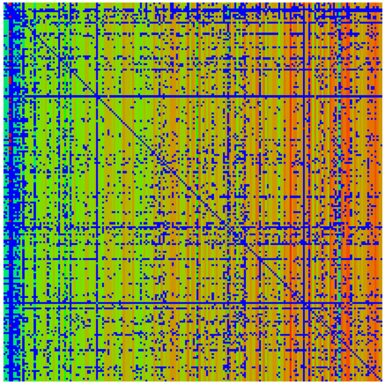
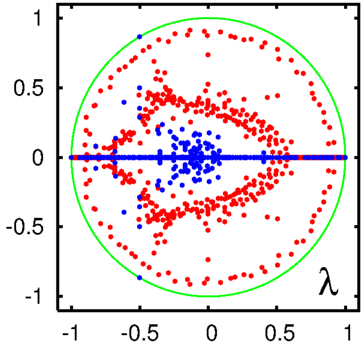




Google matrix of directed networks

Klaus Frahm

Quantware MIPS Center

Université Paul Sabatier

Laboratoire de Physique Théorique, UMR 5152, IRSAMC, CNRS

supported by EC FET Open project NADINE

Trento Workshop, Spectral properties of complex networks, Trento, 24 July 2012

Perron-Frobenius operators

discrete ***Markov process***:

$$p_i(t+1) = \sum_j G_{ij}\, p_j(t)$$

with probabilities $p_i(t) \geq 0$ and the ***Perron-Frobenius*** matrix G such that:

$$\sum_i G_{ij} = 1 \quad , \quad G_{ij} \geq 0 \; .$$

For any vector v :

$$\Rightarrow \qquad \|Gv\|_1 \leq \|v\|_1$$

$\Rightarrow$ complex eigenvalues $|\lambda_j| \leq 1$ and (at least) one eigenvalue $\lambda_1 = 1$ and its right eigenvector P is the stationary distribution:

$$P = \lim_{t\to\infty} p(t)$$

provided λ_1 is <u>not degenerate</u> !

Google matrix for directed networks

Define the ***adjacency matrix*** A by $A_{ij} = 1$ if there is a link from the node j to i in the network (of size N) and $A_{ij} = 0$ otherwise. Let $S_{ij} = A_{ij} / \sum_i A_{ij}$ and $S_{ij} = 1/N$ if $\sum_i A_{ij} = 0$ (dangling nodes). S is of Perron-Frobenius type but for many networks the eigenvalue $\lambda_1 = 1$ is highly degenerate [$\Rightarrow$ convergence problem to arrive at the stationary limit of $p(t+1) = S\, p(t)$].

Therefore define the ***Google matrix***:

$$G(\alpha) = \alpha S + (1-\alpha)\, \frac{1}{N}\, e e^T$$

where $e = (1, \ldots, 1)^T$ and $\alpha = 0.85$ is a typical damping factor. Here there is unique eigenvector for $\lambda_1 = 1$ called the ***PageRank*** P and the convergence goes with α^t.

(***CheiRank*** P^* by replacing: $A \to A^* = A^T$).

Ulam Method

(Ermann, Shepelyansky (2010), KF, Shepelyansky (2010))

to construct a ***Perron-Frobenius matrix*** as discrete approximation for the PF operator of dynamical systems with mixed phase space:

- Subdivide phase space in discrete cells.

- Iterate (for a very long time) a classical trajectory and attribute a new number to each new cell which is entered for the first time. At the same time count the number of transitions from cell i to cell j ($\Rightarrow\ n_{ji}$).

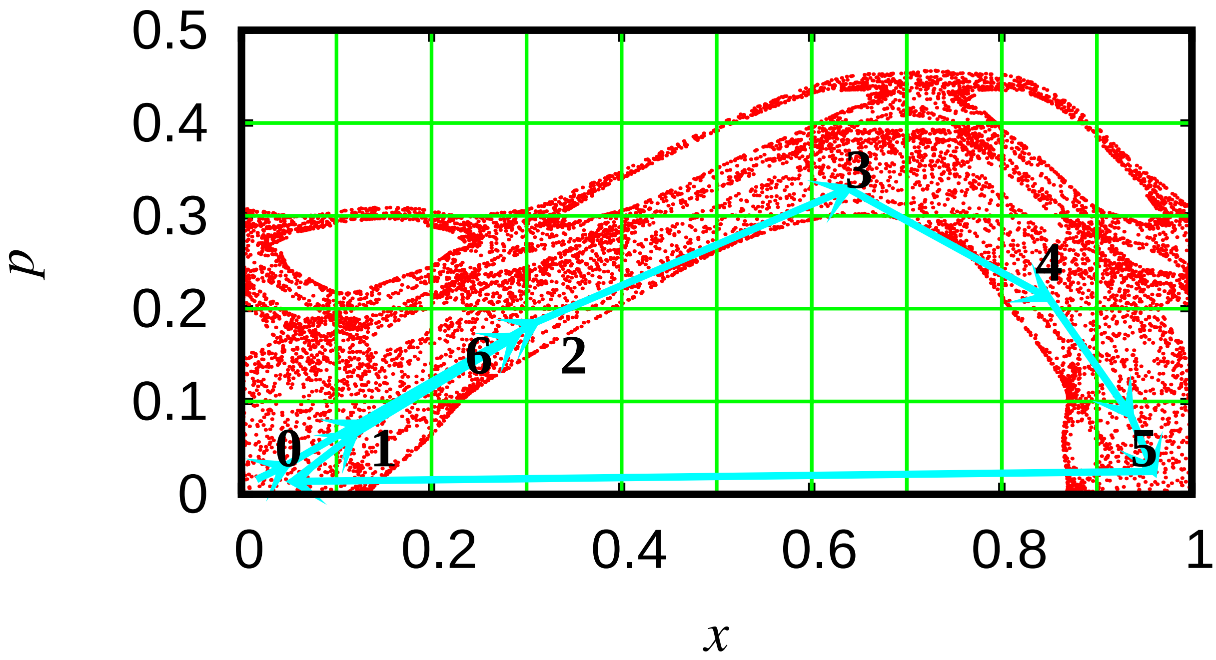
- $\Rightarrow$ The matrix

$$G_{ji} = \frac{n_{ji}}{\sum_l n_{li}}$$

is of Perron-Frobenius type : $G_{ji} \geq 0$, $\sum_j G_{ji} = 1$.

Chirikov Standard map

$$p_{n+1} \;=\; p_n + \frac{k}{2\pi}\sin(2\pi\, x_n)$$
$$x_{n+1} \;=\; x_n + p_{n+1} \quad , \quad k = k_c = 0.971635406$$



Arnoldi method

to (partly) diagonalize large sparse non-symmetric $d \times d$ matrices:

- choose an initial normalized vector ξ_0 (random or "otherwise")

- determine the ***Krylov space*** of dimension n (typically: $1 \ll n \ll d$) spanned by the vectors: $\xi_0,\ G\,\xi_0,\ \ldots,\ G^{n-1}\xi_0$

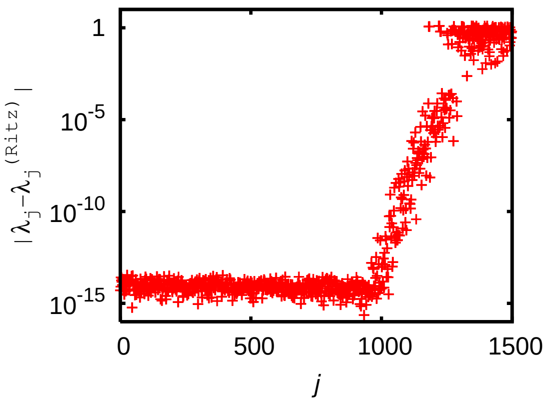
- determine by ***Gram-Schmidt*** orthogonalization an orthonormal basis $\{\xi_0,\ \ldots,\xi_{n-1}\}$ and the representation of G in this basis:

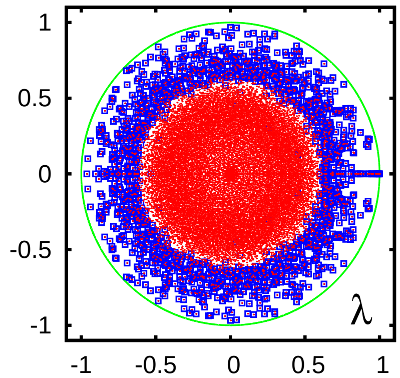
$$G\,\xi_k = \sum_{j=0}^{k+1} H_{jk}\,\xi_j$$

- diagonalize the ***Arnoldi matrix*** H which has ***Hessenberg*** form:

$$H = \begin{pmatrix} * & * & \cdots & * & * \\ * & * & \cdots & * & * \\ 0 & * & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & * & * \\ \hline 0 & 0 & \cdots & 0 & * \end{pmatrix}$$

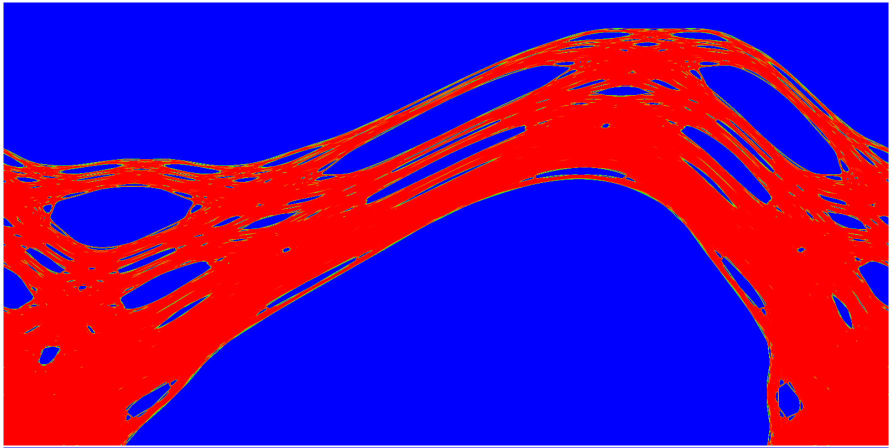
which provides the ***Ritz eigenvalues*** that are very good aproximations to the “largest” eigenvalues of A.





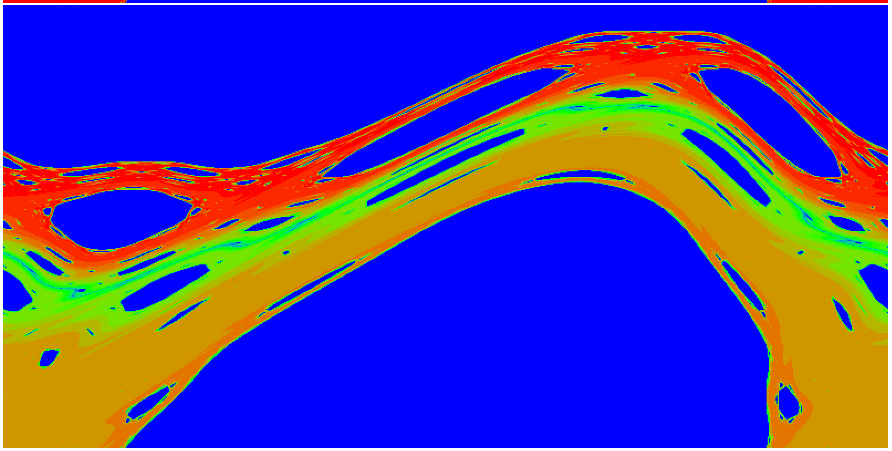
$M \times M/2$ cells, $M = 280$, $d = 16609$, $n = 1500$

Eigenvectors



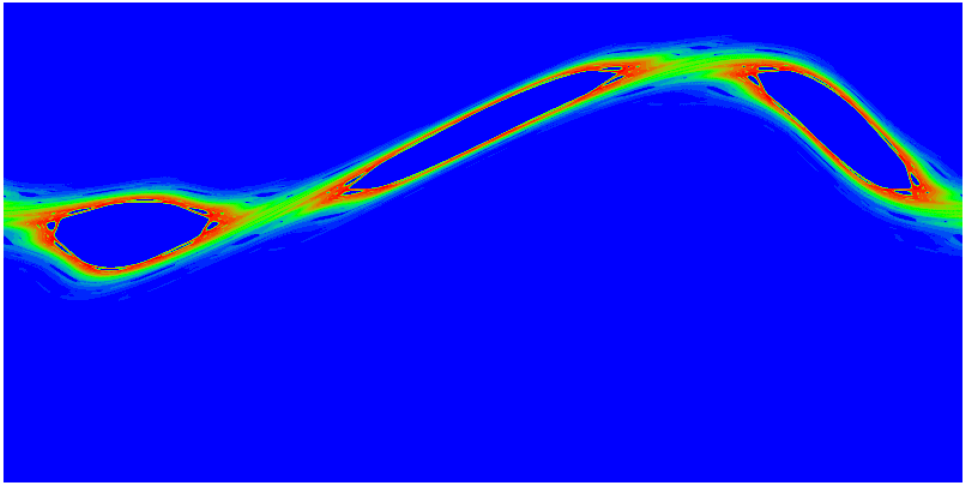
$\lambda_1 = 1$

$M = 1600$
$d = 494964$
$n = 3000$



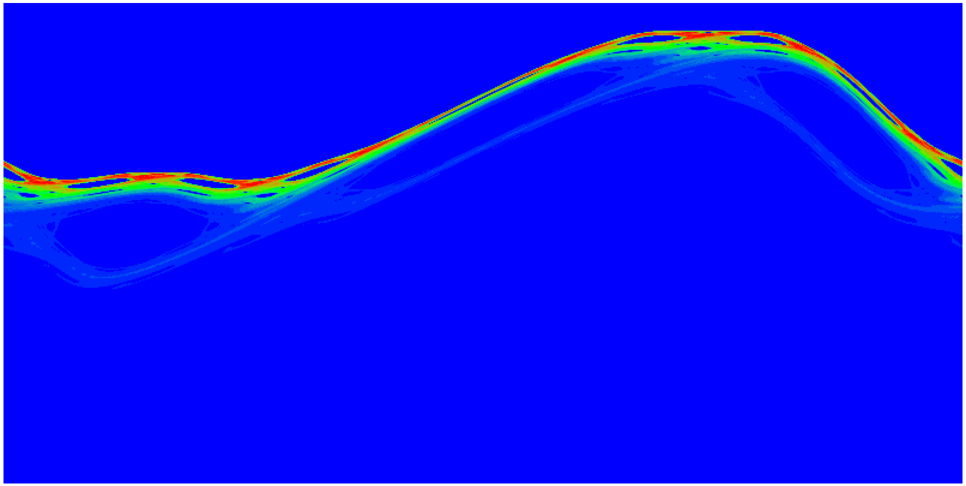
$\lambda_1 = 0.99980431$

$M = 800$
$d = 127282$
$n = 2000$



λ_6
-0.49699831
$+i\, 0.86089756$
$\approx |\lambda_6|\, e^{i\, 2\pi/3}$

$M = 800$
$d = 127282$
$n = 2000$



$\lambda_{19} =$
-0.71213331
$+i\, 0.67961609$
$\approx |\lambda_{19}|\, e^{i\, 2\pi(3/8)}$

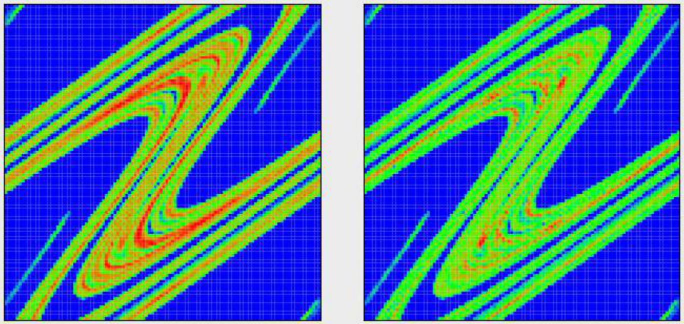
$M = 800$
$d = 127282$
$n = 2000$

Ulam method for dissipative systems

(Ermann, Shepelyansky (2010))

Scattering

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

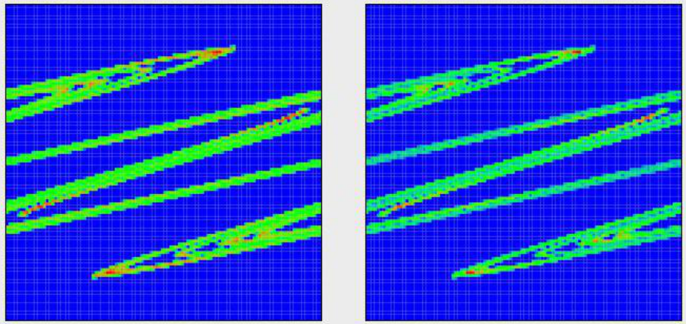


$N = 110 \times 110,\ K = 7, a = 2$

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

Dissipation

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



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

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

Fractal Weyl law

$N_\gamma =$ number of Gamow eigenstates that have escape rates $\gamma_j = -2\ln|\lambda_j|$ in a finite bandwidth $0 \le \gamma_j \le \gamma_b$.

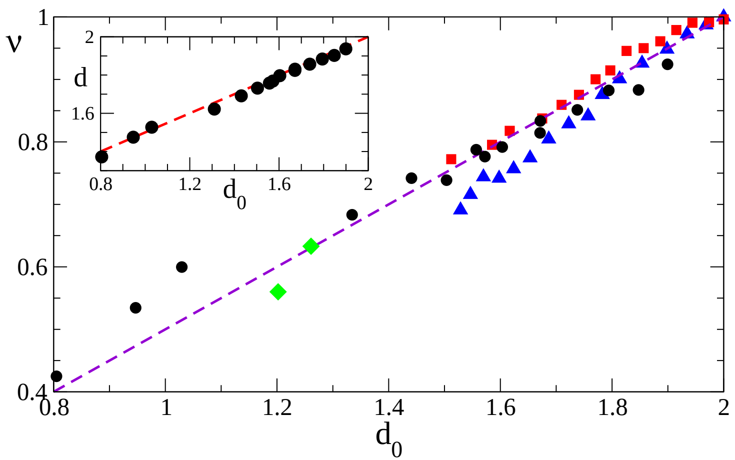
Fractal Weyl law for open quantum systems :

(e.g. Shepelyansky (2008))

$N_\gamma \propto N^{d-1} \propto \hbar^{-(d-1)}$ where d is a fractal dimension of a strange invariant set formed by orbits non-escaping in the future.

Fractal Weyl law for Ulam networks : $N_\gamma \propto N^\nu \propto N^{d_0/2}$

(Ermann, Shepelyansky (2010))



d_0 = dimension of invariant set of strange repeller (formed by orbits non-escaping in the future **and** in the past).

$\nu = d_0/2$

$d =$ dimension of orbits non-escaping in the future

$d = d_0/2 + 1$ (inset)

University networks

(KF, Georgeot, Shepelyansky (2011))

In realistic WWW networks invariant subspaces of nodes create large degeneracies of λ_1 (or λ_2 if $\alpha < 1$) which is very problematic for the Arnoldi method.

Therefore determine the ***invariant subspaces*** as follows:

Let $N_c = bN$ a certain fraction of the network size N (e.g. $b = 0.1$).

- For a given initial node i_0 determine a sequence of node sets S_n by $S_0 = \{i_0\}$ and S_{n+1} is the set containing all nodes of S_n and those which can be reached by a link from a node in S_n.

- If $S_n = S_{n+1}$ with at most N_c elements for some $n \Rightarrow S_n$ is an ***invariant subspace***.

- If for some n the set S_n contains a dangling node (connected by construction to any other node) or if S_n contains more than N_c elements $\Rightarrow i_0$ is identified as a node belonging to the ***core space*** (space of nodes not belonging to an invariant subspace).
- Repeat the procedure for every network node as potential initial node except for those nodes which are already identified as subspace nodes. If for some n the set S_n contains a previously found core space node $\Rightarrow i_0$ also belongs to the core space.
- Merge all subspaces with common members. In this way one obtains a decomposition of the network in many ***separate subspaces*** with N_s nodes and a "big" ***core space***.

This procedure can be efficiently implemented as a computer program. It turns out that for most networks the exact choice of b is not important (e.g. $b = 0.1$ or $b = 0.9$) as long as $b = \mathcal{O}(1)$. Note that a core space node may have a link to an invariant subspace but a subspace node may not have a link to another subspace or the core space.

The decomposition in subspaces and a core space implies a block structure of the matrix S:

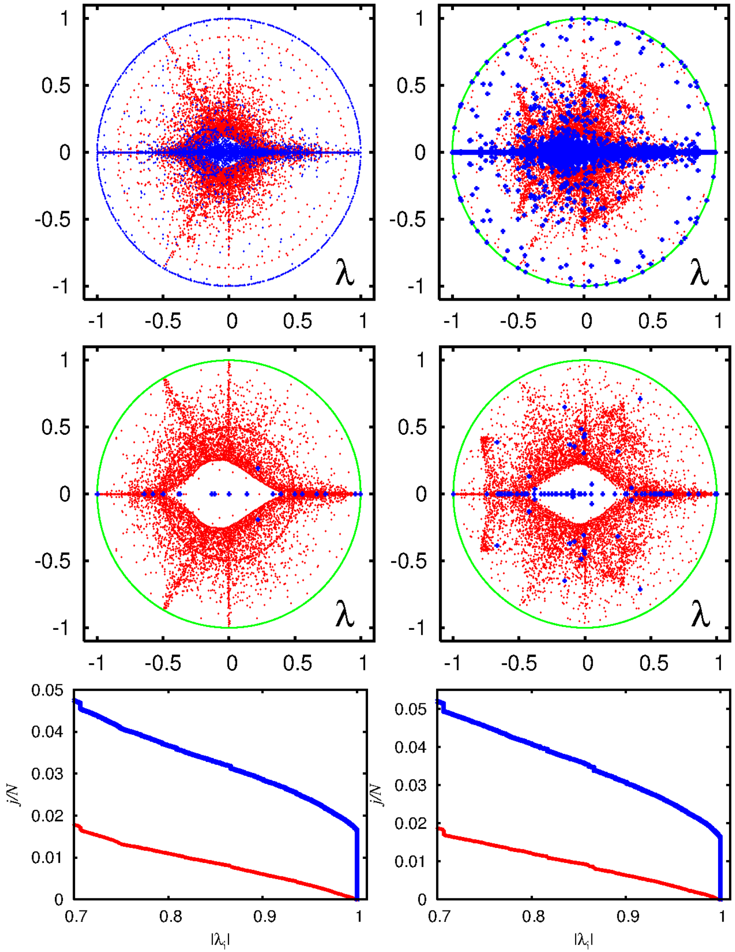
$$S = \begin{pmatrix} S_{ss} & S_{sc} \\ 0 & S_{cc} \end{pmatrix}$$

where S_{ss} is block diagonal according to the subspaces. The subspace blocks of S_{ss} are all matrices of PF type with at least one eigenvalue $\lambda_1 = 1$ explaining the high degeneracies.

To determine the spectrum of S apply:

- Exact (or Arnoldi) diagonalization on each subspace.

- The Arnoldi method to S_{cc} to determine the largest core space eigenvalues λ_j (note: $|\lambda_j| < 1$). The largest eigenvalues of S_{cc} are no longer degenerate but other degeneracies are possible (e.g. $\lambda_j = 0.9$ for Wikipedia).



Cambridge 2006 (left), $N = 212710$, $N_s = 48239$

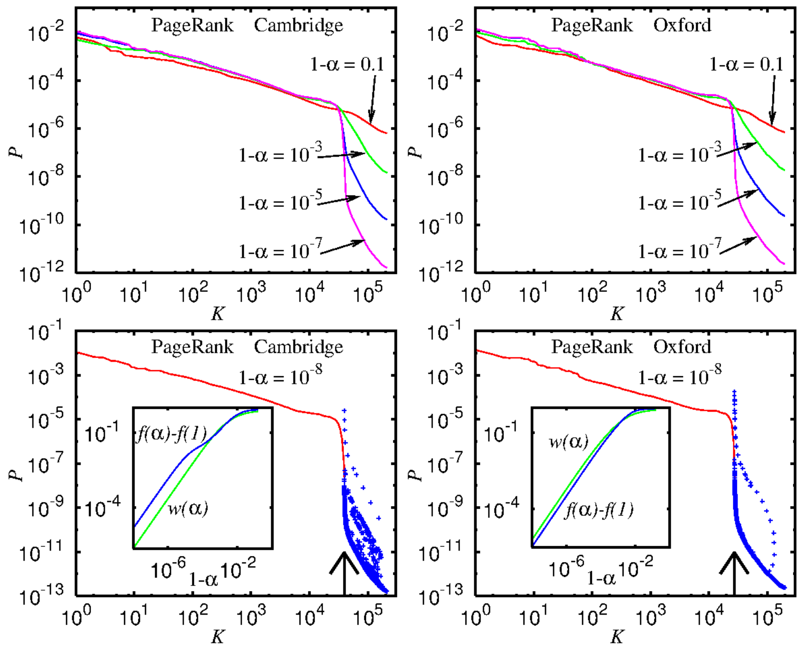
Oxford 2006 (right), $N = 200823$, $N_s = 30579$

Spectrum of S (upper panels), S^* (middle panels) and dependence of rescaled level number on $|\lambda_j|$ (lower panels).

Blue: subspace eigenvalues

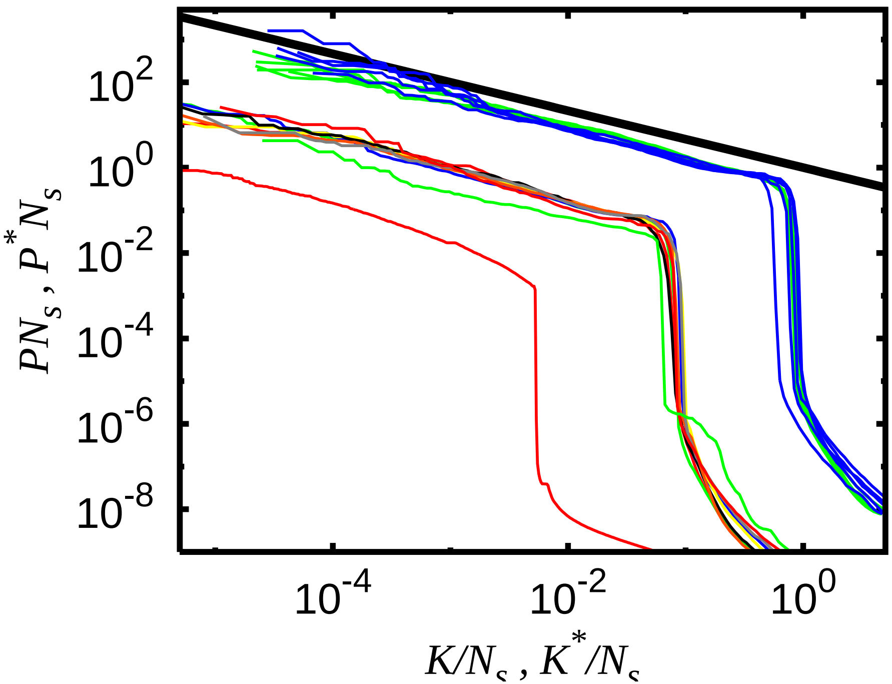
Red: core space eigenvalues (with Arnoldi dimension $n_A = 20000$)

PageRank for $\alpha \to 1$:



$$P = \underbrace{\sum_{\lambda_j=1} c_j\,\psi_j}_{\text{subspace contributions}} + \sum_{\lambda_j\neq 1} \frac{1-\alpha}{(1-\alpha)+\alpha(1-\lambda_j)}\, c_j\,\psi_j\ .$$

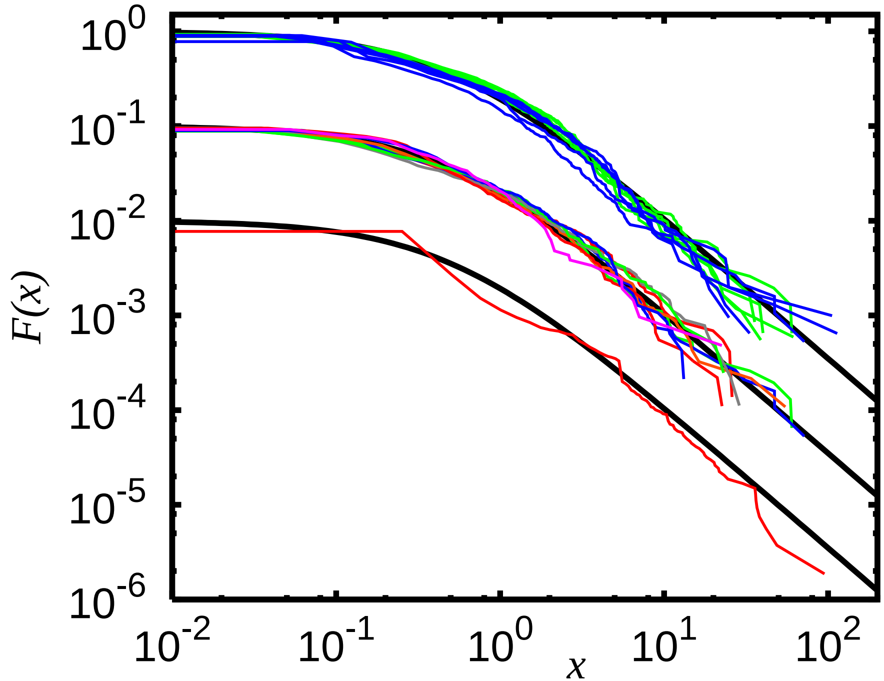
Rescaled PageRank at $\alpha = 1 - 10^{-8}$:



Top: Cambridge, Oxford 2002-2006; middle: other universities; bottom: Wikipedia*; black line $\propto K^{-2/3}$; N_s = sum of all subspace dimensions.

Distribution of dimensions of invariant subspaces

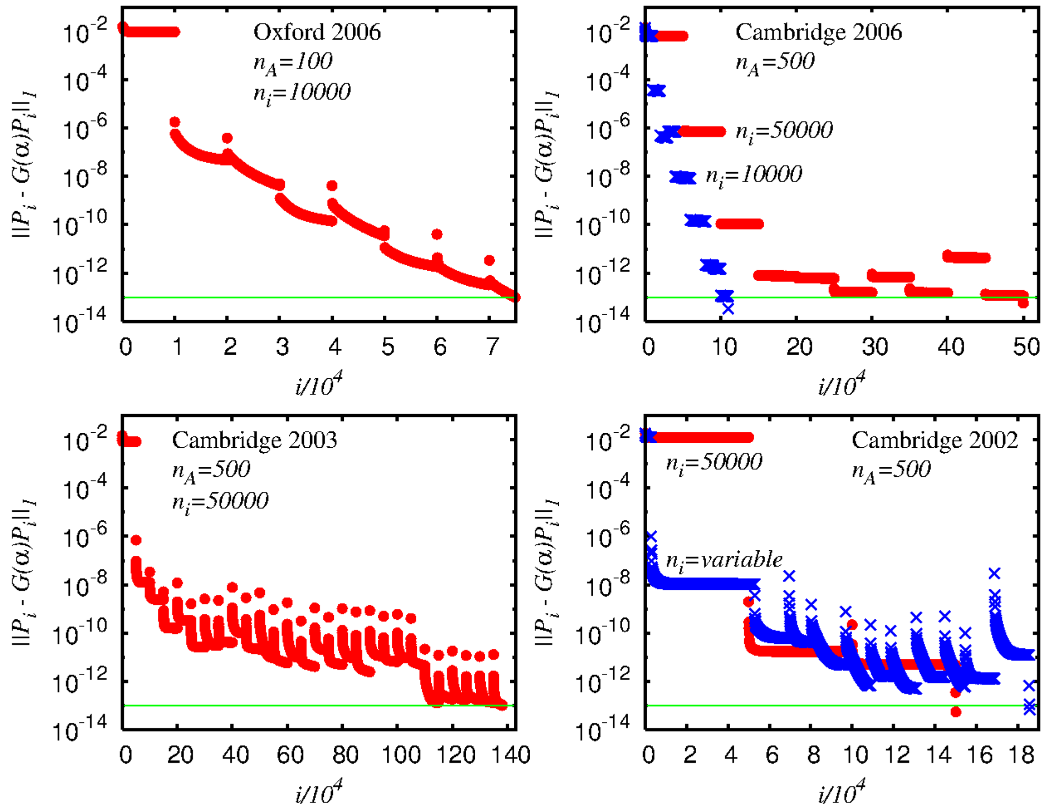
$F(x) =$ fraction of invariant subspaces with dimension larger than $x\langle d\rangle$ where $\langle d\rangle =$ average subspace dimension.



Top: Cambridge, Oxford 2002-2006; middle: other universities; bottom: Wikipedia*; black line: $F(x) = 1/(1+2x)^{3/2}$.

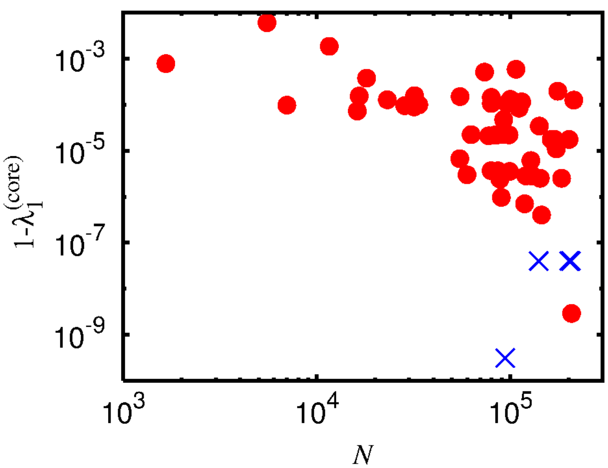
Numerical PageRank method for $\alpha \to 1$

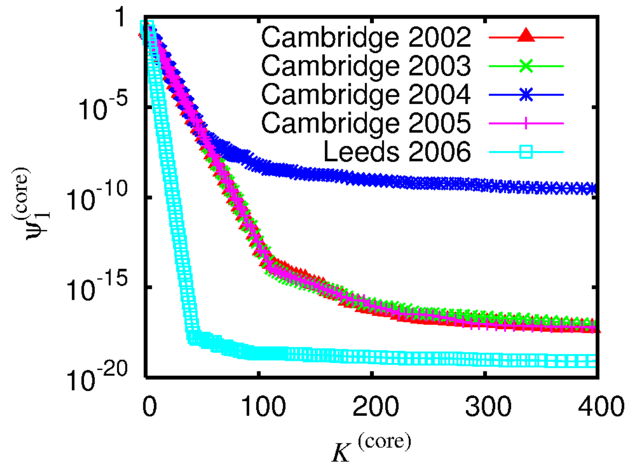
Combination of power method and Arnoldi diagonalization :



Here: $\alpha = 1 - 10^{-8}$

Core space gap and quasi-subspaces





Left: Core space gap $1 - \lambda_1^{(core)}$ vs N for certain british universities.

Red dots for gap $> 10^{-9}$; blue crosses (moved up by 10^9) for gap $< 10^{-16}$.

Right: first core space eigenvecteur for universities with gap $< 10^{-16}$ or gap $= 2.91 \times 10^{-9}$ for Cambridge 2004.

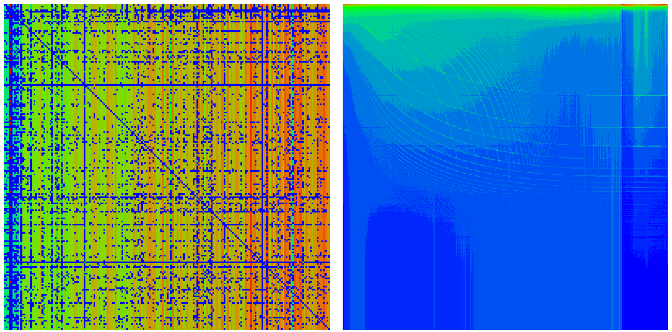
Core space gaps $< 10^{-16}$ correspond to ***quasi-subspaces*** where it takes quite many "iterations" to reach a dangling node.

Twitter network

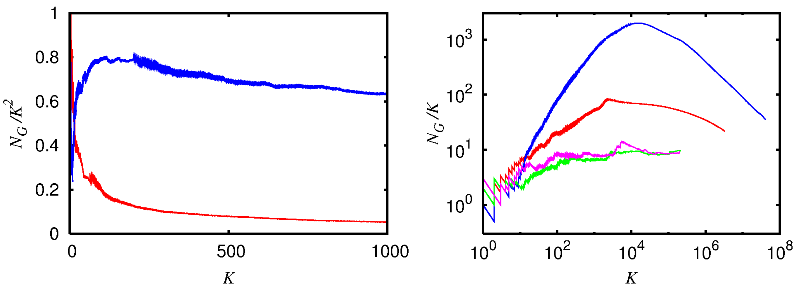
(KF, Shepelyansky (2012), preprint)

Twitter 2009 : $N = 41652230$ nodes, $N_\ell = 1468365182$ network links.

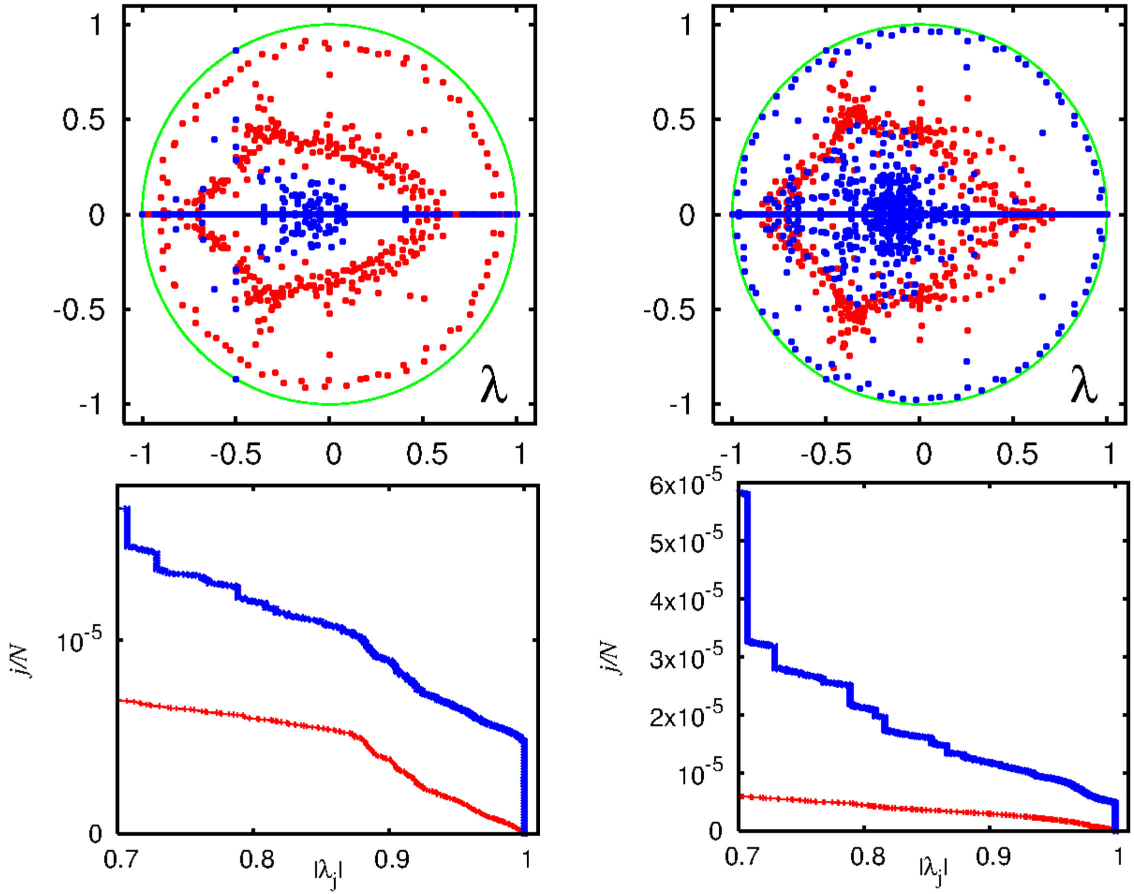
Matrix structure in K-rank order:



Number N_G of non-empty matrix elements in $K \times K$-square:

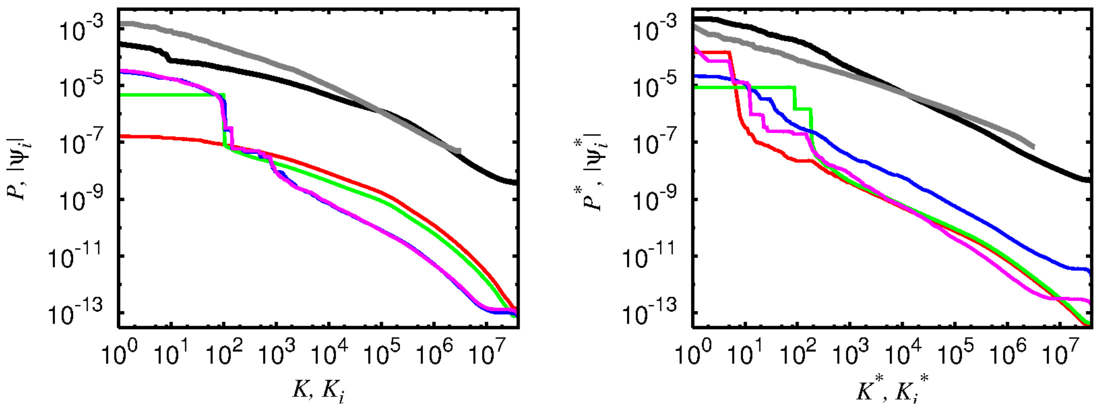


Spectrum

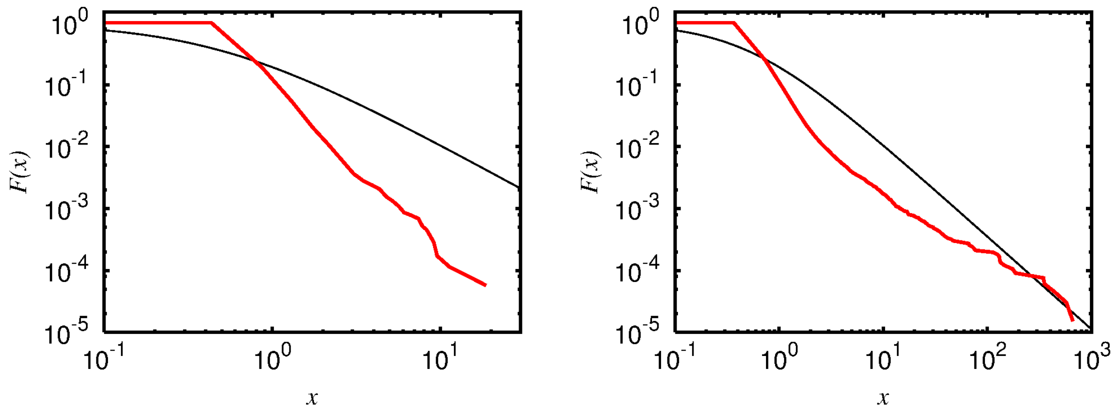


$n_A = 640 \quad \Rightarrow \quad 250$ GB of RAM memory.

PageRank, CheiRank, eigenvectors



Subspace distribution

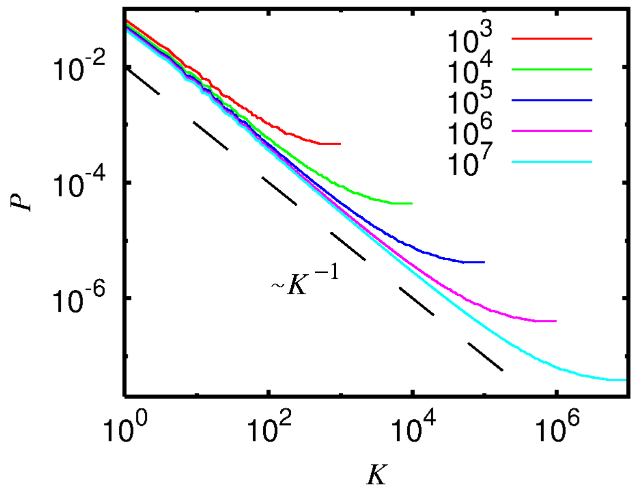


Black line: $F(x) = 1/(1+2x)^{3/2}$.

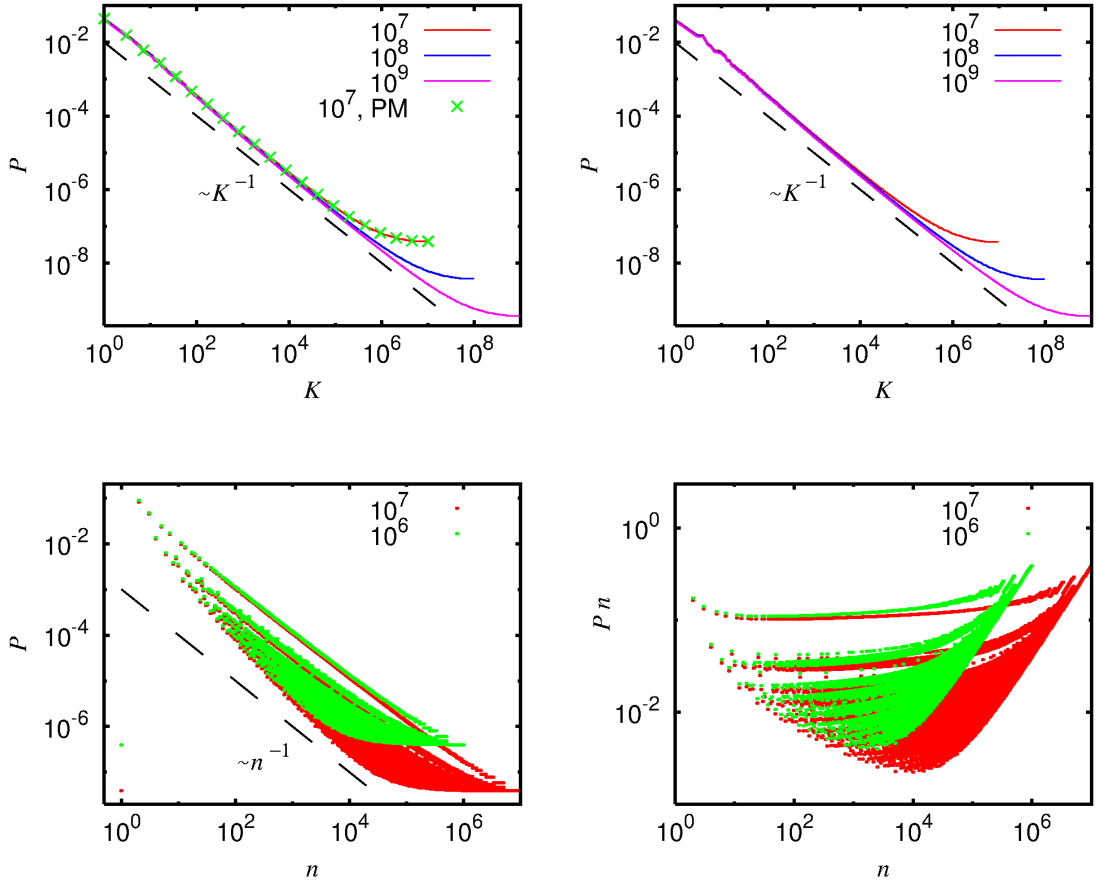
Integer network

(KF, Chepelianskii, Shepelyansky (2012), preprint)

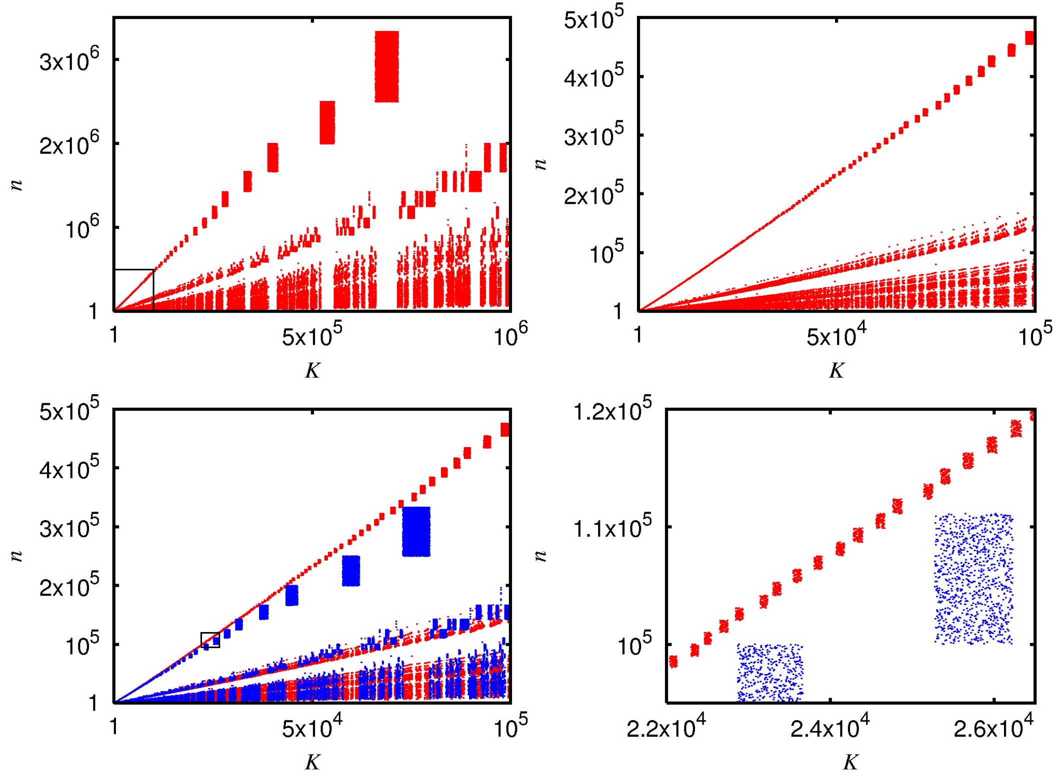
Consider the integers $n \in \{1, \ldots, N\}$ and construct an adjacency matrix by $A_{mn} = k$ where k is the largest integer such that m^k is a divisor of n if $1 < m < n$ and $A_{mn} = 0$ if $m = 1$ or $m = n$ (note $A_{mn} = k = 0$ if m is not a divisor of n). Construct S and G in the usual way from A.



PageRank



Dependence of n on K-index



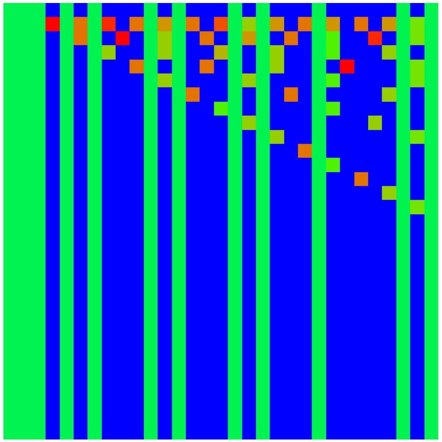
red: $N = 10^7$

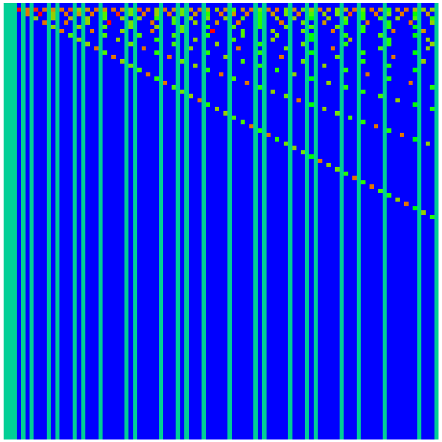
blue: $N = 10^6$

"New order" of integers: $K = 1, 2, \ldots, 32 \Rightarrow n = 2, 3, 5, 7, 4, 11, 13, 17, 6, 19, 9, 23, 29, 8, 31, 10, 37, 41, 43, 14, 47, 15, 53, 59, 61, 25, 67, 12, 71, 73, 22, 21.$

Semi-analytical determination of spectrum, PageRank and eigenvectors

Matrix structure:





$$\boxed{S = S_0 + v\, d^T}$$

where $v = e/N$, $d_j = 1$ for dangling nodes (primes and 1) and $d_j = 0$ otherwise. S_0 is the pure link matrix which is ***nil-potent***:

$$\boxed{S_0^l = 0}$$

with $l = [\log_2(N)] \ll N$.

Let ψ be an eigenvector of S with eigenvalue λ and $C = d^T\psi$.

- If $C = 0 \;\Rightarrow\; \psi$ eigenvector of $S_0 \;\Rightarrow\; \lambda = 0$ since S_0 nil-potent.

- If $C \neq 0 \;\Rightarrow\; \lambda \neq 0$ since the equation $S_0\psi = -C\,v$ does not have a solution $\Rightarrow\; \lambda\mathbf{1} - S_0$ invertible.

$$\Rightarrow\; \psi = C\,(\lambda\mathbf{1} - S_0)^{-1}\,v = \frac{C}{\lambda}\sum_{j=0}^{l-1}\left(\frac{S_0}{\lambda}\right)^j\,v \quad .$$

$$\text{From } \lambda^l = (d^T\psi/C)\lambda^l \;\Rightarrow\; \boxed{\mathcal{P}_r(\lambda) = 0}$$

with the ***reduced polynomial*** of degree $l = [\log_2(N)]$:

$$\boxed{\mathcal{P}_r(\lambda) = \lambda^l - \sum_{j=0}^{l-1}\lambda^{l-1-j}\,c_j = 0 \quad , \quad c_j = d^T\,S_0^j\,v\;.}$$

$\Rightarrow$ at most l eigenvalues $\lambda \neq 0$ which can be numerically determined as the zeros of $\mathcal{P}_r(\lambda)$. (Note: $l \leq 29$ for $N \leq 10^9$).

Furthermore for $\lambda = 1 \ \Rightarrow$ PageRank:

$$\boxed{P = C \sum_{j=0}^{l-1} S_0^j \, v \; , \; C = d^T P \; .}$$

The subspace of $\lambda \neq 0$ is represented by the vectors $v^{(j)} = S_0^j v$ for $j = 0, \ldots, l-1$

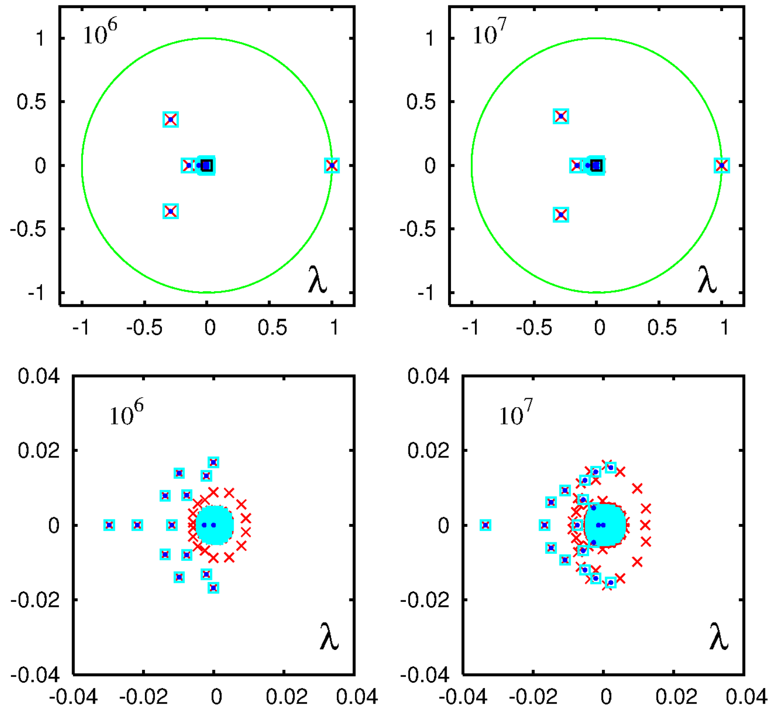
$$\Rightarrow \quad S\, v^{(j)} = c_j\, v^{(0)} + v^{(j+1)} = \sum_{k=0}^{l-1} \bar{S}_{k+1,j+1}\, v^{(k)}$$

"Small" $l \times l$-representation matrix :

$$\bar{S} = \begin{pmatrix} c_0 & c_1 & \cdots & c_{l-2} & c_{l-1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \quad , \quad \bar{P} = C \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

with $P = \sum_j \bar{P}_{j+1}\, v^{(j)} = C \sum_j v^{(j)}$ and due to sum rule: $\sum_j c_j = 1$.

Spectrum I

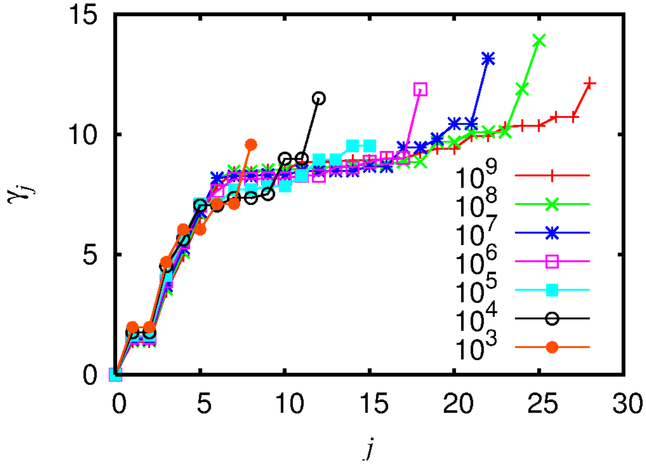


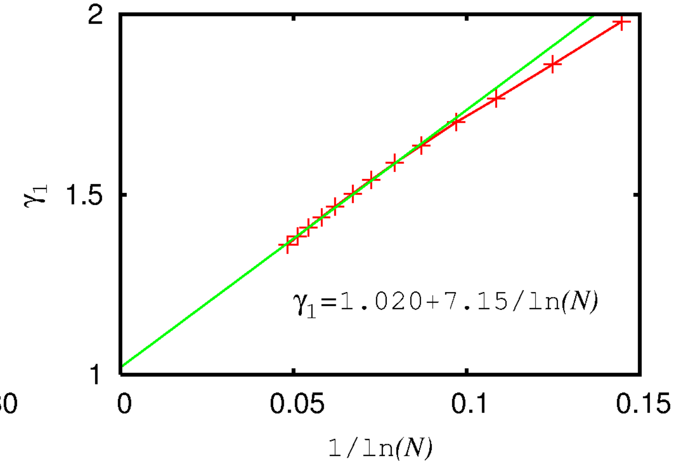
blue dots: semi-analytical eigenvalues as zeros from $\mathcal{P}_r(\lambda)$ (or eigenvalues of $\bar{S}$).

red crosses: Arnoldi method with random initial vector and $n_A = 1000$.

light blue boxes: Arnoldi method with constant initial vector $v = e/N$ and $n_A = 1000$.

Spectrum II





$$\gamma_j = -2 \ln |\lambda_j|$$

Large N limit of γ_1 with the scaling parameter: $1/\ln(N)$.

Note:

$$c_0 = d^T v = \frac{1}{N} \sum_{j=1}^{N} d_j = \frac{1 + \pi(N)}{N} \approx \frac{1}{\ln(N)}$$

where $\pi(N)$ is the number of primes below N.

References

1. D. L. Shepelyansky ***Fractal Weyl law for quantum fractal eigenstates***, Phys. Rev. E **77**, p.015202(R) (2008).
2. L. Ermann and D. L. Shepelyansky, ***Ulam method and fractal Weyl law for Perron-Frobenius operators***, Eur. Phys. J. B **75**, 299 (2010).
3. K. M. Frahm and D. L. Shepelyansky, ***Ulam method for the Chirikov standard map***, Eur. Phys. J. B **76**, 57 (2010).
4. K. M. Frahm, B. Georgeot and D. L. Shepelyansky, ***Universal emergence of PageRank***, J. Phys. A: Math. Theor. **44**, 465101 (2011).
5. K. M. Frahm, A. D. Chepelianskii and D. L. Shepelyansky, ***PageRank of integers***, arxiv:1205.6343[cs.IR] (2012).
6. K. M. Frahm and D. L. Shepelyansky, ***Google matrix of Twitter***, arxiv:1207.3414[cs.SI] (2012).