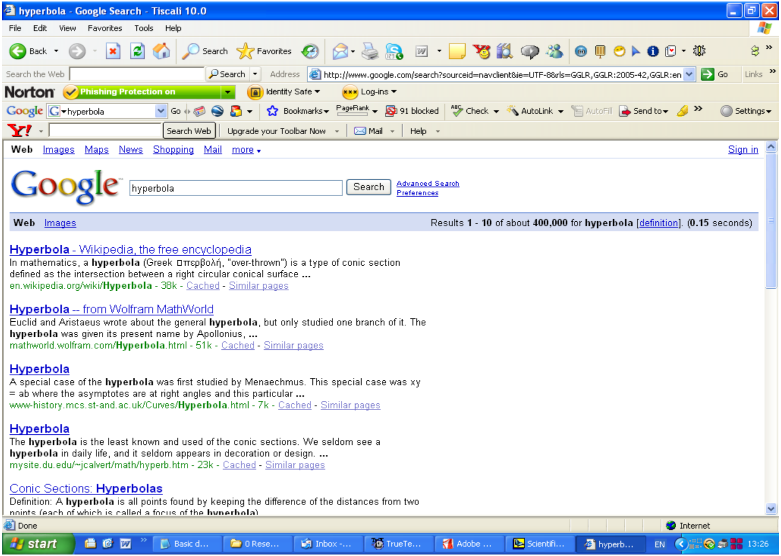
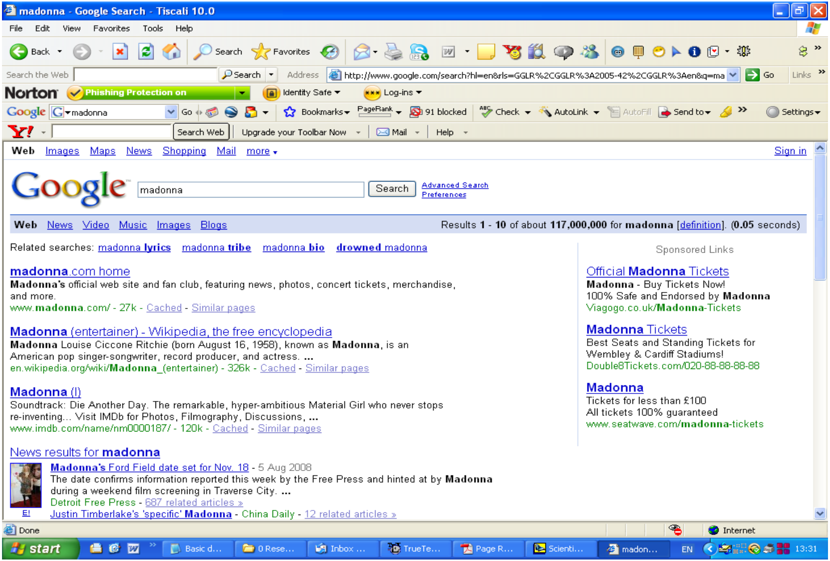
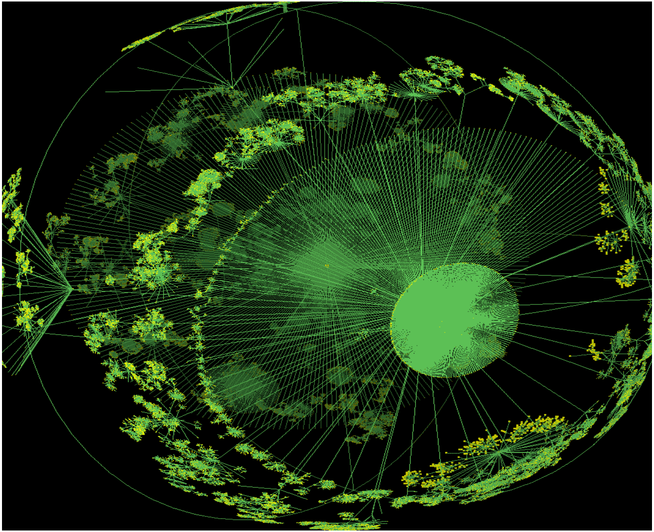
Google: An Application of Linear Algebra (The mathematics of a great success)

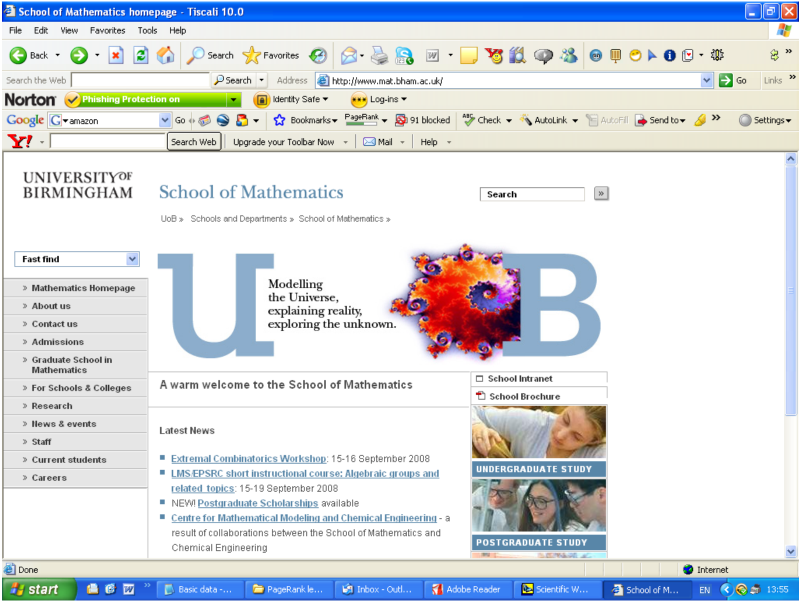
Peter Butkovic

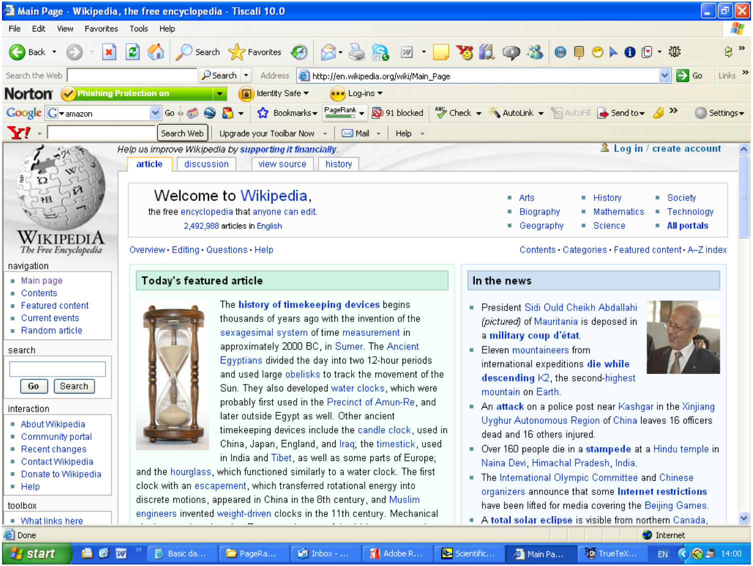

hyperbola - Google Search - Tiscali 10.0
File Edit View Favorites Tools Help
Back Search Favorites
Search the Web Search Address http://www.google.com/search?sourceid=navclient&ie=UTF-8&rls=GGLR,GGLR:2005-42,GGLR:en Go Links
Norton Phishing Protection on Identity Safe Log-ins
Google hyperbola Go Bookmarks PageRank 91 blocked Check AutoLink AutoFill Send to Settings
Y! Search Web Upgrade your Toolbar Now Mail Help
Web Images Maps News Shopping Mail more Sign in
Google
hyperbola Search Advanced Search Preferences
Web Images Results 1 - 10 of about 400,000 for hyperbola [definition]. (0.15 seconds)
Hyperbola - Wikipedia, the free encyclopedia
In mathematics, a hyperbola (Greek ὑπερβολή, "over-thrown") is a type of conic section defined as the intersection between a right circular conical surface ...
en.wikipedia.org/wiki/Hyperbola - 38k - Cached - Similar pages
Hyperbola -- from Wolfram MathWorld
Euclid and Aristaeus wrote about the general hyperbola, but only studied one branch of it. The hyperbola was given its present name by Apollonius, ...
mathworld.wolfram.com/Hyperbola.html - 51k - Cached - Similar pages
Hyperbola
A special case of the hyperbola was first studied by Menaechmus. This special case was xy = ab where the asymptotes are at right angles and this particular ...
www-history.mcs.st-and.ac.uk/Curves/Hyperbola.html - 7k - Cached - Similar pages
Hyperbola
The hyperbola is the least known and used of the conic sections. We seldom see a hyperbola in daily life, and it seldom appears in decoration or design. ...
mysite.du.edu/~jcalvert/math/hyperb.htm - 23k - Cached - Similar pages
Conic Sections: Hyperbolas
Definition: A hyperbola is all points found by keeping the difference of the distances from two
Done Internet
start Basic d... 0 Rese... Inbox -... TrueTe... Adobe ... Scientifi... hyperb... EN 13:26


madonna - Google Search - Tiscali 10.0
Web Images Maps News Shopping Mail more
Sign in
Google
madonna
Search
Advanced Search
Preferences
Web News Video Music Images Blogs
Results 1 - 10 of about 117,000,000 for madonna [definition]. (0.05 seconds)
Related searches: madonna lyrics madonna tribe madonna bio drowned madonna
madonna.com home
Madonna's official web site and fan club, featuring news, photos, concert tickets, merchandise, and more.
www.madonna.com/ - 27k - Cached - Similar pages
Madonna (entertainer) - Wikipedia, the free encyclopedia
Madonna Louise Ciccone Ritchie (born August 16, 1958), known as Madonna, is an American pop singer-songwriter, record producer, and actress. ...
en.wikipedia.org/wiki/Madonna_(entertainer) - 326k - Cached - Similar pages
Madonna (I)
Soundtrack: Die Another Day. The remarkable, hyper-ambitious Material Girl who never stops re-inventing... Visit IMDb for Photos, Filmography, Discussions, ...
www.imdb.com/name/nm0000187/ - 120k - Cached - Similar pages
News results for madonna
Madonna's Ford Field date set for Nov. 18 - 5 Aug 2008
The date confirms information reported this week by the Free Press and hinted at by Madonna during a weekend film screening in Traverse City. ...
Detroit Free Press - 687 related articles »
Justin Timberlake's 'specific' Madonna - China Daily - 12 related articles »
Sponsored Links
Official Madonna Tickets
Madonna - Buy Tickets Now!
100% Safe and Endorsed by Madonna
Viagogo.co.uk/Madonna-Tickets
Madonna Tickets
Best Seats and Standing Tickets for Wembley & Cardiff Stadiums!
Double8Tickets.com/020-88-88-88-88
Madonna
Tickets for less than £100
All tickets 100% guaranteed
www.seatwave.com/madonna-tickets
Done
Internet
13:31



Hyun's Map of the Web


School of Mathematics homepage - Tiscali 10.0
File Edit View Favorites Tools Help
http://www.mat.bham.ac.uk/
UNIVERSITY OF BIRMINGHAM
School of Mathematics
Search
UoB » Schools and Departments » School of Mathematics »
Fast find
Mathematics Homepage
About us
Contact us
Admissions
Graduate School in Mathematics
For Schools & Colleges
Research
News & events
Staff
Current students
Careers
Modelling the Universe, explaining reality, exploring the unknown.
A warm welcome to the School of Mathematics
School Intranet
School Brochure
Latest News
Extremal Combinatorics Workshop: 15-16 September 2008
LMS/EPSRC short instructional course: Algebraic groups and related topics: 15-19 September 2008
NEW! Postgraduate Scholarships available
Centre for Mathematical Modeling and Chemical Engineering - a result of collaborations between the School of Mathematics and Chemical Engineering
UNDERGRADUATE STUDY
POSTGRADUATE STUDY
Done
Internet
start
13:55


Main Page - Wikipedia, the free encyclopedia - Tiscali 10.0
File Edit View Favorites Tools Help
Back Search Favorites
Search the Web Search Address http://en.wikipedia.org/wiki/Main_Page Go Links
Norton Phishing Protection on Identity Safe Log-ins
Google amazon Go Bookmarks PageRank 91 blocked Check AutoLink AutoFill Send to Settings
Search Web Upgrade your Toolbar Now Mail Help
Help us improve Wikipedia by supporting it financially.
Log in / create account
article discussion view source history
WIKIPEDIA
The Free Encyclopedia
navigation
Main page
Contents
Featured content
Current events
Random article
search
Go Search
interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Donate to Wikipedia
Help
toolbox
What links here
Welcome to Wikipedia,
the free encyclopedia that anyone can edit.
2,492,988 articles in English
Arts History Society
Biography Mathematics Technology
Geography Science All portals
Overview · Editing · Questions · Help
Contents · Categories · Featured content · A–Z index
Today's featured article
The history of timekeeping devices begins thousands of years ago with the invention of the sexagesimal system of time measurement in approximately 2000 BC, in Sumer. The Ancient Egyptians divided the day into two 12-hour periods and used large obelisks to track the movement of the Sun. They also developed water clocks, which were probably first used in the Precinct of Amun-Re, and later outside Egypt as well. Other ancient timekeeping devices include the candle clock, used in China, Japan, England, and Iraq; the timestick, used in India and Tibet, as well as some parts of Europe; and the hourglass, which functioned similarly to a water clock. The first clock with an escapement, which transferred rotational energy into discrete motions, appeared in China in the 8th century, and Muslim engineers invented weight-driven clocks in the 11th century. Mechanical
In the news
President Sidi Ould Cheikh Abdallahi (pictured) of Mauritania is deposed in a military coup d'état.
Eleven mountaineers from international expeditions die while descending K2, the second-highest mountain on Earth.
An attack on a police post near Kashgar in the Xinjiang Uyghur Autonomous Region of China leaves 16 officers dead and 16 others injured.
Over 160 people die in a stampede at a Hindu temple in Naina Devi, Himachal Pradesh, India.
The International Olympic Committee and Chinese organizers announce that some Internet restrictions have been lifted for media covering the Beijing Games.
A total solar eclipse is visible from northern Canada,
Done
Internet
start Basic da... PageRa... Inbox - ... Adobe R... Scientific... Main Pa... TrueTeX... EN 14:00

- Google (~googol $= 10^{100}$)

- Google (~googol $= 10^{100}$)
- Founders of Google: Larry Page and Sergey Brin

- Google (~googol $= 10^{100}$)
- Founders of Google: Larry Page and Sergey Brin
- 1995: Research students at Stanford

- Google (~googol $= 10^{100}$)
- Founders of Google: Larry Page and Sergey Brin
- 1995: Research students at Stanford
- 1996: Started a student project on search engines

- Google (~googol $= 10^{100}$)
- Founders of Google: Larry Page and Sergey Brin
- 1995: Research students at Stanford
- 1996: Started a student project on search engines
- 1998: Google incorporates as a company (initial investment: \$1.1million) and files patent for PageRank

- 2000: Selling advertisements began

- 2000: Selling advertisements began
- 2001: Patent granted

- 2000: Selling advertisements began
- 2001: Patent granted
- 2004: Total capital reached $23billion

- 2000: Selling advertisements began
- 2001: Patent granted
- 2004: Total capital reached \$23billion
- 2007: Google web search engine covers 53% of the market (Yahoo 20%)

- 2000: Selling advertisements began
- 2001: Patent granted
- 2004: Total capital reached \$23billion
- 2007: Google web search engine covers 53% of the market (Yahoo 20%)
- 2008: Google uses 450,000 servers in 25 locations around the world to index $>$ 8 billion websites

KEY IDEAS

- Each webpage is assigned a measure of importance ... PageRank

KEY IDEAS

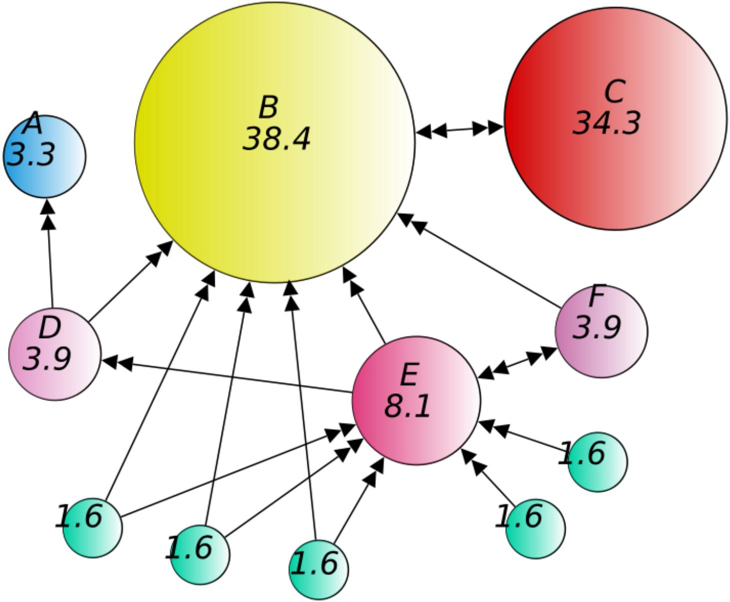
- Each webpage is assigned a measure of importance ... PageRank
- A hyperlink is a "recommendation": A webpage is important if other webpages point to it (via hyperlinks)

KEY IDEAS

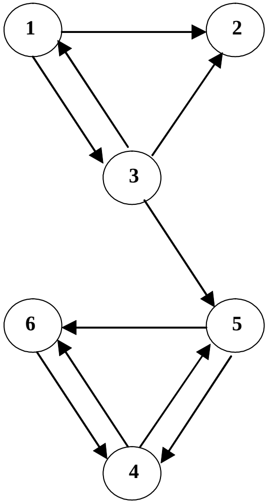
- Each webpage is assigned a measure of importance ... PageRank
- A hyperlink is a "recommendation": A webpage is important if other webpages point to it (via hyperlinks)
- The PageRank of the recommender matters (the higher the better)

KEY IDEAS

- Each webpage is assigned a measure of importance ... PageRank
- A hyperlink is a "recommendation": A webpage is important if other webpages point to it (via hyperlinks)
- The PageRank of the recommender matters (the higher the better)
- The generosity of the recommender matters (the higher the worse)


A
3.3
B
38.4
C
34.3
D
3.9
E
8.1
F
3.9
1.6
1.6
1.6
1.6
1.6

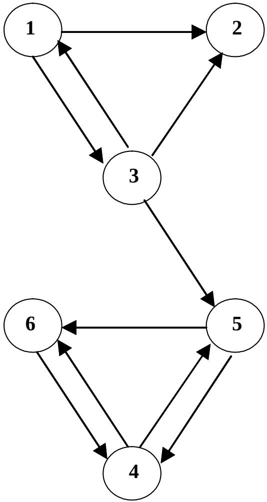
A.N.Langville and C.D.Meyer: Google's PageRank and Beyond (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$

Node 2 - *dangling*

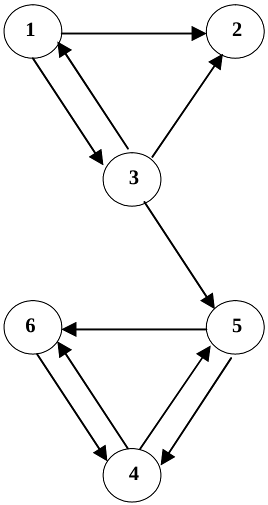
A.N.Langville and C.D.Meyer: *Google's PageRank and Beyond* (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$
- $r(P_2) = \frac{1}{2} r(P_1) + \frac{1}{3} r(P_3)$

Node 2 - *dangling*

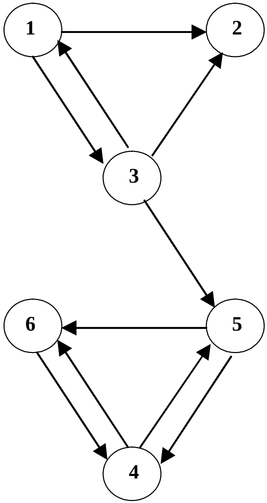
A.N.Langville and C.D.Meyer: Google's PageRank and Beyond (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$
- $r(P_2) = \frac{1}{2} r(P_1) + \frac{1}{3} r(P_3)$
- $r(P_3) = \frac{1}{2} r(P_1)$

Node 2 - *dangling*

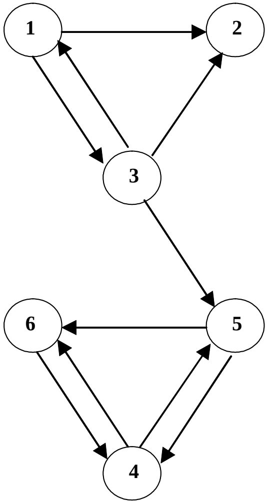
A.N.Langville and C.D.Meyer: Google's PageRank and Beyond (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$
- $r(P_2) = \frac{1}{2} r(P_1) + \frac{1}{3} r(P_3)$
- $r(P_3) = \frac{1}{2} r(P_1)$
- $r(P_4) = \frac{1}{2} r(P_5) + r(P_6)$

Node 2 - *dangling*

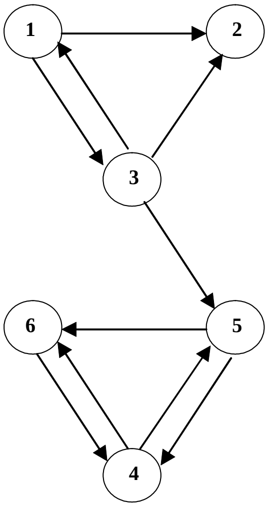
A.N.Langville and C.D.Meyer: Google's PageRank and Beyond (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$
- $r(P_2) = \frac{1}{2} r(P_1) + \frac{1}{3} r(P_3)$
- $r(P_3) = \frac{1}{2} r(P_1)$
- $r(P_4) = \frac{1}{2} r(P_5) + r(P_6)$
- $r(P_5) = \frac{1}{3} r(P_3) + \frac{1}{2} r(P_4)$

Node 2 - *dangling*

A.N.Langville and C.D.Meyer: Google's PageRank and Beyond (PUP 2006)



- $r(P_1) = \frac{1}{3} r(P_3)$
- $r(P_2) = \frac{1}{2} r(P_1) + \frac{1}{3} r(P_3)$
- $r(P_3) = \frac{1}{2} r(P_1)$
- $r(P_4) = \frac{1}{2} r(P_5) + r(P_6)$
- $r(P_5) = \frac{1}{3} r(P_3) + \frac{1}{2} r(P_4)$
- $r(P_6) = \frac{1}{2} r(P_4) + \frac{1}{2} r(P_5)$

Node 2 - *dangling*

- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots\right)$

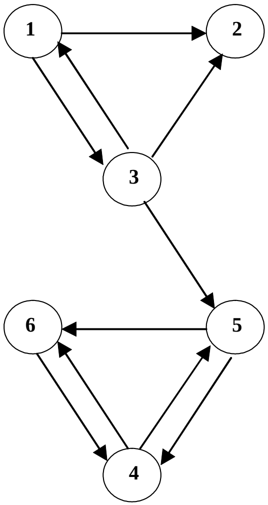
- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots\right)$
- $r_0 \longrightarrow r_1$

- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots\right)$
- $r_0 \longrightarrow r_1$
- $r_1 \longrightarrow r_2$

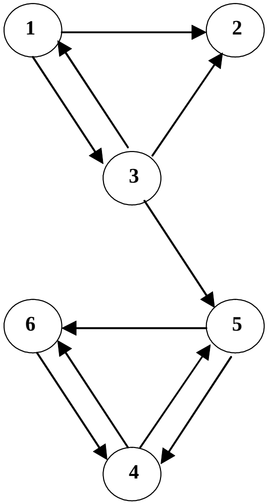
- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots\right)$
- $r_0 \longrightarrow r_1$
- $r_1 \longrightarrow r_2$
- ...

- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \dots\right)$
- $r_0 \longrightarrow r_1$
- $r_1 \longrightarrow r_2$
- ...
- $r_k \longrightarrow r_{k+1}$

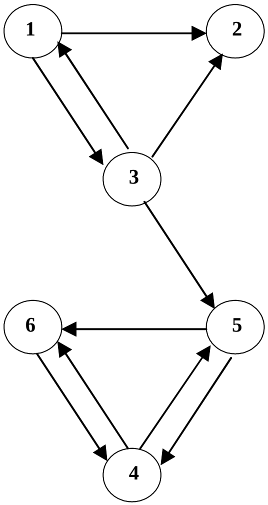
- $r_0 = \left(\frac{1}{n}, \frac{1}{n}, \ldots\right)$
- $r_0 \longrightarrow r_1$
- $r_1 \longrightarrow r_2$
- $\ldots$
- $r_k \longrightarrow r_{k+1}$
- $\ldots$



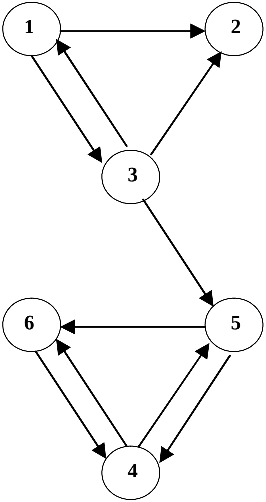
- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$



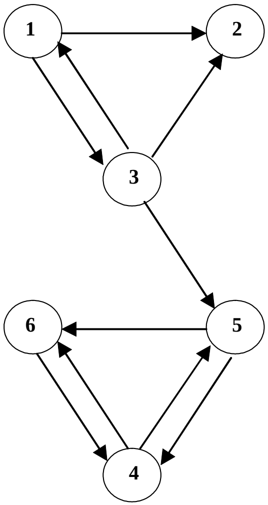
- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1\left(P_1\right) = \frac{1}{3} r_0\left(P_3\right) = \frac{1}{18}$



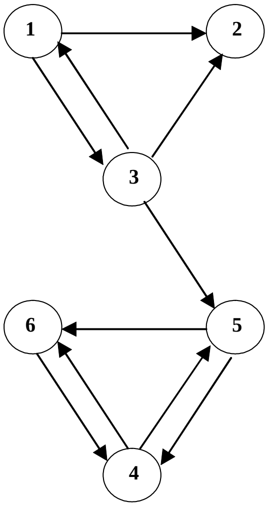
- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1(P_1) = \frac{1}{3} r_0(P_3) = \frac{1}{18}$
- $r_1(P_2) = \frac{1}{2} r_0(P_1) + \frac{1}{3} r_0(P_3) = \frac{5}{36}$



- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1(P_1) = \frac{1}{3} r_0(P_3) = \frac{1}{18}$
- $r_1(P_2) =$
  $\frac{1}{2} r_0(P_1) + \frac{1}{3} r_0(P_3) = \frac{5}{36}$
- $r_1(P_3) = \frac{1}{2} r_0(P_1) = \frac{1}{12}$

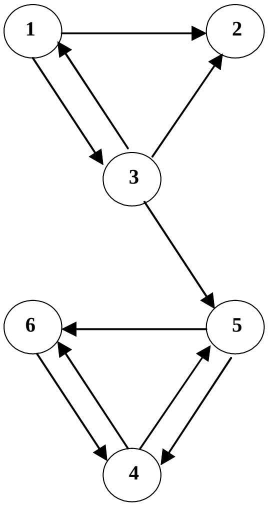


- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1\left(P_1\right) = \frac{1}{3} r_0\left(P_3\right) = \frac{1}{18}$
- $r_1\left(P_2\right) = \frac{1}{2} r_0\left(P_1\right) + \frac{1}{3} r_0\left(P_3\right) = \frac{5}{36}$
- $r_1\left(P_3\right) = \frac{1}{2} r_0\left(P_1\right) = \frac{1}{12}$
- $r_1\left(P_4\right) = \frac{1}{2} r_0\left(P_5\right) + r_0\left(P_6\right) = \frac{1}{4}$



- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1\left(P_1\right) = \frac{1}{3} r_0\left(P_3\right) = \frac{1}{18}$

- $r_1\left(P_2\right) = \frac{1}{2} r_0\left(P_1\right) + \frac{1}{3} r_0\left(P_3\right) = \frac{5}{36}$
- $r_1\left(P_3\right) = \frac{1}{2} r_0\left(P_1\right) = \frac{1}{12}$
- $r_1\left(P_4\right) = \frac{1}{2} r_0\left(P_5\right) + r_0\left(P_6\right) = \frac{1}{4}$
- $r_1\left(P_5\right) = \frac{1}{3} r_0\left(P_3\right) + \frac{1}{2} r_0\left(P_4\right) = \frac{5}{36}$



- $r_0 = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$
- $r_1(P_1) = \frac{1}{3} r_0(P_3) = \frac{1}{18}$
- $r_1(P_2) = \frac{1}{2} r_0(P_1) + \frac{1}{3} r_0(P_3) = \frac{5}{36}$
- $r_1(P_3) = \frac{1}{2} r_0(P_1) = \frac{1}{12}$
- $r_1(P_4) = \frac{1}{2} r_0(P_5) + r_0(P_6) = \frac{1}{4}$
- $r_1(P_5) = \frac{1}{3} r_0(P_3) + \frac{1}{2} r_0(P_4) = \frac{5}{36}$
- $r_1(P_6) = \frac{1}{2} r_0(P_4) + \frac{1}{2} r_0(P_5) = \frac{1}{6}$

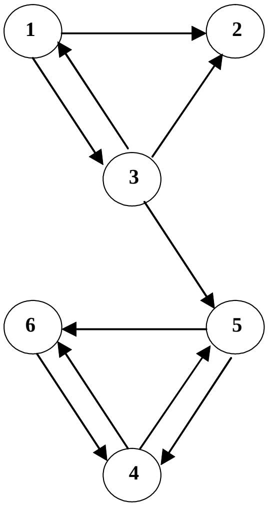
	r_0	r_1
P_1	$1/6$	$1/18$
P_2	$1/6$	$5/36$
P_3	$1/6$	$1/12$
P_4	$1/6$	$1/4$
P_5	$1/6$	$5/36$
P_6	$1/6$	$1/6$

	r_0	r_1	rank
P_1	$1/6$	$1/18$	6
P_2	$1/6$	$5/36$	$3-4$
P_3	$1/6$	$1/12$	5
P_4	$1/6$	$1/4$	1
P_5	$1/6$	$5/36$	$3-4$
P_6	$1/6$	$1/6$	2

	r_0	r_1	rank	r_2
P_1	$1/6$	$1/18$	6	$1/36$
P_2	$1/6$	$5/36$	$3-4$	$1/18$
P_3	$1/6$	$1/12$	5	$1/36$
P_4	$1/6$	$1/4$	1	$17/72$
P_5	$1/6$	$5/36$	$3-4$	$11/72$
P_6	$1/6$	$1/6$	2	$14/72$

	r_0	r_1	rank	r_2	rank
P_1	$1/6$	$1/18$	6	$1/36$	$5-6$
P_2	$1/6$	$5/36$	$3-4$	$1/18$	4
P_3	$1/6$	$1/12$	5	$1/36$	$5-6$
P_4	$1/6$	$1/4$	1	$17/72$	1
P_5	$1/6$	$5/36$	$3-4$	$11/72$	3
P_6	$1/6$	$1/6$	2	$14/72$	2

	r_0	r_1	rank	r_2	rank
P_1	$1/6$	$1/18$	6	$1/36$	$5-6$
P_2	$1/6$	$5/36$	$3-4$	$1/18$	4
P_3	$1/6$	$1/12$	5	$1/36$	$5-6$
P_4	$1/6$	$1/4$	1	$17/72$	1
P_5	$1/6$	$5/36$	$3-4$	$11/72$	3
P_6	$1/6$	$1/6$	2	$14/72$	2



Hyperlink matrix H :

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
P_2	0	0	0	0	0	0
P_3	$\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{1}{3}$	0
P_4	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
P_5	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
P_6	0	0	0	1	0	0

Stochastic matrix: Every row is ≥ 0 and sums to 1.
H could be a stochastic matrix if it was not for the rows corresponding to the dangling nodes

- $$\underbrace{\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)}_{r_0}\cdot\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

$$=\underbrace{\left(\frac{1}{18},\frac{5}{36},\frac{1}{12},\frac{1}{4},\frac{5}{36},\frac{1}{6}\right)}_{r_1}$$

- $$\underbrace{\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)}_{r_0}\cdot\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$
  $$=\underbrace{\left(\frac{1}{18},\frac{5}{36},\frac{1}{12},\frac{1}{4},\frac{5}{36},\frac{1}{6}\right)}_{r_1}$$
- $$r_1 = r_0 \cdot H$$

- $$\underbrace{\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)}_{r_0}\cdot\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$
$$= \underbrace{\left(\frac{1}{18},\frac{5}{36},\frac{1}{12},\frac{1}{4},\frac{5}{36},\frac{1}{6}\right)}_{r_1}$$
- $$r_1 = r_0 \cdot H$$
- Similarly:
$$r_2 = r_1 \cdot H = (r_0 \cdot H) \cdot H = r_0 \cdot H^2$$

- $$r_3 = r_2 \cdot H = ... = r_0 \cdot H^3$$

- $$r_3 = r_2 \cdot H = ... = r_0 \cdot H^3$$
- In general:
$$r_k = r_{k-1} \cdot H = ... = r_0 \cdot H^k$$

- $$r_3 = r_2 \cdot H = \ldots = r_0 \cdot H^3$$
- In general:
  $$r_k = r_{k-1} \cdot H = \ldots = r_0 \cdot H^k$$
- ... "Power Method"

QUESTIONS

- Will this power method converge? If not what conditions must be satisfied?

QUESTIONS

- Will this power method converge? If not what conditions must be satisfied?
- If it converges, will it do so to a vector meaningful for page ranking?

QUESTIONS

- Will this power method converge? If not what conditions must be satisfied?
- If it converges, will it do so to a vector meaningful for page ranking?
- Does the convergence depend on the starting vector?

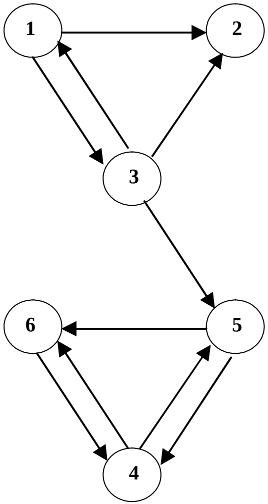
QUESTIONS

- Will this power method converge? If not what conditions must be satisfied?
- If it converges, will it do so to a vector meaningful for page ranking?
- Does the convergence depend on the starting vector?
- Does the limit depend on the starting vector?

QUESTIONS

- Will this power method converge? If not what conditions must be satisfied?
- If it converges, will it do so to a vector meaningful for page ranking?
- Does the convergence depend on the starting vector?
- Does the limit depend on the starting vector?
- *If it converges, how long does it take?*

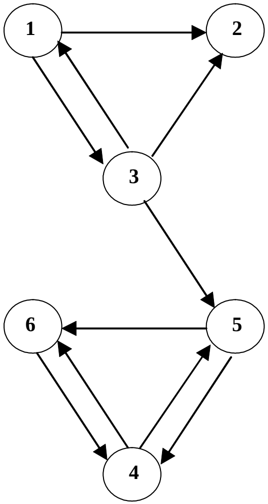
Page and Brin first had to deal with a number of problems:
Rank sink pages



-

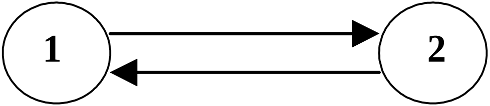
	r_0	r_1	...	r_{13}
P_1	1/6	1/18	...	0
P_2	1/6	5/36	...	0
P_3	1/6	1/12	...	0
P_4	1/6	1/4	...	2/3
P_5	1/6	5/36	...	1/3
P_6	1/6	1/6	...	1/5

Page and Brin first had to deal with a number of problems:
Rank sink pages

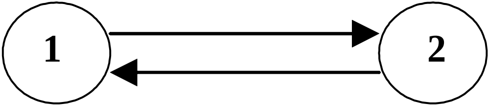


	r_0	r_1	...	r_{13}
P_1	1/6	1/18	...	0
P_2	1/6	5/36	...	0
P_3	1/6	1/12	...	0
P_4	1/6	1/4	...	2/3
P_5	1/6	5/36	...	1/3
P_6	1/6	1/6	...	1/5
- Nodes 4,5,6 are a *link farm*.

- Another problem: *Cycles*

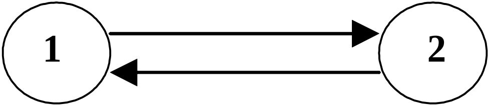


- Another problem: *Cycles*



- $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

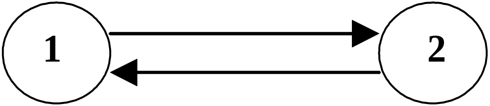
- Another problem: *Cycles*



- $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

- $(1,0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (0,1)$

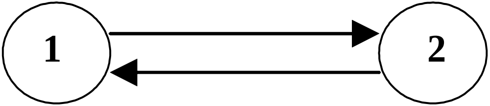
- Another problem: *Cycles*



- $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

- $(1,0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (0,1)$
- $(0,1)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (1,0)$

- Another problem: *Cycles*



- $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
- $(1,0)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (0,1)$
- $(0,1)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (1,0)$
- Flip-flop $\longrightarrow$ no convergence

Elements of the Markov chains theory:

- If H is a stochastic matrix and x_0 a stochastic vector then the sequence

$$\{x_0, x_1 = x_0 H, x_2 = x_1 H, x_3 = x_2 H, ...\}$$

is called a *Markov chain*.

Elements of the Markov chains theory:

- If H is a stochastic matrix and x_0 a stochastic vector then the sequence

$$\{x_0, x_1 = x_0 H, x_2 = x_1 H, x_3 = x_2 H, ...\}$$

  is called a *Markov chain*.
- H is called the *transition probability matrix*.

Theorem (Markov, 1906)

If H is a **positive** *transition probability matrix of a Markov chain then this chain converges to a unique positive vector (called stationary vector) independently of the starting vector.*

- Brin and Page: Adjustments to the basic model using the concept of a *random surfer*.

- Brin and Page: Adjustments to the basic model using the concept of a *random surfer*.
- **Adjustment 1** (*stochasticity*): Zero rows in H (corresponding to the dangling nodes) are replaced by

$$\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

The new matrix is S.

- Brin and Page: Adjustments to the basic model using the concept of a *random surfer*.
- **Adjustment 1** (*stochasticity*): Zero rows in H (corresponding to the dangling nodes) are replaced by

$$\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

The new matrix is S.
- Our example:

$$H \to S = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

- Brin and Page: Adjustments to the basic model using the concept of a *random surfer*.
- **Adjustment 1** (*stochasticity*): Zero rows in H (corresponding to the dangling nodes) are replaced by

$$\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

  The new matrix is S.
- Our example:

$$H \to S = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

- S is stochastic! (But not yet positive)

- **Adjustment 2** (*primitivity*): The random surfer (RS) from time to time moves to a new webpage without using a hyperlink (he "teleports") and then he resumes hyperlink surfing again.

- **Adjustment 2** (*primitivity*): The random surfer (RS) from time to time moves to a new webpage without using a hyperlink (he "teleports") and then he resumes hyperlink surfing again.
- $$S \longrightarrow G = \alpha S + (1 - \alpha) E$$

- **Adjustment 2** (*primitivity*): The random surfer (RS) from time to time moves to a new webpage without using a hyperlink (he "teleports") and then he resumes hyperlink surfing again.
-
$$S \rightarrow G = \alpha S + (1-\alpha) E$$
- where
$$E = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

- **Adjustment 2** (*primitivity*): The random surfer (RS) from time to time moves to a new webpage without using a hyperlink (he "teleports") and then he resumes hyperlink surfing again.
- $$S \rightarrow G = \alpha S + (1-\alpha) E$$
- where
$$E = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$
- and $\alpha \in (0,1)$ controls the proportion of the time RS follows the *hyperlinks* as opposed to *teleportation*

- $G = \alpha S + (1 - \alpha) E$ is called the *Google matrix*: it is stochastic and positive!

- $G = \alpha S + (1 - \alpha) E$ is called the *Google matrix*: it is stochastic and positive!
- Hence any Markov chain with G is guaranteed to converge to a unique positive vector, independently of the starting vector.

- $G = \alpha S + (1-\alpha) E$ is called the *Google matrix*: it is stochastic and positive!
- Hence any Markov chain with G is guaranteed to converge to a unique positive vector, independently of the starting vector.
- Actually used $\alpha \approx .85$

- Our example for $\alpha = 0.9$:

- Our example for $\alpha = 0.9$:
- $G = 0.9S + 0.1E = \frac{1}{60}\begin{pmatrix} 1 & 28 & 28 & 1 & 1 & 1 \\ 10 & 10 & 10 & 10 & 10 & 10 \\ 19 & 19 & 1 & 1 & 19 & 1 \\ 1 & 1 & 1 & 1 & 28 & 28 \\ 1 & 1 & 1 & 28 & 1 & 28 \\ 1 & 1 & 1 & 55 & 1 & 1 \end{pmatrix}$

- Google's PageRank vector is the stationary vector of G which is

$$(.03721, .05396, .04151, .3751, .206, .2802)$$

- Google's PageRank vector is the stationary vector of G which is
$$(.03721, .05396, .04151, .3751, .206, .2802)$$
- Random surfer spends 3.721% of their time on P_1, etc...

- Google's PageRank vector is the stationary vector of G which is
$$(.03721, .05396, .04151, .3751, .206, .2802)$$
- Random surfer spends 3.721% of their time on P_1, etc...
- The importance ranking therefore is: $P_4, P_6, P_5, P_2, P_3, P_1$.

- Will this iterative process converge? YES

- Will this iterative process converge? YES
- If it converges, will it do so to a vector meaningful for page ranking? YES

- Will this iterative process converge? YES
- If it converges, will it do so to a vector meaningful for page ranking? YES
- Does the convergence depend on the starting vector? NO

- Will this iterative process converge? YES
- If it converges, will it do so to a vector meaningful for page ranking? YES
- Does the convergence depend on the starting vector? NO
- Does the limit depend on the starting vector? NO

- Will this iterative process converge? YES
- If it converges, will it do so to a vector meaningful for page ranking? YES
- Does the convergence depend on the starting vector? NO
- Does the limit depend on the starting vector? NO
- *If it converges, how long does it take?*

- $r_k \cdot G = r_{k+1}$

- $r_k \cdot G = r_{k+1}$
- If $r_k \rightarrow r$ then $r \cdot G = r$

- $r_k \cdot G = r_{k+1}$
- If $r_k \to r$ then $r \cdot G = r$
- In general: $x \cdot M = \lambda x$

- $r_k \cdot G = r_{k+1}$
- If $r_k \longrightarrow r$ then $r \cdot G = r$
- In general: $x \cdot M = \lambda x$
- λ ... *eigenvalue* of M

- $r_k \cdot G = r_{k+1}$
- If $r_k \to r$ then $r \cdot G = r$
- In general: $x \cdot M = \lambda x$
- λ ... *eigenvalue* of M
- x ... *eigenvector* of M (if $x \neq 0$) corresponding to the eigenvalue λ

- $r_k \cdot G = r_{k+1}$
- If $r_k \to r$ then $r \cdot G = r$
- In general: $x \cdot M = \lambda x$
- λ ... *eigenvalue* of M
- x ... *eigenvector* of M (if $x \neq 0$) corresponding to the eigenvalue λ
- The stationary point of any Markov chain with transition matrix G is an eigenvector of G corresponding to the eigenvalue 1

Perron-Frobenius theory of Linear Algebra solves the eigenvector-eigenvalue problem for non-negative matrices.

Theorem (Perron, 1912)

If G is a positive, stochastic matrix then $\lambda = 1$ is an eigenvalue of G and G has a unique positive eigenvector (up to multiples).

- If G is a Google matrix - the power method needs roughly 50 iterations

- If G is a Google matrix - the power method needs roughly 50 iterations
- In each iteration $\approx n$ arithmetic operations

- If G is a Google matrix - the power method needs roughly 50 iterations
- In each iteration $\approx n$ arithmetic operations
- Total: $\approx 50n$ operations (recall $n > 8$ billion)

- If G is a Google matrix - the power method needs roughly 50 iterations
- In each iteration $\approx n$ arithmetic operations
- Total: $\approx 50n$ operations (recall $n > 8$ billion)
- CONCLUSION: The power method with Google matrix is *very* fast!

THANK YOU

Next session in this room at 12.00:

"Careers, degrees and mathematics"