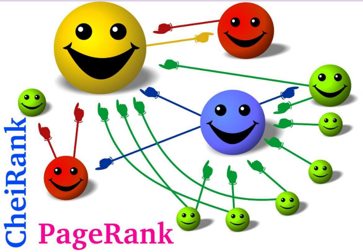
CheiRank versus PageRank

Dima Shepelyansky (CNRS, Toulouse)
www.quantware.ups-tlse.fr/dima

with Alexei Chepelianskii (Cambridge Univ.)
Leonardo Ermann, Klaus Frahm, Bertrand Georgeot (CNRS, Toulouse)

Through mechanisms still only partially understood,
Google finds instantaneously the page you need.

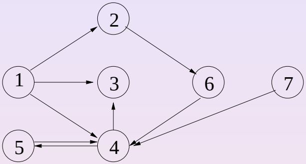


Story started in 1998 (now $N \sim 10^{11}$ nodes)

S. Brin and L. Page, Computer Networks and ISDN Systems **30**,107 (1998).

Markov chains and Directed networks

Weighted adjacency matrix



$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

For a directed network with N nodes the adjacency matrix $\mathbf{A}$ is defined as $A_{ij} = 1$ if there is a link from node j to node i and $A_{ij} = 0$ otherwise. The weighted adjacency matrix is

$$S_{ij} = A_{ij} / \sum_k A_{kj}$$

In addition the elements of columns with only zeros elements are replaced by $1/N$.

How Google works

Google Matrix and Computation of PageRank

$\mathbf{p} = \mathbf{S}\mathbf{p} \Rightarrow \mathbf{p}$= stationary vector of **S**; can be computed by iteration of **S**.

To remove convergence problems:

- Replace columns of 0 (dangling nodes) by $\frac{1}{N}$:

  In our example, $\mathbf{S} = \begin{pmatrix} 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 1 & 1 & 1 \\ 0 & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \end{pmatrix}$.

- To remove degeneracies of $\lambda = 1$, replace **S** by **Google matrix**

  $\mathbf{G} = \alpha \mathbf{S} + (1-\alpha)\frac{\mathbf{E}}{N}$; $Gp = \lambda p$ => Perron-Frobenius operator

- α models a random surfer with a random jump after approximately 6 clicks (usually $\alpha = 0.85$); PageRank vector => p at $\lambda = 1$ ($\sum_j p_j = 1$).

- CheiRank: $\mathbf{S}^*$ with inverted link directions proposed at A.D.Chepelianskii arXiv:1003.5455 (2010)

Models of real networks

Real networks are characterized by:

- **small world property**: average distance between 2 nodes $\sim \log N$
- **scale-free property**: distribution of the number of incoming or outcoming links $P(k) \sim k^{-\nu}$ $(\nu \sim 2.1(in); 2.7(out))$

Can be explained by a twofold mechanism:

- Constant growth: new nodes appear regularly and are attached to the network
- Preferential attachment: nodes are preferentially linked to already highly connected vertices.

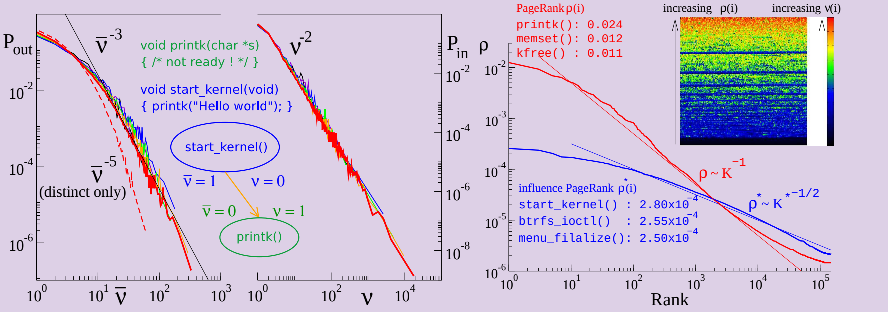
PageRank vector for large WWW (proportional to ingoing links):

- $p_j \sim 1/j^{\beta}$, where j is the ordered index
- number of nodes N_n with PageRank p scales as $N_n \sim 1/p^{\nu}$ with numerical values $\nu = 1 + 1/\beta \approx 2.1$ and $\beta \approx 0.9$

CheiRank vector is proportional to outgoing links $\beta \approx 0.6$

Linux Kernel Network

Procedure call network for Linux

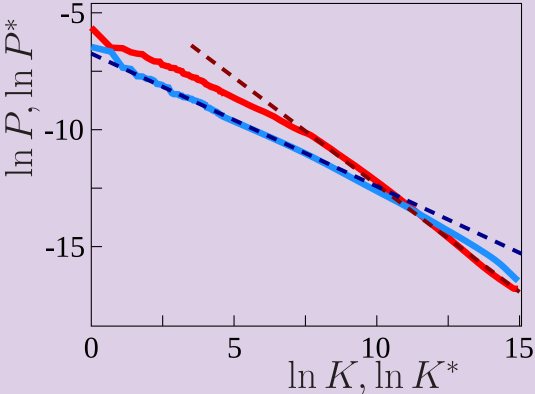


Links distribution (left); PageRank and inverse PageRank (CheiRank) distribution (right) for Linux versions up to 2.6.32 with $N = 285509$ ($\rho \sim 1/j^{\beta}, \beta = 1/(\nu - 1)$).

[Chepelianskii arxiv:1003.5455]

Two-dimensional ranking of Wikipedia articles

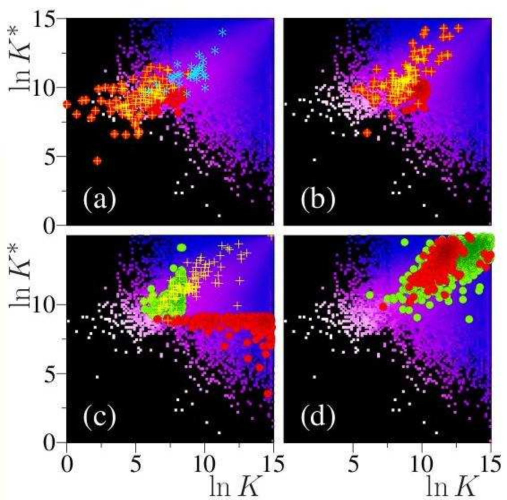
Wikipedia English articles $N = 3282257$ dated Aug 18, 2009



Dependence of probability of PagRank P (red) and CheiRank P^* (blue) on corresponding rank indexes K, K^*; lines show slopes $\beta = 1/(\nu - 1)$ with $\beta = 0.92; 0.57$ respectively for $\nu = 2.09; 2.76$.

[Zhirov, Zhirov, DS arxiv:1006.4270]

Two-dimensional ranking of Wikipedia articles



Density distribution in plane of PageRank and CheiRank indexes ($\ln K$, $\ln K^*$): (a)100 top countries from 2DRank (red), 100 top from SJR (yellow), 30 Dow-Jones companies (cyan); (b)100 top universities from 2DRank (red) and Shanghai (yellow); (c)100 top personalities from PageRank (green), CheiRank (red) and Hart book (yellow); (d)758 physicists (green) and 193 Nobel laureates (red).

Wikipedia ranking of physicists

Physicists:

PageRank: 1. Aristotle, 2. Albert Einstein, 3. Isaac Newton, 4. Thomas Edison, 5. Benjamin Franklin, 6. Gottfried Leibniz, 7. Avicenna, 8. Carl Friedrich Gauss, 9. Galileo Galilei, 10. Nikola Tesla.

2DRank: 1. Albert Einstein, 2. Nikola Tesla, 3. Benjamin Franklin, 4. Avicenna, 5. Isaac Newton, 6. Thomas Edison, 7. Stephen Hawking, 8. Gottfried Leibniz, 9. Richard Feynman, 10. Aristotle.

CheiRank: 1. Hubert Reeves, 2. Shen Kuo, 3. Stephen Hawking, 4. Nikola Tesla, 5. Albert Einstein, 6. Arthur Stanley Eddington, 7. Richard Feynman, 8. John Joseph Montgomery, 9. Josiah Willard Gibbs, 10. Heinrich Hertz.

Nobel laureates:

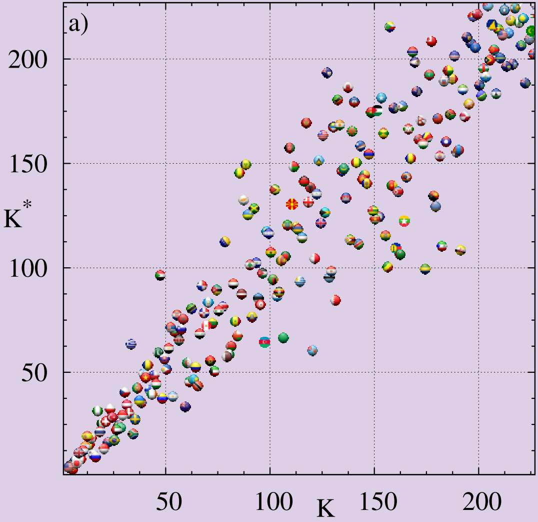
PageRank: 1. Albert Einstein, 2. Enrico Fermi, 3. Richard Feynman, 4. Max Planck, 5. Guglielmo Marconi, 6. Werner Heisenberg, 7. Marie Curie, 8. Niels Bohr, 9. Paul Dirac, 10. J.J.Thomson.

2DRank: 1. Albert Einstein, 2. Richard Feynman, 3. Werner Heisenberg, 4. Enrico Fermi, 5. Max Born, 6. Marie Curie, 7. Wolfgang Pauli, 8. Max Planck, 9. Eugene Wigner, 10. Paul Dirac.

CheiRank: 1. Albert Einstein, 2. Richard Feynman, 3. Werner Heisenberg, 4. Brian David Josephson, 5. Abdus Salam, 6. C.V.Raman, 7. Peter Debye, 8. Enrico Fermi, 9. Wolfgang Pauli, 10. Steven Weinberg.

Ranking of World Trade

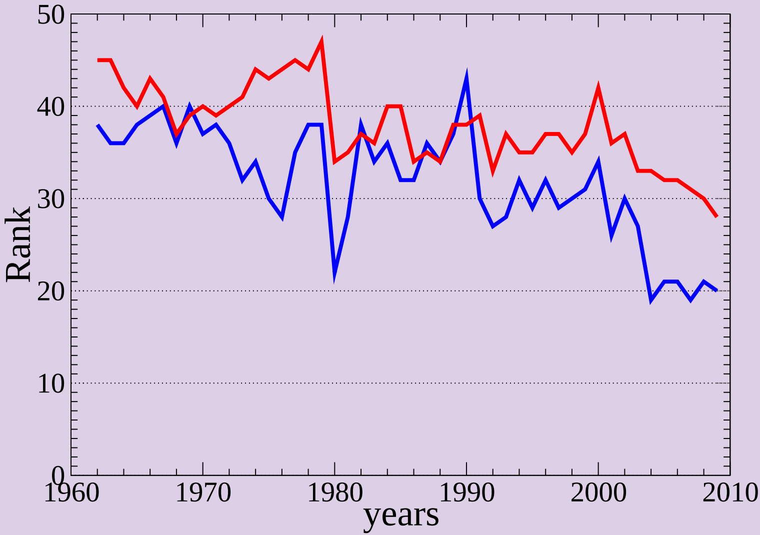
2008: All commodities



L.Ermann, DS arXiv:1105.5027

Rank of Poland

PageRank K and CheiRank K^*

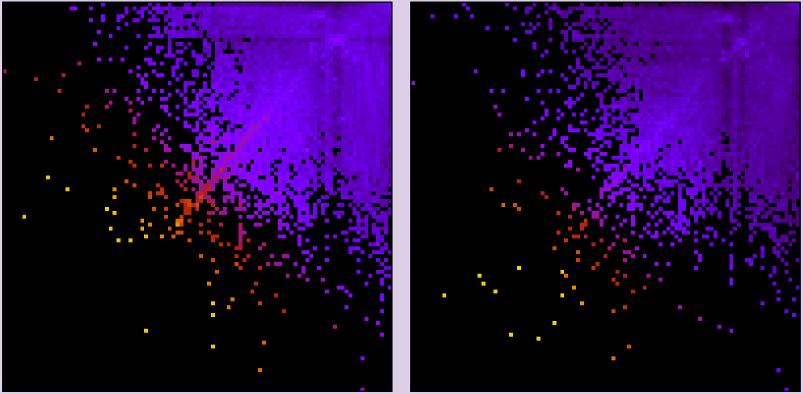


Rank table 2008 (74% of countries of G20)

Table 1. Top 20 ranking for *all commodities* – 2008.

Ran	K	K^*	K_2	$\tilde{K}$	$\tilde{K}^*$
1	USA	China	USA	USA	China
2	Germany	USA	China	Germany	Germany
3	China	Germany	Germany	China	USA
4	France	Japan	Japan	France	Japan
5	Japan	France	France	Japan	France
6	UK	Italy	Italy	UK	Netherlands
7	Italy	Russian Fed.	UK	Netherlands	Italy
8	Netherlands	Rep. of Korea	Netherlands	Italy	Russian Fed.
9	India	UK	India	Belgium	UK
10	Spain	Netherlands	Rep. of Korea	Canada	Belgium
11	Belgium	Singapore	Belgium	Spain	Canada
12	Canada	India	Russian Fed.	Rep. of Korea	Rep. of Korea
13	Rep. of Korea	Belgium	Canada	Russian Fed.	Mexico
14	Russian Fed.	Australia	Spain	Mexico	Saudi Arabia
15	Nigeria	Brazil	Singapore	Singapore	Singapore
16	Thailand	Canada	Thailand	India	Spain
17	Mexico	Spain	Australia	Poland	Malaysia
18	Singapore	South Africa	Brazil	Switzerland	Brazil
19	Switzerland	Thailand	Mexico	Turkey	India
20	Australia	U. Arab Emir.	U. Arab Emir.	Brazil	Switzerland

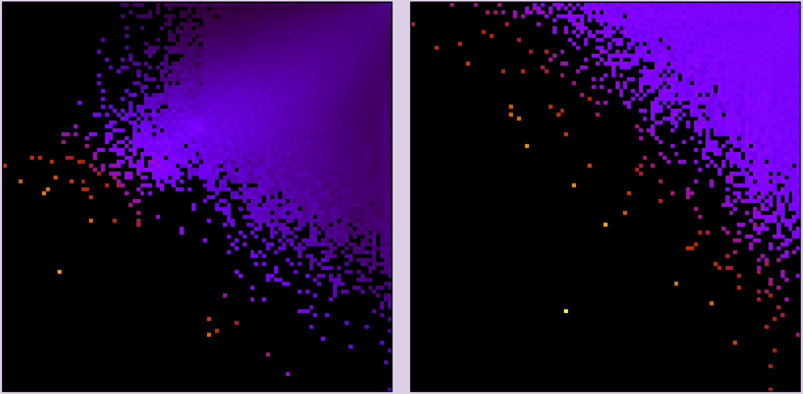
Towards two-dimensional search engine



Node density distribution in PageRank ($log_N K$) and CheRank ($log_N K^*$) plane for Univ. Cambridge (left) and Oxford (right) in 2006 ($N \approx 200000$)

[Ermann, Chepelianskii, DS in preparation]

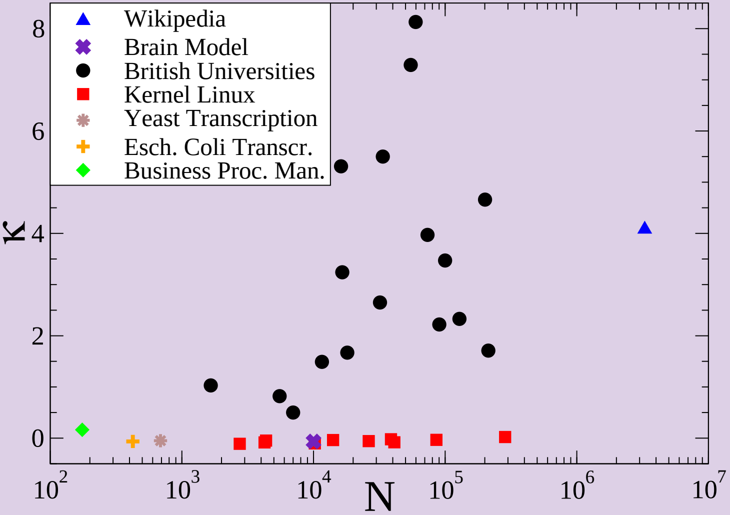
Towards two-dimensional search engine



Node density distribution in PageRank ($log_N K$) and CheRank ($log_N K^*$) plane for Wikipedia (left) and Linux 2.4 (right)

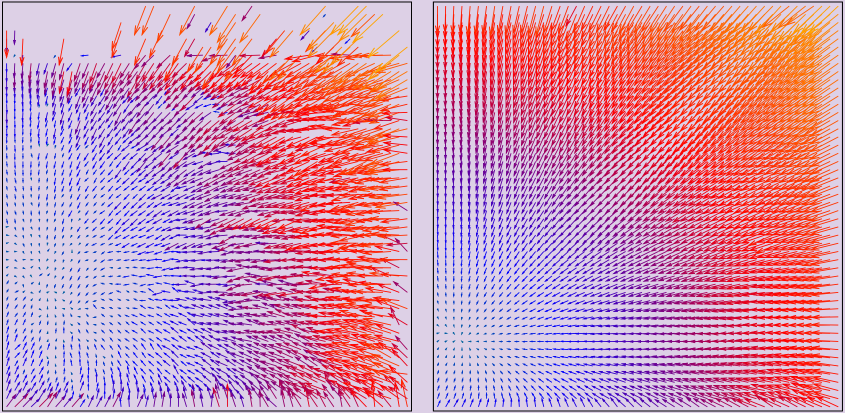
[Ermann, Chepelianskii, DS in preparation]

Correlator of PageRank and CheiRank



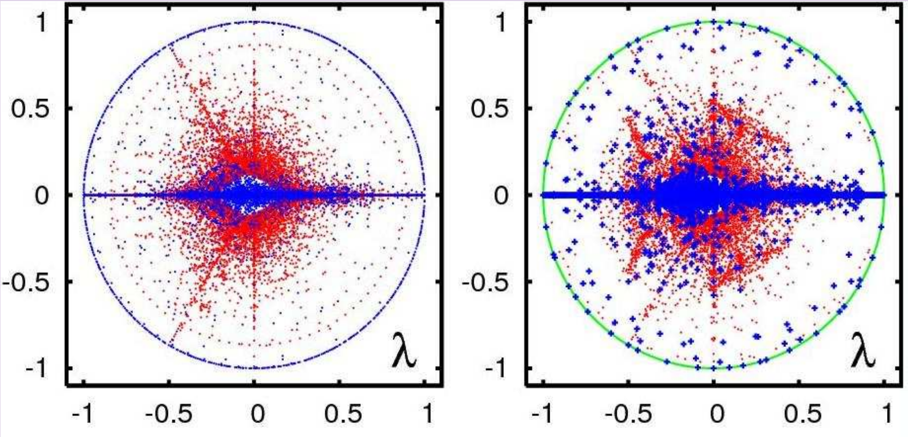
$\kappa = N \sum_i P(K(i))P(K^*(i)) - 1$ [Ermann, Chepelianskii, DS in preparation]

Flows in two dimensions



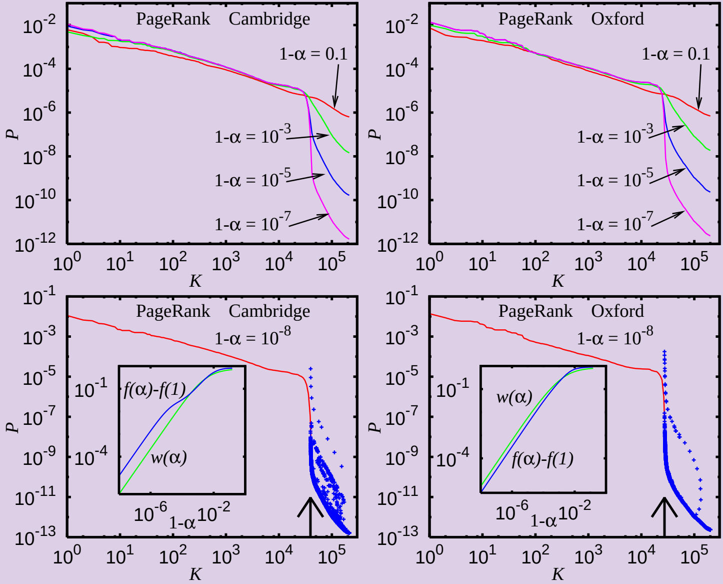
$N \times N$ plane of (K, K^*) for Univ. of Cambridge (2006) and Wikipedia [Ermann, Chepelianskii, DS in preparation]

Absence of spectral gap in real WWW

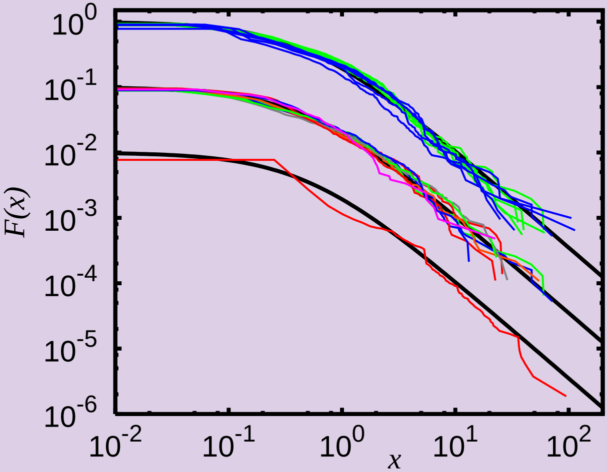


Spectrum of Google matrix for Univ. of Cambridge (left) and Oxford (right) in 2006 ($N \approx 200000$, $\alpha = 1$). [Frahm, Georgeot, DS arXiv:1105.1062]

Universal Emergence of PageRank

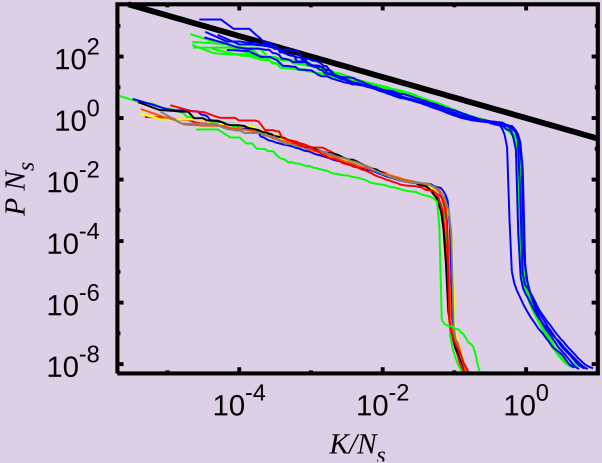


Invariant subspaces size distribution



$F(x)$ integrated number of invariant subspaces with size larger that d/d_0; $x = d/d_0$, d_0 is average size of subspaces (top: Cambridge, Oxford 2002-2006; middle: all others; bottom: Wikipedia CheiRank). Curve: $F(x) = 1/(1+2x)^{3/2}$.

PageRank at $\alpha \to 1$



Top: Cambridge, Oxford 2002-2006; bottom: all others ($\alpha = 1 - 10^{-8}$).

$P = \frac{1-\alpha}{1-\alpha S}\,\frac{1}{N}\,e$; $P = \sum_{\lambda_j=1} c_j\,\psi_j + \sum_{\lambda_j\neq 1} \frac{1-\alpha}{(1-\alpha)+\alpha(1-\lambda_j)}\,c_j\,\psi_j$

Google Matrix Applications

practically to everything ....

more data at

http://www.quantware.ups-tlse.fr/QWLIB/2drankwikipedia/ .../tradecheirank/