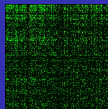
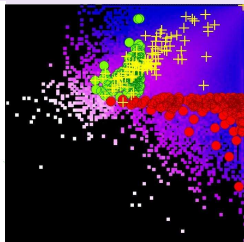
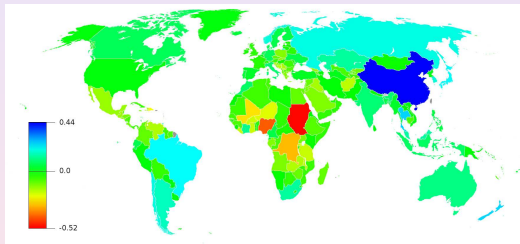


Google matrix analysis of Markov chains

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



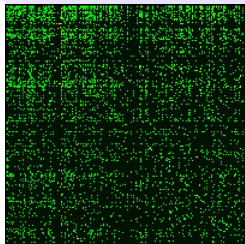
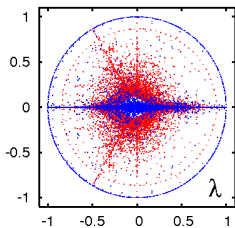
with L.Ermann (CNEA TANDAR), A.Chepelianskii (Orsay), K.Frahm (LPT),
J.Lages (U Besancon)+MASTODONS partners; thanks to UN COMTRADE



- * Markov (1906) → Brin and Page (1998)
- * Google matrix of directed networks (brief introduction)
- * Applications: Wikipedia, world trade network (UN COMTRADE + WTO Geneve) ...

Support: MASTODONS CNRS project APLIGOOGL; Refs =>
www.quantware.ups-tlse.fr/FETNADINE/ + Rev. Mod. Phys. 87, 1261 (2015)

(1906) Markov vs Wigner (1955)



1945: Nuclear physics → Wigner (1955) → Random Matrix Theory

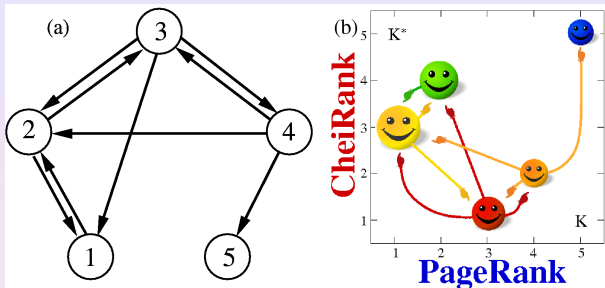
1991: WWW, small world social networks → Markov (1906) → Google matrix

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)

Google matrix construction rules

Markov chains (1906) and Directed networks



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 construction rules

Google Matrix and Computation of PageRank

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

To remove convergence problems:

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

$$\mathbf{S} = \begin{pmatrix} 0 & 1/2 & 1/3 & 0 & 1/5 \\ 1 & 0 & 1/3 & 1/3 & 1/5 \\ 0 & 1/2 & 0 & 1/3 & 1/5 \\ 0 & 0 & 1/3 & 0 & 1/5 \\ 0 & 0 & 0 & 1/3 & 1/5 \end{pmatrix} \quad \mathbf{S}^* = \begin{pmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/3 & 0 & 1 & 0 \\ 0 & 1/3 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

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

$$\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \frac{\mathbf{E}}{N}; \quad \mathbf{G}\mathbf{P} = \lambda \mathbf{P} \Rightarrow \text{Perron-Frobenius operator}$$

- α models a random surfer with a random jump after approximately 6 clicks (usually $\alpha = 0.85$); **PageRank vector** $\Rightarrow \mathbf{P}$ at $\lambda = 1$ ($\sum_j P_j = 1$).
- **CheiRank vector \mathbf{P}^*** : $\mathbf{G}^* = \alpha \mathbf{S}^* + (1 - \alpha) \frac{\mathbf{E}}{N}$, $\mathbf{G}^* \mathbf{P}^* = \mathbf{P}^*$
(\mathbf{S}^* with inverted link directions)
Fogaras (2003) ... Chepelianskii arXiv:1003.5455 (2010) ...

Real directed networks => small world

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 $\rho(k) \sim k^{-\nu}$

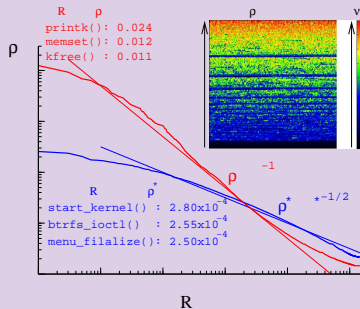
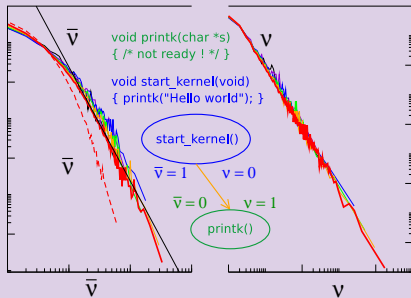
PageRank vector for large WWW:

- $P(K) \sim 1/K^\beta$, 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)

Linux Kernel Network

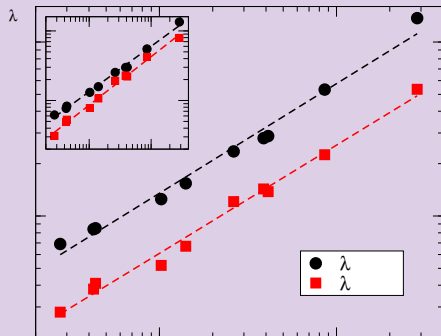
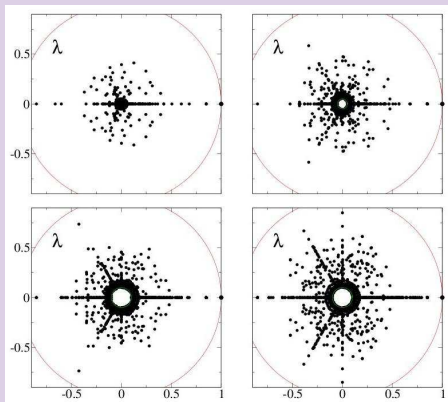
Procedure call network for Linux



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)

Fractal Weyl law for Linux Network

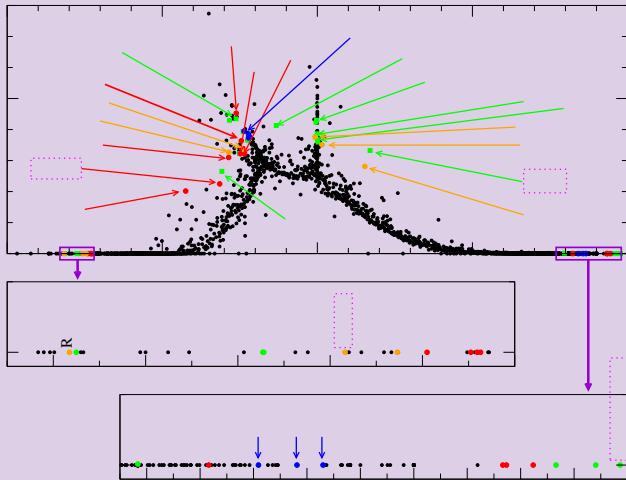


(left) Spectrum of Google matrix;

(right) Number of states N_λ with $|\lambda| > 0.1; 0.25$ vs. N , lines show $N_\lambda \sim N^\nu$ with $\nu \approx 0.65$; various Linux Kernel versions

fractal dimension of Linux network: $d \approx 1.3$
(Ermann, Chepelianskii, DS EPJB (2011))

Wikipedia spectrum and eigenstates



Spectrum S of EN Wikipedia, Aug 2009, $N = 3282257$. Eigenvalues-communities are labeled by most repeated words following word counting of first 1000 nodes.

(Ermann, Frahm, DS 2013)

Reduced Google matrix

A selected network of interest with $N_r < N$ nodes called reduced network.
Block structure of G matrix:

$$G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix}$$

with s index for scattering network $N_s = N - N_r$.
Reduced G_R matrix

$$G_R P_r = P_r \quad , \quad G_R = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1} G_{sr}$$

Useful expansion

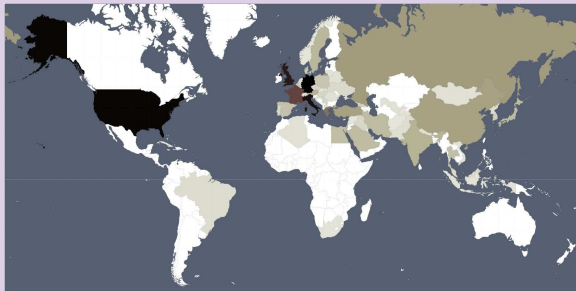
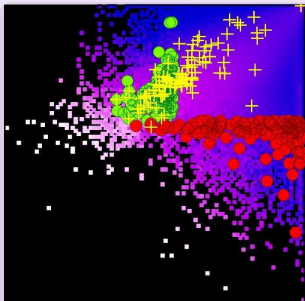
$$(\mathbf{1} - G_{ss})^{-1} = P_c \frac{1}{1 - \lambda_c} + Q_c \sum_{l=0}^{\infty} \bar{G}_{ss}^l$$

with projector $P_c = \psi_R \psi_L^T$ on eigenstate of maximal eigenvalue λ_c of G_{ss} , the complementary projector $Q_c = \mathbf{1} - P_c$ and $\bar{G}_{ss} = Q_c G_{ss} Q_c$.

K.Frahm, DS arxiv:1602.02394 + ...

Top historical figures of 24 Wikipedia editions

2DRanking of Wikipedia articles; top 100 historical figures;
comparison with historical studies of M.Hart (37 and 43 percent overlap)
35 centuries and all countries by birth place; 17 millions wiki-articles

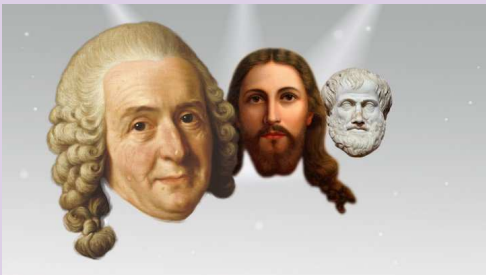


A.Zhirov, O.Zhirov, DLS EPJB (2010); Y.-H.Eom, P.Aragon, D.Laniado,
A.Kaltenbrunner, S.Vigna, DLS arXiv2014 - PLoS ONE (2015)

Russia: Peter the Great, Alexander Pushkin, Joseph Stalin

Top historical figures of 24 Wikipedia editions

Top global PageRank historical figures: Carl Linnaeus, Jesus, Aristotle ...



theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Jobs

Comment is free

And the winner of Wikipedia's influence list is ... an 18th century botanist. Hear hear

Carl Linnaeus is hardly a household name, but the Swedish doctor who created a global naming system for species deserves this accolade



Patrick Barkham

theguardian.com, Friday 13 June 2014 09:00 BST

Jump to comments (51)

Media highlights: The Guardian, The Independent, The Washington Post, France24, EC CORDIS

==>Uppsala Universitet: "Carl Linnaeus ranked most influential person of all time" ... (about 20 countries)

Competitors: MIT Pantheon project <http://pantheon.media.mit.edu> (2014);
Stony-Brook NY <http://www.whoisbigger.com/> (2014)

Multiproduct world trade networks

UN COMTRADE data sets: up to 227 countries, 61 products (up to about 5000 products), about 50 years (no transections between products)

==> talk of **Leonardo Ermann**

World network of economic activities: countries $N_c = 57 + 1$, activity sectors $N_s = 37$ from OECD-WTO; years 1995, 2000, 2005, 2008, 2009 (with **V.Kandiah and H.Escaith (WTO Geneve)**)

Google matrix approach => Democratic (equal) treatment of all countries (rich and poor) in agreement with the UN principles

Treatment of products according to their weight in trade (introduction of preferential vector in G)

Charcateristics: country trade balance in a given year

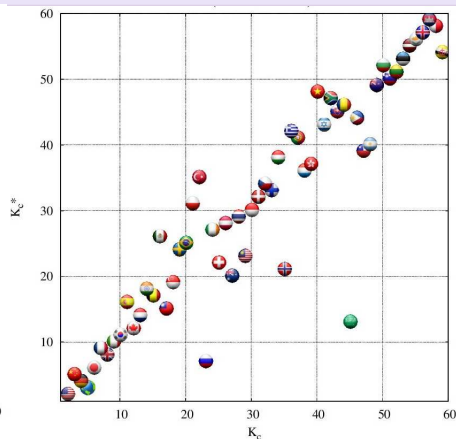
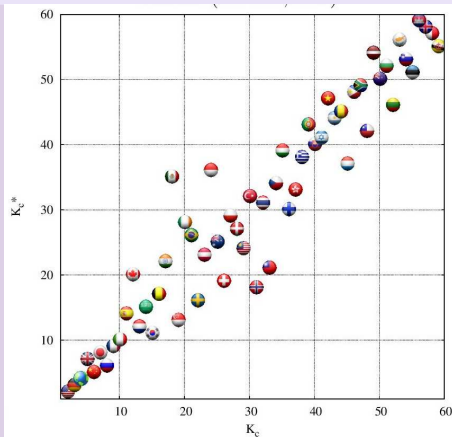
$B_c = (P_c^* - P_c) / ((P_c^* + P_c))$; $dB_c/d\delta_p$ sencitivity to product (e.g. petroleum price); $dB_c/d\sigma_{c'}$ sencitivity to labor cost of a given country

Import is like PageRank; Export is like CheiRank

Ranking plane of WNEA of OECD-WTO (2008)

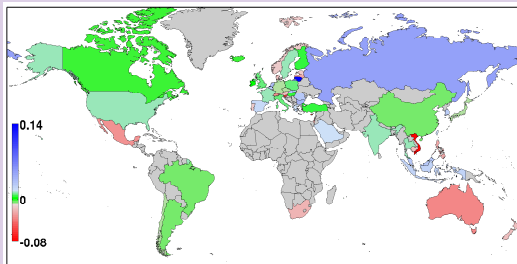
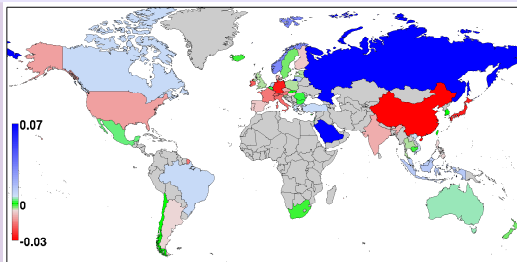
Left: PageRank-CheiRank plane of countries (K_C, K_C^*)

Right: Import-Export plane of countries



WNEA sensitivity to petroleum price (2008)

Top: PageRank-CheiRank; Bottom: Import-Export ($dB_C/d\delta_7$)



WBW: Towards bank financial network control

World Bank Web:

EU challenge (only 6000 or 2000 nodes for all FED USA or Germany!)

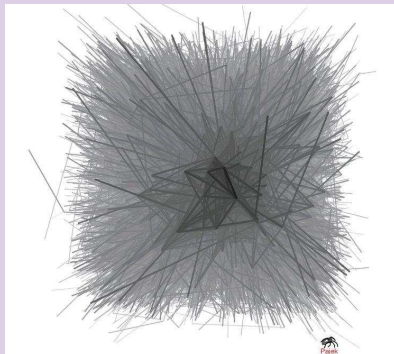
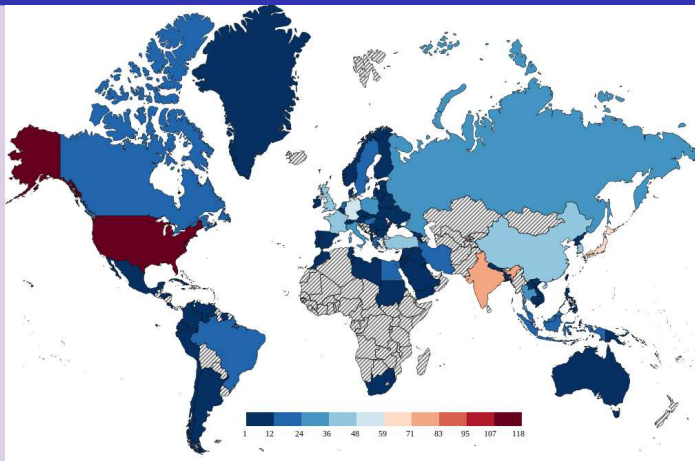


Fig. 1. Fedwire interbank payment network. First day of Sample. 6600 nodes and over 70,000 undirected links [39].

K.Soramäki *et al.*, *The topology of interbank payment flows*, *Physica A* **379**, 317 (2007); R.Garratt *et al.* WP 2008-42, Bank of Canada, WP 413 Bank of England (2011); B.Craig, G. von Peter N 12/2010 Deutsche Bundesbank (2010)

Wikipedia Ranking of World Universities (24 wiki)



Top universities of 24wiki (total 1025; top 100 overlap 0.62 WRWU/Shanghai) ;
J.Lages, A.Patt, DLS EPJB (2016)

Top PageRank: U Cambridge, U Oxford, Harvard U

Top 10 Russia: MoscowGU, SPetersburgGU, KazanU, Bauman, MFTI,
HighSchoolEconm, SPeterPolytech, TomskGU, MIPHI, NskGU