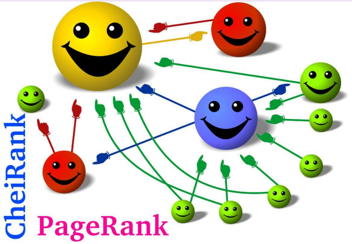
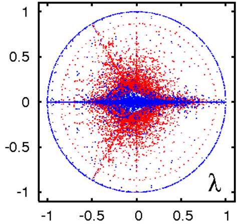
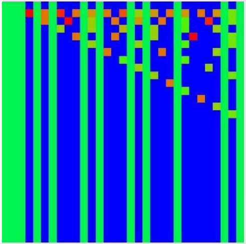
Google matrix of Markov chains

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

with A.Chepelianskii, L.Ermann, K.Frahm, B.Georgeot, O.Zhirov
support → EC FET Open grant NADINE





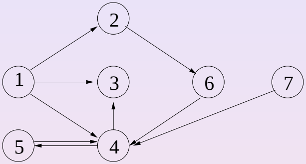


1906: Markov chains → WWW (1991) → Google matrix (1998)
1945: Nuclear physics → Wigner (1955) → Random Matrix Theory
→ PageRank of integers (in preparation 2012)

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

How Google works

Markov chains (1906) 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 $1/N$.

Google Matrix and Computation of PageRank

P $=$ **SP** $\Rightarrow$ **P**= stationary vector of **S**; can be computed by iteration of **S**.

To remove convergence problems:

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

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 1 & 1 & 1 \\ 0 & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \end{pmatrix}; \mathbf{S}^* = \begin{pmatrix} \frac{1}{7} & 1 & \frac{1}{2} & \frac{1}{4} & 0 & 0 & \frac{1}{7} \\ \frac{1}{7} & 0 & 0 & 0 & 0 & 1 & \frac{1}{7} \\ \frac{1}{7} & 0 & 0 & 0 & 0 & 0 & \frac{1}{7} \\ \frac{1}{7} & 0 & \frac{1}{2} & 0 & 1 & 0 & \frac{1}{7} \\ \frac{1}{7} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{7} \\ \frac{1}{7} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{7} \\ \frac{1}{7} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{7} \end{pmatrix}.$$

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

  $\mathbf{G} = \alpha \mathbf{S} + (1-\alpha)\frac{\mathbf{E}}{N}$; $\quad GP = \lambda P \quad$ => 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 vector P^*: $G^* = \alpha \mathbf{S}^* + (1-\alpha)\frac{\mathbf{E}}{N}$, $G^* P^* = P^*$
  ($\mathbf{S}^*$ with inverted link directions)
  Fogaras (2003) ... Chepelianskii arXiv:1003.5455 (2010) ...

Real directed 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 ingoing or outgoing links $\rho(k) \sim k^{-\nu}$

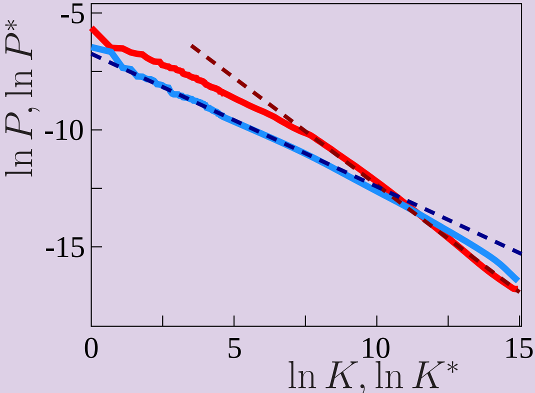
PageRank vector for large WWW:

- $P(K) \sim 1/K^{\beta}$, where K is the ordered rank 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 $\beta \approx 0.9$.
- PageRank $P(K)$ on average is proportional to the number of ingoing links
- CheiRank $P^*(K^*) \sim 1/K^{*\beta}$ on average is proportional to the number of outgoing links ($\nu \approx 2.7; \beta = 1/(\nu - 1) \approx 0.6$)
- WWW at present: $\sim 10^{11}$ web pages

Donato *et al.* EPJB **38**, 239 (2004)

Wikipedia ranking of human knowledge

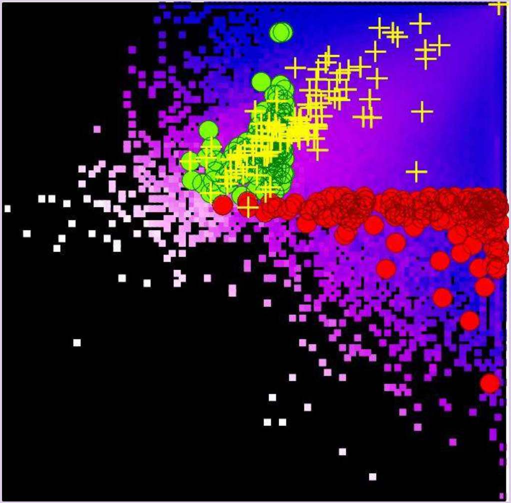
Wikipedia English articles $N = 3282257$ dated Aug 18, 2009



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 EPJB **77**, 523 (2010)]

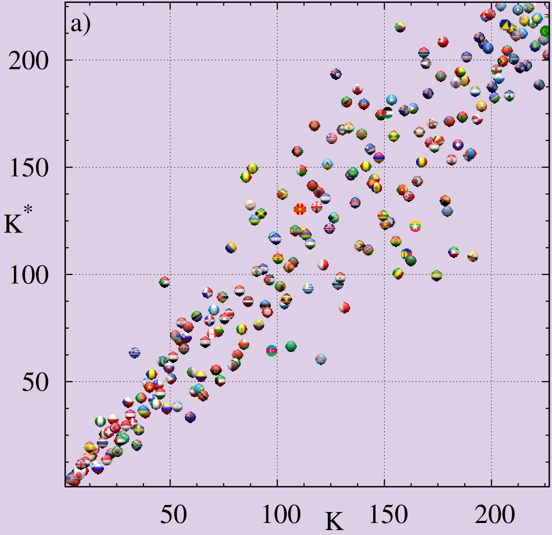
Two-dimensional ranking of Wikipedia articles



Density distribution in plane of PageRank and CheiRank indexes $(\ln K, \ln K^*)$: 100 top personalities from PageRank (green), CheiRank (red) and Hart book (yellow)

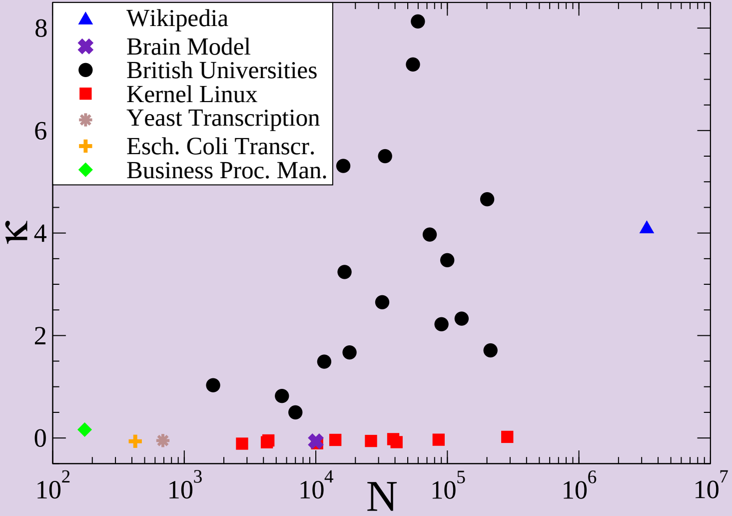
Ranking of World Trade

UN COMTRADE database 2008: All commodities



Ermann, DS arxiv:1103.5027 (2011)

Correlator of PageRank and CheiRank



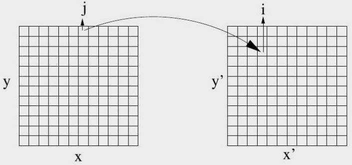
$\kappa = N \sum_i P(K(i))P^*(K^*(i)) - 1$ [Chepelianskii, Ermann, DS arxiv:1106.6215]

Ulam networks

Ulam conjecture (method) for discrete approximant of Perron-Frobenius operator of dynamical systems

Discretized phase-space:
Adjacency matrix $\mathbf{A} = P(j \rightarrow i)$

$N = N_x \times N_y$ cells.



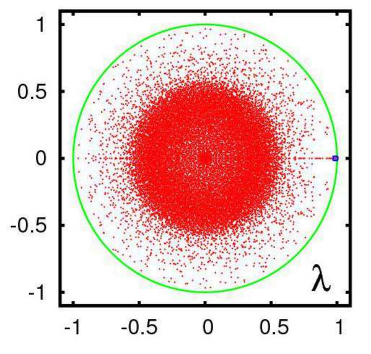
N_c: traj. from cell j
N_i: traj. to call i

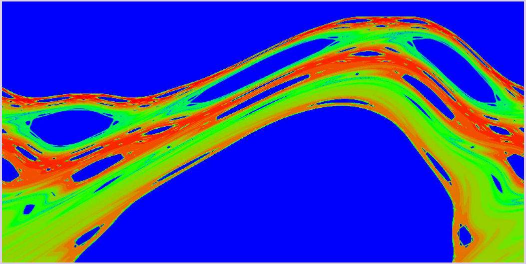
$$\begin{cases} \mathbf{A}_{i,j} = N_i / N_c \\ \sum_i \mathbf{A}_{i,j} = 1 \quad \text{(closed systems)} \end{cases}$$

S.M.Ulam, *A Collection of mathematical problems*, Interscience, **8**, 73 N.Y. (1960)
A rigorous prove for hyperbolic maps:
T.-Y.Li J.Approx. Theory **17**, 177 (1976)
Related works:
Z. Kovacs and T. Tel, Phys. Rev. A 40, 4641 (1989)
M.Blank, G.Keller, and C.Liverani, Nonlinearity **15**, 1905 (2002)
D.Terhesiu and G.Froyland, Nonlinearity **21**, 1953 (2008)

Links to Markov chains: $\infty\infty\infty\infty\infty\infty\infty\infty\infty\infty$
Contre-example: Hamiltonian systems with invariant curves, e.g. the Chirikov standard map: noise, induced by coarse-graining, destroys the KAM curves and gives homogeneous ergodic eigenvector at $\lambda = 1$

Ulam method for the Chirikov standard map



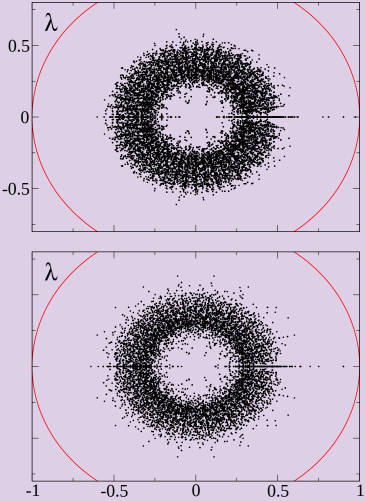


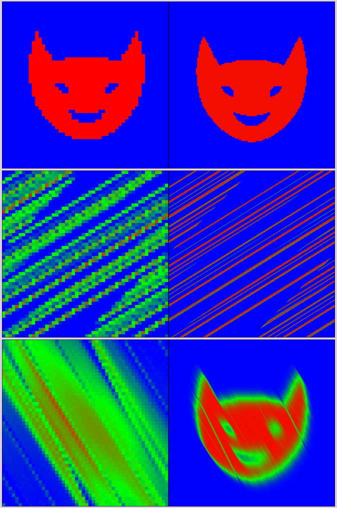
Left: spectrum $G\psi = \lambda\psi$, $M \times M/2$ cells; $M = 280$, $N_d = 16609$, exact and Arnoldi method for matrix diagonalization; generalized Ulam method of one trajectory.

Right: modulus of eigenstate of $\lambda_2 = 0.99878...$, $M = 1600$, $N_d = 494964$. Here $K = K_G$

(Frahm, DS (2010))

Arnold cat map, Ulam method and time reversal



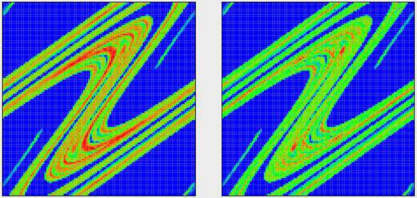


Left: spectrum $L = 4$, $N_x = 43$ (left); $L = 8$, $N_x = 51$ (right).
Right: time revesal at $N_x = 51; 751$ (left; right), $L = 8$; top to bottom time $t = 0$, $t = t_r = 4$, $t = 2t_r = 8$. (Ermann, DS Physica D **241**, 514 (2012))

Ulam method for dissipative systems

Scattering

$$\begin{cases} \bar{y} = y + K\sin(x + y/2) \\ \bar{x} = x + (y + \bar{y})/2 \pmod{2\pi} \end{cases}$$

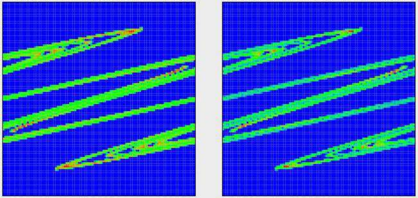


$N = 110 \times 110,\ K = 7, a = 2$

$\lambda_1 = 0.756 \quad \lambda_3 = -0.01 + i0.513$

Dissipation

$$\begin{cases} \bar{y} = \eta y + K\sin x \\ \bar{x} = x + \bar{y} \pmod{2\pi} \end{cases}$$



$N = 110 \times 110,\ K = 7, \eta = 0.3$

$\lambda_1 = 1 \quad \lambda_3 = -0.258 + i0.445$

(Ermann, DS (2010))

Fractal Weyl law

invented for open quantum systems, quantum chaotic scattering:
the number of Gamow eigenstates N_γ, that have escape rates γ in a finite bandwidth $0 \le \gamma \le \gamma_b$, scales as

$$N_\gamma \propto \hbar^{-\nu}, \quad \nu = d/2$$

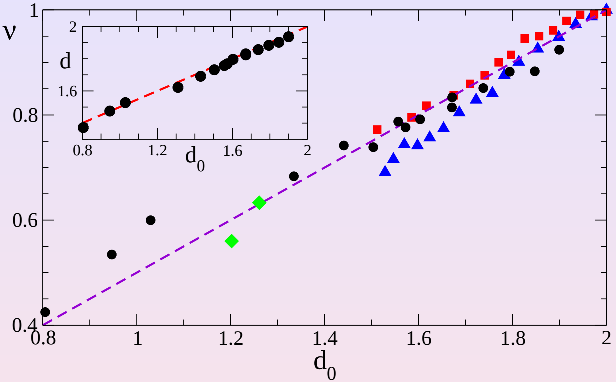
where d is a fractal dimension of a strange invariant set formed by obits non-escaping in the future and in the past

References:
J.Sjostrand, Duke Math. J. **60**, 1 (1990)
M.Zworski, Not. Am. Math. Soc. **46**, 319 (1999)
W.T.Lu, S.Sridhar and M.Zworski, Phys. Rev. Lett. **91**, 154101 (2003)
S.Nonnenmacher and M.Zworski, Commun. Math. Phys. **269**, 311 (2007)

Quantum Chirikov standard map with absorption
F.Borgonovi, I.Guarneri, DLS, Phys. Rev. A **43**, 4517 (1991)
DLS, Phys. Rev. E **77**, 015202(R) (2008)

Perron-Frobenius operators?

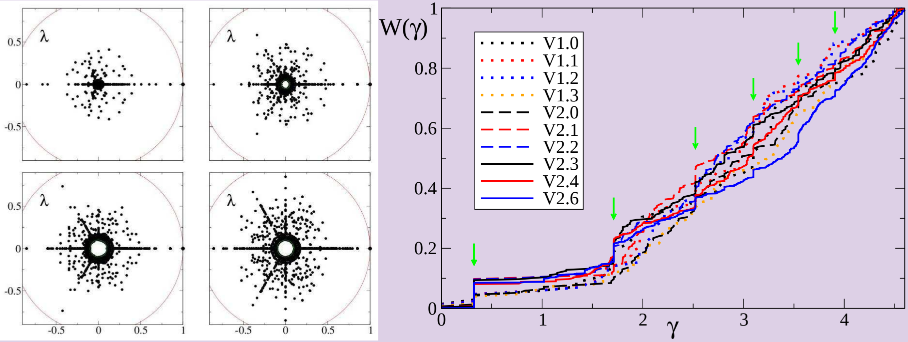
Fractal Weyl law for Ulam networks



Fractal Weyl law for three different models with dimension d_0 of invariant set. The fractal Weyl exponent ν is shown as a function of fractal dimension d_0 of the strange repeller in model 1 and strange attractor in model 2 and Henon map; dashed line shows the theory dependence $\nu = d_0/2$. Inset shows relation between the fractal dimension d of trajectories nonescaping in future and the fractal inv-set dimension d_0 for model 1; dashed line is $d = d_0/2 + 1$. (Ermann, DS (2010))

Fractal Weyl law for Linux Network

Sjöstrand Duke Math J. 60, 1 (1990), Zworski *et al.* PRL 91, 154101 (2003) $\rightarrow$ quantum chaotic scattering;
Ermann, DS EPJB 75, 299 (2010)$\rightarrow$ Perron-Frobenius operators

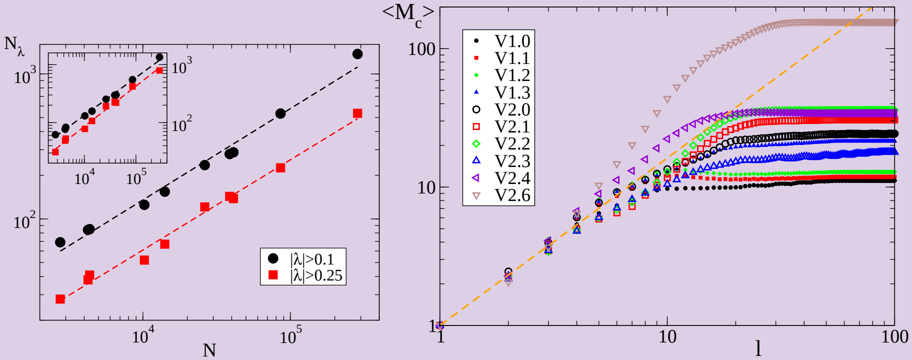


Spectrum of Google matrix (left); integrated density of states for relaxation rate $\gamma = -2 \ln |\lambda|$ (right) for Linux versions, $\alpha = 0.85$.

(Chepelianskii arxiv:1003.5455; Ermann, Chepelianskii, DS (2011))

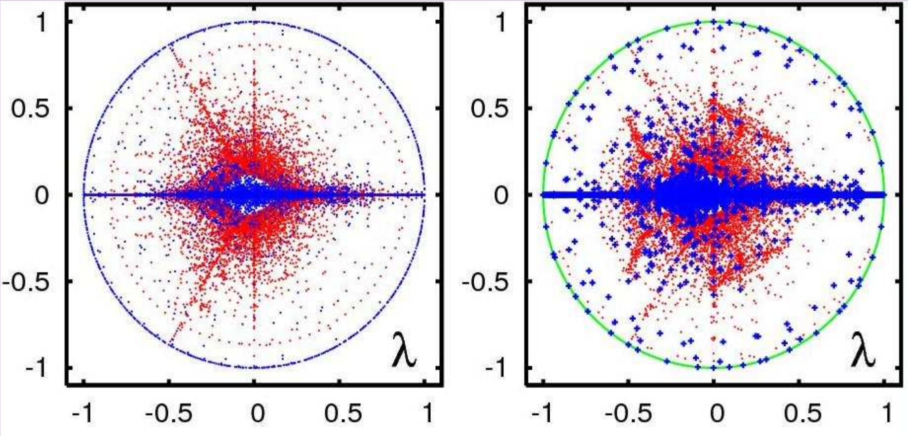
Fractal Weyl law for Linux Network

Number of states $N_\lambda \sim N^\nu, \quad \nu = d/2 \quad (N \sim 1/\hbar^{d/2})$



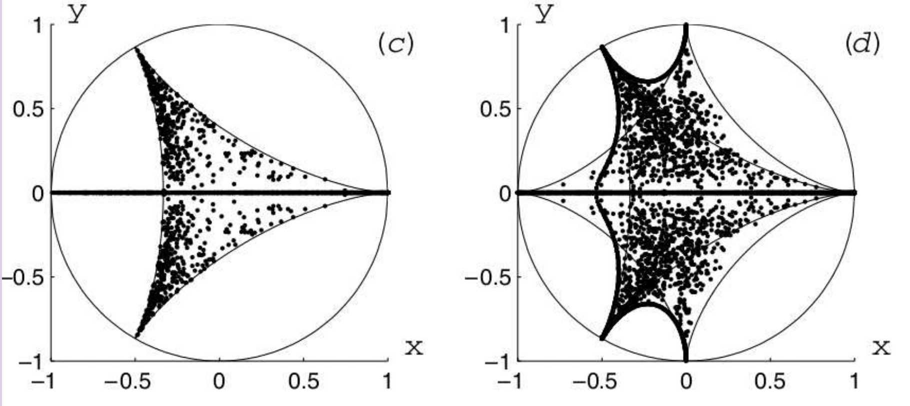
Number of states N_λ with $|\lambda| > 0.1; 0.25$ vs. N, lines show $N_\lambda \sim N^\nu$ with $\nu \approx 0.65$ (left); average mass $< M_c >$ (number of nodes) as a functon of network distance l, line shows the power law for fractal dimension $< M_c > \sim l^d$ with $d \approx 1.3$ (right).

Spectrum of UK University networks



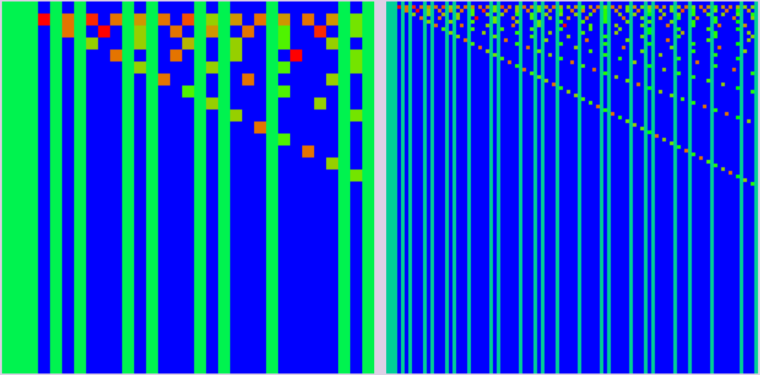
Arnoldi method: Spectrum of Google matrix for Univ. of Cambridge (left) and Oxford (right) in 2006; 20% at $\lambda = 1$ ($N \approx 200000$, $\alpha = 1$). [Frahm, Georgeot, DS arxiv:1105.1062 (2011)]

Spectrum of random orthostochastic matrices



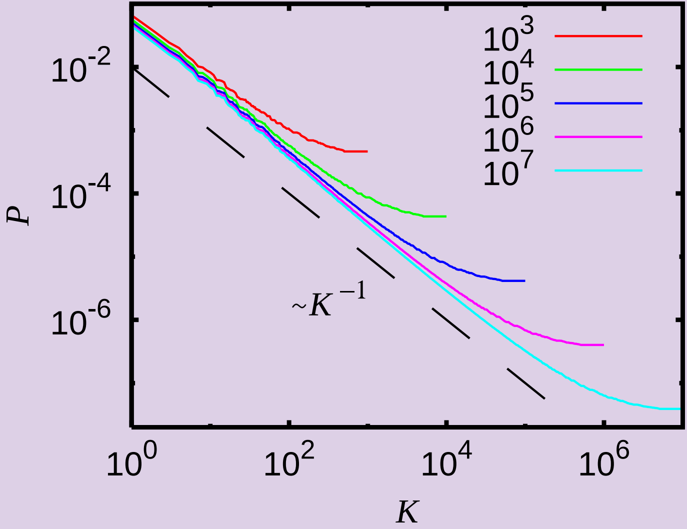
Spectrum $N = 3$ (left), 4 (right) [K.Zyczkowski *et al.* J.Phys. A **36**, 3425 (2003)]

Google matrix of integers



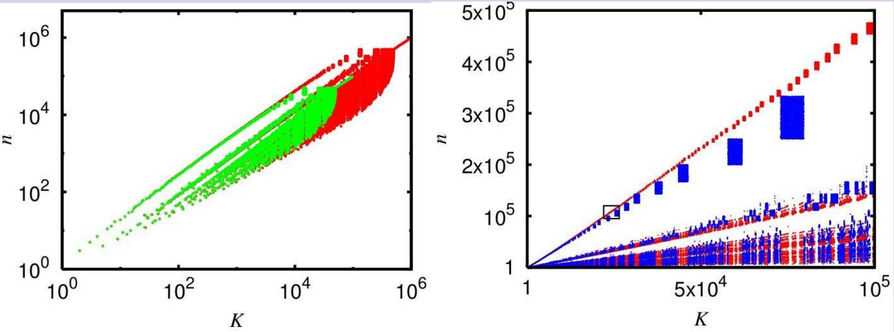
$N = 31, 101 \rightarrow$ link from n to m if m is divisor of n (multiple count); $\alpha = 1$
[Frahm, Chepelianksii, DS (in preparation)]

PageRank of integers



PageRank vs its index K for various matrix size

PageRank order of integers

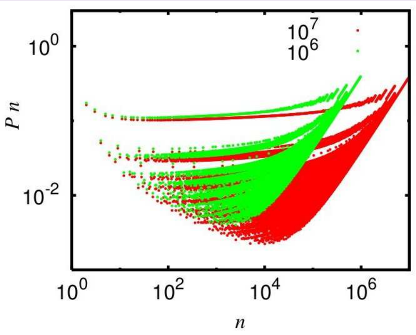


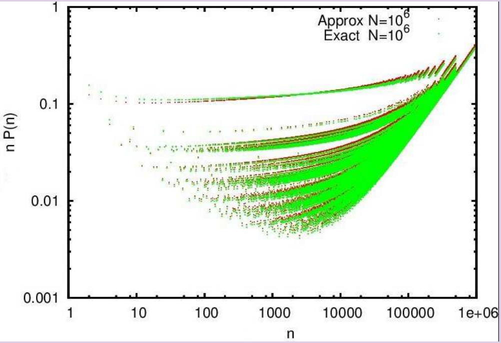
integer number n vs. PageRank index K at $N = 10^6; 10^7$

Eigenvalues: $|\lambda_{1,2}| \approx 0.48...$, $|\lambda_3| \approx 0.1...$

Superconvergence

Plateaus of PageRank of integers





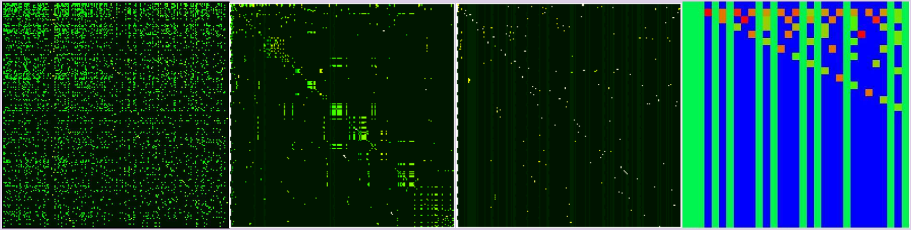
An iterative solution can be constructed:

$$P_N(n) = \sigma_N \left(1 + \sum_{m_1 \geq 2}^{m_1 n \leq N} \frac{1}{d(m_1 n) - 2} + \sum_{m_1 \geq 2}^{m_1 n \leq N} \sum_{m_2 \geq 2}^{m_2 m_1 n \leq N} \frac{1}{d(m_1 n) - 2} \frac{1}{d(m_2 m_1 n) - 2} + \ldots \right)$$

where σ_N is determined from normalization $\sum_n P_N(n) = 1$; with $d(x)$ the number of divisors of number x; $d(2) = 2$, $d(3) = 2$ and $d(4) = 3$...

Google Matrix Gallery

matrices of Wikipedia, U Cambridge 2006, Linux, integers ($K, K' \leq 200$)



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

References:

R1. S.Brin and L.Page, *The anatomy of a large-scale hypertextual Web search engine*, Comp. Networks ISDN Systems **30**, 107 (1998)
R2. A.A. Markov, *Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga*, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, **15** (1906) 135 (in Russian) [English trans.: *Extension of the limit theorems of probability theory to a sum of variables connected in a chain* reprinted in Appendix B of: R.A. Howard *Dynamic Probabilistic Systems*, volume 1: *Markov models*, Dover Publ. (2007)].
R3. D.Austin, *How Google Finds Your Needle in the Web's Haystack*. AMS Feature Columns, http://www.ams.org/samplings/feature-column/fcarc-pagerank (2008)
R4. Wikipedia articles *PageRank*, *CheiRank*, *Google matrix* (2008-2011)
R5. D.Fogaras, *Where to start browsing the web?*, Lect. Notes Computer Sci. **2877**, 65 (2003)
R6. V.Hrisitidis, H.Hwang and Y.Papakonstantinou, *Authority-based keyword search in databases*, ACM Trans. Database Syst. **33**, 1 (2008)
R7. A.D.Chepelianksii, *Towards physical laws for software architecture* arXiv:1003.5455[cs.SE] (2010)
R8. A.O.Zhirov, O.V.Zhirov and D.L.Shepelyansky, *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77**, 523 (2010)
R9. S.M. Ulam, *A Collection of mathematical problems*, Vol. 8 of Interscience tracs in pure and applied mathematics, Interscience, New York, p. 73 (1960).

References (continued):

R10. K.M.Frahm and D.L.Shepelyansky, *Ulam method for the Chirikov standard map*, Eur. Phys. J. B **76**, 57 (2010)

R11. L.Ermann and D.L.Shepelyansky, *Ulam method and fractal Weyl law for Perron-Frobenius operators*, Eur. Phys. J. B **75**, 299 (2010)

R12. L.Ermann, A.D.Chepelianskii and D.L.Shepelyansky, *Fractal Weyl law for Linux Kernel Architecture*, Eur. Phys. J. B **79**, 115 (2011)

R13. L.Ermann and D.L.Shepelyansky, *Google matrix of the world trade network*, arxiv:1103.5027 (2011)

R14. L.Ermann, A.D.Chepelianskii and D.L.Shepelyansky, *Towards two-dimensional search engines*, arxiv:1106.6215[cs.IR] (2011)

R15. K.M.Frahm, B.Georgeot and D.L.Shepelyansky, *Universal emergence of PageRank*, arxiv:1105.1062[cs.IR] (2011)

R16. L.Ermann and D.L.Shepelyansky, *The Arnold cam map, the Ulam method and time reversal*, Physica D **241**, 514 (2012)

Books, reviews:

B1. A. M. Langville and C. D. Meyer, *Google's PageRank and beyond: the science of search engine rankings*, Princeton University Press, Princeton (2006)

B2. M. Brin and G. Stuck, *Introduction to dynamical systems*, Cambridge Univ. Press, Cambridge, UK (2002).

B3. E. Ott, *Chaos in dynamical systems*, Cambridge Univ. Press, Cambridge (1993).