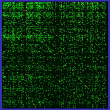
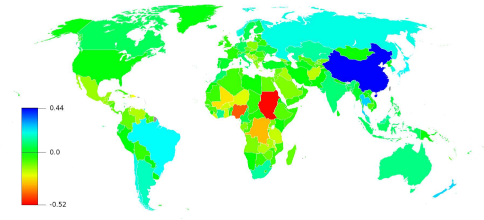
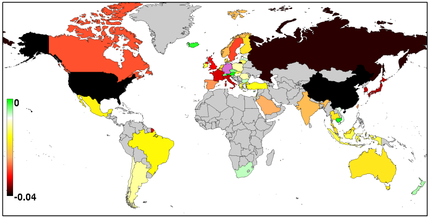
Introduction to Google matrix and world network of economic activities



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

with L.Ermann (CNEA TANDAR), K.Frahm (LPT), V.Kandiah (LPT), H.Escaith (WTO Geneve); thanks to UN COMTRADE

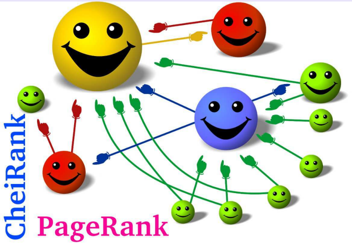


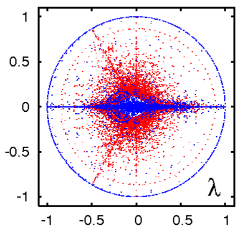


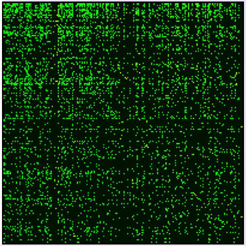
* Markov (1906) $\rightarrow$ Brin and Page (1998)
* Google matrix of directed networks (brief introduction)
* Applications: multiproduct world trade network (UN COMTRADE + OECD-WTO)

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

(1906) Markov vs Wigner (1955)







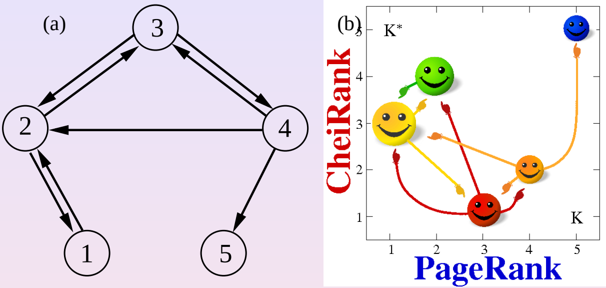
1945: Nuclear physics → Wigner (1955)→ Random Matrix Theory
1991: WWW, small world social networks → Markov (1906) → Google matrix

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 **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

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

To remove convergence problems:

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

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

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

  $\mathbf{G} = \alpha \mathbf{S} + (1-\alpha)\frac{\mathbf{E}}{N}$; $\quad GP = \lambda P \quad$ => 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)

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

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

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

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)

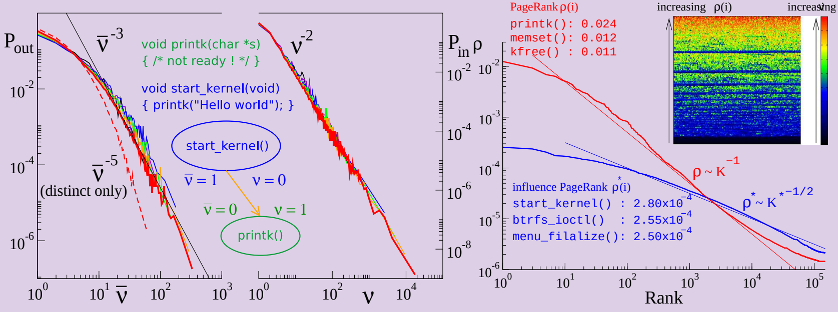
Resonances in quantum chaotic scattering:
three disks, quantum maps with absorption

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

Linux kernel network $d = 1.3$, $N \leq 285509$;
Phys. Rev. up to 2009 $d \approx 1$, $N = 460422$

Linux Kernel Network

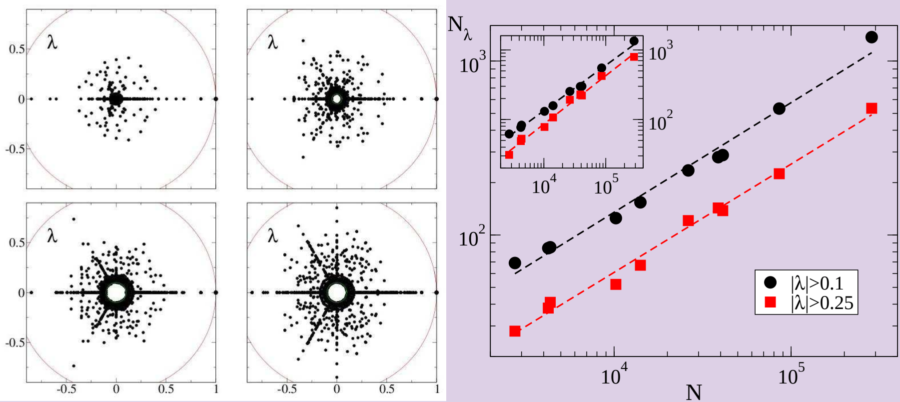
Procedure call network for Linux



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))

Anderson transition on directed networks

Anderson (1958) metal-insulator transition for electron transport in disordered solids

$H = \epsilon_n \psi_n + V(\psi_{na+1} + \psi_{n-1}) = E\psi_n \; ; \quad -W/2 < \epsilon_n < W/2$

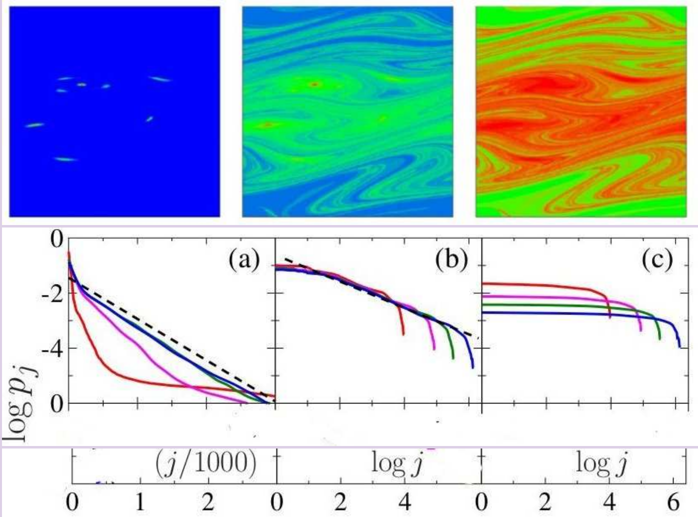
In dimensions $d = 1, 2$ all eigenstates are exponentially localized, insulating phase. At $d = 3$ for $W > 16.5V$ all eigenstates are exponentially localized, for $W < 16.5V$ there are metalic delocalized states, mobility edge, metalic phase

Random Matrix Theory - RMT (Wigner (1955)) for Hermitian and unitary matrices (quantum chaos, many-body quantum systems, quantum computers)

Google matrix, Markov chains, Perron-Frobenium operators:
=> complex spectrum of eigenvalues; new field of research
Can we have the Anderson transition for Google matrix? All the world would go blind if PageRank is delocalized What are good RMT models of Google matrix? Subspaces and core

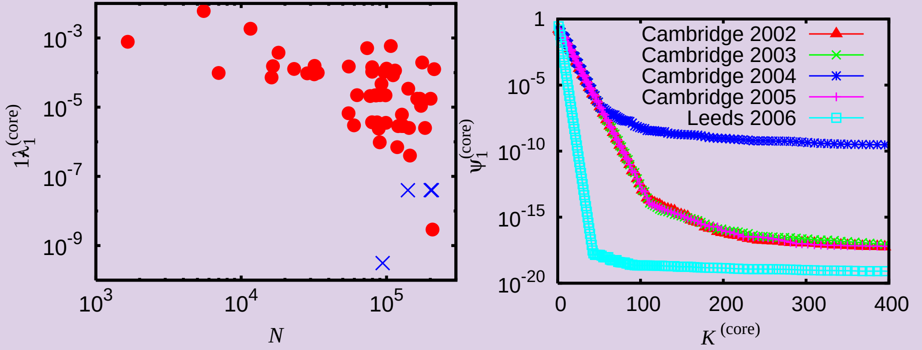
$$\mathbf{S} = \begin{pmatrix} \mathbf{S_{ss}} & \mathbf{S_{sc}} \\ 0 & \mathbf{S_{cc}} \end{pmatrix}$$

Anderson delocalization of PageRank ?



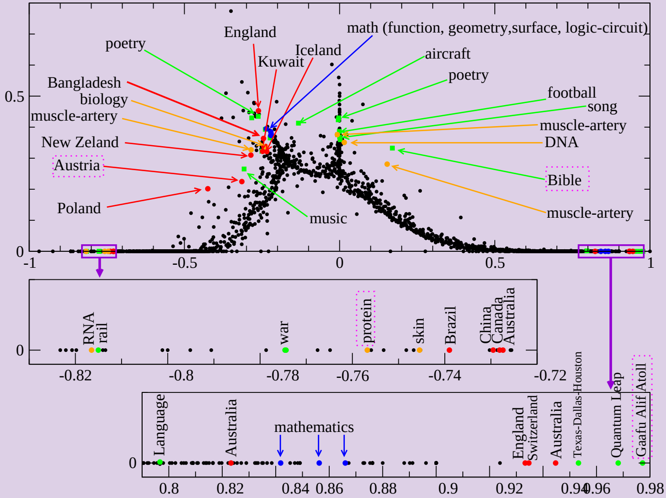
Ulam network of dynamical map $\alpha = 1; 0.95; 0.85$

Gap of core space at $\alpha = 1$



(Left) Gap vs N for universities Glasgow, Cambridge, Oxford, Edinburgh, UCL, Manchester, Leeds, Bristol and Birkbeck (2002-2006) and Bath,Hull,Keele,Kent,Nottingham, Aberdeen, Sussex, Birmingham, East Anglia, Cardiff, York (2006). Red dots are for gap $> 10^{-9}$ and blue crosses (moved up by 10^9) are for Cambridge 2002, 2003 and 2005 and Leeds 2006 with gap $< 10^{-16}$; point at $2.91 \cdot 10^{-9}$ is Cambridge 2004. (Right) First core eigenstates

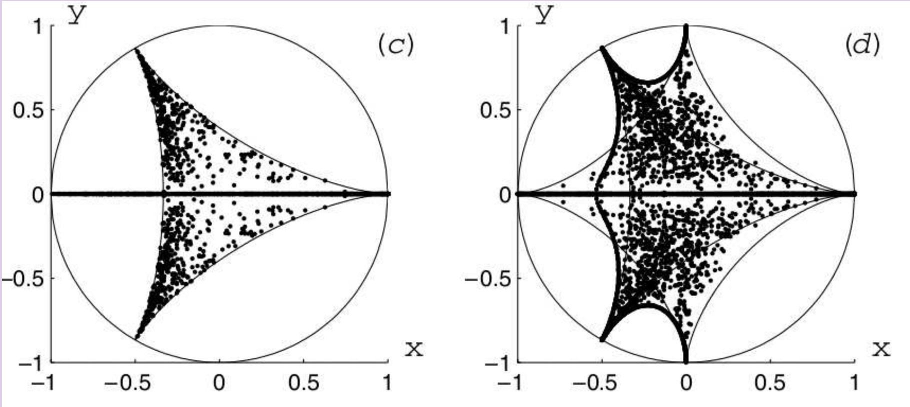
Wikipedia spectrum and eigenstates



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)

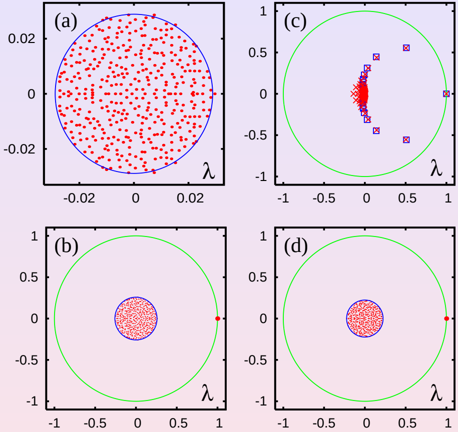
Spectrum of random orthostochastic matrices



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

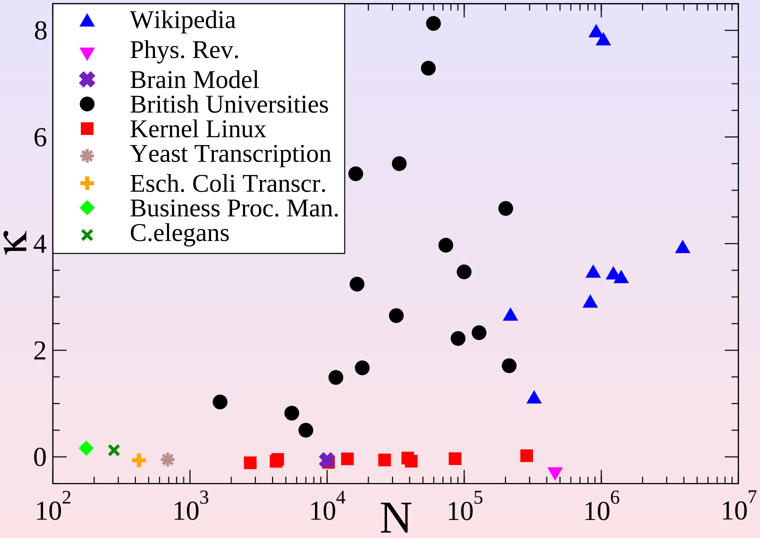
Random Matrix Models of directed networks

random matrix elements of G with sum equat unity in each column ($N = 400$)



(a) N positive random elements with unit sum in each column;
(c) triangular matrix with random elements;
(b),(d) $Q = 20$ nonzero elements in each column
- blue circle is theory with radius $\sim 1/\sqrt{N}, 1/\sqrt{Q}$

Correlator of PageRank and CheiRank



$$\kappa = N \sum P(K_i) P^*(K_i) - 1$$

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

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

Useful expansion

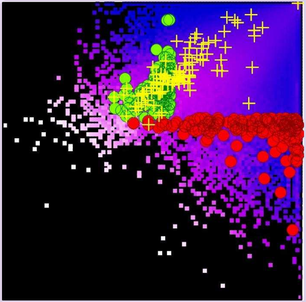
$$(\mathbf{1} - G_{ss})^{-1} \quad = \quad \mathcal{P}_c \frac{1}{1 - \lambda_c} + \mathcal{Q}_c \sum_{l=0}^{\infty} \bar{G}_{ss}^{l}$$

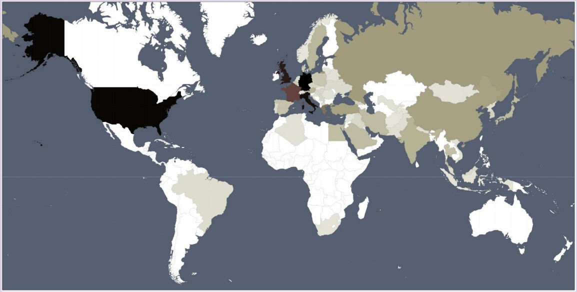
with projector $\mathcal{P}_c = \psi_R \psi_L^T$ on eigenstate of maximal eigenvalue λ_c of G_{ss}, the complementary projector $\mathcal{Q}_c = \mathbf{1} - \mathcal{P}_c$ and $\bar{G}_{ss} = \mathcal{Q}_c G_{ss} \mathcal{Q}_c$.

K.Frahm, DS arxiv:1602.02394 and WTN examples => L.Ermann talk

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 centures and all countries by birth place; 17 millions wiki-articles

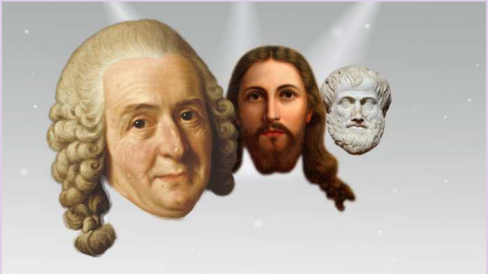




A.Zhirov, O.Zhirov, DLS EPJB (2010); Y.-H.Eom, K.M.Frahm, A.Benczur, DLS EPJB (2013); Y.-H.Eom, DLS PLoS ONE (2013), Y.-H.Eom,P.Aragon, D.Laniado, A.Kaltenbrunner, S.Vigna, DLS arXiv2014 - PLoS ONE (2015)

Top historical figures of 24 Wikipedia editions

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



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

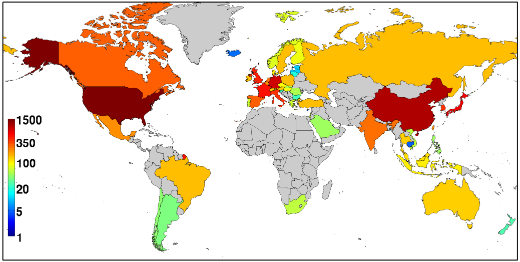
Treatment of products according to their weight in trade (introduction of preferential vector in G)

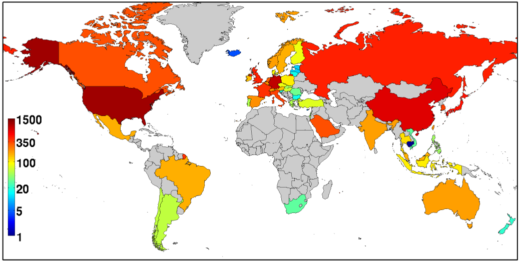
Charcateristics: country trade balance in a given year $B_c = (P^*_c - P_c)/((P^*_c + P_c)$; $dB_c/d\delta_p$ sencitivity to product (e.g. petroleum price); $dB_c/d\sigma_{c'}$ sensitivity to labor cost of a given country

Import is like PageRank; Export is like CheiRank

Import-Export of WNEA of OECD-WTO (2008)

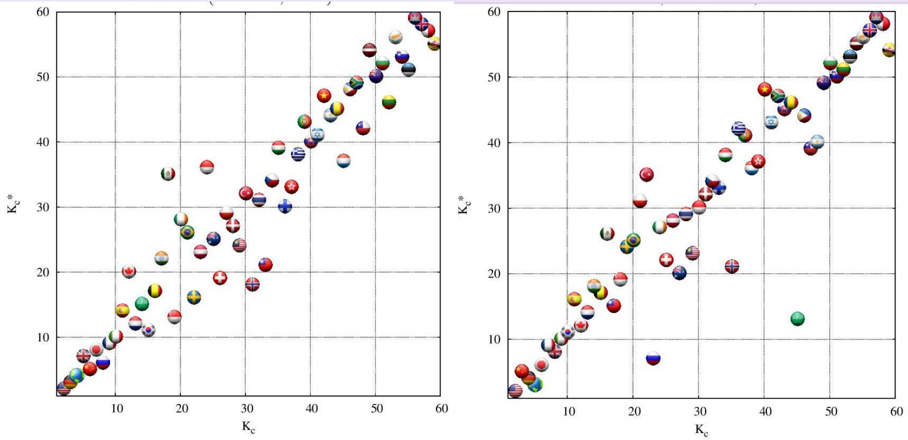
Top: country import in billions of USD,
bottom: export; gray color marks ROW (rest of the world)





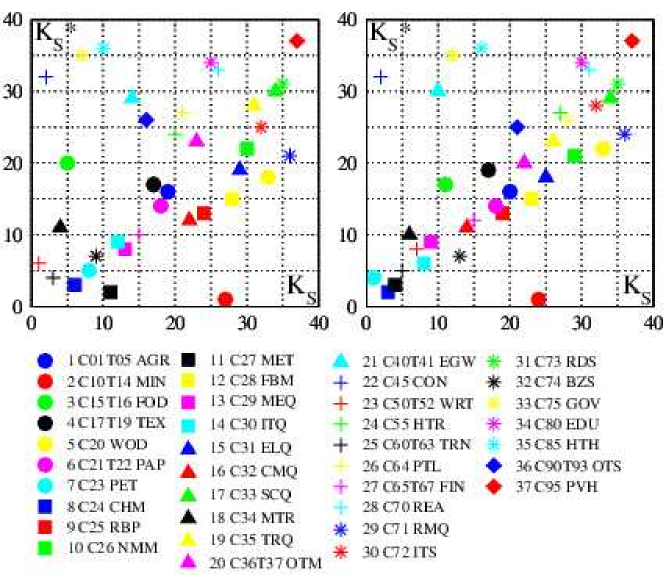
Ranking plane of WNEA of OECD-WTO (2008)

Left: PageRank-CheiRank plane of countries (K_c, K_c^*)
Right: Import-Export plane of countries



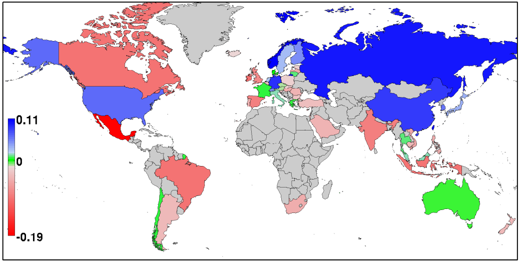
Ranking plane of WNEA sectors (2008)

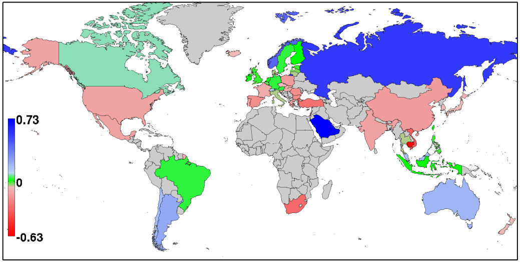
Left: PageRank-CheiRank plane of sectors (K_s, K_s^*); Right: Import-Export



WNEA world map of trade balance (2008)

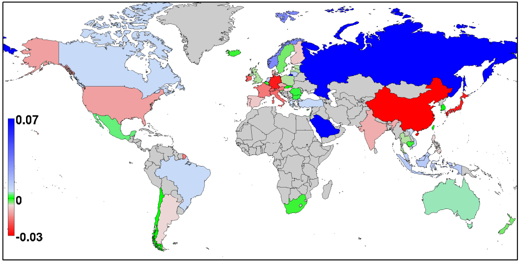
Top: PageRank-CheiRank balance; Bottom: Import-Export balance

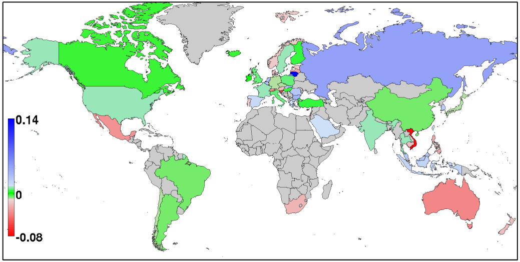




WNEA sensitivity to petroleum price (2008)

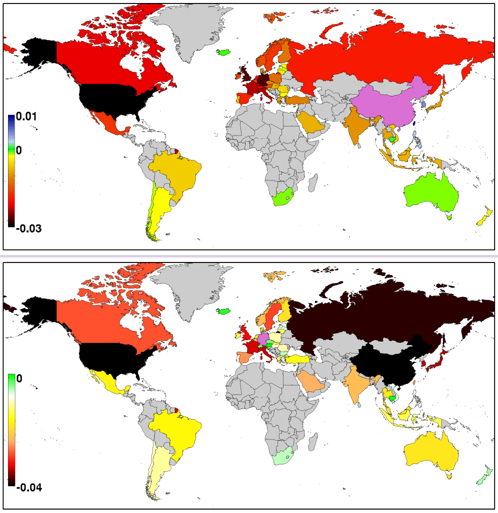
Top: PageRank-CheiRank; Bottom: Import-Export ($dB_c/d\delta_7$)



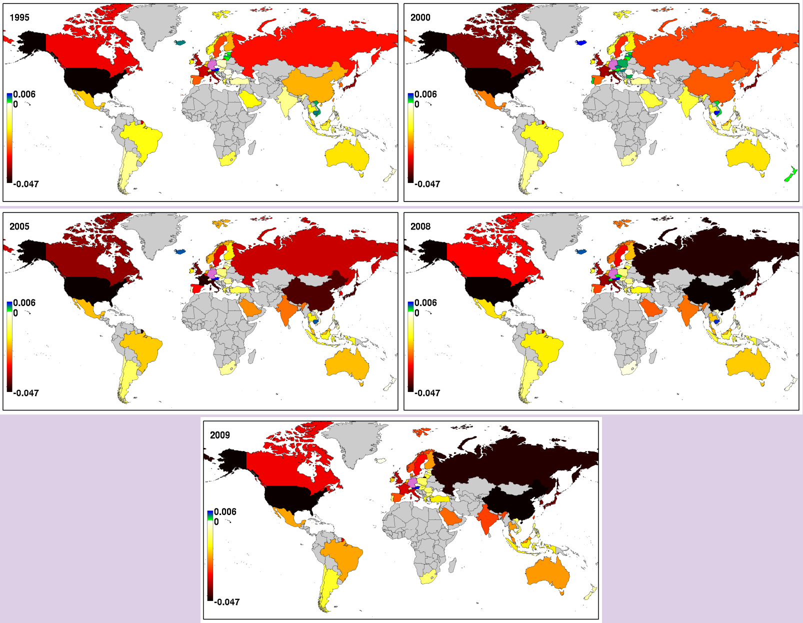


WNEA of OECD-WTO (2008)

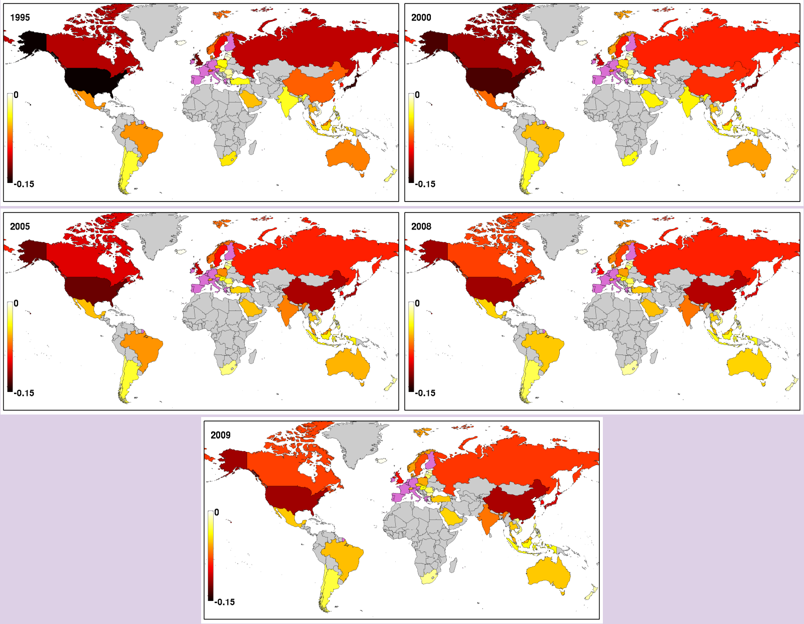
World network of economic activities: countries $N_c = 58$, activity sectors $N_s = 37$ (with V.Kandiah and H.Escaith (WTO Geneve)); $dB_c/d\sigma_{c'}$ sensitivity to labor cost of China (top), Germany (bottom)



WNEA-WTO-OECD of DEU labor cost



WNEA EUROZONE labor cost



WBW: Towards bank financial network control

World Bank Web:
EU challenge (only 6000 or 2000 nodes for all FED USA or Germany!)

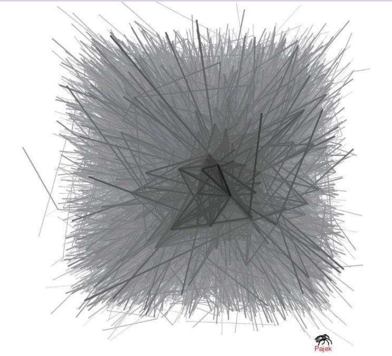


Fig. 1. Fedwire interbank payment network. First day of Sample. 6600 nodes and over 70,000 undirected links [39].

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)