The Mathematics of Google

A. BACKGROUND

1989 World Wide Web
Tim Bernors-Lee (CERN)

1993 Mosaic Web Browser
Marc Andreeson (NCSA-UIC)

1994 Lycos Search Engine
Michael Mauldin (CMU)

1998- Google Search Engine
Larry Page, Sergey Brin
(Stanford)

Research paper:
"The Page-Rank Citation Ranking:
Bringing Order to the Web"

LESSONS?

wikipedia.org
google.com

GOOGLE:

NUMBERS & FACTS

- indexes 3 billion pages
- clean design, fast searches
  relevant results
- 250 million queries/day
  - 3000/sec.
- runs on 10000+ Linux computers
  Red Hat
- 1/2 of $2 billion/yr search
  engine business

1 MATHEMATICAL IDEA

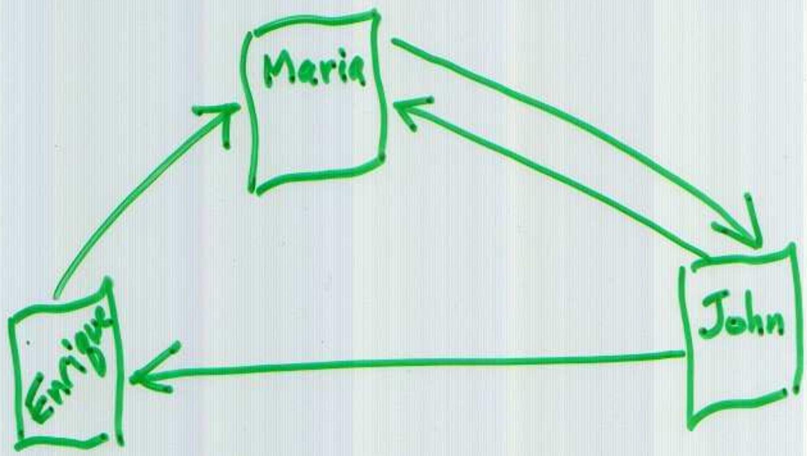
⇓

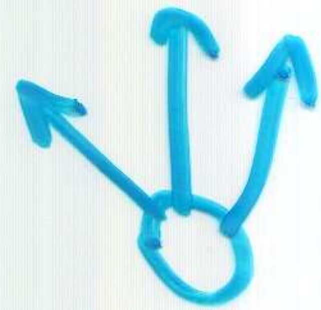
NEW COMPANY

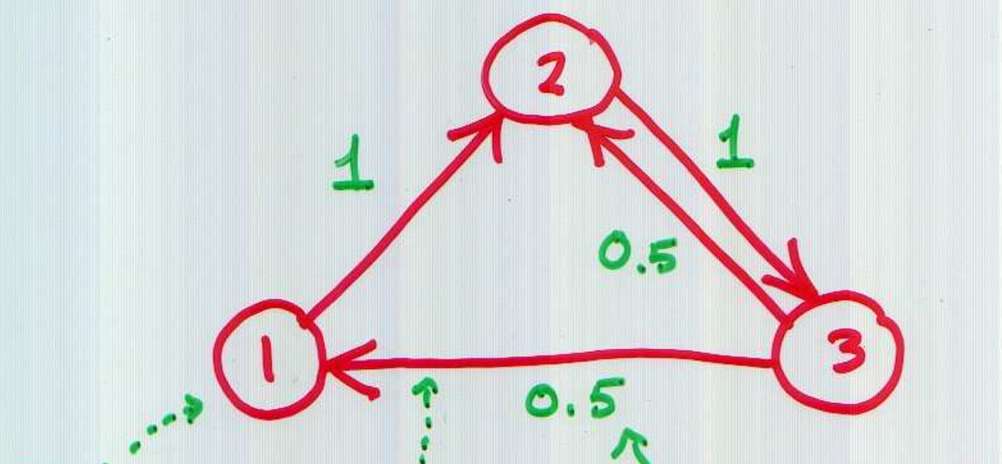
Another Example: RSA Paper
(MIT)

B. The Mathematics

The web is a directed graph.







in Link Table

1	3
2	1, 3
3	2

Graph constructed by a web crawler.

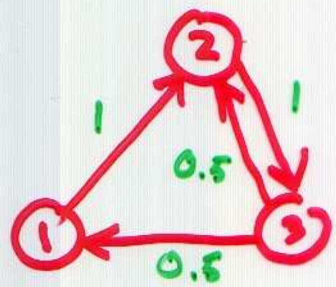
Goal: assign **rank** to each page (node). Rank = measure of interest

Rank of i-th node = x_i

$x_i = w_5 x_5 + w_7 x_7$

(*) $\boxed{x_i = \text{weighted sum of the ranks of the nodes linked to } x_i}$

Example



$$\begin{cases} x_1 = 0.5\, x_3 \\ x_2 = x_1 + 0.5\, x_3 \\ x_3 = x_2 \end{cases}$$

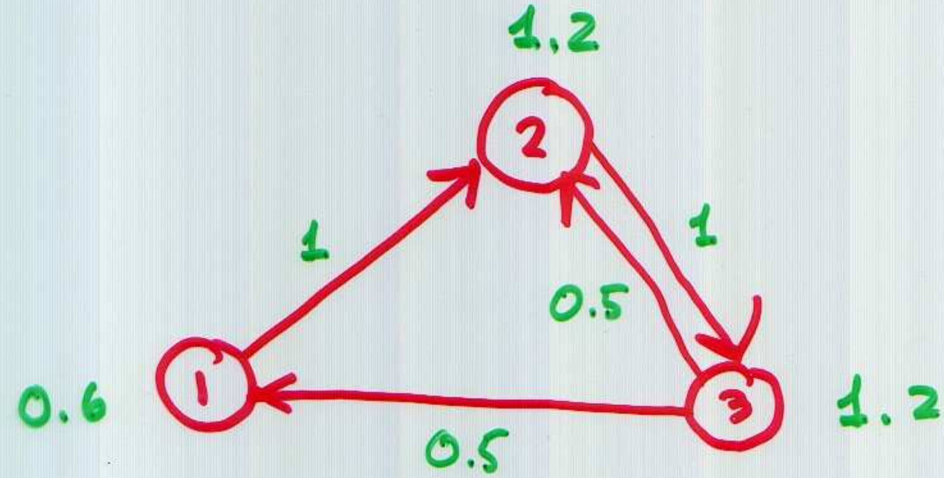
Solution: $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$

Note: we can rescale solution

$x = (0.2, 0.4, 0.4)$ sum 1

$x = (0.6, 1.2, 1.2)$ sum 3

Why? Equation (*) is homogeneous.



- Rank of node depends on rank of nodes linked to it ... on entire web (??)

- RANDOM SURFER MODEL

C. How to compute Page Rank

Rewrite system

$$
\begin{aligned}
x_1 &= 0 \cdot x_1 + 0 \cdot x_2 + 0.5\, x_3 \\
x_2 &= 1 \cdot x_1 + 0 \cdot x_2 + 0.5\, x_3 \\
x_3 &= 0 \cdot x_1 + 1 \cdot x_2 + 0\, x_3
\end{aligned}
$$

$$
\underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{\text{vector } x} = \underbrace{\begin{bmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}}_{\substack{\text{column} \\ \text{sums} = 1}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}
$$

matrix T

$$
(**) \qquad \boxed{\; x = Tx \;}
$$

Note: x is an eigenvector of T with eigenvalue 1

$$
Tx = cx
$$

eigenvector

eigenvalue

Iterative Computation

seed x

$$x' = \begin{bmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.5 \\ 1.0 \end{bmatrix}$$

$$x'' = \begin{bmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1.5 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \\ 1.5 \end{bmatrix}$$

$$x''' = \begin{bmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1.0 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 1.25 \\ 1.0 \end{bmatrix}$$

$$\boxed{x^{(\infty)} = \lim_{n \to \infty} x^{(n)}}$$

1 0.5
1
1

$$x^{(\infty)} = \begin{bmatrix} 0.6 \\ 1.2 \\ 1.2 \end{bmatrix}$$

GRAPH G1

0:	1.00	1.00	1.00	3.00
1:	0.50	1.50	1.00	3.00
2:	0.50	1.00	1.50	3.00
3:	0.75	1.25	1.00	3.00
4:	0.50	1.25	1.25	3.00
5:	0.62	1.12	1.25	3.00
6:	0.62	1.25	1.12	3.00
7:	0.56	1.19	1.25	3.00
8:	0.62	1.19	1.19	3.00
9:	0.59	1.22	1.19	3.00
10:	0.59	1.19	1.22	3.00
11:	0.61	1.20	1.19	3.00
12:	0.59	1.20	1.20	3.00
13:	0.60	1.20	1.20	3.00
14:	0.60	1.20	1.20	3.00
15:	0.60	1.20	1.20	3.00
16:	0.60	1.20	1.20	3.00
17:	0.60	1.20	1.20	3.00
18:	0.60	1.20	1.20	3.00
19:	0.60	1.20	1.20	3.00

Python

D. The Google Equations, v.2

$$\boxed{x = (1-\delta)s + \delta T x}$$

When $\delta = 1$ this gives $x = Tx$.

δ = damping factor $(0 < \delta \leq 1)$

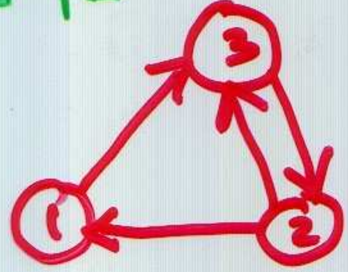
s = "rank source" = $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Why the change?

- Ensures convergence to $x^{(\infty)}$
- Makes convergence more rapid

$f(x) = (1-\delta)s + \delta T x$ is a contraction mapping for $0 \leq \delta < 1$

Example



$$\delta = 0.95$$

$$s = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad T = \begin{pmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}$$

$$x' = 0.05 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 0.95 \begin{pmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.05 \\ 0.05 \\ 0.05 \end{pmatrix} + 0.95 \begin{pmatrix} 0.5 \\ 1.5 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.525 \\ 1.475 \\ 1 \end{pmatrix}$$

This very close to the old x':

$$x'_{old} = \begin{pmatrix} 0.5 \\ 1.5 \\ 1 \end{pmatrix}$$

Graph G1 again

damping = 1.0					damping = 0.95				
0:	1.00	1.00	1.00	3.00	0:	1.00	1.00	1.00	3.00
1:	0.50	1.50	1.00	3.00	1:	0.53	1.47	1.00	3.00
2:	0.50	1.00	1.50	3.00	2:	0.53	1.02	1.45	3.00
3:	0.75	1.25	1.00	3.00	3:	0.74	1.24	1.02	3.00
4:	0.50	1.25	1.25	3.00	4:	0.54	1.24	1.23	3.00
5:	0.62	1.12	1.25	3.00	5:	0.63	1.14	1.23	3.00
6:	0.62	1.25	1.12	3.00	6:	0.63	1.23	1.13	3.00
7:	0.56	1.19	1.25	3.00	7:	0.59	1.19	1.22	3.00
8:	0.62	1.19	1.19	3.00	8:	0.63	1.19	1.18	3.00
9:	0.59	1.22	1.19	3.00	9:	0.61	1.21	1.18	3.00
10:	0.59	1.19	1.22	3.00	10:	0.61	1.19	1.20	3.00
11:	0.61	1.20	1.19	3.00	11:	0.62	1.20	1.18	3.00
12:	0.59	1.20	1.20	3.00	12:	0.61	1.20	1.19	3.00
13:	0.60	1.20	1.20	3.00	13:	0.62	1.20	1.19	3.00
14:	0.60	1.20	1.20	3.00	14:	0.62	1.20	1.19	3.00
15:	0.60	1.20	1.20	3.00	15:	0.61	1.20	1.19	3.00
16:	0.60	1.20	1.20	3.00	16:	0.61	1.20	1.19	3.00
17:	0.60	1.20	1.20	3.00	17:	0.61	1.20	1.19	3.00
18:	0.60	1.20	1.20	3.00	18:	0.61	1.20	1.19	3.00
19:	0.60	1.20	1.20	3.00	19:	0.61	1.20	1.19	3.00

Graph G1 one more time

damping = 1.0

```
 0: 1.00 1.00 1.00     3.00
 1: 0.50 1.50 1.00     3.00
 2: 0.50 1.00 1.50     3.00
 3: 0.75 1.25 1.00     3.00
 4: 0.50 1.25 1.25     3.00
 5: 0.62 1.12 1.25     3.00
 6: 0.62 1.25 1.12     3.00
 7: 0.56 1.19 1.25     3.00
 8: 0.62 1.19 1.19     3.00
 9: 0.59 1.22 1.19     3.00
10: 0.59 1.19 1.22     3.00
11: 0.61 1.20 1.19     3.00
12: 0.59 1.20 1.20     3.00
13: 0.60 1.20 1.20     3.00
14: 0.60 1.20 1.20     3.00
15: 0.60 1.20 1.20     3.00
16: 0.60 1.20 1.20     3.00
17: 0.60 1.20 1.20     3.00
18: 0.60 1.20 1.20     3.00
19: 0.60 1.20 1.20     3.00
```

damping = 0.85

```
 0: 1.00 1.00 1.00     3.00
 1: 0.57 1.42 1.00     3.00
 2: 0.57 1.06 1.36     3.00
 3: 0.73 1.22 1.05     3.00
 4: 0.60 1.22 1.18     3.00
 5: 0.65 1.16 1.18     3.00
 6: 0.65 1.21 1.14     3.00
 7: 0.63 1.19 1.18     3.00
 8: 0.65 1.19 1.16     3.00
 9: 0.64 1.20 1.16     3.00
10: 0.64 1.19 1.17     3.00
11: 0.65 1.19 1.16     3.00
12: 0.64 1.19 1.16     3.00
13: 0.64 1.19 1.16     3.00
14: 0.64 1.19 1.16     3.00
15: 0.64 1.19 1.16     3.00
16: 0.64 1.19 1.16     3.00
17: 0.64 1.19 1.16     3.00
18: 0.64 1.19 1.16     3.00
19: 0.64 1.19 1.16     3.00
```

Graph G1 one more time

damping = 1.0

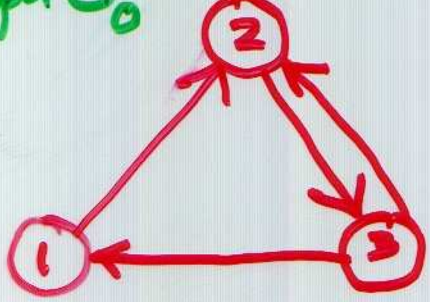
0:	1.00	1.00	1.00	3.00
1:	0.50	1.50	1.00	3.00
2:	0.50	1.00	1.50	3.00
3:	0.75	1.25	1.00	3.00
4:	0.50	1.25	1.25	3.00
5:	0.62	1.12	1.25	3.00
6:	0.62	1.25	1.12	3.00
7:	0.56	1.19	1.25	3.00
8:	0.62	1.19	1.19	3.00
9:	0.59	1.22	1.19	3.00
10:	0.59	1.19	1.22	3.00
11:	0.61	1.20	1.19	3.00
12:	0.59	1.20	1.20	3.00
13:	0.60	1.20	1.20	3.00
14:	0.60	1.20	1.20	3.00
15:	0.60	1.20	1.20	3.00
16:	0.60	1.20	1.20	3.00
17:	0.60	1.20	1.20	3.00
18:	0.60	1.20	1.20	3.00
19:	0.60	1.20	1.20	3.00

damping = 0.85

0:	1.00	1.00	1.00	3.00
1:	0.57	1.42	1.00	3.00
2:	0.57	1.06	1.36	3.00
3:	0.73	1.22	1.05	3.00
4:	0.60	1.22	1.18	3.00
5:	0.65	1.16	1.18	3.00
6:	0.65	1.21	1.14	3.00
7:	0.63	1.19	1.18	3.00
8:	0.65	1.19	1.16	3.00
9:	0.64	1.20	1.16	3.00
10:	0.64	1.19	1.17	3.00
11:	0.65	1.19	1.16	3.00
12:	0.64	1.19	1.16	3.00
13:	0.64	1.19	1.16	3.00
14:	0.64	1.19	1.16	3.00
15:	0.64	1.19	1.16	3.00
16:	0.64	1.19	1.16	3.00
17:	0.64	1.19	1.16	3.00
18:	0.64	1.19	1.16	3.00
19:	0.64	1.19	1.16	3.00

E. Interlude: Adjacency Matrix

graph G_0:



$$\approx \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

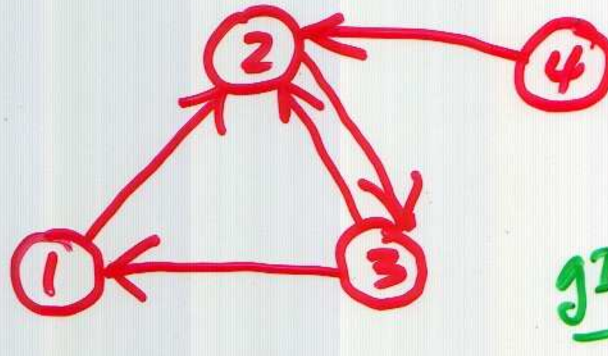
$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 0 & 0.5 \\ 1 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

normalize columns

$$A \Rightarrow T$$

adjacency matrix → transition matrix

$\boxed{\text{What is the adjacency matrix ?}}$



g1b

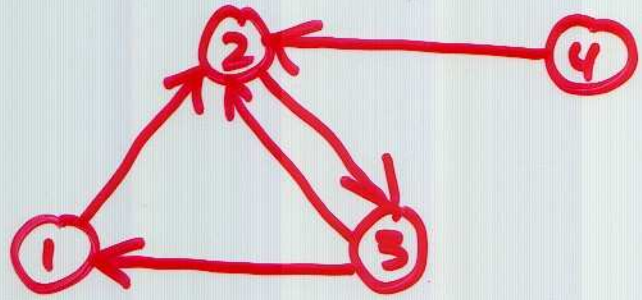
13

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$$

Problem 1

Consider the graph G_1:

G_1



$g^1 6$

Find:

a. the adjacency matrix A

b. the transition matrix T

c. the vectors x', x''
use damping factor 0.85

d. the vector $x^{(\infty)}$
(you will need to write a computer program for this)

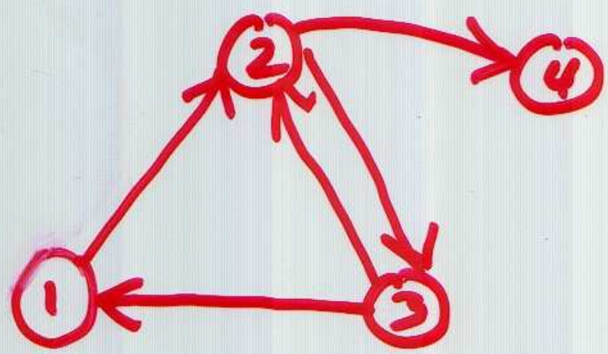
Discuss:

How the page Rank differs from that of graph G_0

Problem 2

Let graph G_2 be the one obtained by reversing the arrow coming out of node 4:

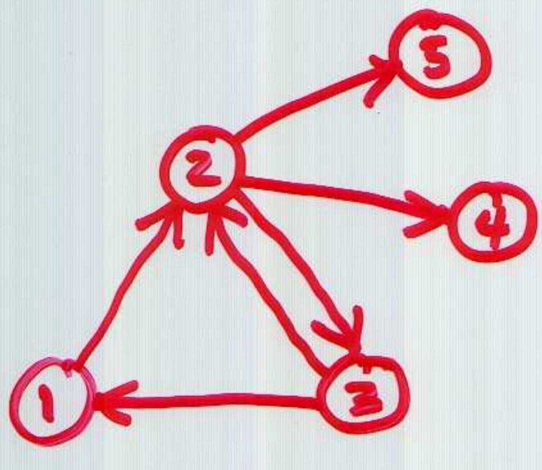
G_2



(dangling link)

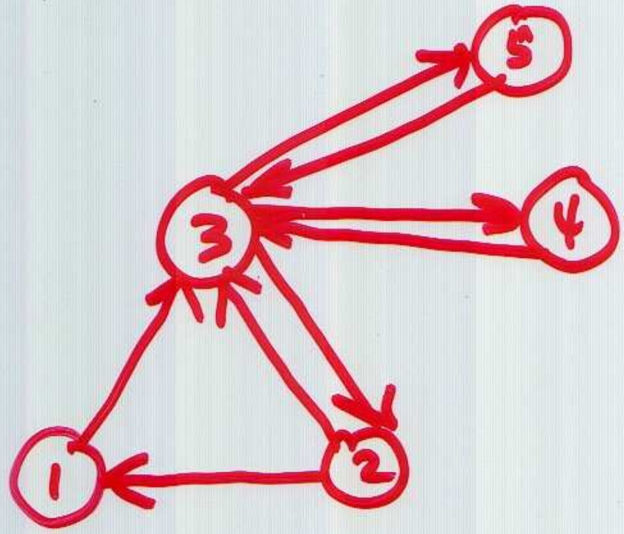
Proceed as in problem 1. (Compare the results for G_1 and G_2).

Problem 3. Let's add more pages to the "home page" at node 2:

$G_3 =$ 

Proceed as in problem 2. (Compare the results for G_2 and G_3).

Problem 4. Now let's link the pages 4 and 5 back to the home page at 2:

$G_4 =$ 

Proceed as in problem 3.

(Compare the results for G_3 and G_4).

GOOGLE: REMARKS

1. Power Method

$$Av_i = \lambda_i v_i \quad \underline{\text{eigenvectors}}$$

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

$$Av = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \cdots + c_n \lambda_n v_n$$

$$A^2 v = c_1 \lambda_1^2 v_1 + c_2 \lambda_2^2 v_2 + \cdots + c_n \lambda_n^2 v_n$$

$$\vdots$$

$$A^k v = c_1 \underbrace{\lambda_1}_{=1}^k v_1 + c_2 \underbrace{\lambda_2^k}_{|\lambda_2|<1} v_2 + \cdots + c_n \underbrace{\lambda_n^k}_{|\lambda_n|<1} v_n$$

λ_1 = dominant eigenvalue

$$A^k v = c_1 \lambda^k v_1 + \underline{\underline{\text{much smaller stuff}}}$$

If $\lambda_1 = 1$:

$$A^k v_1 = \underline{c_1 v_1} + \underline{\text{stuff}}$$

2. Stochastic Matrices

$$\begin{array}{c|c} \text{probability} & \text{state} \\ \hline \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} & \begin{matrix} 1 \\ 2 \end{matrix} \end{array}$$

- v = probability vector:
  entries between 0 & 1
  sum of entries = 1.
- Matrix A: <u>when is Av a probability vector if v is?</u>
  - If entries of A are between 0 & 1 and <u>column</u>sums are 1.

- <u>Eigenvalues</u>

  Let $|v| = |v_1| + |v_2| + \dots + |v_n|$ for a vector v (L_1 norm)

  Let A be a stochastic matrix.

  Then $\boxed{|Av| \leqq |v|}$

  <u>Corollary</u>: If λ is an eigenvalue of A, then $|\lambda| \leqq 1$

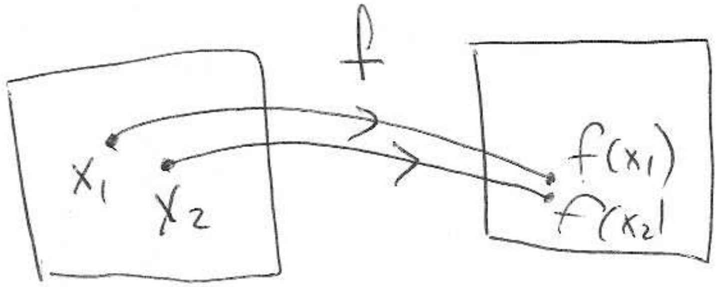
- <u>Special case</u> (Perron's Thm 1907) If (and A is stochastic) all entries of A are positive
  then $\lambda = 1$ is an eigenvalue with multiplicity 1:
  all other eigenvalues μ satisfy $|\mu| < 1$.

3. FIXED POINTS $\boxed{f(x) = x}$

- PERRON-FROBENIUS

  2nd Eigenvalue controls ...

- CONTRACTION MAPPINGS



$$\mathrm{dist}(f(x_1), f(x_2)) \leq C\,\mathrm{dist}(x_1, x_2)$$

$$\boxed{C < 1}$$

Thm: Every contraction mapping has a unique fixed point.

1. Google Map

$$f(x) = cAx + (1-c)u$$

$$\boxed{0 < c < 1}$$

all 1's

$$|f(x) - f(y)| = c\,|A(x-y)|$$

$|Av| \leq |v|$

$$\boxed{|f(x) - f(y)| \leq c\,|x-y|}$$

$0 < c < 1$

Because A is stochastic

So f contracts —
rapidly for $c = 0.85$