ICME Fundamentals of Data Science: Introduction to Linear Algebra

Day 1: Fundamentals of vectors, matrices and their operations motivated by ranking and searching.

- Vectors: vector multiplication (inner product), cosine rule, vector norms
- Matrices: matrix-vector multiplication, matrix-vector systems, matrix-matrix multiplication, matrices as operators, orthogonal matrices, inverses and determinants, and a bit about eigenvalues and eigenvectors

Day 2: Diving deeper with as ultimate goal the SVD and linear regression.

- Matrices: Singular value decomposition, least squares, normal equations, QR decomposition

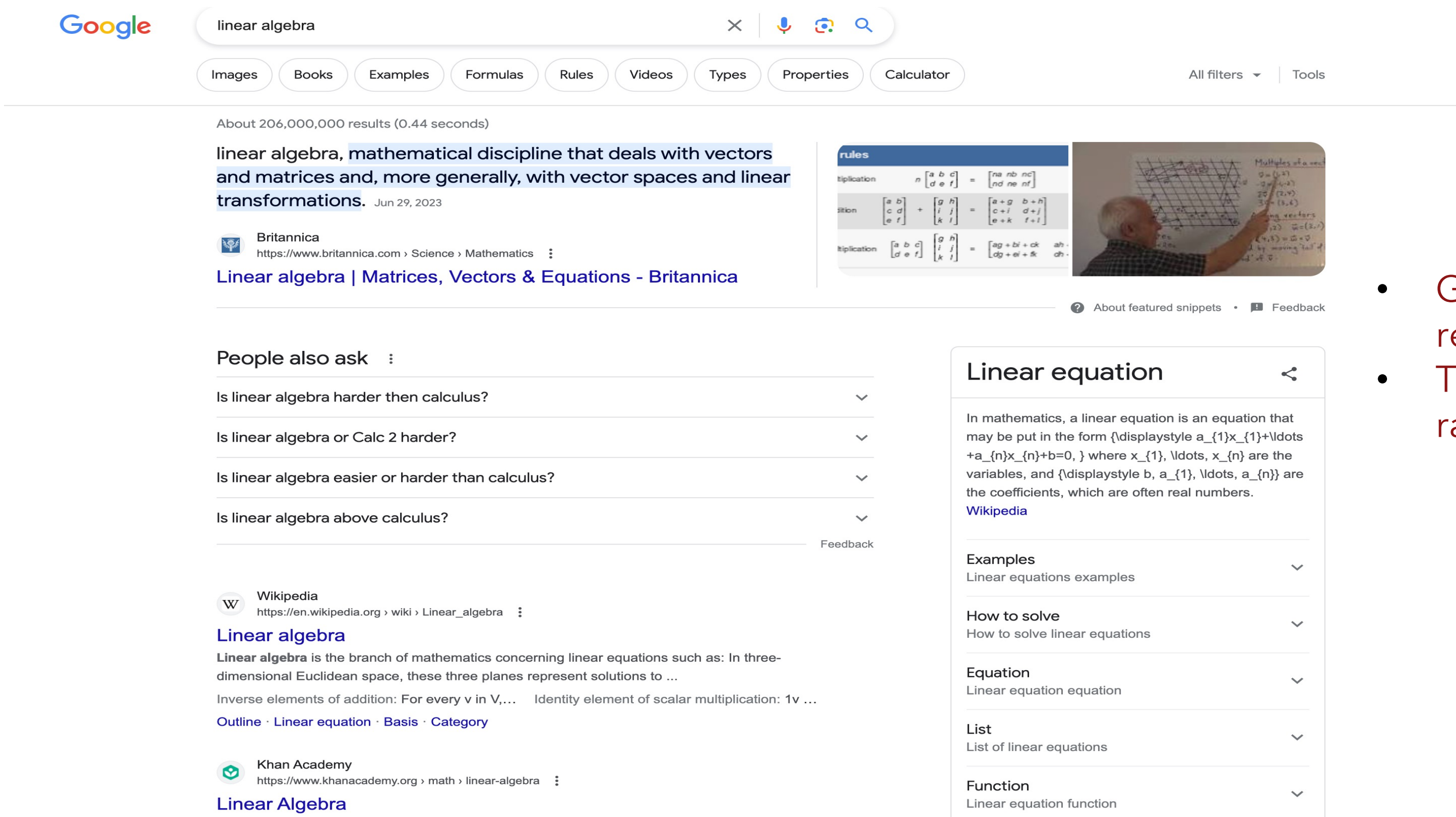
Approximate Schedule

Day 1	
$8 - 8{:}15$	Introduction
$8{:}15 - 9$	New material
$9 - 9{:}30$	Practice
$9{:}30 - 9{:}45$	Break
$9{:}45 - 10{:}30$	New material
$10{:}30 - 11$	Practice

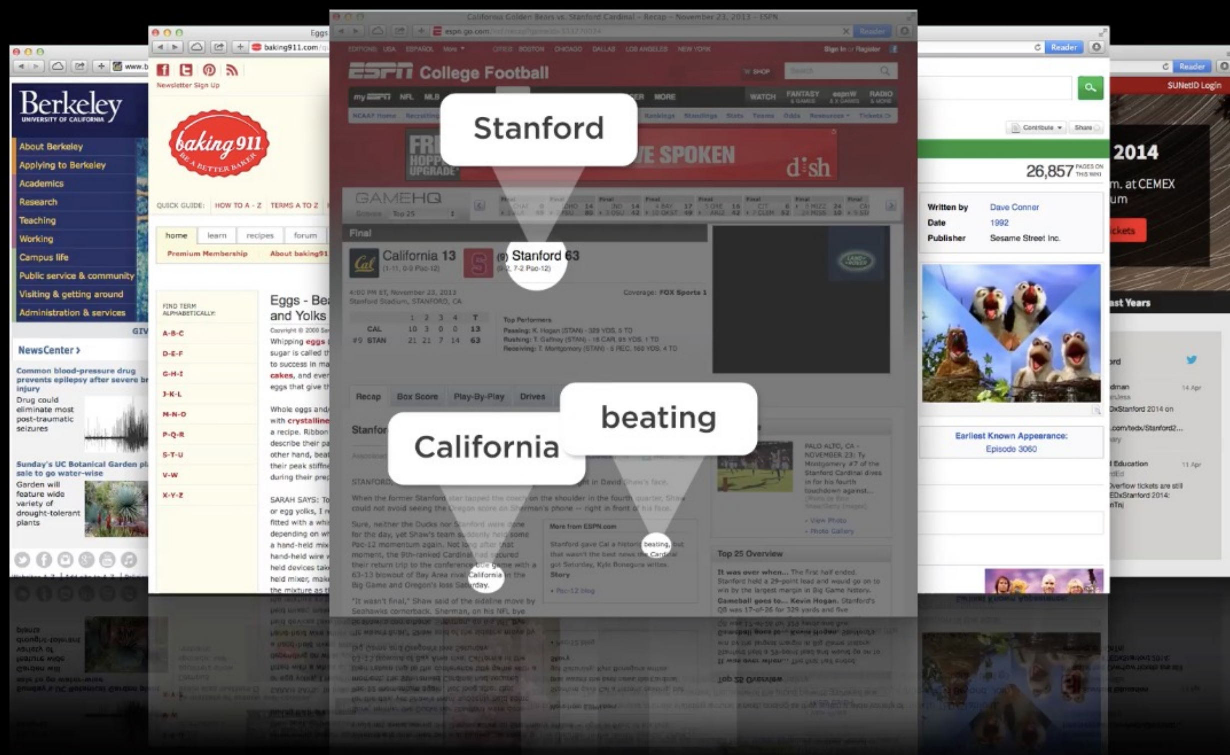
Day 2	
$8 - 9$	New material
$9 - 9{:}30$	Practice
$9{:}30 - 9{:}45$	Break
$9{:}45 - 10{:}30$	New material
$10{:}30 - 10{:}50$	Practice
$10{:}50 - 11$	Closing

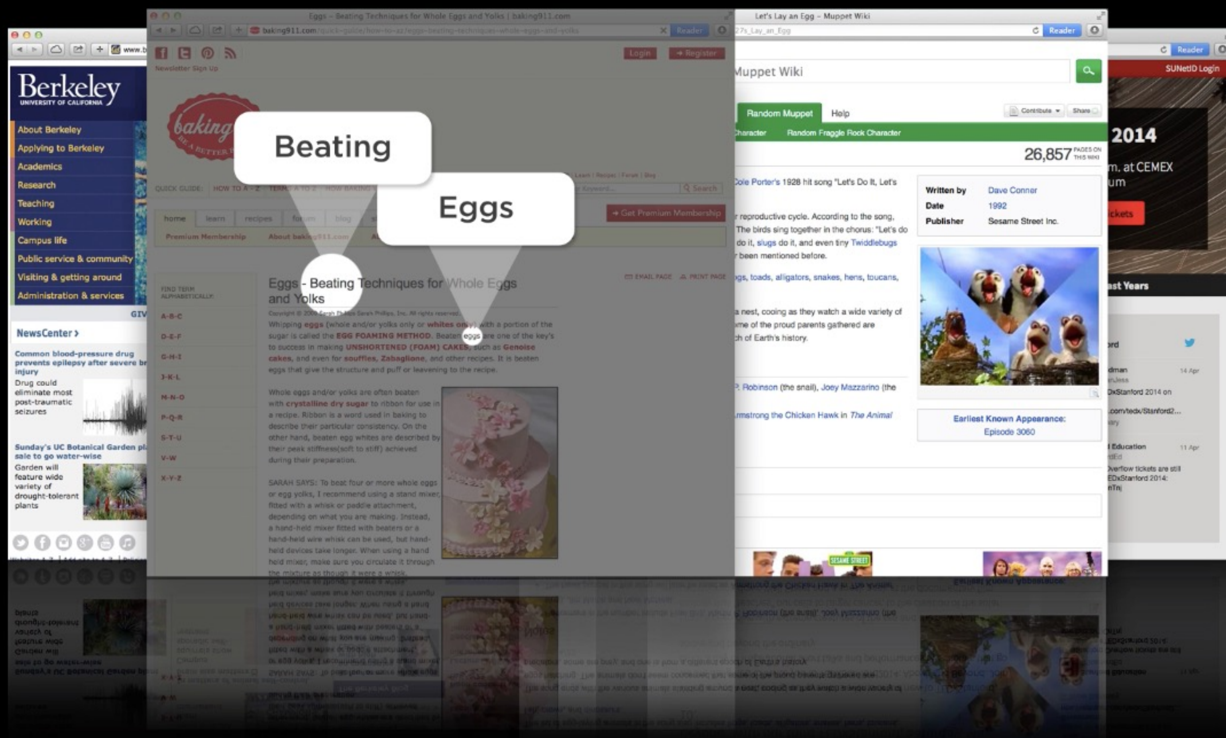
Slides after this are used throughout the short course

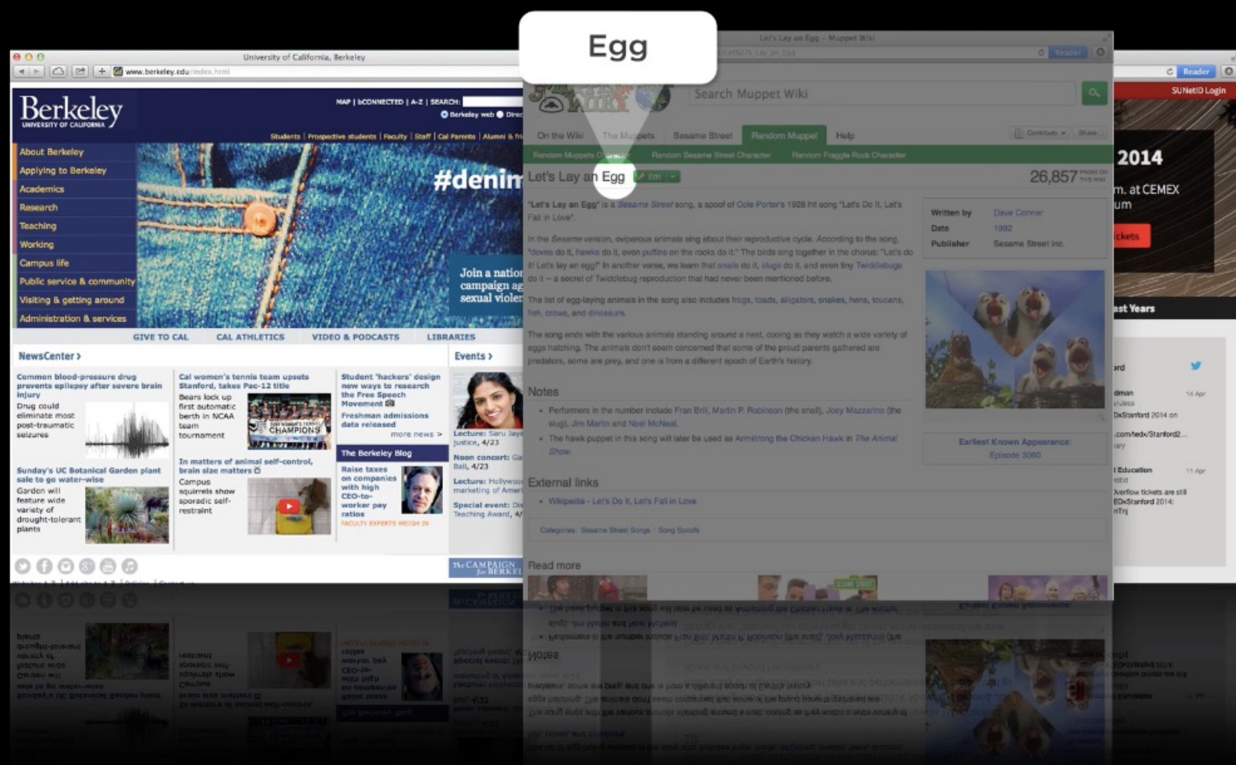
Google Search Results

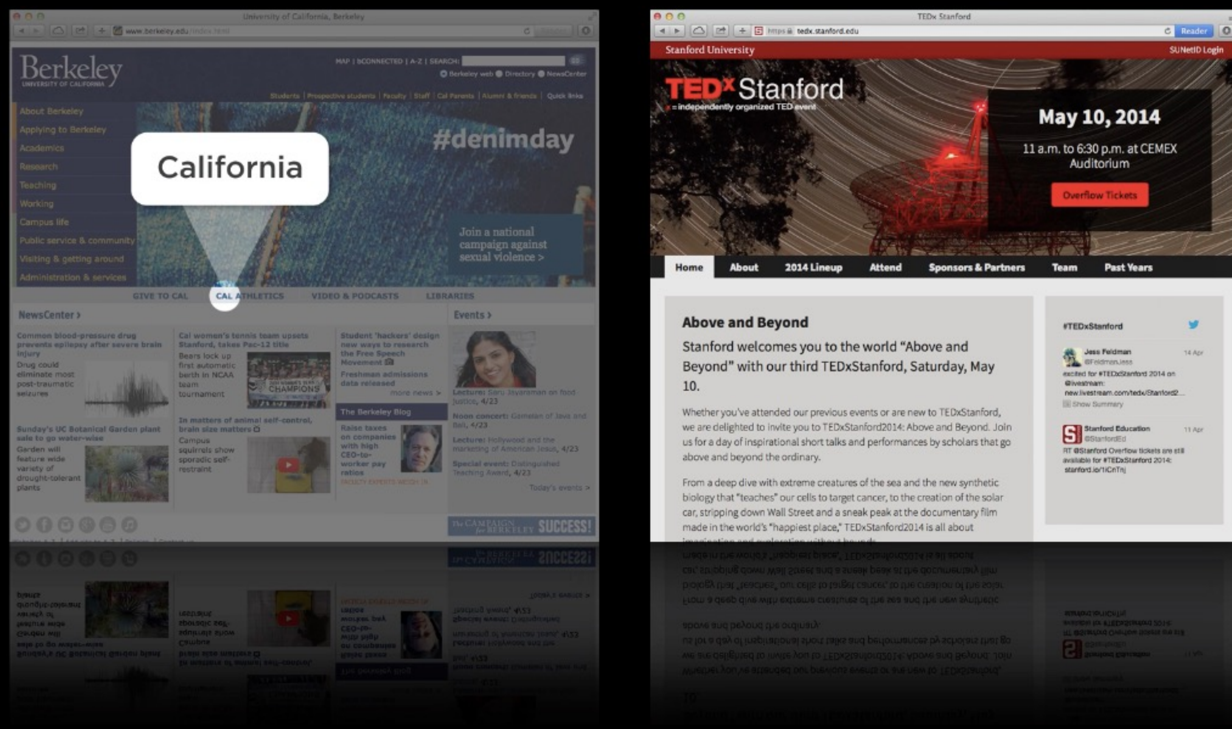


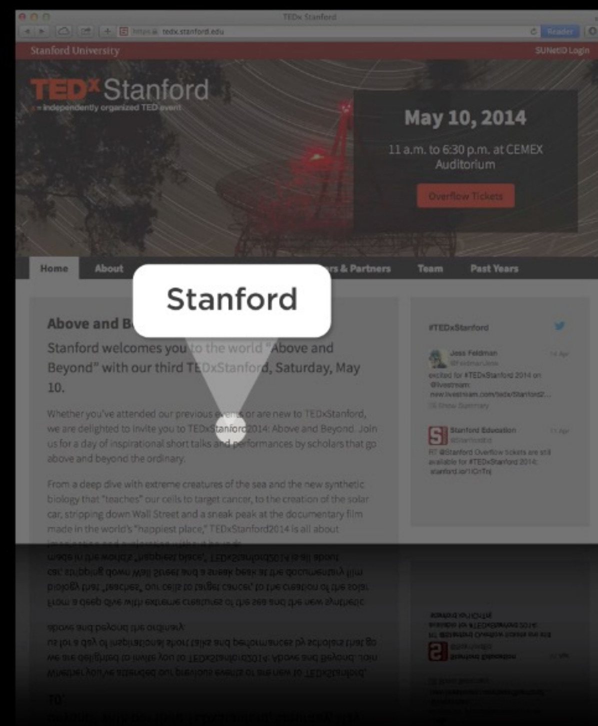
- Google returns exact and related matches
- The resulting pages are ranked in some order


Stanford
beating
California


Beating
Eggs
Berkeley
Eggs - Beating Techniques for Whole Eggs and Yolks
Muppet Wiki


Egg
Berkeley
UNIVERSITY OF CALIFORNIA
#denim
Search Muppet Wiki
Let's Lay an Egg
26,857
2014


Berkeley
UNIVERSITY OF CALIFORNIA
California
#denimday
Join a national campaign against sexual violence >
Stanford University
SUNetID Login
TEDx Stanford
x = independently organized TED event
May 10, 2014
11 a.m. to 6:30 p.m. at CEMEX Auditorium
Overflow Tickets
Home
About
2014 Lineup
Attend
Sponsors & Partners
Team
Past Years
Above and Beyond
Stanford welcomes you to the world "Above and Beyond" with our third TEDxStanford, Saturday, May 10.
Whether you've attended our previous events or are new to TEDxStanford, we are delighted to invite you to TEDxStanford2014: Above and Beyond. Join us for a day of inspirational short talks and performances by scholars that go above and beyond the ordinary.
From a deep dive with extreme creatures of the sea and the new synthetic biology that "teaches" our cells to target cancer, to the creation of the solar car, stripping down Wall Street and a sneak peak at the documentary film made in the world's "happiest place," TEDxStanford2014 is all about
#TEDxStanford



Stanford University
SUNetID Login
TEDx Stanford
May 10, 2014
11 a.m. to 6:30 p.m. at CEMEX Auditorium
Overflow Tickets
Home
About
Team
Past Years
Stanford
Above and B
Stanford welcomes you to the world "Above and Beyond" with our third TEDxStanford, Saturday, May 10.
#TEDxStanford

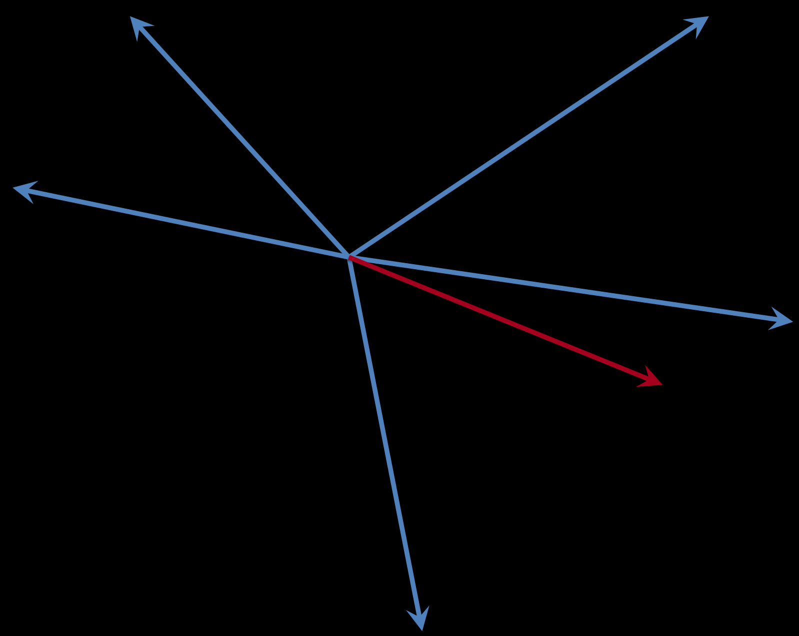
words \ website	Page 1	Page 2	Page 3	Page 4	Page 5
Stanford	1	1	0	0	0
Beating	1	0	0	1	0
CAL	1	0	1	0	0
Eggs	0	0	0	1	1

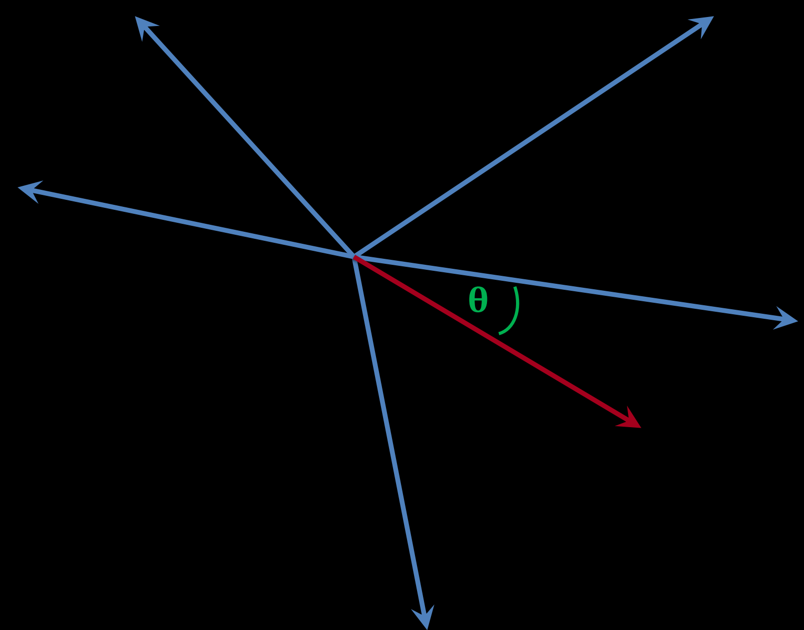
How to search for the phrase “Stanford eggs”

words \ website	Page 1	Page 2	Page 3	Page 4	Page 5
Stanford	1	1	0	0	0
Beating	1	0	0	1	0
CAL	1	0	1	0	0
Eggs	0	0	0	1	1

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

How would you determine which of the page vectors are closest?

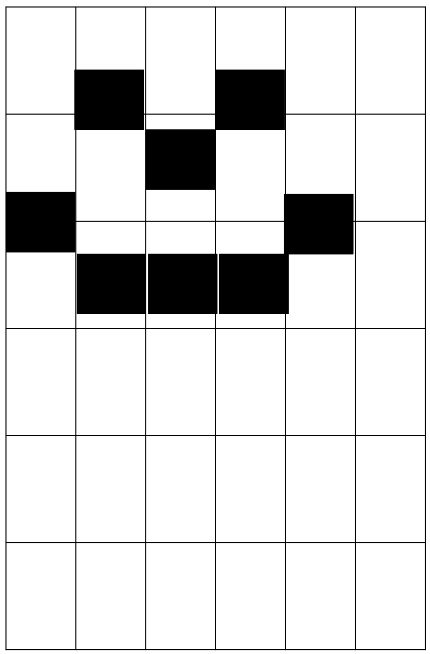



θ

words \ website	Page 1	Page 2	Page 3	Page 4	Page 5
Stanford	1	1	0	0	0
Beating	1	0	0	1	0
CAL	1	0	1	0	0
Eggs	0	0	0	1	1

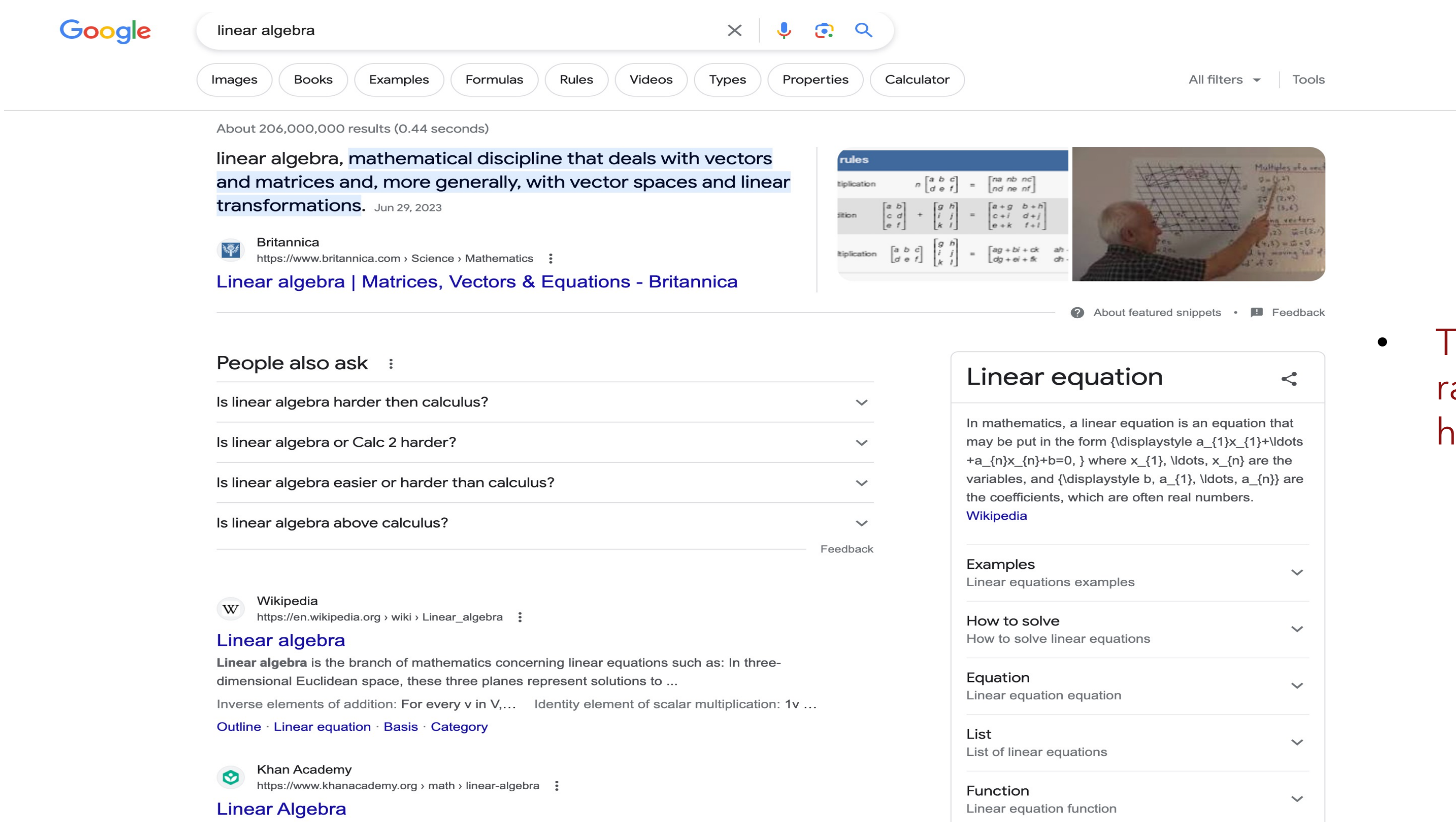
1	1	0	0	0
1	0	0	1	0
1	0	1	0	0
0	0	0	1	1

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



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

Google Search Results



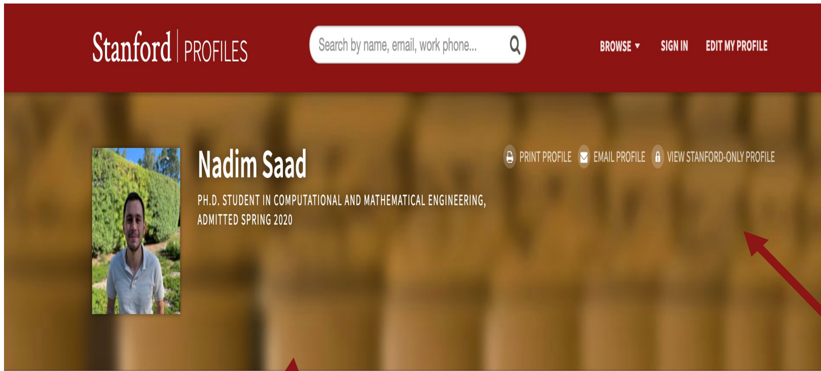
- The web pages are ranked in some order, how would you do it?

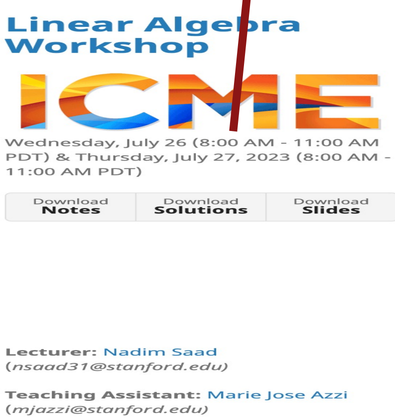
Ranking depends on many factors

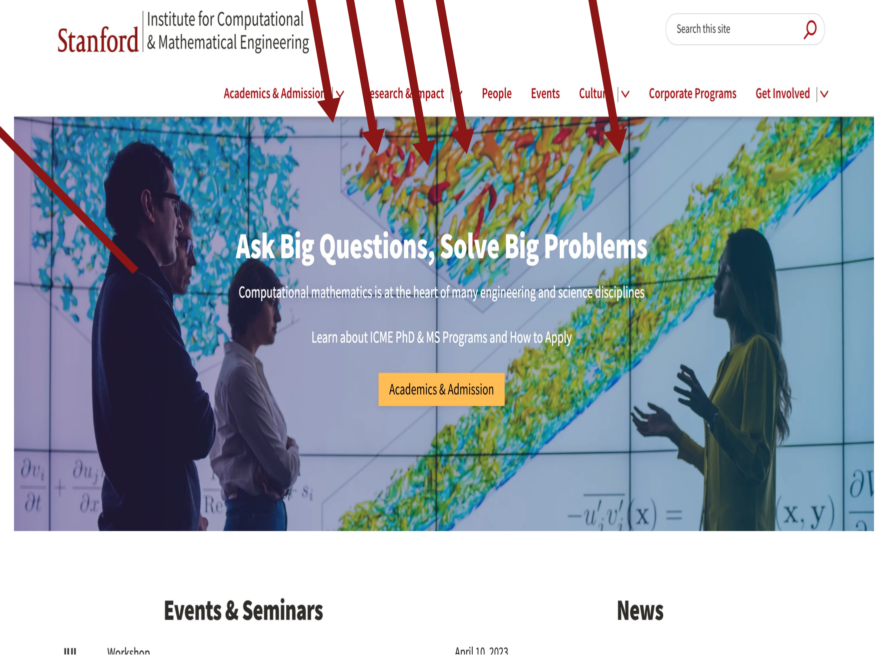
- Traditional page rank algorithm (measure of importance of a page)
- Number of visits
- Age
- Recent edits
- …

Initial PageRank Algorithm

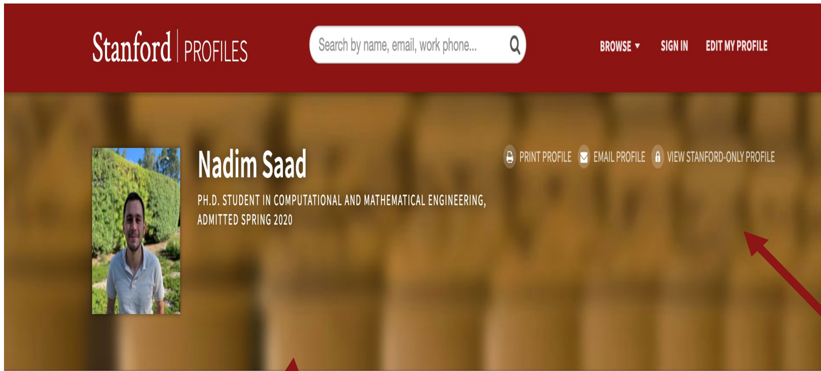
Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd "The PageRank Citation Ranking: Bringing Order to the Web" Technical Report, Stanford InfoLab, 1999

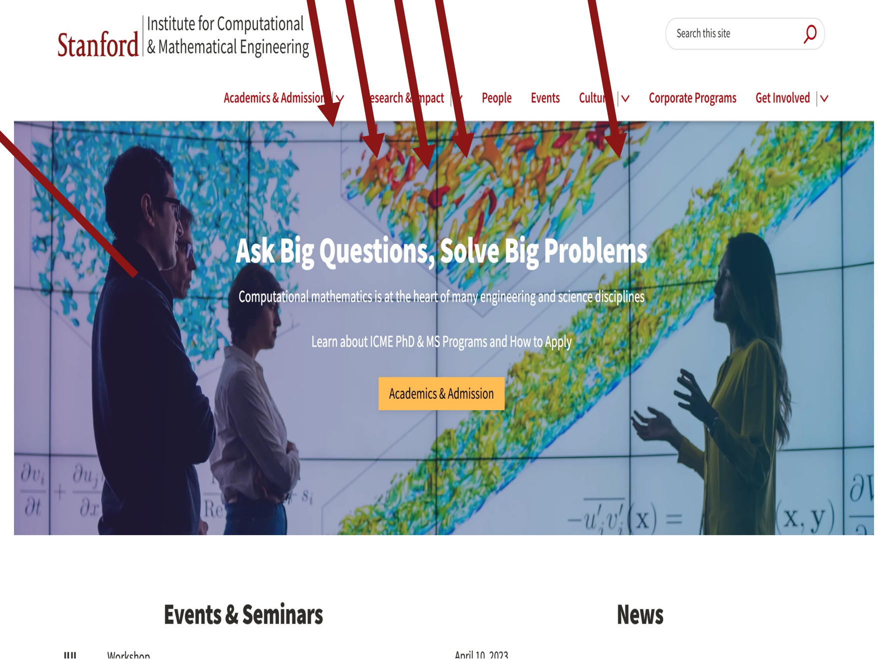


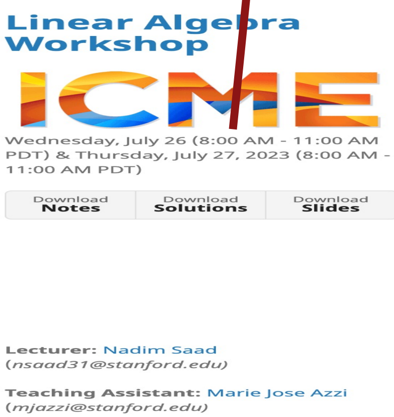




The more the incoming links, the higher the page rank





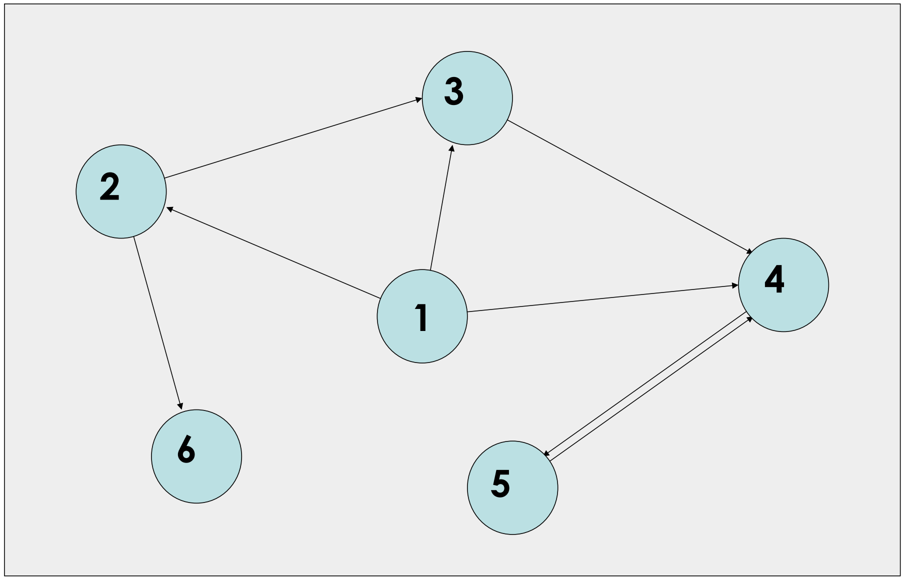


Links coming in from an import page have more weight

Rank propagation

Rank of page j denoted by x_j

$x_1 = 0$



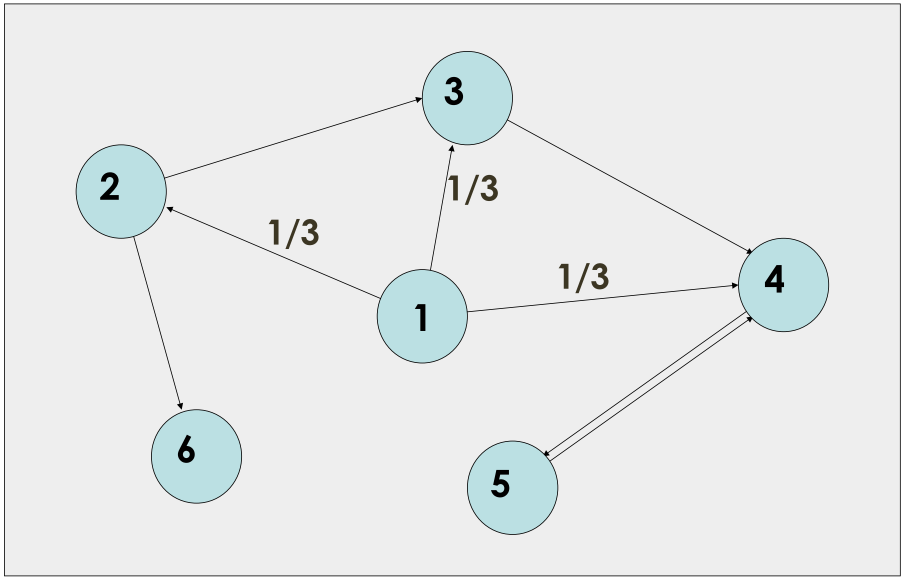
A very small internet "graph"

Rank propagation

Rank of page j denoted by x_j

$$x_1 = 0$$

$$x_2 = \frac{1}{3} x_1$$

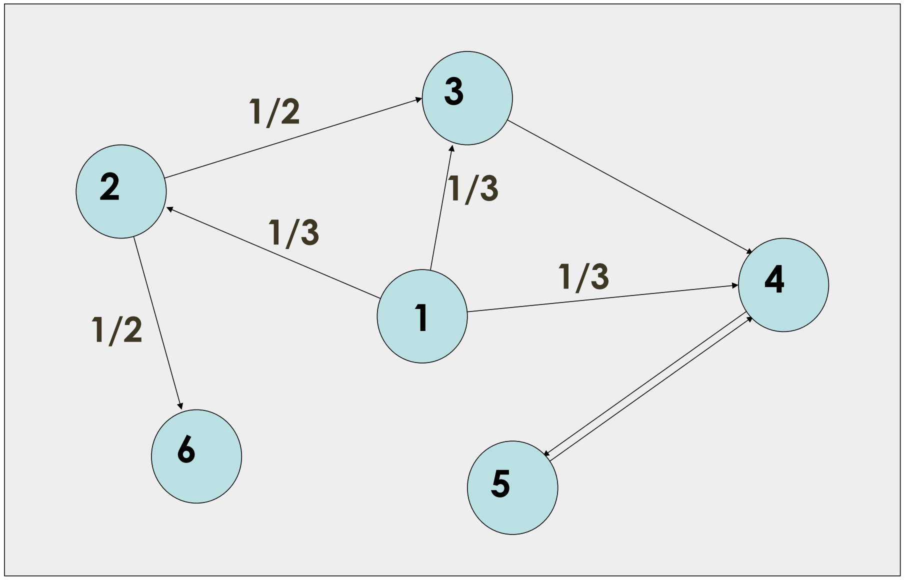


A very small internet "graph"

Rank propagation

Rank of page j denoted by x_j

$$x_1 = 0$$
$$x_2 = \frac{1}{3}x_1$$
$$x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2$$
$$x_4 = \frac{1}{3}x_1 + x_3 + x_5$$
$$x_5 = x_4$$
$$x_6 = \frac{1}{2}x_2$$



A very small internet "graph"

A COUPLED AND LINEAR SYSTEM OF EQUATIONS

$$x_1 = 0$$

$$x_2 = \frac{1}{3}x_1$$

$$x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2$$

$$x_4 = \frac{1}{3}x_1 + x_3 + x_5$$

$$x_5 = x_4$$

$$x_6 = \frac{1}{2}x_2$$

In matrix-vector notation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

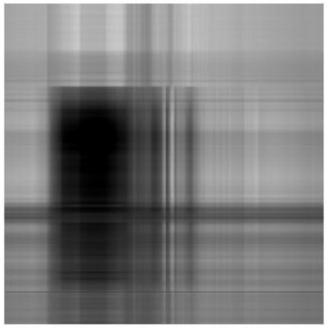
$$x = Px$$

or

$$Px = x$$

SVD for Compression on more than a 5x5 matrix

256x256



keep 2 terms

keep 8 terms

keep 32 terms