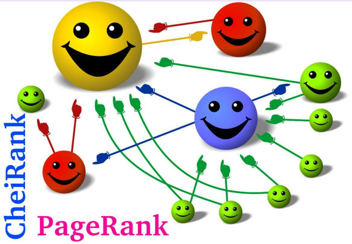
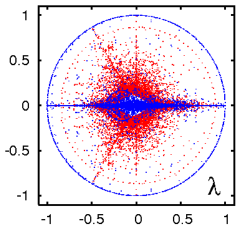
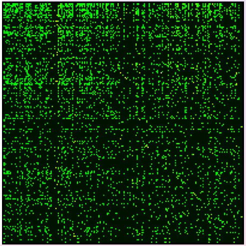
New tools and algorithms for directed network analysis

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

FET Open NADINE project No 288956 (4 partners) =>
http://www.quantware.ups-tlse.fr/FETNADINE/







1945: Nuclear physics → Wigner (1955)→ Random Matrix Theory
1991: WWW, small world social networks → Markov chains (1906) → Google matrix

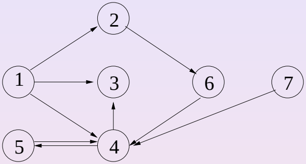
Despite the importance of large-scale search engines on the web, very little academic research has been done on them.
S.Brin and L.Page, Comp. Networks ISDN Systems **30**, 107 (1998)

Workpackages and milestones

WP1: CheiRank versus PageRank, RTD centrality measures and network structure (UTWE)
WP2: Network analysis through Google matrix eigenspectrum and eigenstates (CNRS)
WP3: Applications to voting systems in social networks (UMIL)
WP4: Applications of new tools and algorithms to real-world network structures (MTA_SZTAKI)
WP5: Database development of real-world networks (UMIL)
DM1: Correlation properties of directed networks (WP1.1)
DM2: Statistical characterization of 2DRanking (WP1.2, WP2.1, WP4.3)
DM3: Eigenstate community detection (WP2.2, WP3.1)
DM4: Spam filter protocols (WP4.2)
DM5: Network-specific centrality measures (WP1.1, WP1.3, WP3.1, WP3.2)
M6: Fractal Weyl law properties of networks; M7: Protocols for large-scale network processing; M8: Characterization of multi-product world trade network; M9: Webcrawler development and database collection; M10: Monte Carlo algorithms for centrality measures; M11: Delocalization conditions for Google matrix eigenstates; M12: New protocols for social voting and recommendation; M13: Characterization of ranking of Wikipedia and other networks; M14: Characterization of time-evolving Web structures

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

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

To remove convergence problems:

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

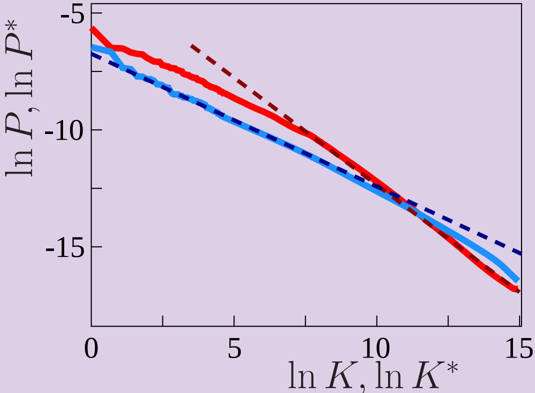
$$\mathbf{S} = \begin{pmatrix} 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 1 & 1 & 1 \\ 0 & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{7} & 0 &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

- To remove degeneracies of $\lambda = 1$, replace $\mathbf{S}$ 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 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) ...

Wikipedia ranking of human knowledge

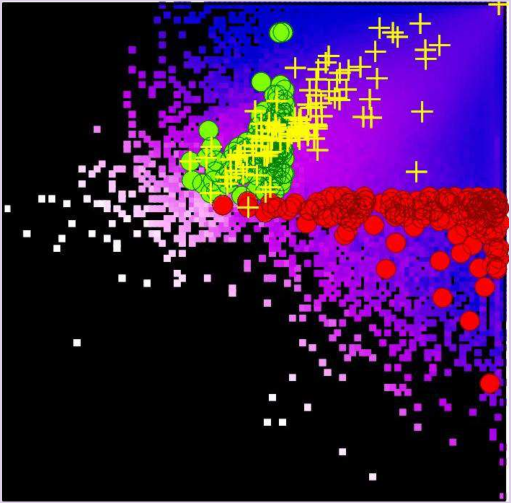
Wikipedia English articles $N = 3282257$ dated Aug 18, 2009



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)

Wikipedia ranking of universities, personalities

Universities:

PageRank: 1. Harvard, 2. Oxford, 3. Cambridge, 4. Columbia, 5. Yale, 6. MIT, 7. Stanford, 8. Berkeley, 9. Princeton, 10. Cornell.

2DRank: 1. Columbia, 2. U. of Florida, 3. Florida State U., 4. Berkeley, 5. Northwestern U., 6. Brown, 7. U. Southern CA, 8. Carnegie Mellon, 9. MIT, 10. U. Michigan.

CheiRank: 1. Columbia, 2. U. of Florida, 3. Florida State U., 4. Brooklyn College, 5. Amherst College, 6. U. of Western Ontario, 7. U. Sheffield, 8. Berkeley, 9. Northwestern U., 10. Northeastern U.

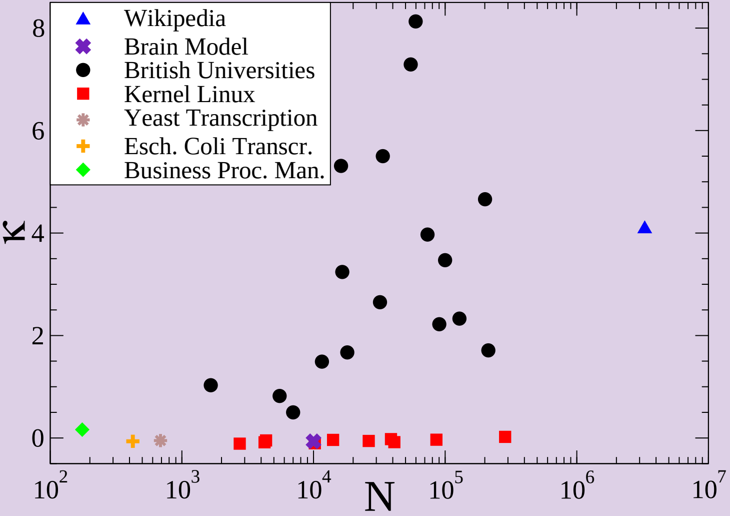
Personalities:

PageRank: 1. Napoleon I of France, 2. George W. Bush, 3. Elizabeth II of the United Kingdom, 4. William Shakespeare, 5. Carl Linnaeus, 6. Adolf Hitler, 7. Aristotle, 8. Bill Clinton, 9. Franklin D. Roosevelt, 10. Ronald Reagan.

2DRank: 1. Michael Jackson, 2. Frank Lloyd Wright, 3. David Bowie, 4. Hillary Rodham Clinton, 5. Charles Darwin, 6. Stephen King, 7. Richard Nixon, 8. Isaac Asimov, 9. Frank Sinatra, 10. Elvis Presley.

CheiRank: 1. Kasey S. Pipes, 2. Roger Calmel, 3. Yury G. Chernavsky, 4. Josh Billings (pitcher), 5. George Lyell, 6. Landon Donovan, 7. Marilyn C. Solvay, 8. Matt Kelley, 9. Johann Georg Hagen, 10. Chikage Oogi.

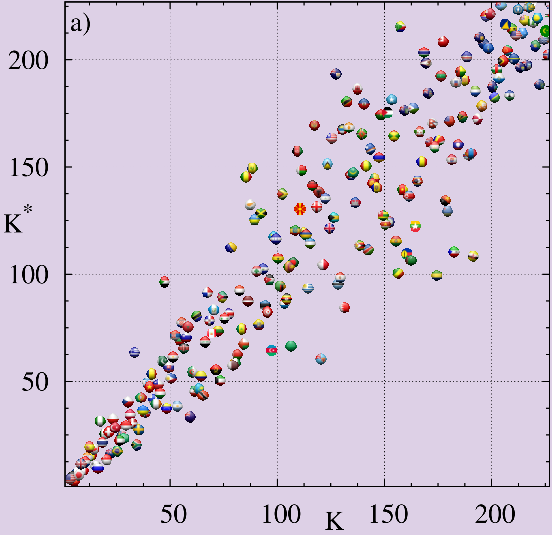
Correlator of PageRank and CheiRank



$\kappa = N \sum_i P(K(i))P^*(K^*(i)) - 1$; $\kappa = -0.278$(Phys. Rev.); 112.6 (Twitter)

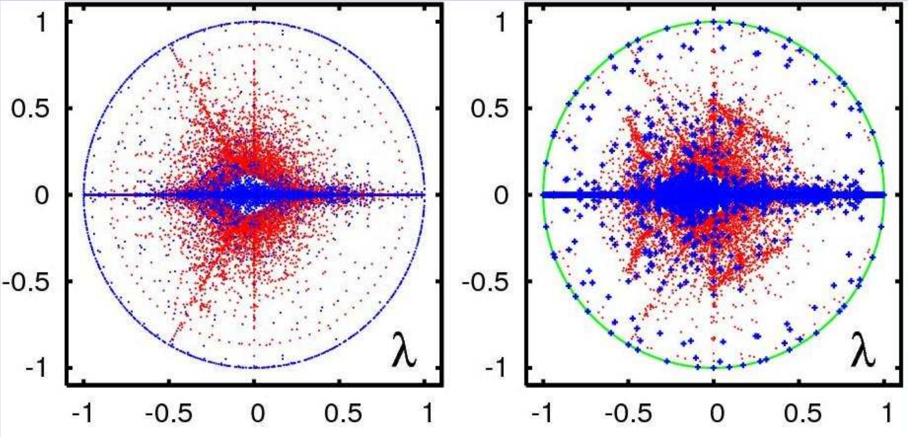
Ranking of World Trade

UN COMTRADE database 2008: All commodities



Ermann, DS arxiv:1103.5027 (2011)

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)]

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 $0 \leq \gamma \leq \gamma_b$, 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

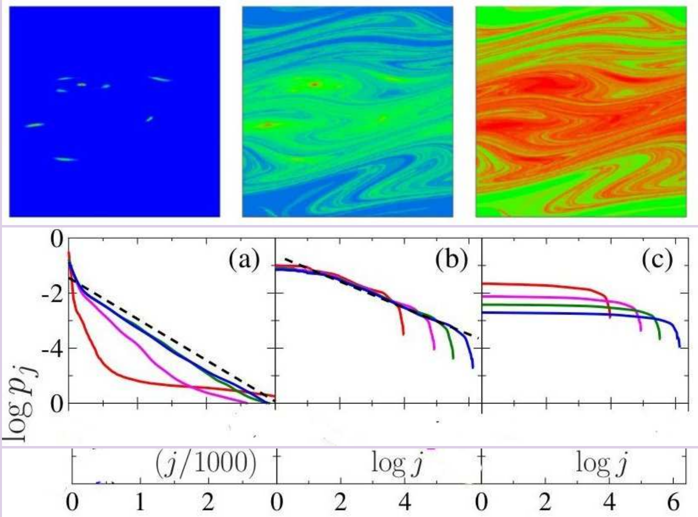
References:
J.Sjostrand, Duke Math. J. **60**, 1 (1990)
M.Zworski, Not. Am. Math. Soc. **46**, 319 (1999)
W.T.Lu, S.Sridhar and M.Zworski, Phys. Rev. Lett. **91**, 154101 (2003)
S.Nonnenmacher and M.Zworski, Commun. Math. Phys. **269**, 311 (2007)

Resonances in quantum chaotic scattering:
three disks, quantum maps with absorption

Perron-Frobenius operators, Ulam method for dynamical maps, Ulam networks, dynamical maps, strange attractors

Linux kernel network $d = 1.3$, $N \leq 285509$

Anderson delocalization of PageRank ?



Ulam network of dynamical map $\alpha = 1; 0.95; 0.85$

Quantware group + partners

Toulouse

Permanent:

Klaus Frahm (PR Univ. P. Sabatier)
Bertrand Georgeot (DR CNRS)

Post-doc: Young-Ho Eom (NADINE)

PhD: Vivek Kandiah (CNRS - Region Midi-Pyrenees)

Expernal collaborators:

Alexei Chepelianskii (U Cambridge)
Leonardo Ermann (CNEA Buenos Aires)
Oleg Zhirov (BINP Russian Academy Novosibirsk)
NOMAO.COM: Estelle Delpech, Samuel Phan

Twente => Nelly Litvak (mathematics)

Budapest => Andras Benczur (computer science)

Milano => Sebastiano Vigna (computer science)

Results: 32 publications, 37 conference presentations, ...