projekt.wifo.uni-mannheim.de/praservergleich/> [World: Deutsch: Zuhause: Verbraucherinformationen: Preisagenturen: Online-Preisvergleiche](#) (0)
4. [University of Mannheim](#) - Offers undergraduate and graduate programs; includes information on study and research at the university, directories of financial aid.
-- <http://www.uni-mannheim.de/> [Reference: Education: Colleges and Universities: Europe: Germany: Baden-Württemberg: University of Mannheim](#) (0)



- crawls through the web using hyperlinks
- makes an index of the words contained in a page
- ranks pages for a search query according to number of occurrences of keywords



- crawls through the web using hyperlinks
- makes an index of the words contained in a page
- ranks pages for a search query according to number of occurrences of keywords



- crawls through the web using hyperlinks
- makes an index of the words contained in a page
- ranks pages for a search query according to number of occurrences of keywords

How can we use this massive amount of information?

- 625.3 million websites
 - Wikipedia is one website and has several millions of pages
- ⇒ we need to rank websites!

How can we use this massive amount of information?

- 625.3 million websites
- Wikipedia is one website and has several millions of pages

⇒ we need to rank websites!

How can we use this massive amount of information?

- 625.3 million websites
 - Wikipedia is one website and has several millions of pages
- ⇒ we need to rank websites!

We all know that:

- humans know what is good for them
 - ✗ machines don't know what's good for humans
 - humans create websites
 - humans will only [link](#) to websites they like
- ⇒ hyperlinks are a quality indicator

We all know that:

- humans know what is good for them
 - ✗ machines don't know what's good for humans
 - humans create websites
 - humans will only [link](#) to websites they like
- ⇒ hyperlinks are a quality indicator

We all know that:

- humans know what is good for them
 - ✗ machines don't know what's good for humans
 - humans create websites
 - humans will only [link](#) to websites they like
- ⇒ hyperlinks are a quality indicator

We all know that:

- humans know what is good for them
 - ✗ machines don't know what's good for humans
 - humans create websites
 - humans will only [link](#) to websites they like
- ⇒ hyperlinks are a quality indicator

We all know that:

- humans know what is good for them
 - ✗ machines don't know what's good for humans
 - humans create websites
 - humans will only [link](#) to websites they like
- ⇒ hyperlinks are a quality indicator

How Could We Use That?

- simply count number of links to a website
- ✗ 10,000 links from only one page
- count number of websites that link to a website
- ✗ quality of the linking website matters
- ✗ total number of links on the source page matters

How Could We Use That?

- simply count number of links to a website
- ✗ 10,000 links from only one page
- count number of websites that link to a website
- ✗ quality of the linking website matters
- ✗ total number of links on the source page matters

How Could We Use That?

- simply count number of links to a website
- ✗ 10,000 links from only one page
- count number of websites that link to a website
- ✗ quality of the linking website matters
- ✗ total number of links on the source page matters

How Could We Use That?

- simply count number of links to a website
- ✗ 10,000 links from only one page
- count number of websites that link to a website
- ✗ quality of the linking website matters
- ✗ total number of links on the source page matters

How Could We Use That?

- simply count number of links to a website
- ✗ 10,000 links from only one page
- count number of websites that link to a website
- ✗ quality of the linking website matters
- ✗ total number of links on the source page matters

A Brilliant Idea



Sergey Brin



Larry Page

- decisions of humans are complicated

- a lot of webpages get visited

⇒ modelize clicks on links as random behaviour

- links are important

- links of page A get less important, if A has many links

- links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modelize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modellize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modelize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modellize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modellize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modelize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

if A links to B then

$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

- decisions of humans are complicated
- a lot of webpages get visited

⇒ modellize clicks on links as random behaviour

- links are important
 - links of page A get less important, if A has many links
 - links of page A get more important, if many link to A

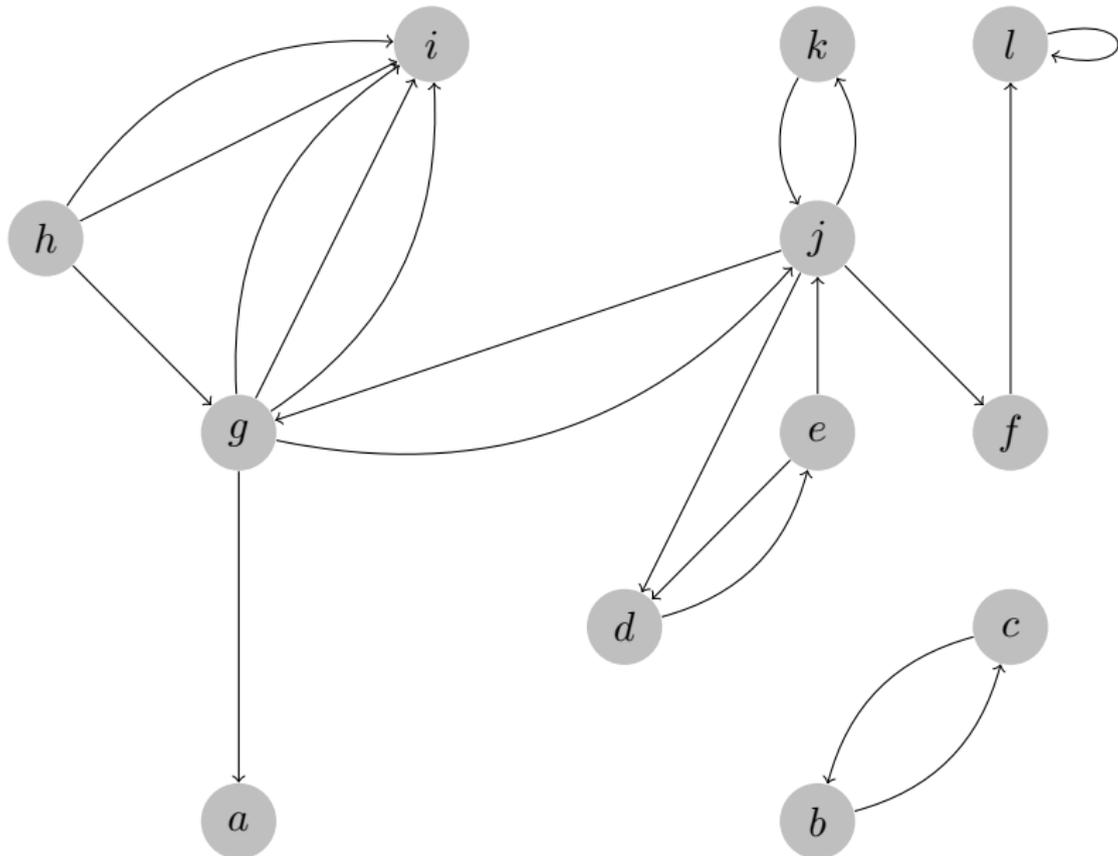
⇒ if B has a link from A, the rank of B increases by $\frac{Rank(A)}{Links(A)}$

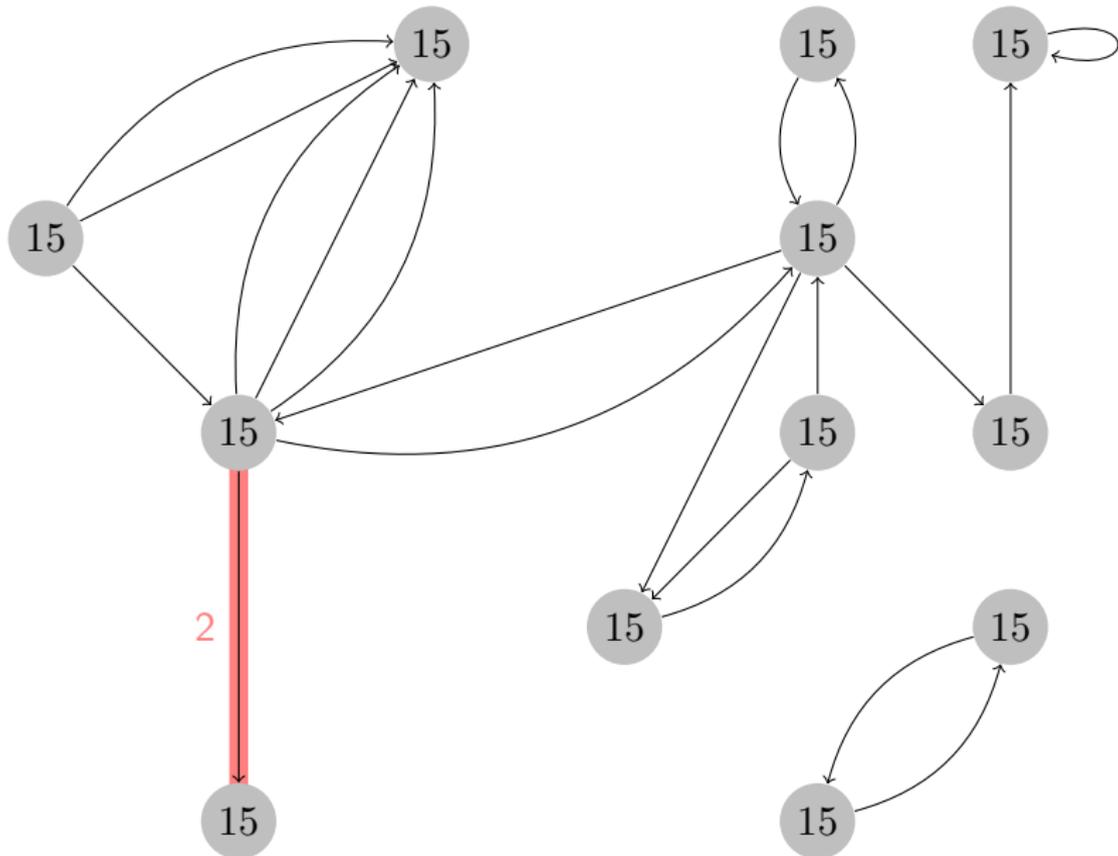
if A links to B then

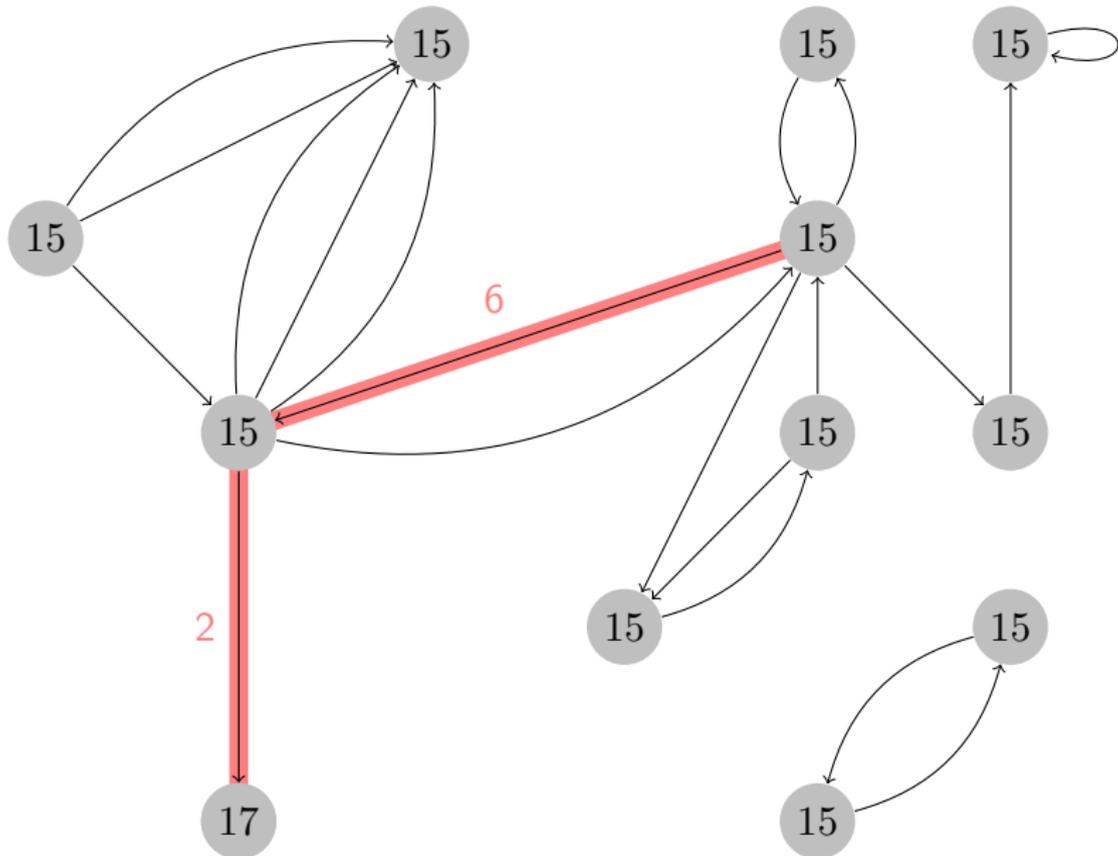
$$Rank(B) += \frac{Rank(A)}{Links(A)}$$

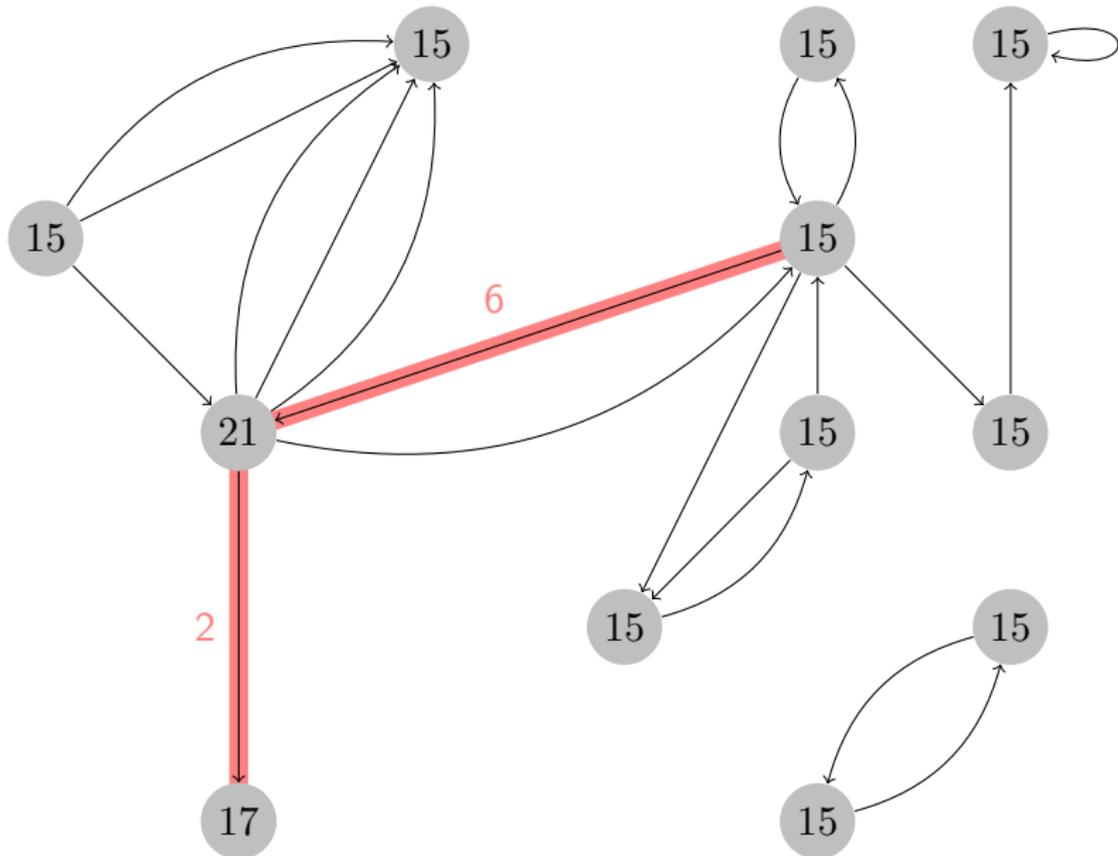
What is PageRank?

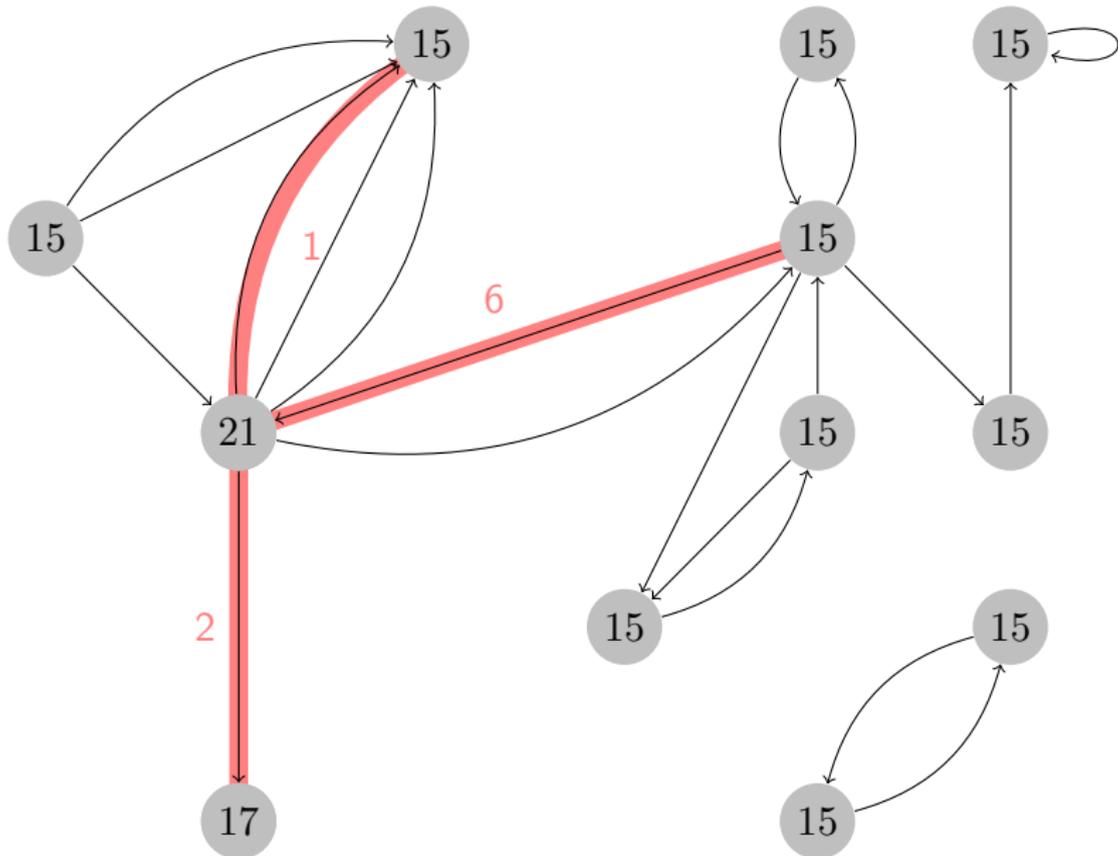
The PageRank algorithm calculates the probability of a randomly clicking user ending up on a given page.

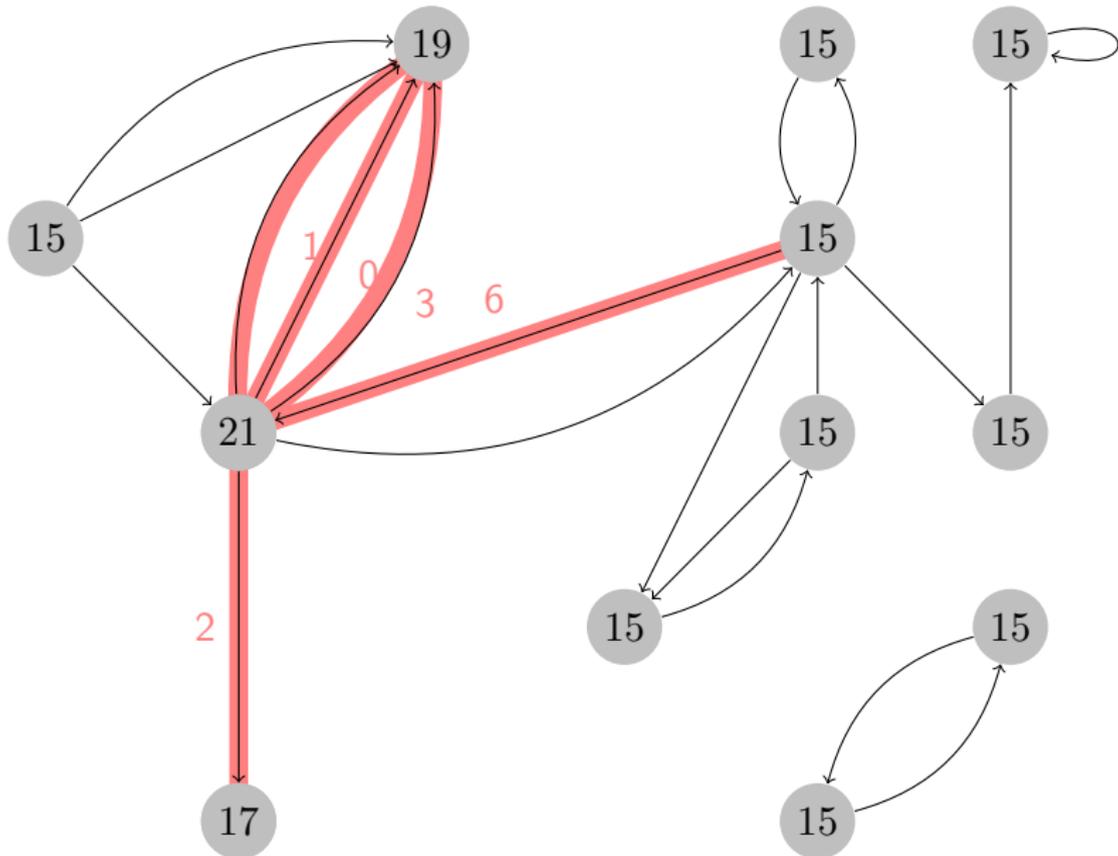












Let x be a web page. Then

- $L(x)$ is the set of websites that link to x
- $C(y)$ is the out-degree of page y
- α is probability of random jump
- N is the total number of websites

$$PR(x) := \alpha \left(\frac{1}{N} \right) + (1 - \alpha) \sum_{y \in L(x)} \frac{PR(y)}{C(y)}$$

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all page  $\in$  web do ▷ calculate pageRank of page  
      page.pageRank =  $q$   
      for all  $y \in L(\textit{page})$  do  
        page.pageRank +=  $\frac{y.\textit{pageRank}}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all page  $\in$  web do                                 $\triangleright$  calculate pageRank of page  
      page.pageRank =  $q$   
      for all  $y \in L(\textit{page})$  do  
        page.pageRank +=  $\frac{y.\textit{pageRank}}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all page  $\in$  web do ▷ calculate pageRank of page  
      page.pageRank =  $q$   
      for all  $y \in L(\textit{page})$  do  
        page.pageRank +=  $\frac{y.\textit{pageRank}}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all page  $\in$  web do                                 $\triangleright$  calculate pageRank of page  
      page.pageRank =  $q$   
      for all  $y \in L(\textit{page})$  do  
        page.pageRank +=  $\frac{y.\textit{pageRank}}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all  $page \in web$  do                                ▷ calculate pageRank of page  
       $page.pageRank = q$   
      for all  $y \in L(page)$  do  
         $page.pageRank += \frac{y.pageRank}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

```
function PAGERANK(Graph web, double  $q = 0.15$ , int iterations)  
  while iterations > 0 do  
    for all page  $\in$  web do                                 $\triangleright$  calculate pageRank of page  
      page.pageRank =  $q$   
      for all  $y \in L(\textit{page})$  do  
        page.pageRank +=  $\frac{y.\textit{pageRank}}{C(y)}$   
      end for  
    end for  
  iterations -= 1
```

Factors of Influence

- language
- location
- social information: what your friends like

Factors of Influence

- language
- location
- social information: what your friends like

- language
- location
- social information: what your friends like

The Filter Bubble

dontbubble.us
www.thefilterbubble.com

The Filter Bubble

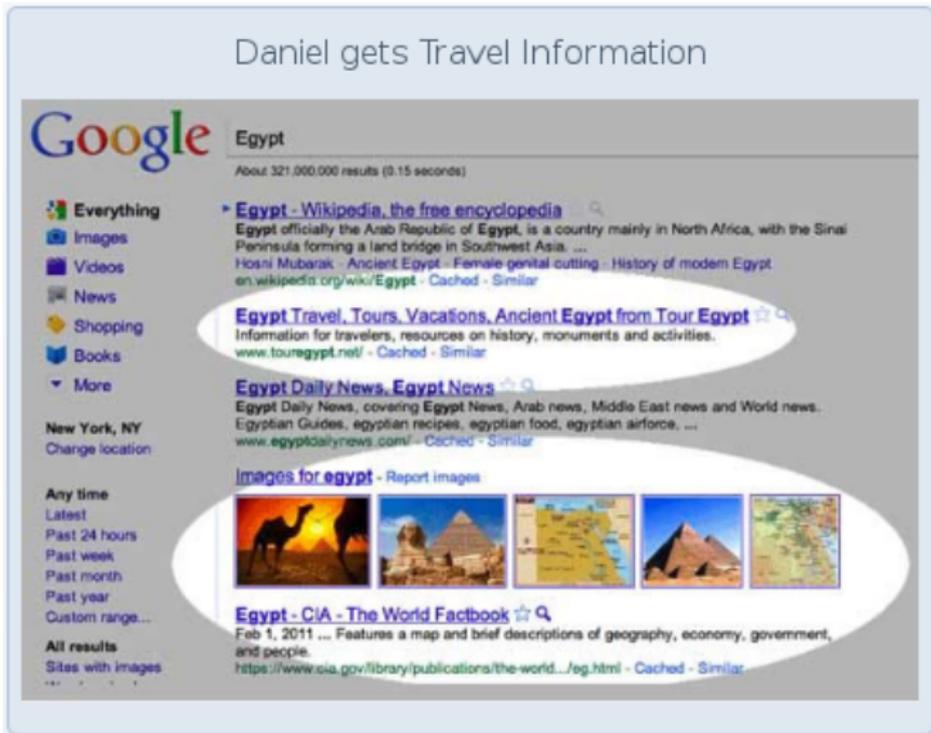
Scott gets Egyptian Protests



The screenshot shows a Google search for "Egypt" with approximately 350,000,000 results. The search results are filtered to show only news and information related to the 2011 Egyptian protests. Two white ovals highlight specific results: one for "Egypt News - The Protests of 2011 - The New York Times" and another for "News for Egypt" which includes an article about Lara Logan's experience in Egypt. The left sidebar shows standard search filters like "Everything", "Images", "Videos", "News", "Shopping", "Realtime", "Books", and "More".

The Filter Bubble

Daniel gets Travel Information



The screenshot shows a Google search interface for the query "Egypt". The search results are filtered to show travel-related information. The search bar shows "Egypt" and the results count is "About 321,000,000 results (0.15 seconds)".

Google Egypt

About 321,000,000 results (0.15 seconds)

Everything
Images
Videos
News
Shopping
Books
More

New York, NY
Change location

Any time
Latest
Past 24 hours
Past week
Past month
Past year
Custom range...

All results
Sites with images

Egypt - Wikipedia, the free encyclopedia
Egypt officially the Arab Republic of **Egypt**, is a country mainly in North Africa, with the Sinai Peninsula forming a land bridge in Southwest Asia. ...
Hosni Mubarak - Ancient Egypt - Female genital cutting - History of modern Egypt
en.wikipedia.org/wiki/Egypt - Cached - Similar

Egypt Travel, Tours, Vacations, Ancient Egypt from Tour Egypt
Information for travelers, resources on history, monuments and activities.
www.touregypt.net/ - Cached - Similar

Egypt Daily News, Egypt News
Egypt Daily News, covering Egypt News, Arab news, Middle East news and World news.
Egyptian Guides, egyptian recipes, egyptian food, egyptian airforce. ...
www.egyptdailynews.com/ - Cached - Similar

Images for egypt - Report images



Egypt - CIA - The World Factbook
Feb 1, 2011 ... Features a map and brief descriptions of geography, economy, government, and people.
<https://www.cia.gov/library/publications/the-world.../eg.html> - Cached - Similar

What You've Learned

- web directories
- web crawler
- graph (nodes, edges)
- random walk (ants)
- PageRank
- read pseudocode
- filter bubble

- [PageRank](#) by Felipe Micaroni Lalli
- screenshots of www.dmoz.org
- [Hyperlink](#) by Bernard Ladenthin
- screenshots of dontbubble.us
- [Sergey Brin](#) by enlewof
- [Larry Page](#) by aweigend

Thanks for Your Attention!

Days 1 - 10

Teach yourself variables, constants, arrays, strings, expressions, statements, functions,...



Days 11 - 21

Teach yourself program flow, pointers, references, classes, objects, inheritance, polymorphism,



Days 22 - 697

Do a lot of recreational programming. Have fun hacking but remember to learn from your mistakes.



Days 698 - 3648

Interact with other programmers. Work on programming projects together. Learn from them.



Days 3649 - 7781

Teach yourself advanced theoretical physics and formulate a consistent theory of quantum gravity.



Days 7782 - 14611

Teach yourself biochemistry, molecular biology, genetics,...



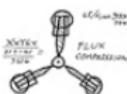
Day 14611

Use knowledge of biology to make an age-reversing potion.



Day 14611

Use knowledge of physics to build flux capacitor and go back in time to day 21.



Day 21

Replace younger self.



As far as I know, this is the easiest way to "Teach Yourself C++ in 21 Days".