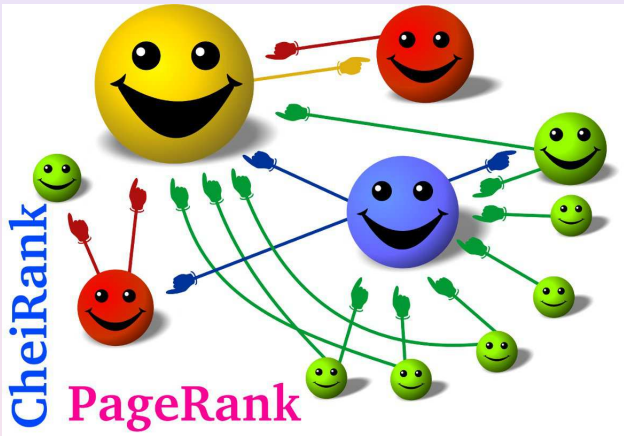
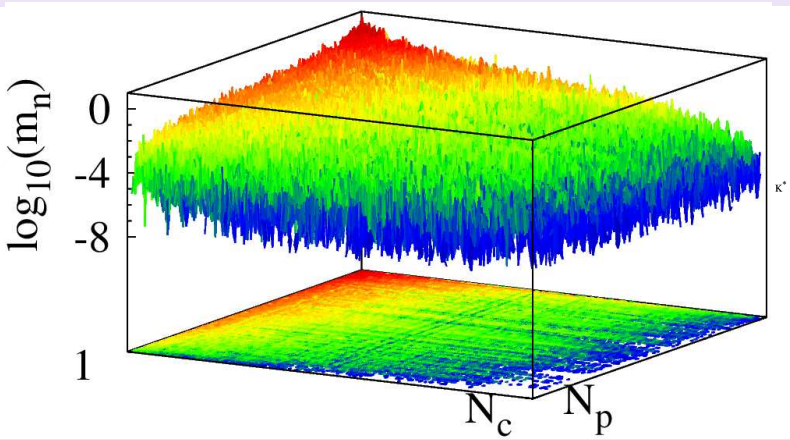
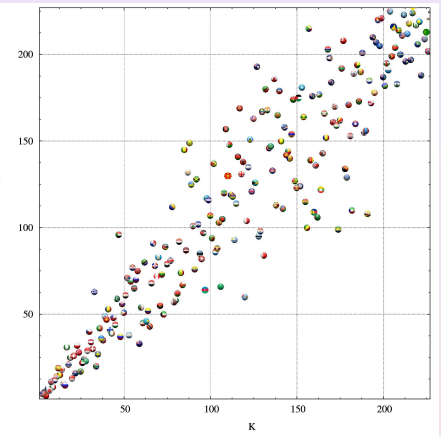
Google matrix analysis of world trade and financial networks

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

EC FET Open NADINE project





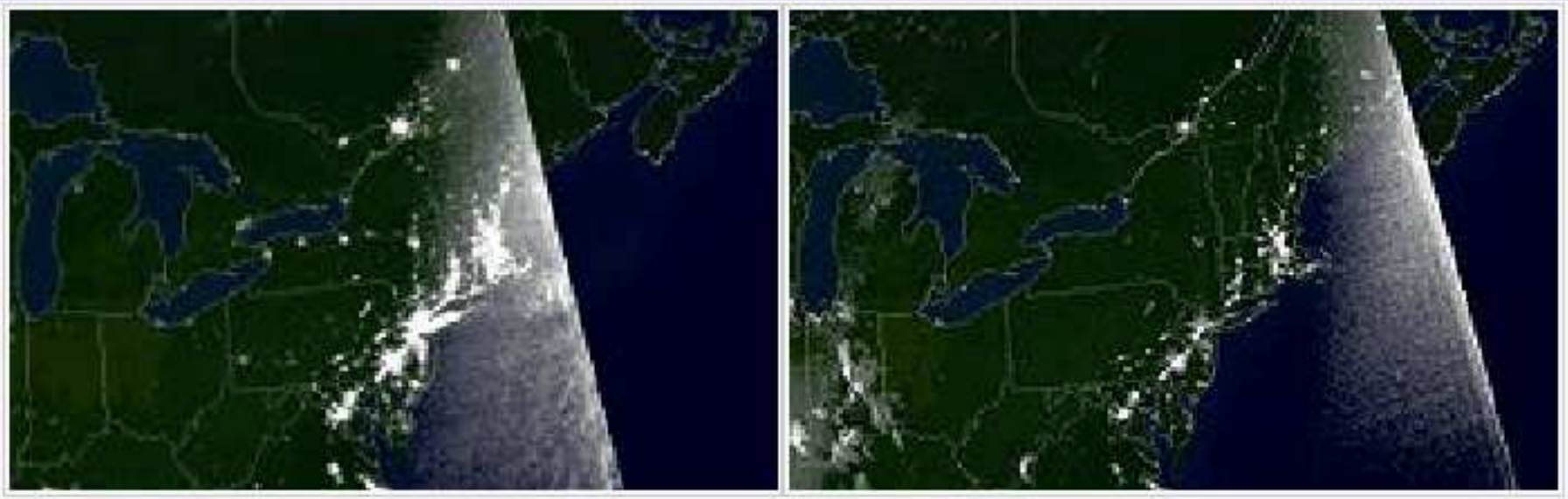


* Quantware group: classical/quantum chaos, dynamical systems, large matrices
* How Google search works, PageRank, CheiRank
* Examples of directed networks: Wikipedia, University networks, DvvaDi search; Ulam networks, Linux Kernel network, fractal Weyl law
* World trade from UN COMTRADE 1962 - 2009: arxiv:1103.5027
  => democratic treatment of all UN countries; ecology analysis arxiv:1201.3584
* Towards ranking of bank financial flows: WWW ==> WBW

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

Monitoring of grids and networks

Any large network requires monitoring ...



NOAA satellite imagery one day before and the night of the blackout.

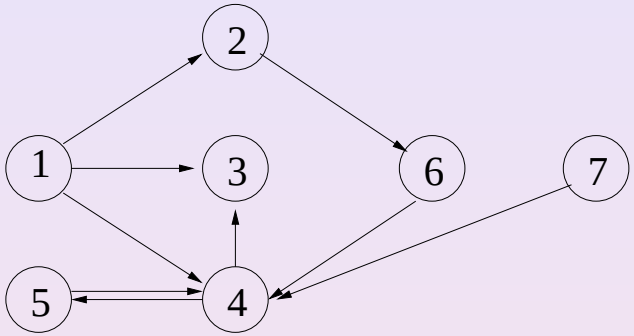
Example of Northeast blackout of electical power grid, Aug 14, 2003.
Wikipedia article “Northeast blackout of 2003”

==> Analysis of network flows:
==> World Wide Web with $\sim 10^{11}$ *sites*
==> project launched at CENR by Tim Berners-Lee, 1991

==> World Bank Web exists (SWIFT ...)

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 $1/N$.

How Google works

Google Matrix and Computation of PageRank

$\mathbf{P} = \mathbf{SP} \Rightarrow$ **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**

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

Real directed networks

Real networks are characterized by:

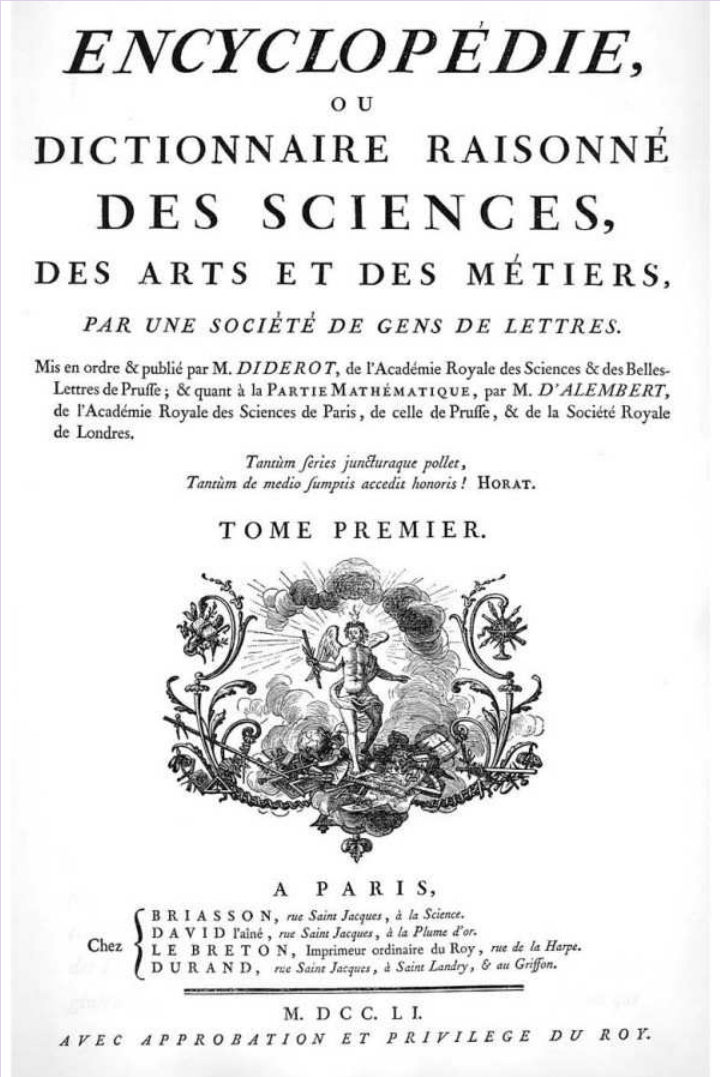
- **small world property**: average distance between 2 nodes $\sim \log N$
- **scale-free property**: distribution of the number of ingoing or outgoing links $\rho(k) \sim k^{-\nu}$

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)

From Encyclopédie (1751) to Wikipedia (2009)



ENCYCLOPÉDIE,

OU

DICTIONNAIRE RAISONNÉ

DES SCIENCES,

DES ARTS ET DES MÉTIERS,

PAR UNE SOCIÉTÉ DE GENS DE LETTRES.

Mis en ordre & publié par M. *DIDEROT*, de l'Académie Royale des Sciences & des Belles-Lettres de Prusse; & quant à la PARTIE MATHÉMATIQUE, par M. *D'ALEMBERT*, de l'Académie Royale des Sciences de Paris, de celle de Prusse, & de la Société Royale de Londres.

Tantùm series juncturaque pollet,
Tantùm de medio sumptis accedit honoris! HORAT.

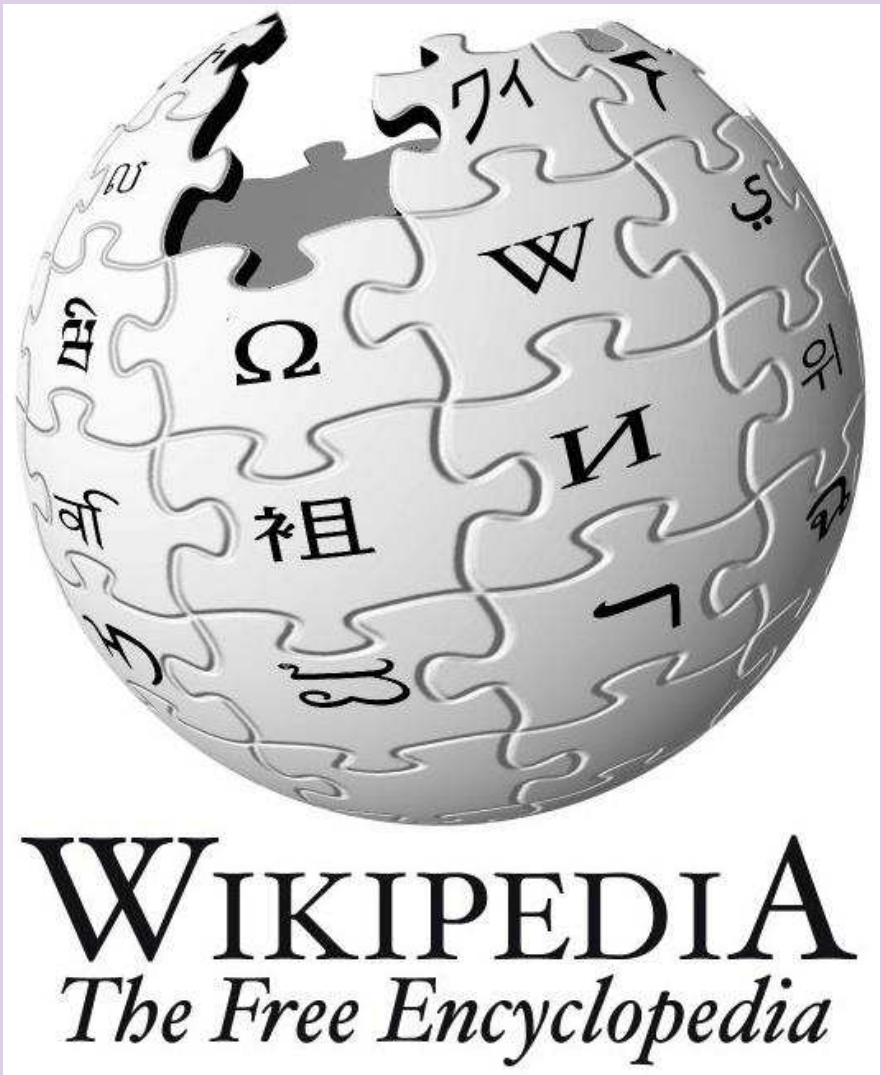
TOME PREMIER.

A PARIS,

Chez { BRIASSON, *rue Saint Jacques, à la Science.*
DAVID l'aîné, *rue Saint Jacques, à la Plume d'or.*
LE BRETON, Imprimeur ordinaire du Roy, *rue de la Harpe.*
DURAND, *rue Saint Jacques, à Saint Landry, & au Griffon.*

M. DCC. LI.

AVEC APPROBATION ET PRIVILEGE DU ROY.

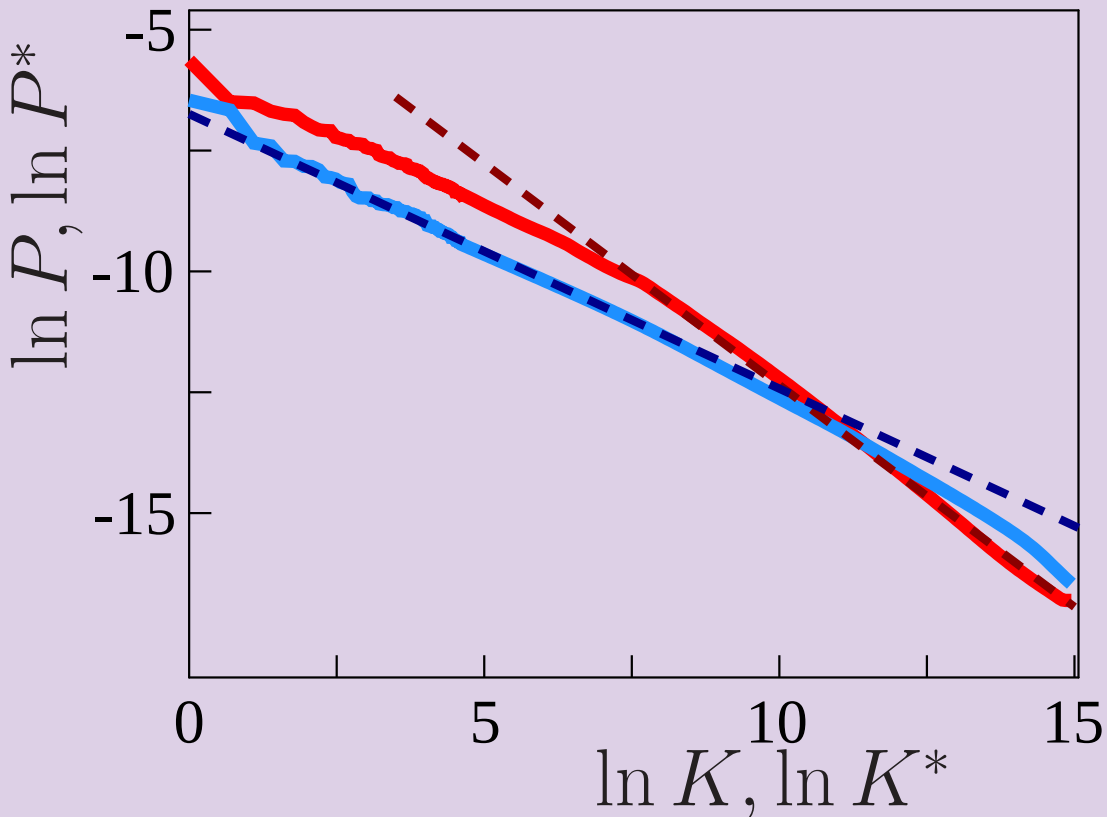


"The library exists *ab aeterno.*"

Jorge Luis Borges *The Library of Babel, Ficciones*

Wikipedia ranking of human knowledge

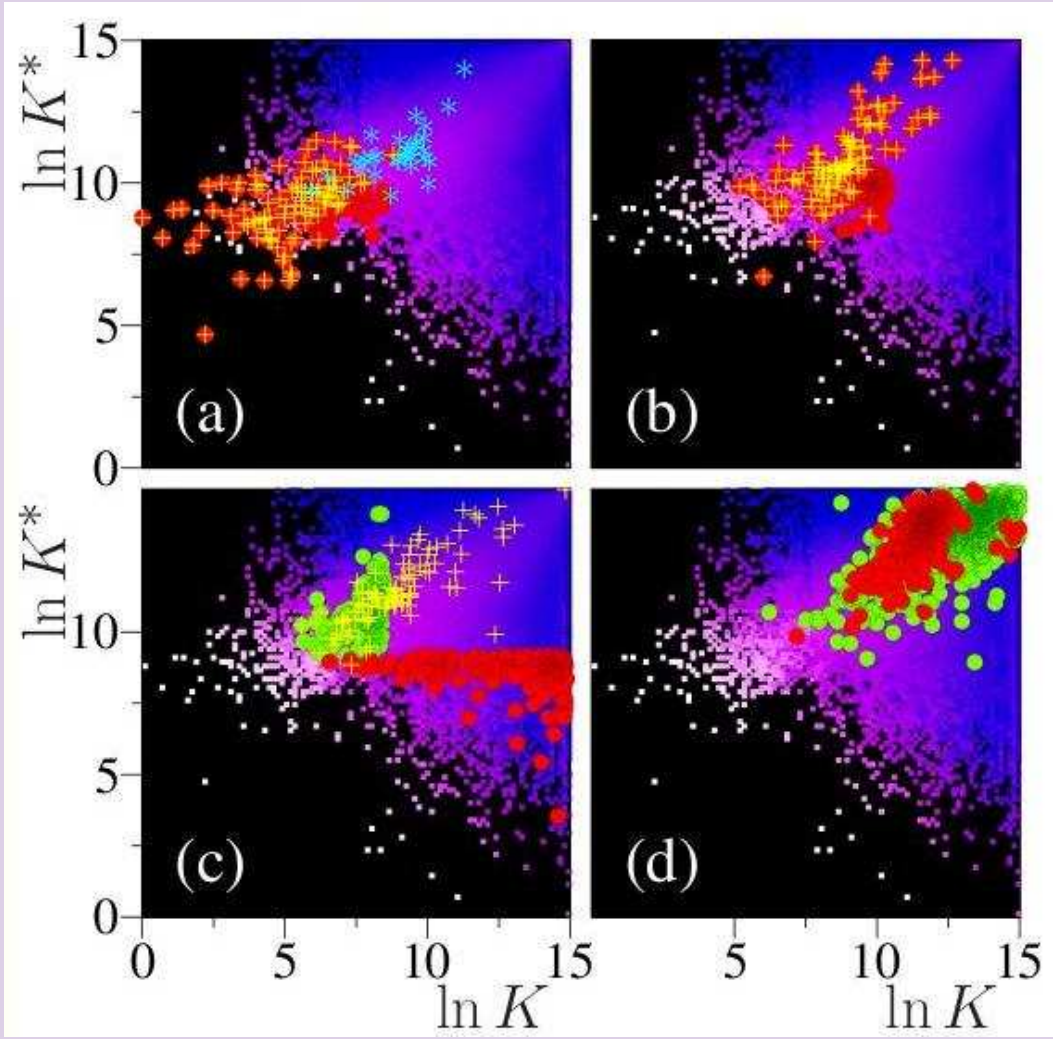
Wikipedia English articles $N = 3282257$ dated Aug 18, 2009



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 ($\ln K$, $\ln K^*$): (a)100 top countries from 2DRank (red), 100 top from SJR (yellow), 30 Dow-Jones companies (cyan); (b)100 top universities from 2DRank (red) and Shanghai (yellow); (c)100 top personalities from PageRank (green), CheiRank (red) and Hart book (yellow); (d)758 physicists (green) and 193 Nobel laureates (red).

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.

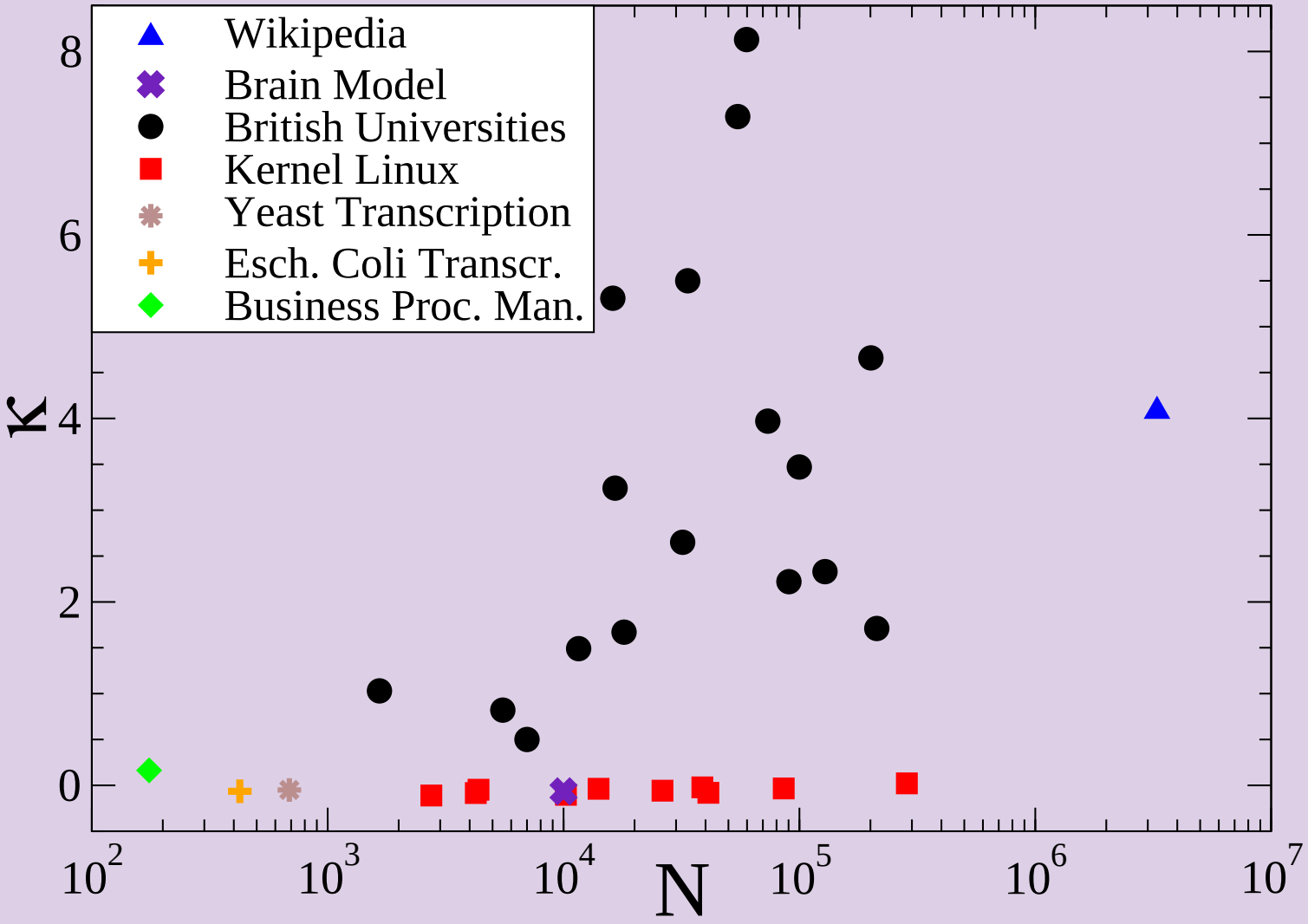
Personalities:

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.

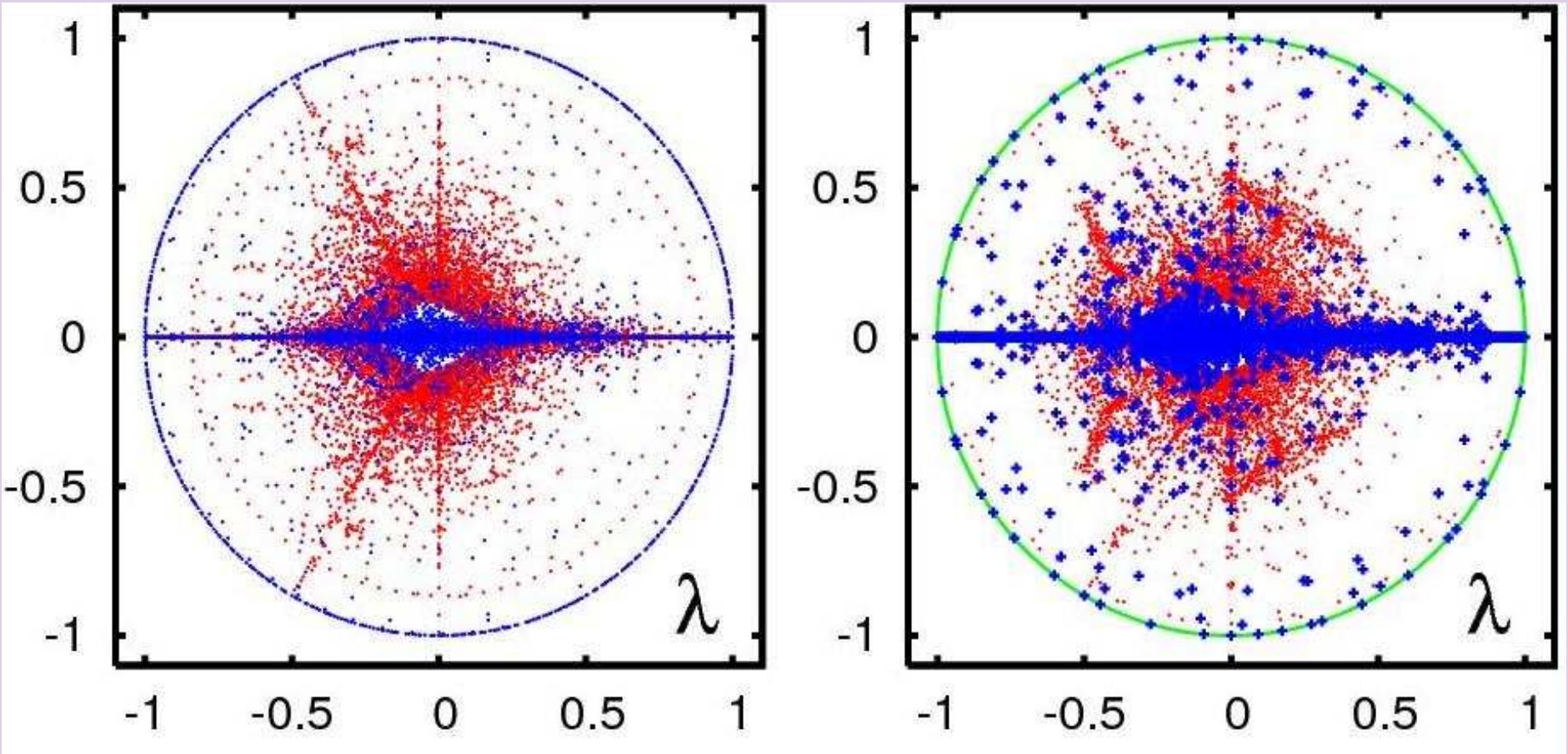
CheiRank: 1. Kasey S. Pipes, 2. Roger Calmel, 3. Yury G. Chernavsky, 4. Josh Billings (pitcher), 5. George Lyell, 6. Landon Donovan, 7. Marilyn C. Solvay, 8. Matt Kelley, 9. Johann Georg Hagen, 10. Chikage Oogi.

Correlator of PageRank and CheiRank



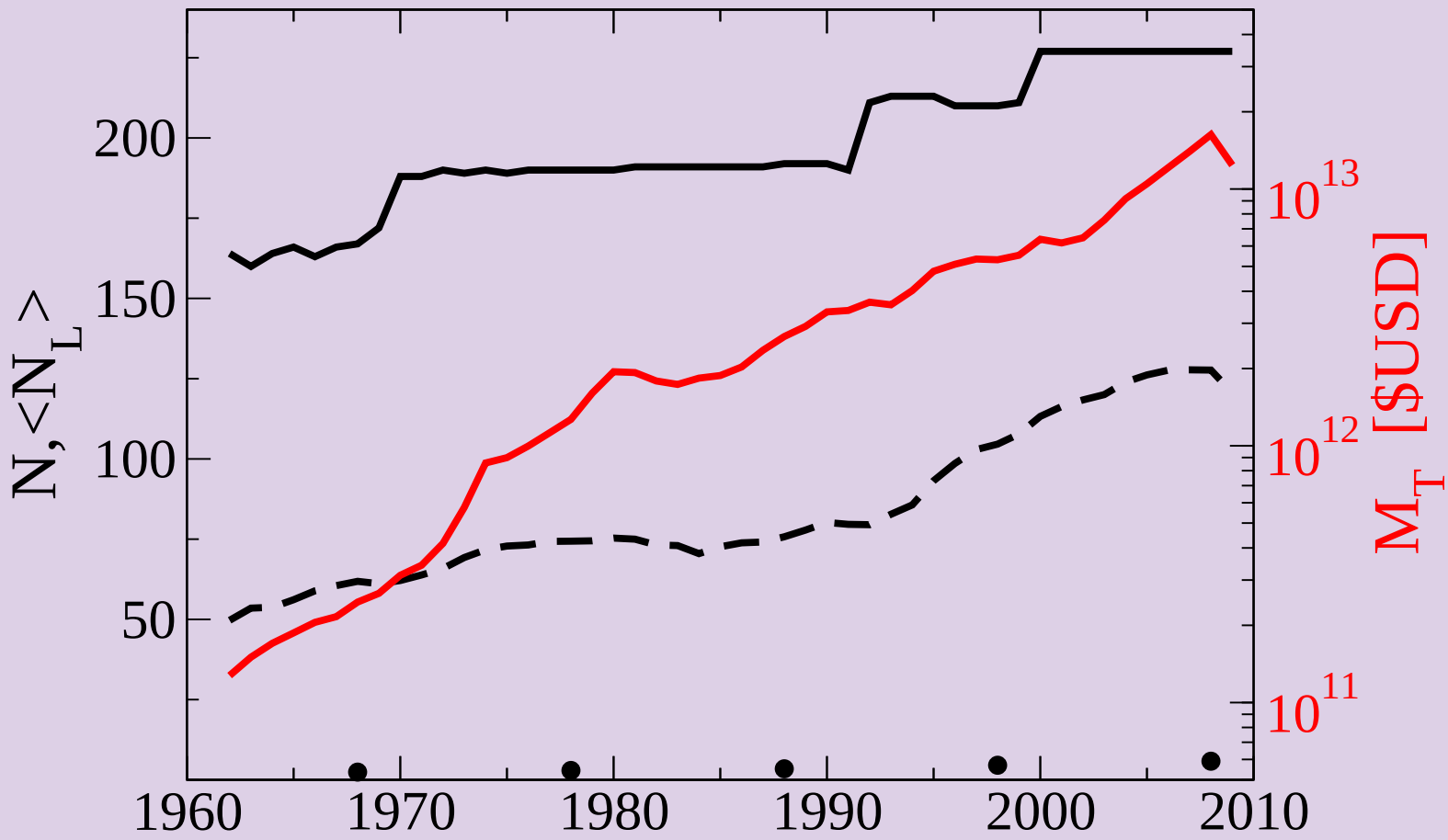
$$\kappa = N \sum_i P(K(i)) P^*(K^*(i)) - 1$$

Spectrum of UK University networks



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

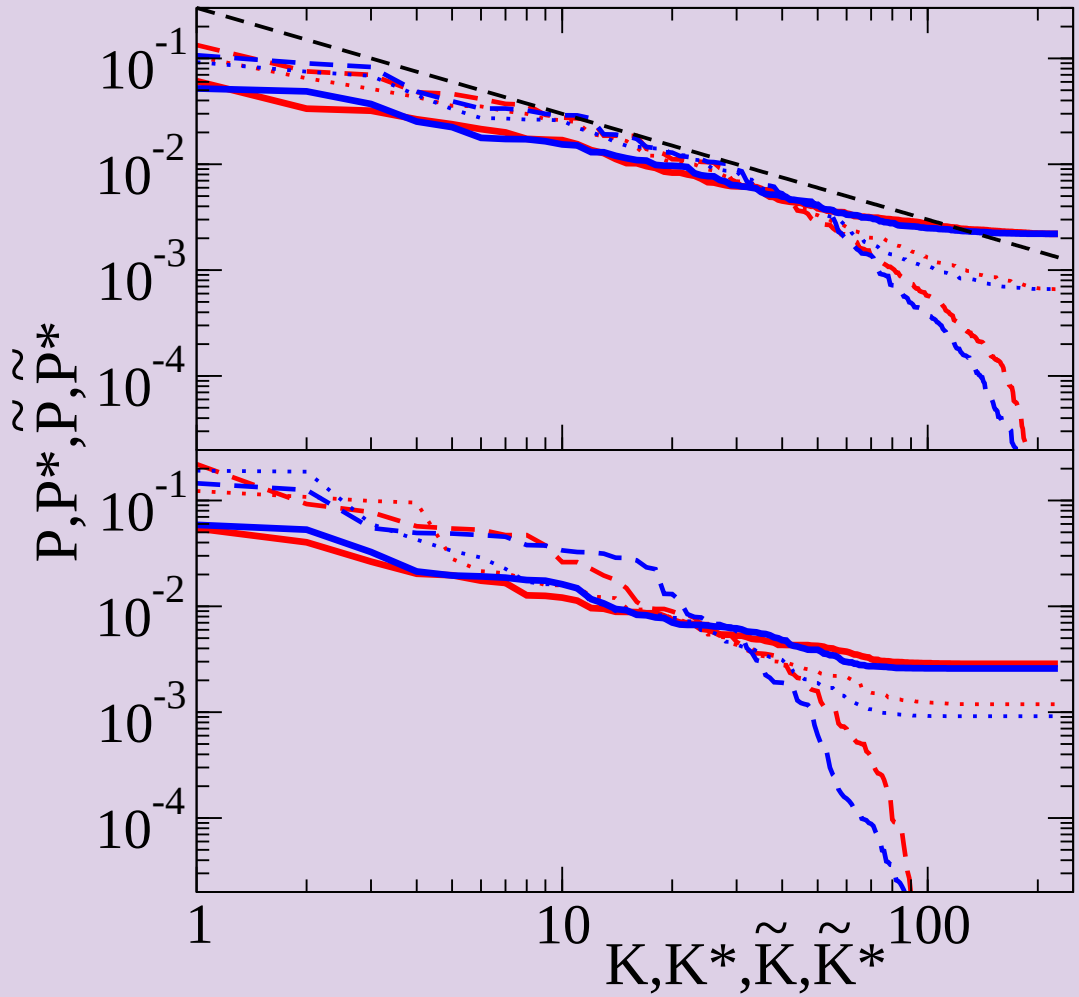
World trade network (WTN) of United Nations COMTRADE 1962-2009



Number of countries (black), links (dashed/points) and mass volume in USD (red)

Leonardo Ermann, DS arxiv:1103.5027 (2011)

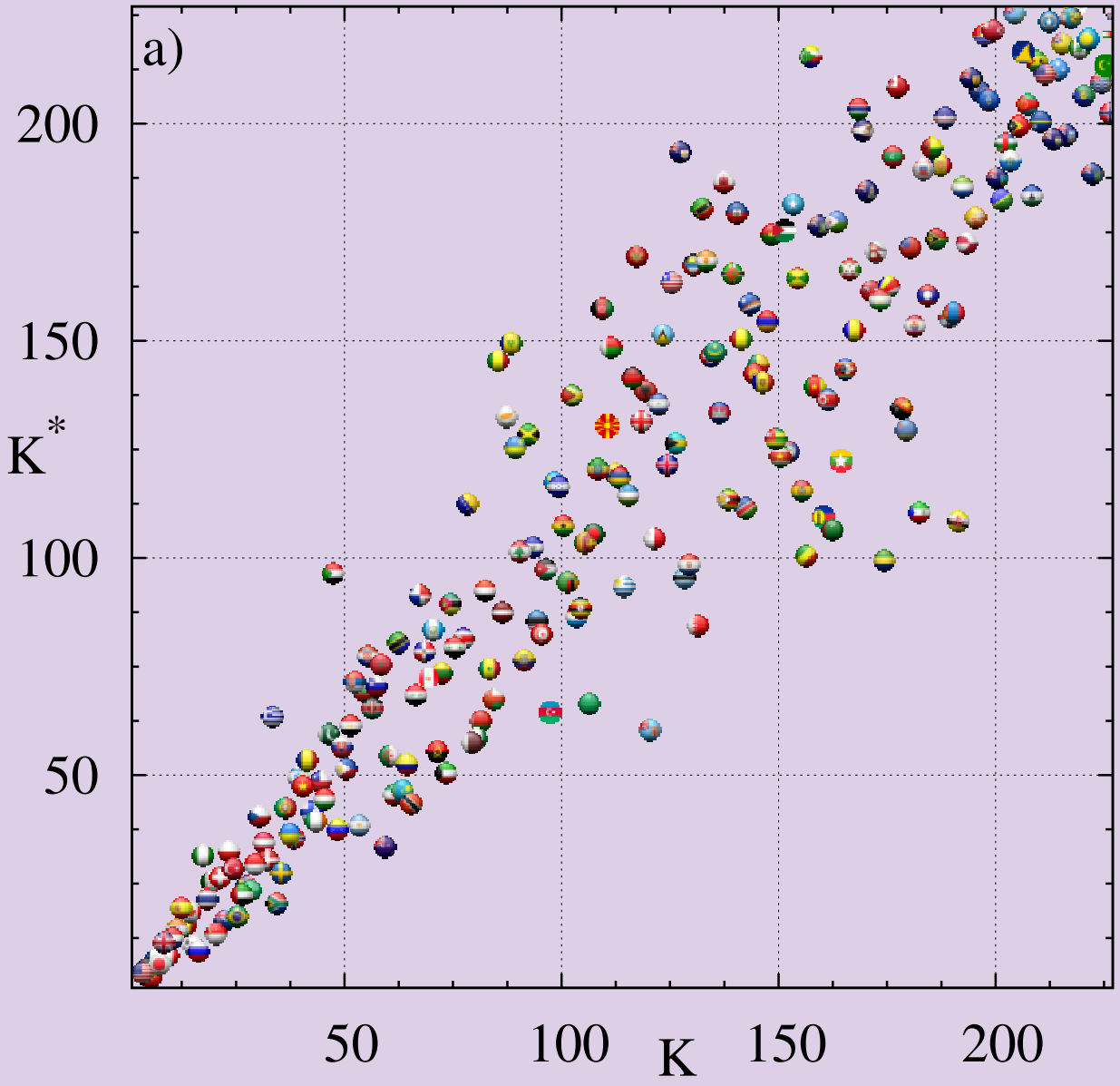
PageRank, CheiRank of World Trade



Year 2008: Probabilities of PageRank $P(K)$ (red), CheiRank $P^*(K^*)$ (blue) for all commodities (top) and crude petroleum (bottom), $\alpha = \mathbf{0.5}; 0.85$ (full/dotted); (dashed curves are for ImportRank, ExportRank); dashed line Zipf law $P \sim 1/K$; 227 countries

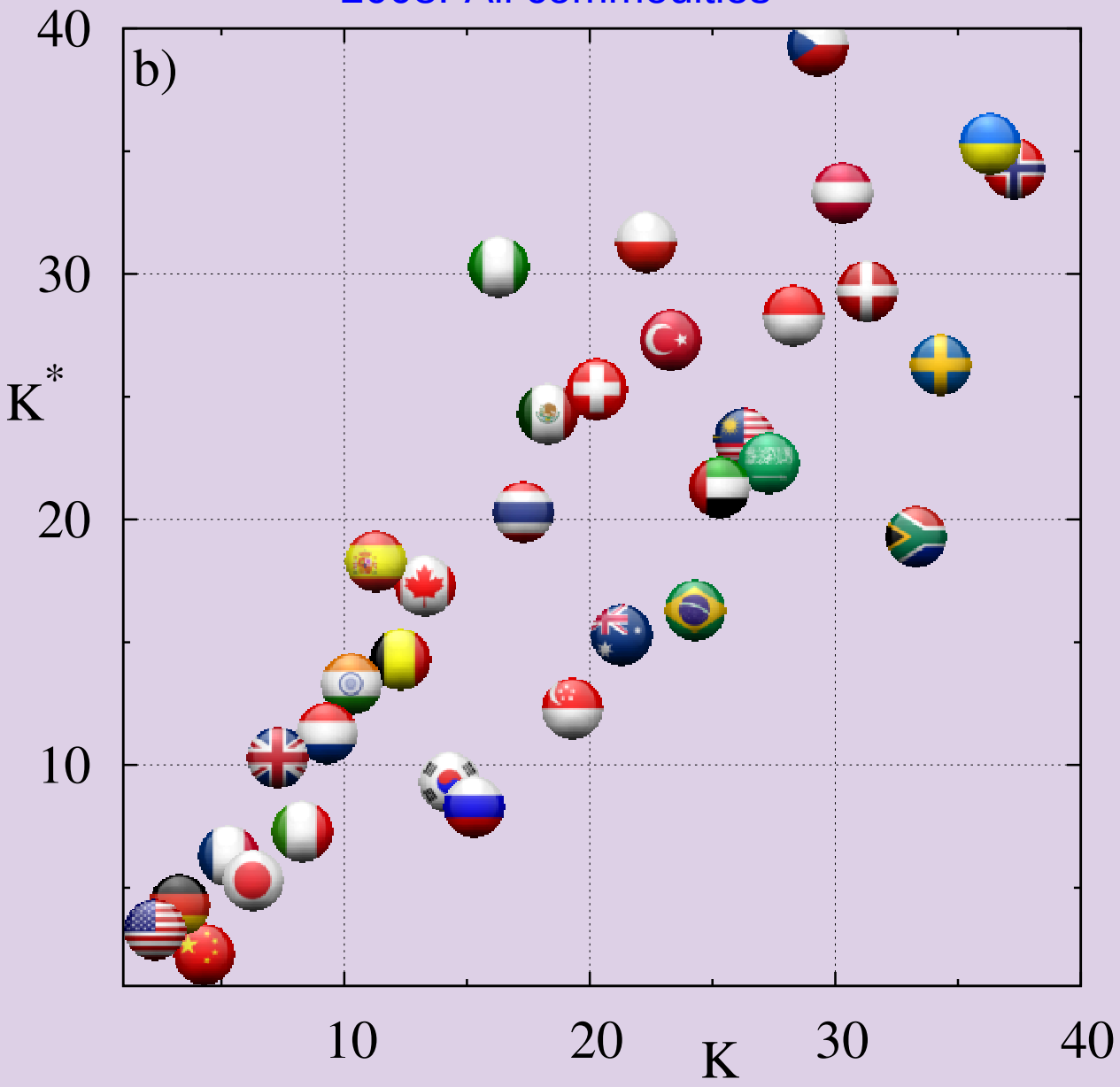
Ranking of World Trade

2008: All commodities



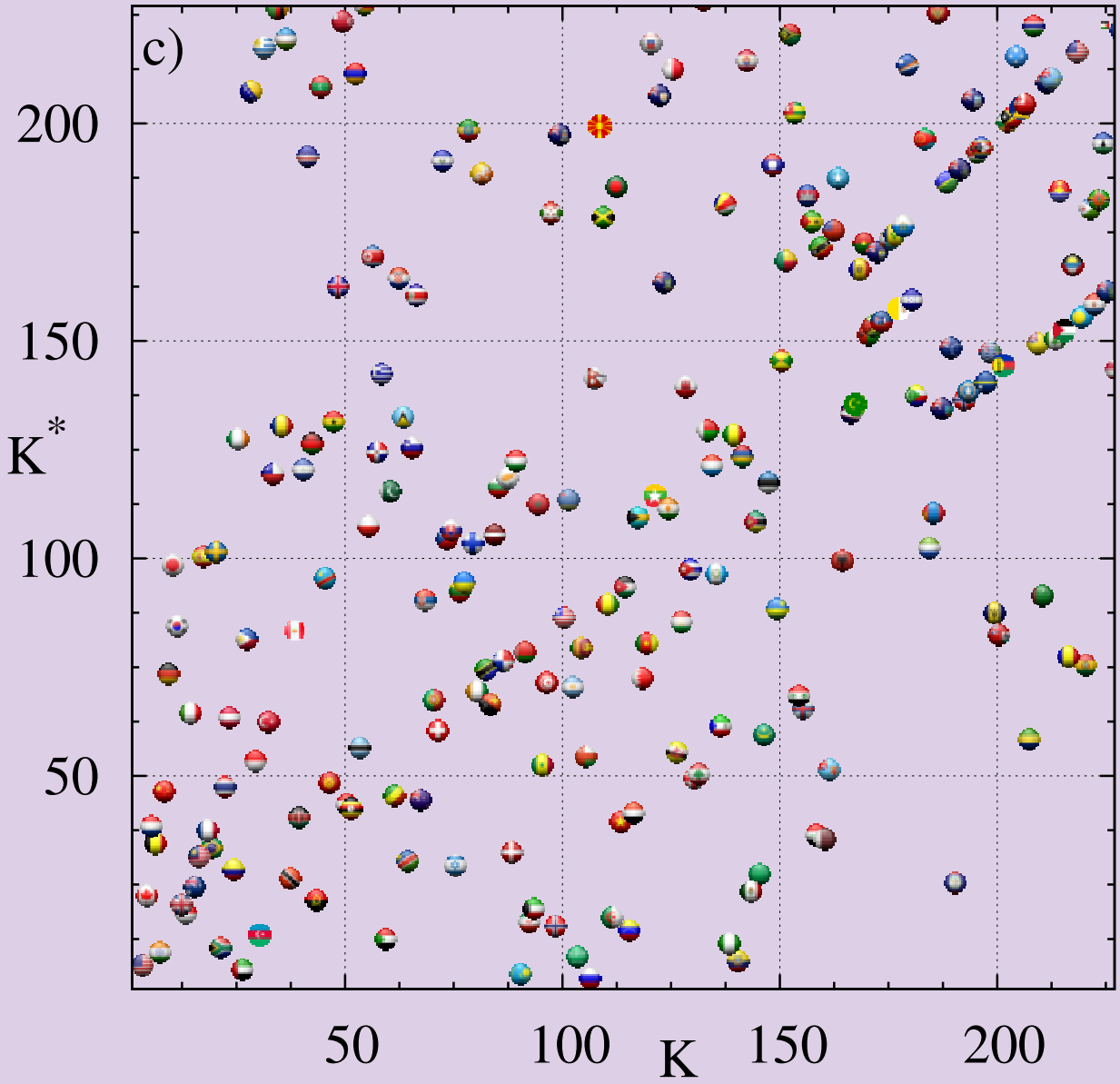
Ranking of World Trade

2008: All commodities

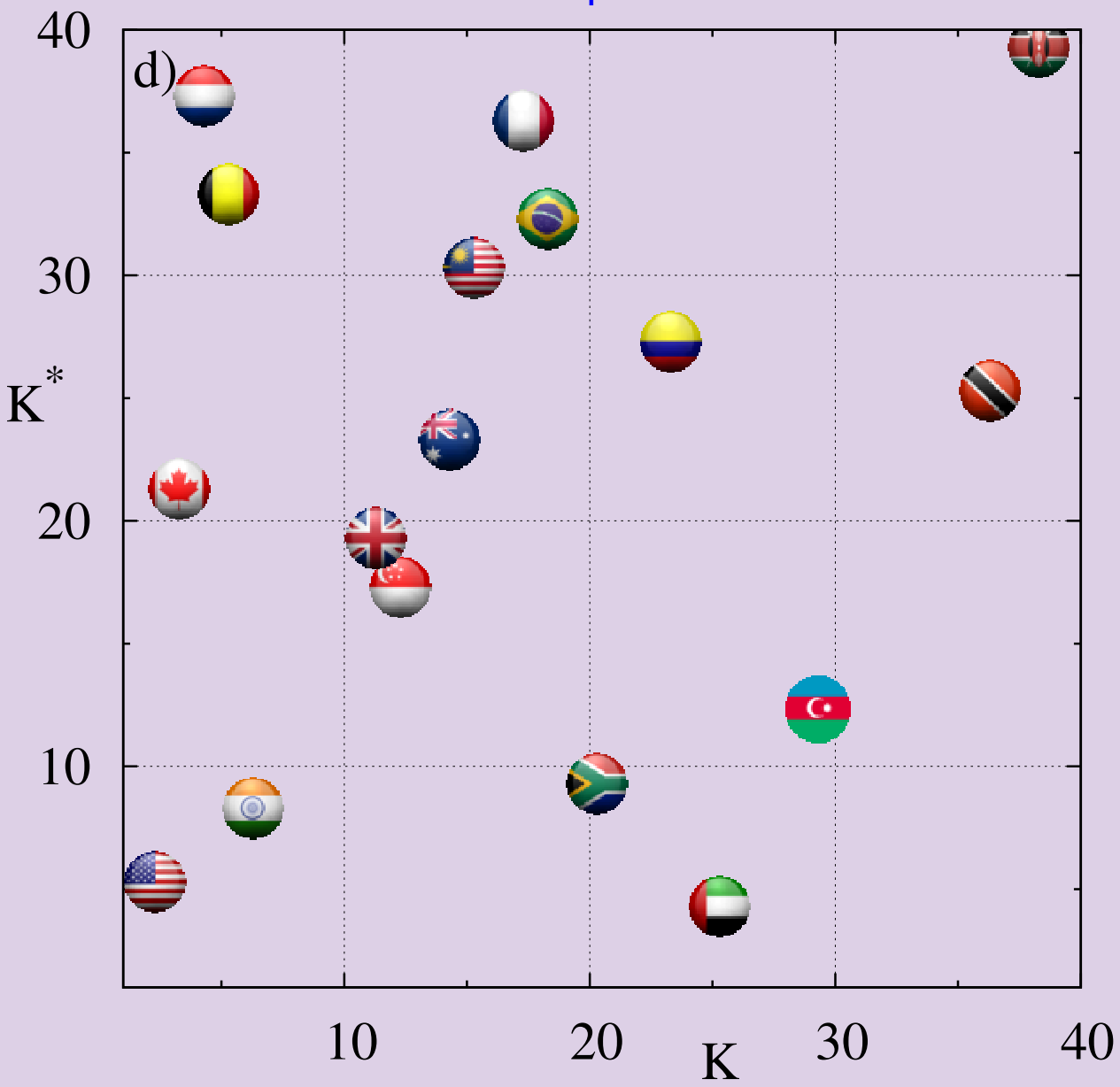


Ranking of World Trade

2008: Crude petroleum

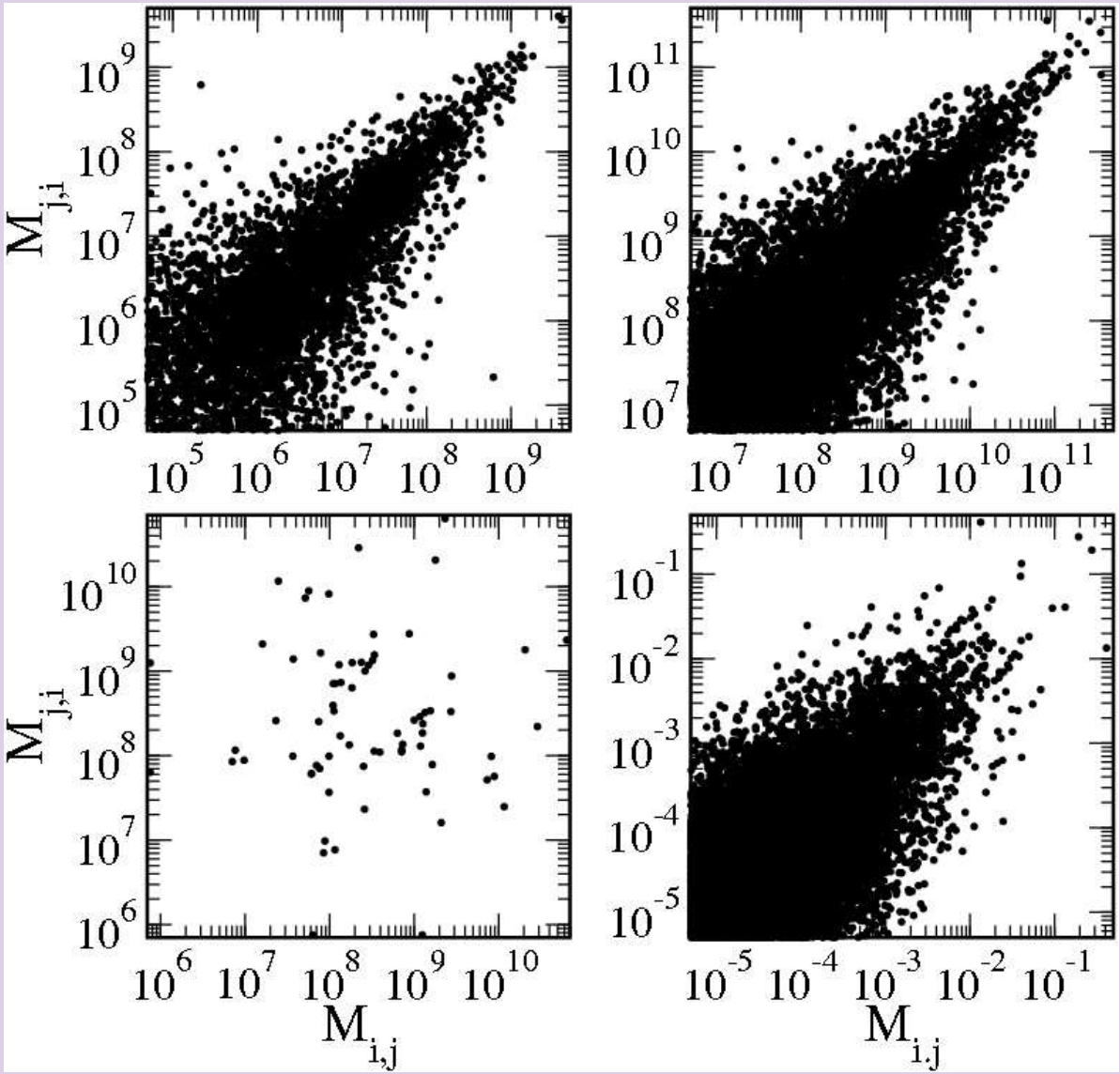


Ranking of World Trade



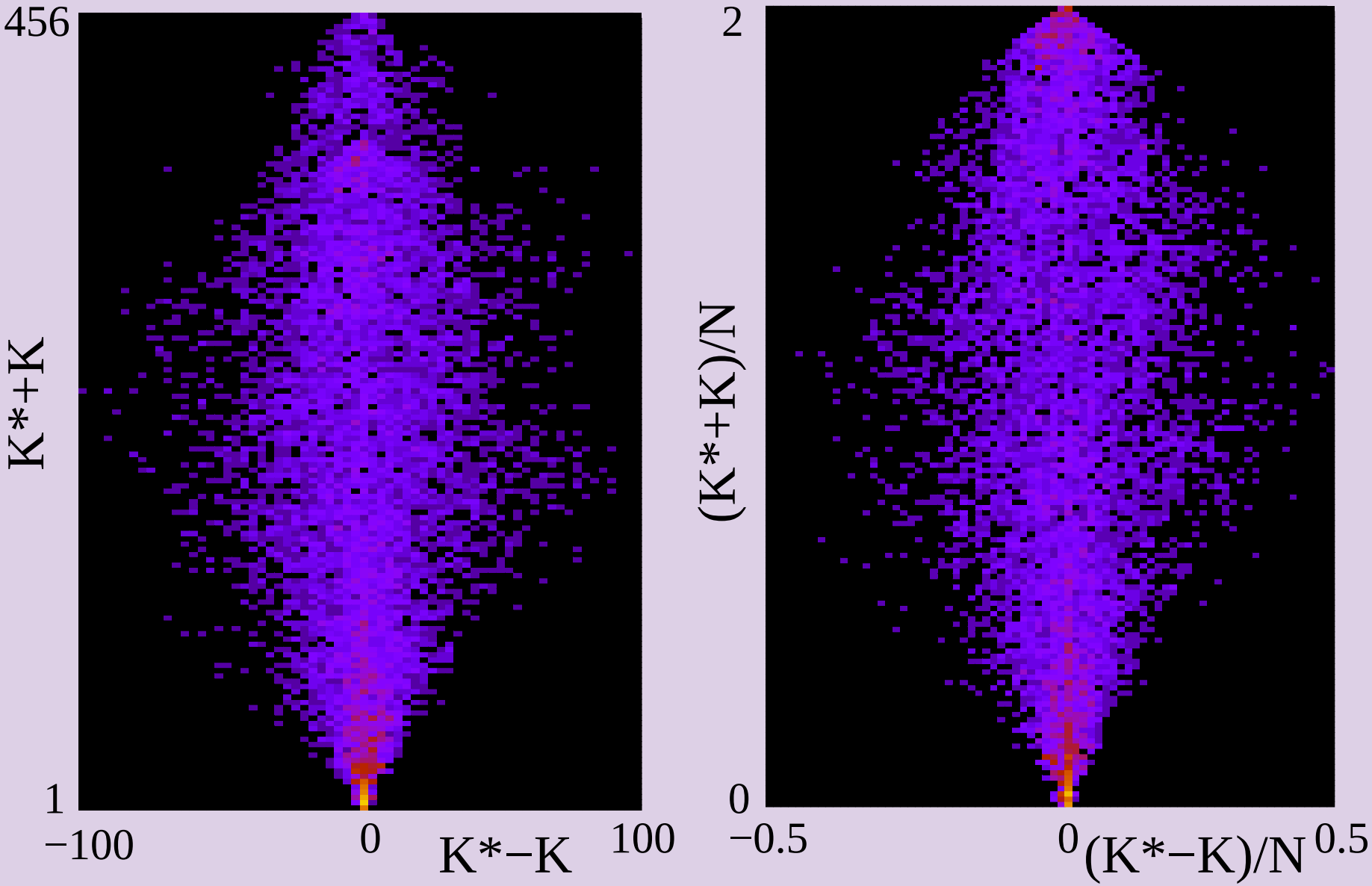
Mass flow on World Trade Network (WTN)

RMT model $M_{ij} = \epsilon_i \epsilon_j / ij$ (all commod. 1962/2008 left/right top; petroleum left bottom; model right bottom)



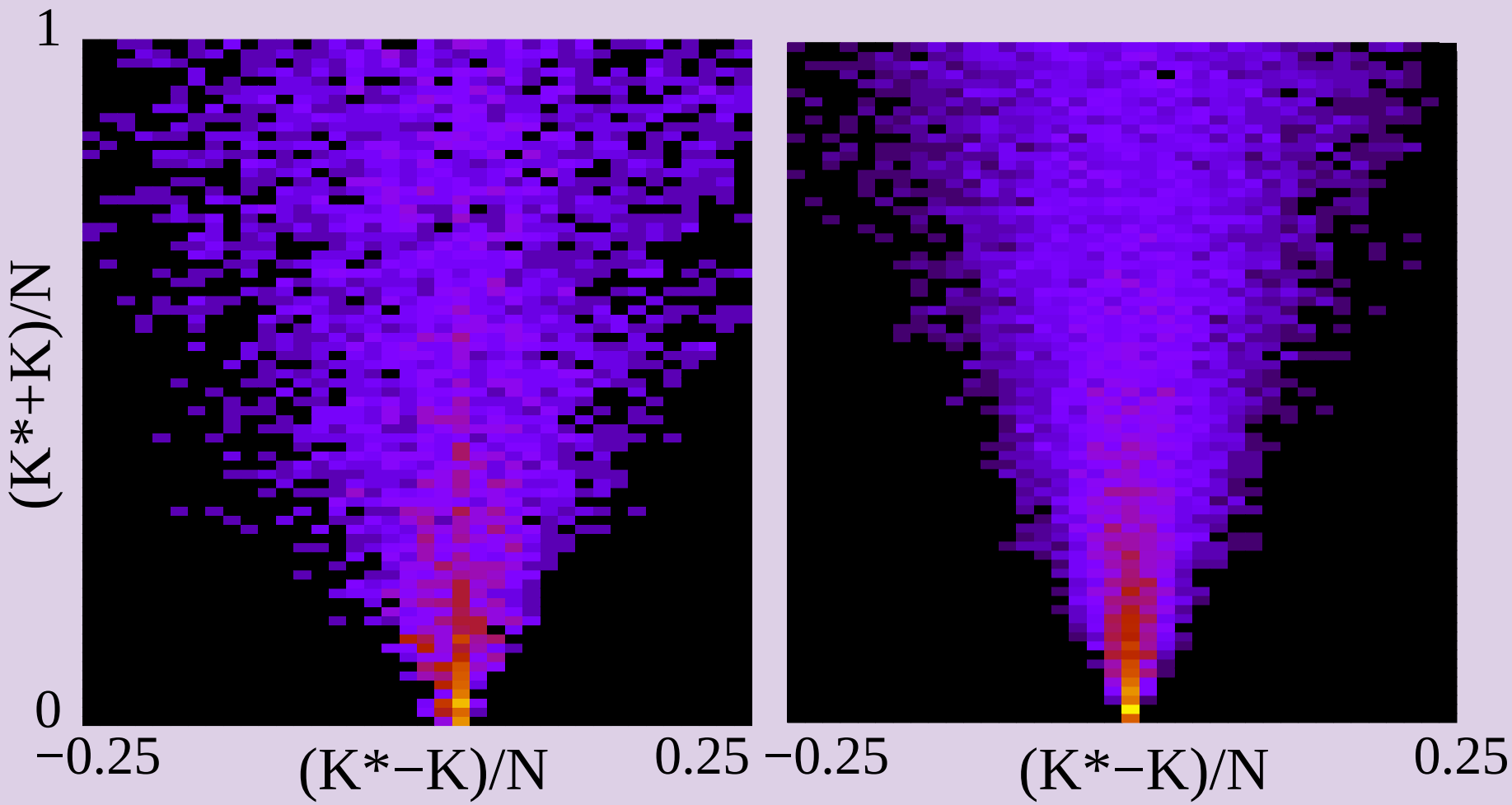
Global distribution for WTN

All commodities 1962-2009



Global distribution for WTN

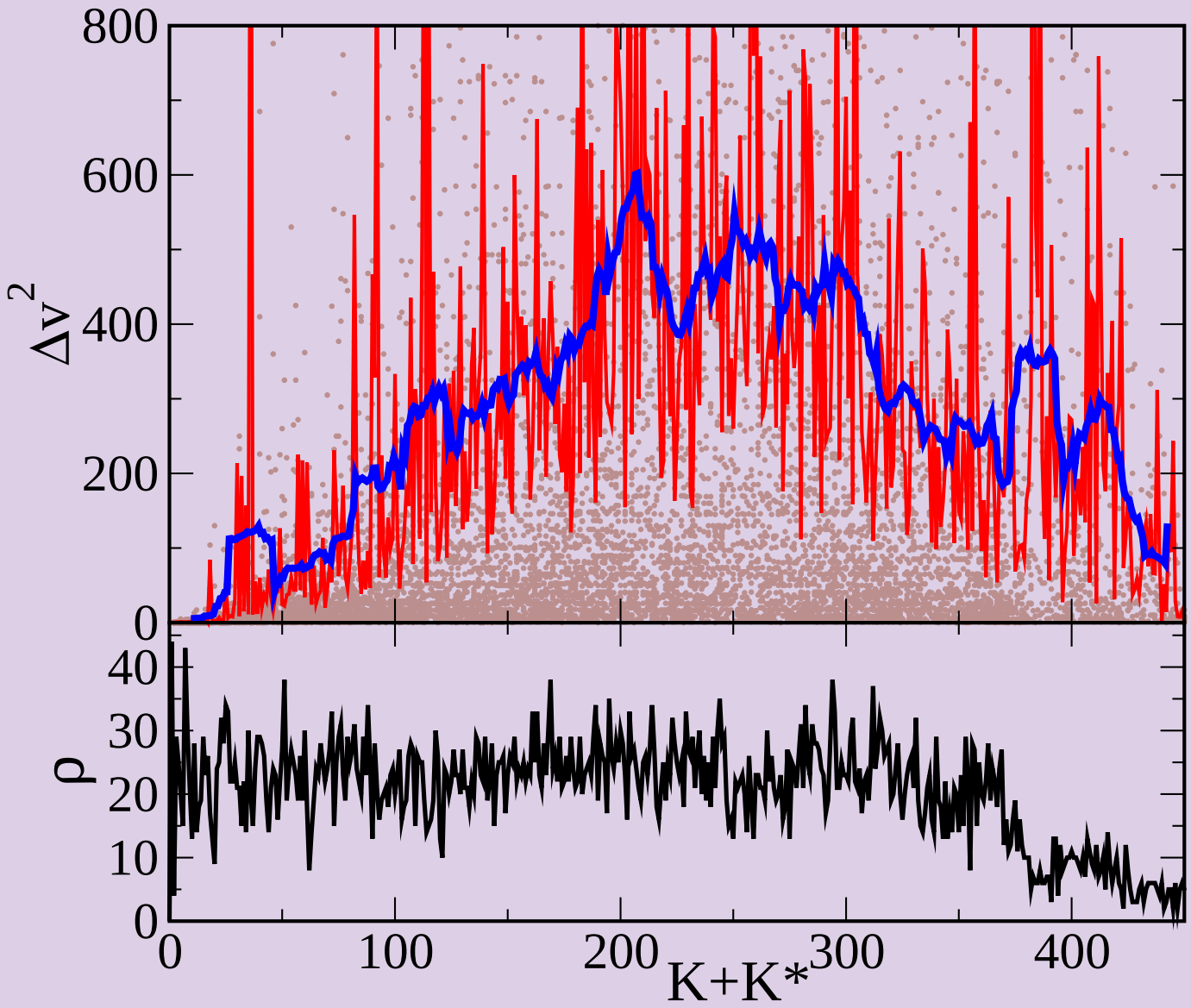
All commodities 1962-2009: left - zoom, right - RMT model



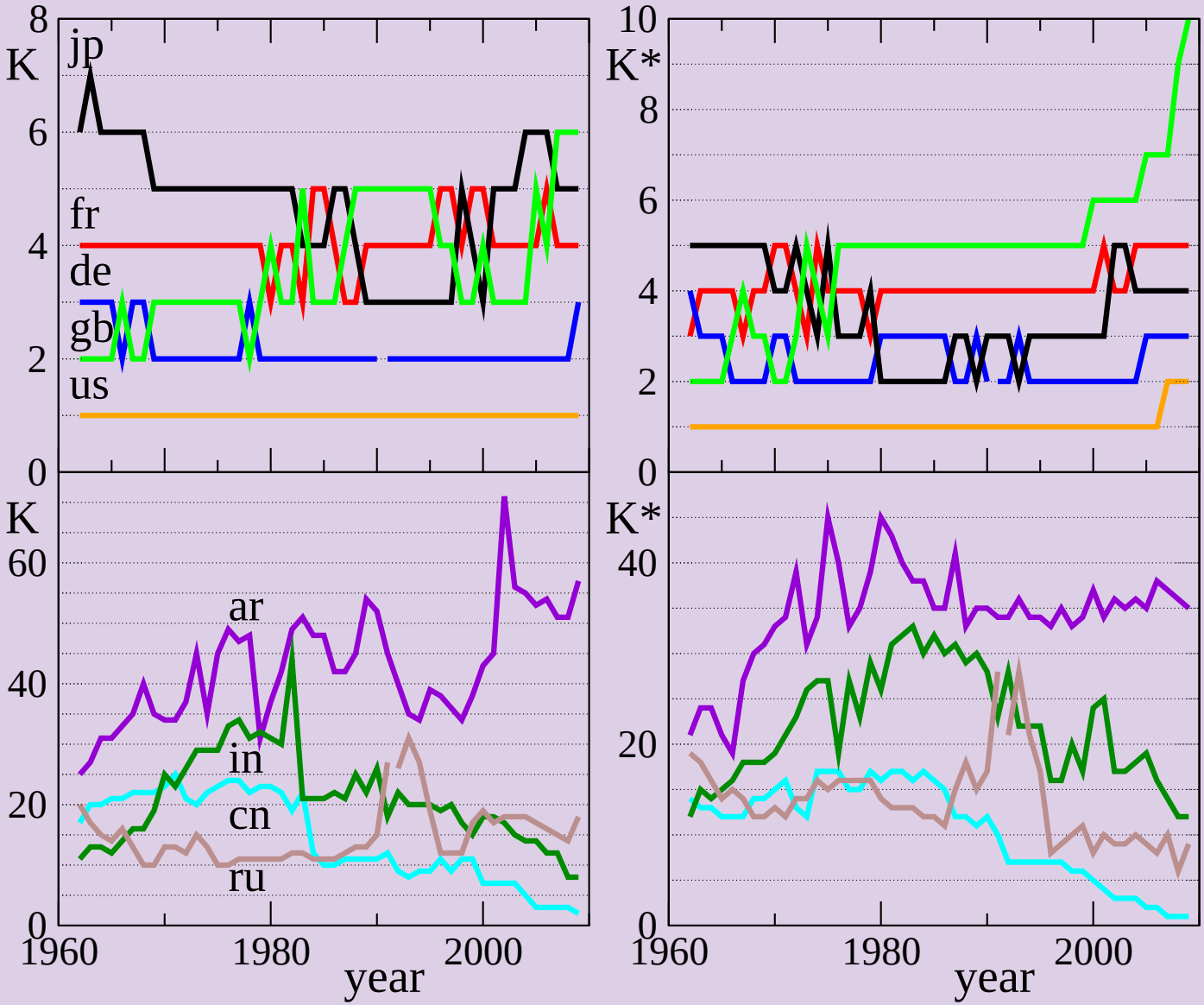
The poor stay poor and the rich stay rich

Velocity fluctuations for WTN

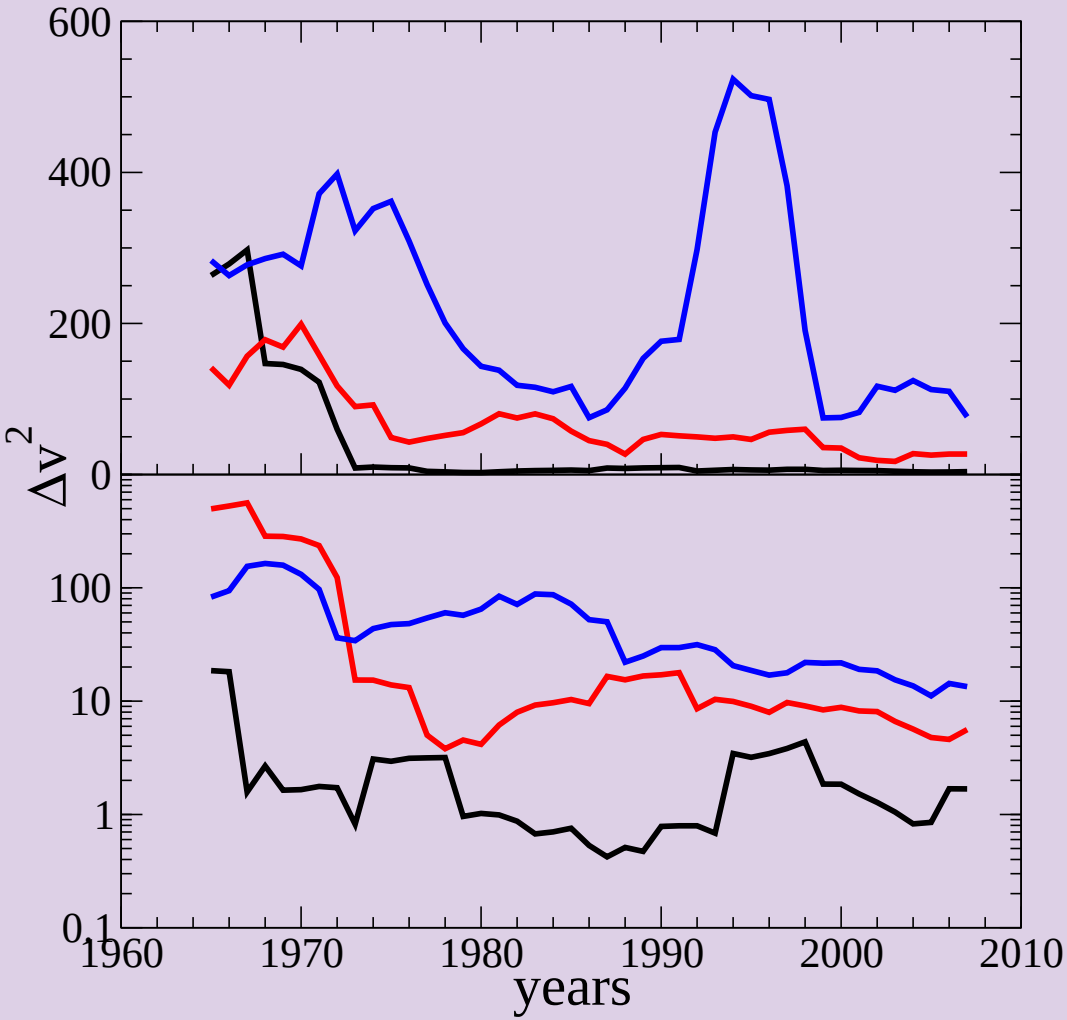
1962-2009: Rank velocity fluctuations $(\Delta v)^2 = (\Delta K)^2 + (\Delta K^*)^2$



Rank evolution in time



Rank evolution in time



Top: $1 \leq K + K^* \leq 40$; $41 \leq K + K^* \leq 80$; $81 \leq K + K^* \leq 120$;

Bottom: $1 \leq K + K^* \leq 20$; $21 \leq K + K^* \leq 40$; $41 \leq K + K^* \leq 60$

Rank table 2008 (74% of countries of G20)

Table 1. Top 20 ranking for *all commodities* – 2008.

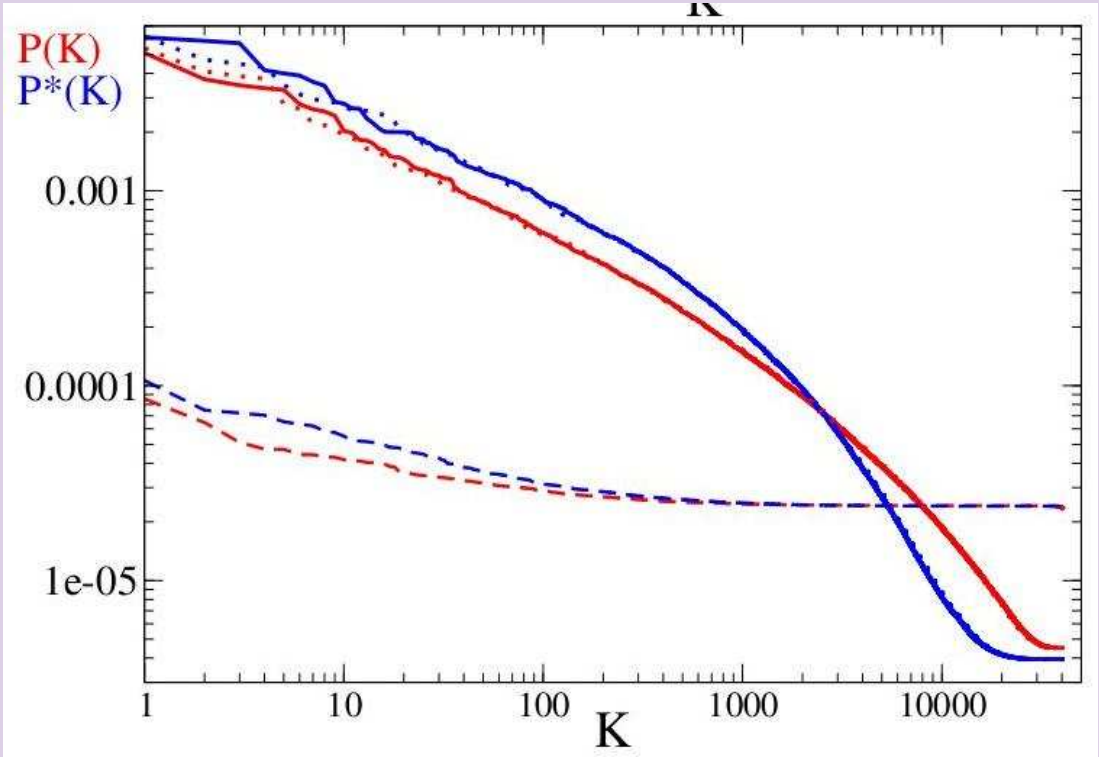
Ran	K	K^*	K_2	$\tilde{K}$	$\tilde{K}^*$
1	USA	China	USA	USA	China
2	Germany	USA	China	Germany	Germany
3	China	Germany	Germany	China	USA
4	France	Japan	Japan	France	Japan
5	Japan	France	France	Japan	France
6	UK	Italy	Italy	UK	Netherlands
7	Italy	Russian Fed.	UK	Netherlands	Italy
8	Netherlands	Rep. of Korea	Netherlands	Italy	Russian Fed.
9	India	UK	India	Belgium	UK
10	Spain	Netherlands	Rep. of Korea	Canada	Belgium
11	Belgium	Singapore	Belgium	Spain	Canada
12	Canada	India	Russian Fed.	Rep. of Korea	Rep. of Korea
13	Rep. of Korea	Belgium	Canada	Russian Fed.	Mexico
14	Russian Fed.	Australia	Spain	Mexico	Saudi Arabia
15	Nigeria	Brazil	Singapore	Singapore	Singapore
16	Thailand	Canada	Thailand	India	Spain
17	Mexico	Spain	Australia	Poland	Malaysia
18	Singapore	South Africa	Brazil	Switzerland	Brazil
19	Switzerland	Thailand	Mexico	Turkey	India
20	Australia	U. Arab Emir.	U. Arab Emir.	Brazil	Switzerland

Rank table 2008

Table 2. Top 20 ranking for *crude petroleum* – 2008.

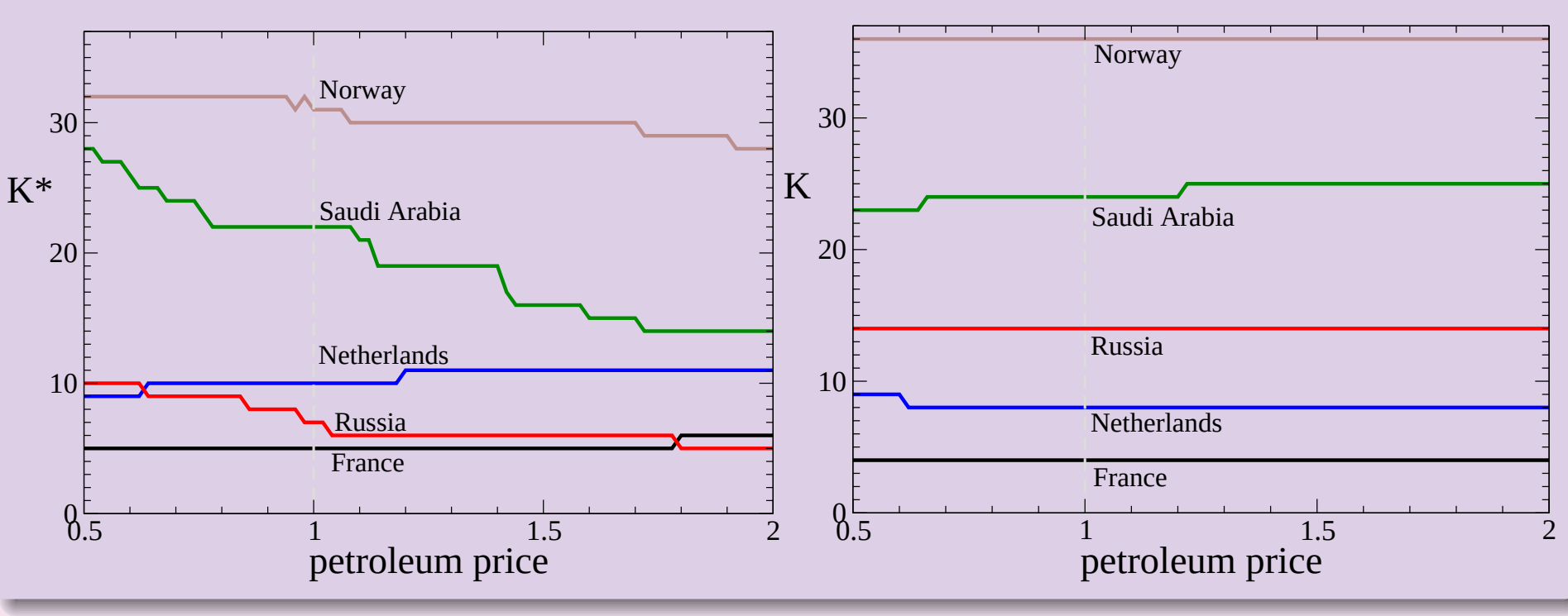
Ran	K	K^*	K_2	$\tilde{K}$	$\tilde{K}^*$
1	USA	Russian Fed.	USA	USA	Saudi Arabia
2	Canada	Kazakhstan	India	Japan	Russian Fed.
3	Netherlands	U. Arab Emir.	Singapore	China	U. Arab Emir.
4	Belgium	USA	UK	Italy	Nigeria
5	India	Ecuador	South Africa	Rep. of Korea	Iran
6	China	Saudi Arabia	Canada	India	Venezuela
7	Germany	India	Australia	Germany	Norway
8	Japan	South Africa	U. Arab Emir.	Netherlands	Canada
9	Rep. of Korea	Nigeria	Colombia	France	Angola
10	UK	Sudan	Azerbaijan	UK	Iraq
11	Singapore	Azerbaijan	Malaysia	Spain	Libya
12	Italy	Venezuela	Brazil	Singapore	Kazakhstan
13	Australia	Norway	Belgium	Canada	Kuwait
14	Malaysia	Iran	Trinidad and Tobago	Thailand	Azerbaijan
15	Spain	Algeria	France	Belgium	Algeria
16	France	Singapore	Netherlands	Brazil	Mexico
17	Brazil	Kuwait	Kenya	Turkey	UK
18	Sweden	UK	Angola	South Africa	Qatar
19	South Africa	Angola	China	Poland	Oman
20	Thailand	Canada	Thailand	Australia	Netherlands

Google matrix of multiproduct trade 2008



CheiRank vs. PageRank for multiproduct trade at $N_p = 182$ for $N_c = 227$ UN COMTRADE countries in 2008; 3 models of product coupling (full, dotted, dashed curves)
L.Ermann, DS (in progress)

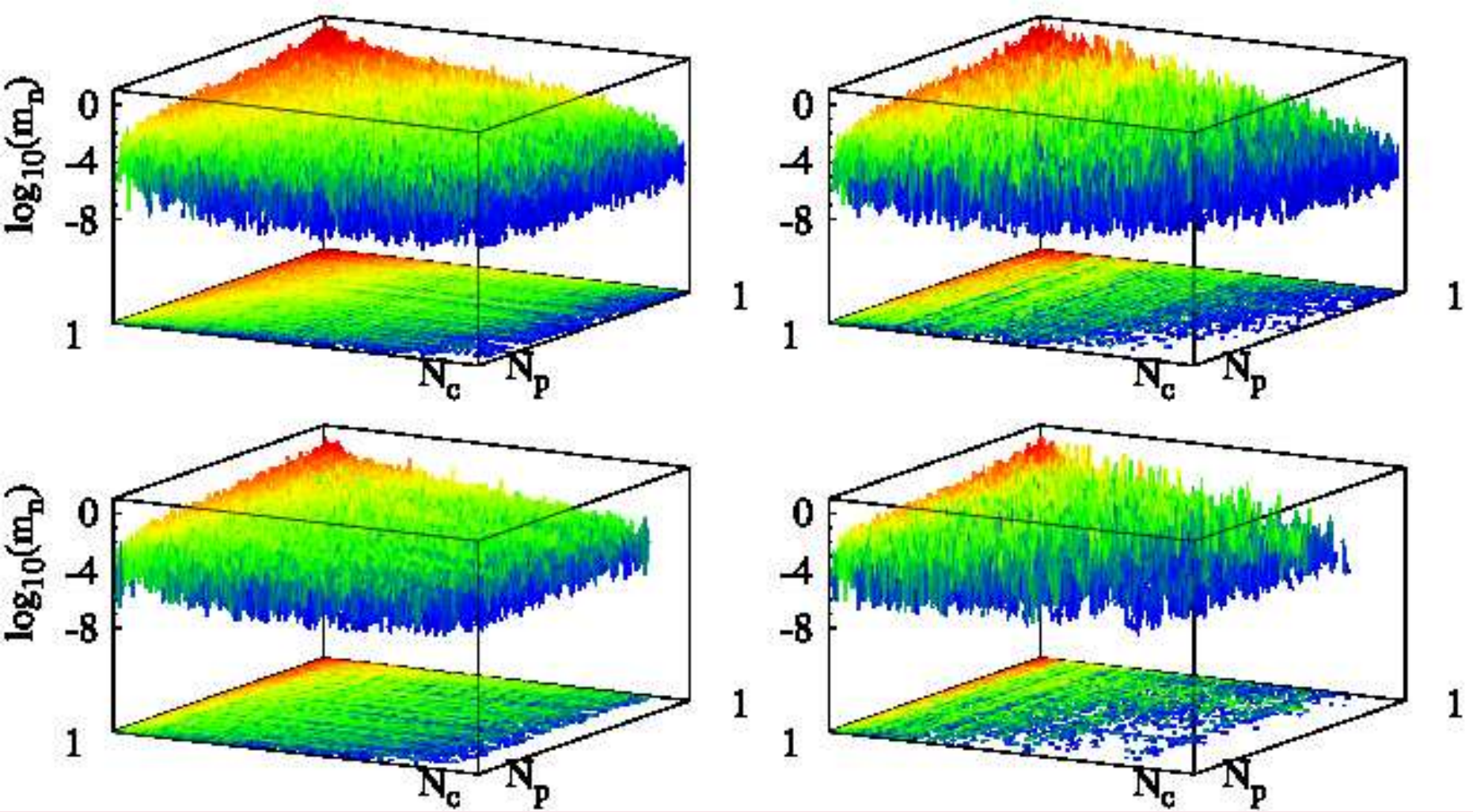
Petroleum price effect on ranking of trade 2008



CheiRank K^*, PageRank K variation with petroleum price in respect to price of 2008

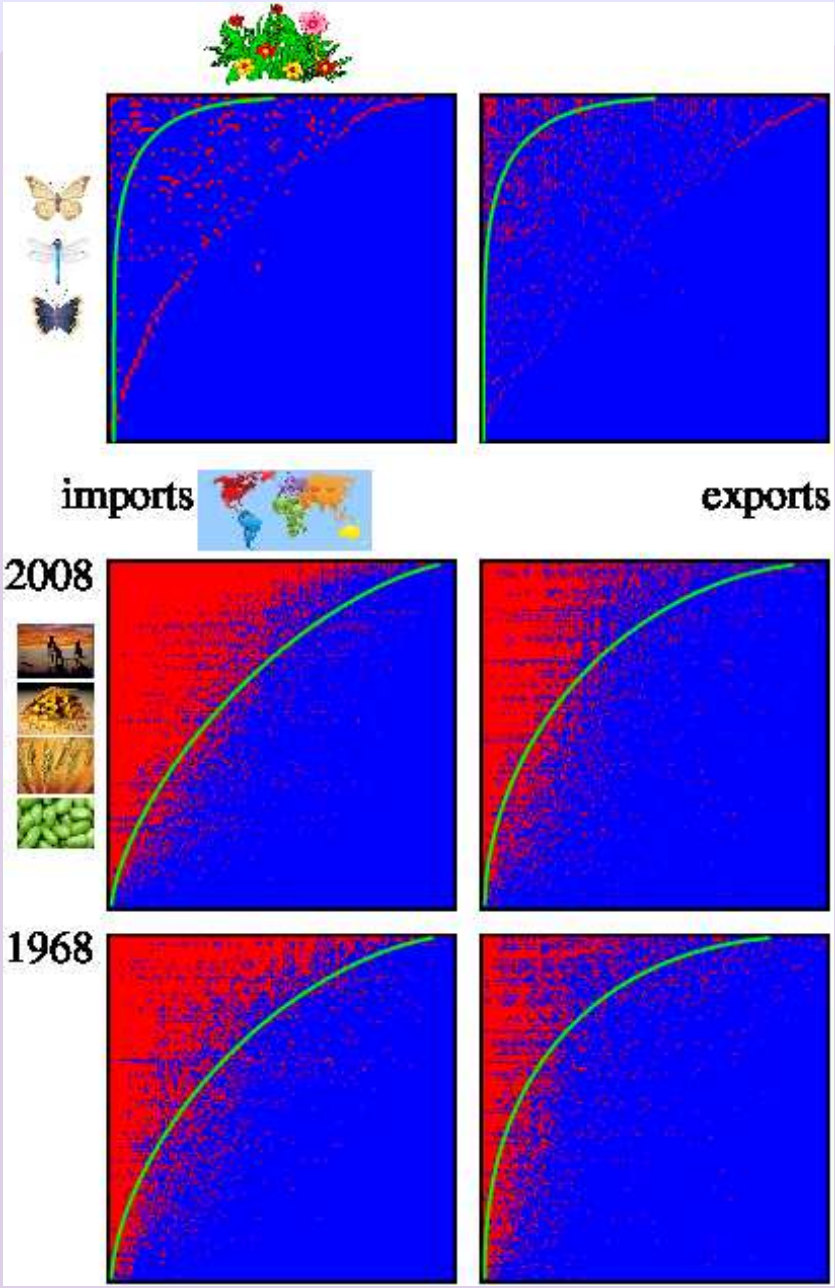
L.Ermann, DS (in progress)

Ecological analysis of world trade



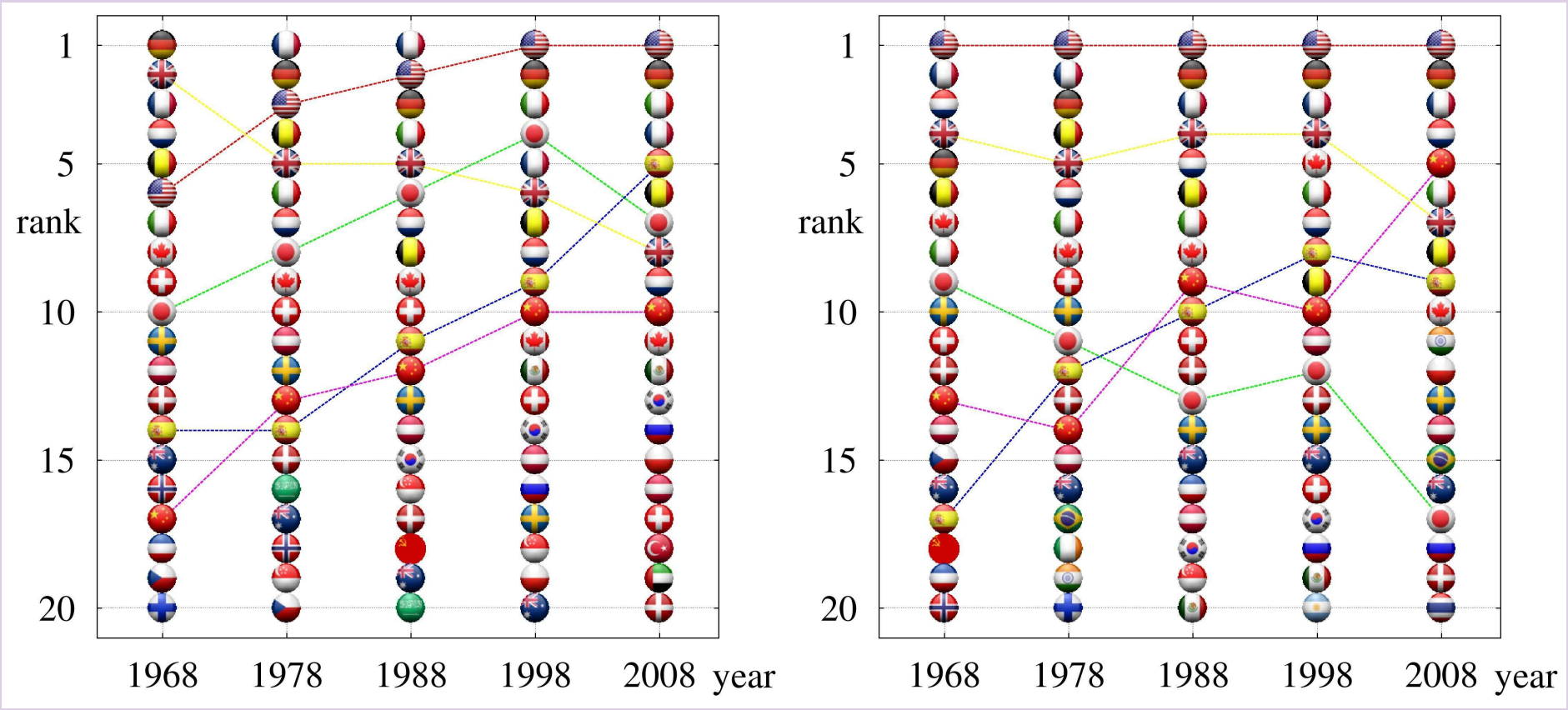
Normalized monetary trade volume: import (left), export (right), 1968 (bottom) and 2008 (top); arxiv:1201.3584; countries/products $N_c = 164, 227 / N_p = 182$; import/export $M^{(i)}_{p,c} = \sum_{c'=1}^{N_c} M^{p}_{c,c'}$ / $M^{(e)}_{p,c} = \sum_{c'=1}^{N_c} M^{p}_{c',c}$; $M_{p,c} > / < \mu => 1/0$

Plants-animals => Countries-products



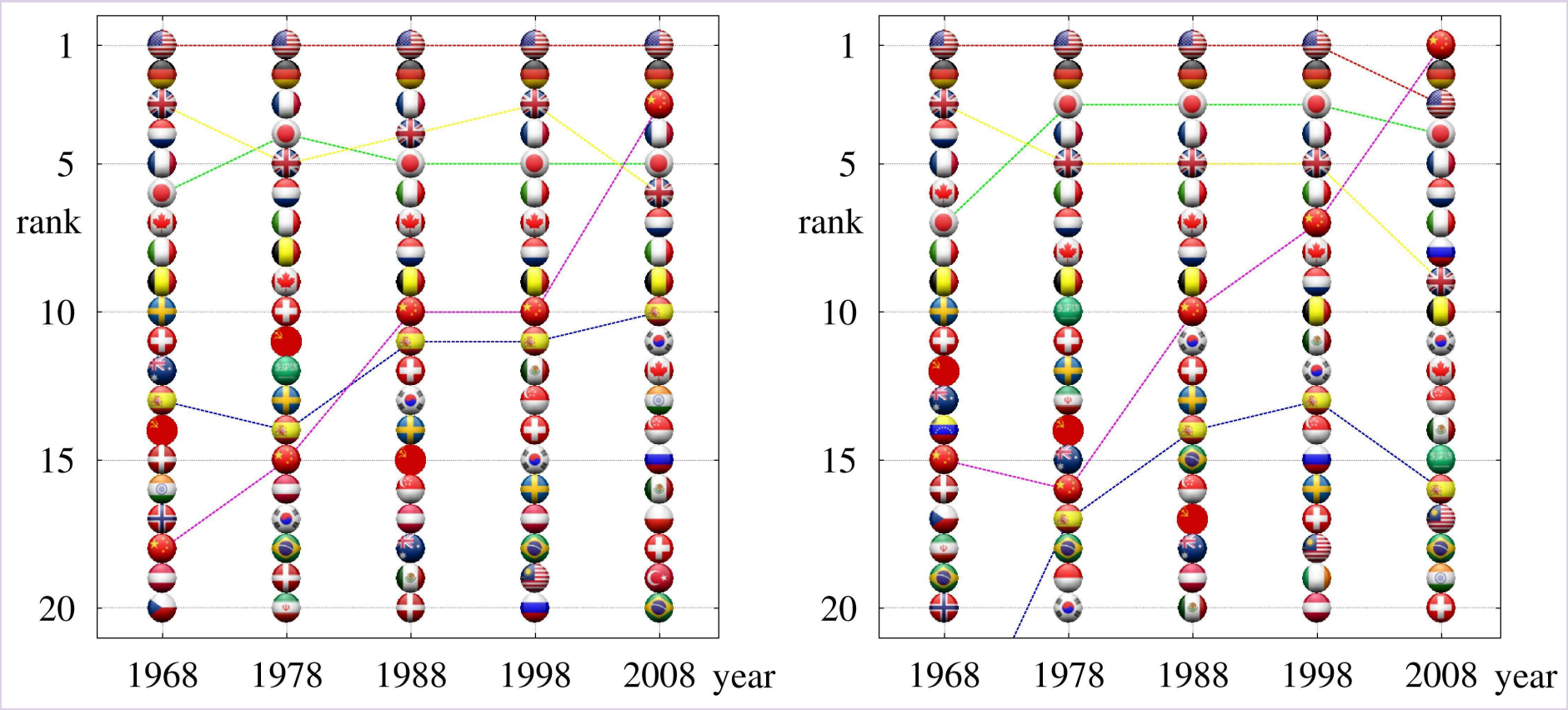
Mutualistic nestedness matrix:
Top: two ecological systems
from J.Bascompte et al.
"The architecture of mutualistic networks minimizes competition and increases biodiversity“
Nature 458, 1018 (2009);
Middle-bottom: WTN data
Nestedness ordering algorithm

Ecological ranking of world trade (countries)



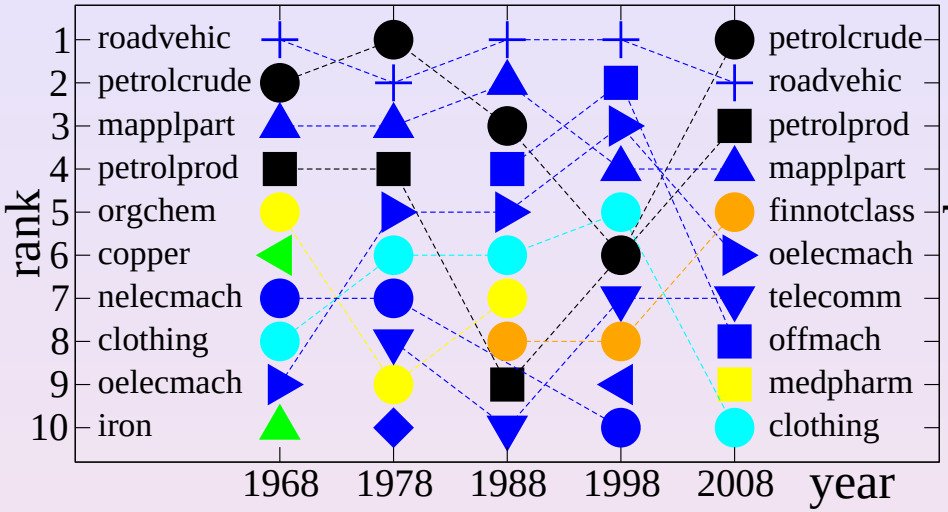
Left: import; Right: export

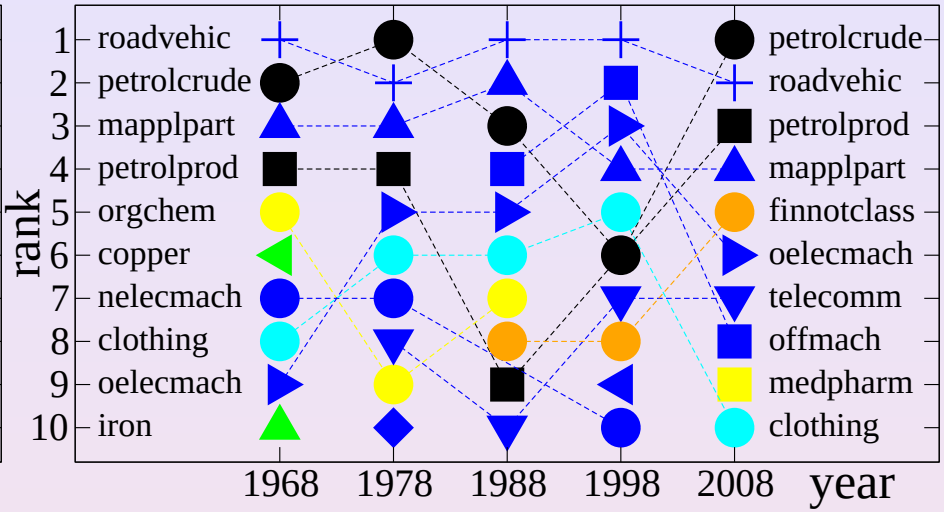
Trade volume ranking of world trade (countries)

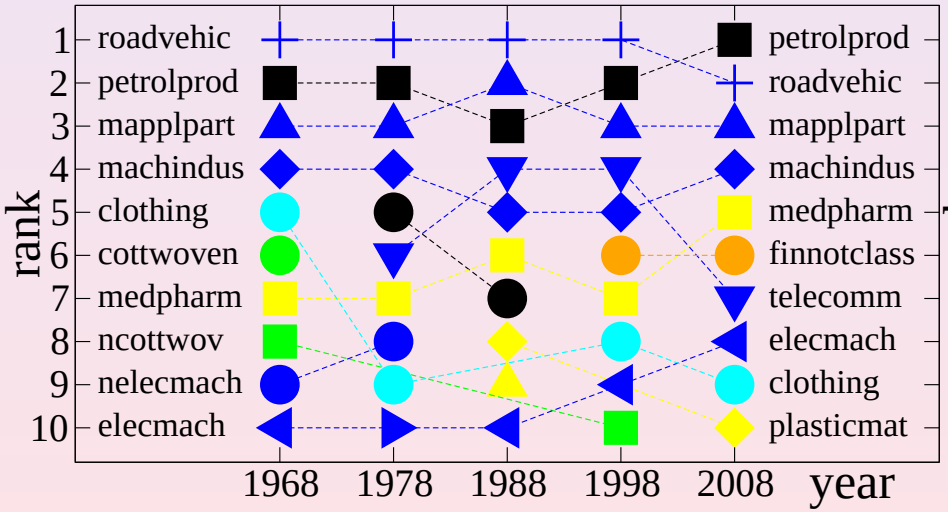


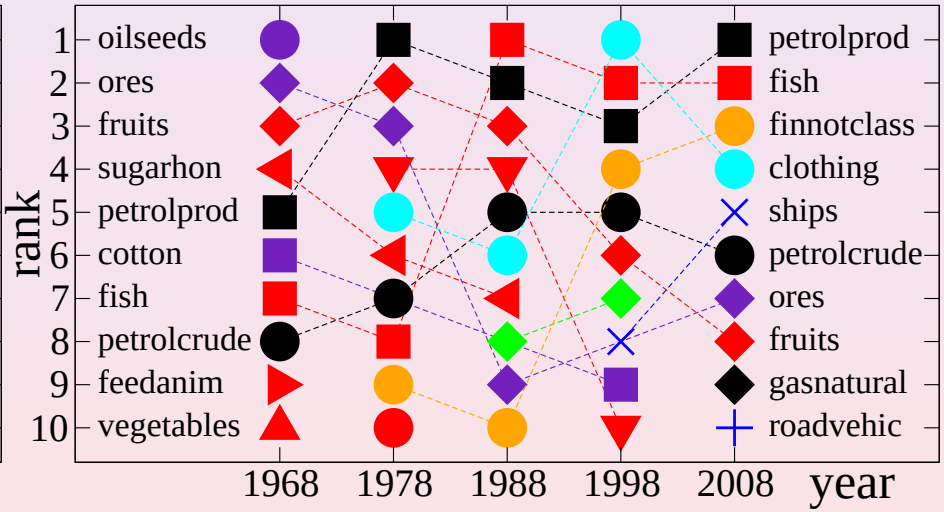
Left: import; Right: export

Ecological ranking of world trade (products)









Top: trade volume; Left bottom: EcoloRank imp; Right bottom: EcoloRank exp

WBW: Towards bank financial network ranking

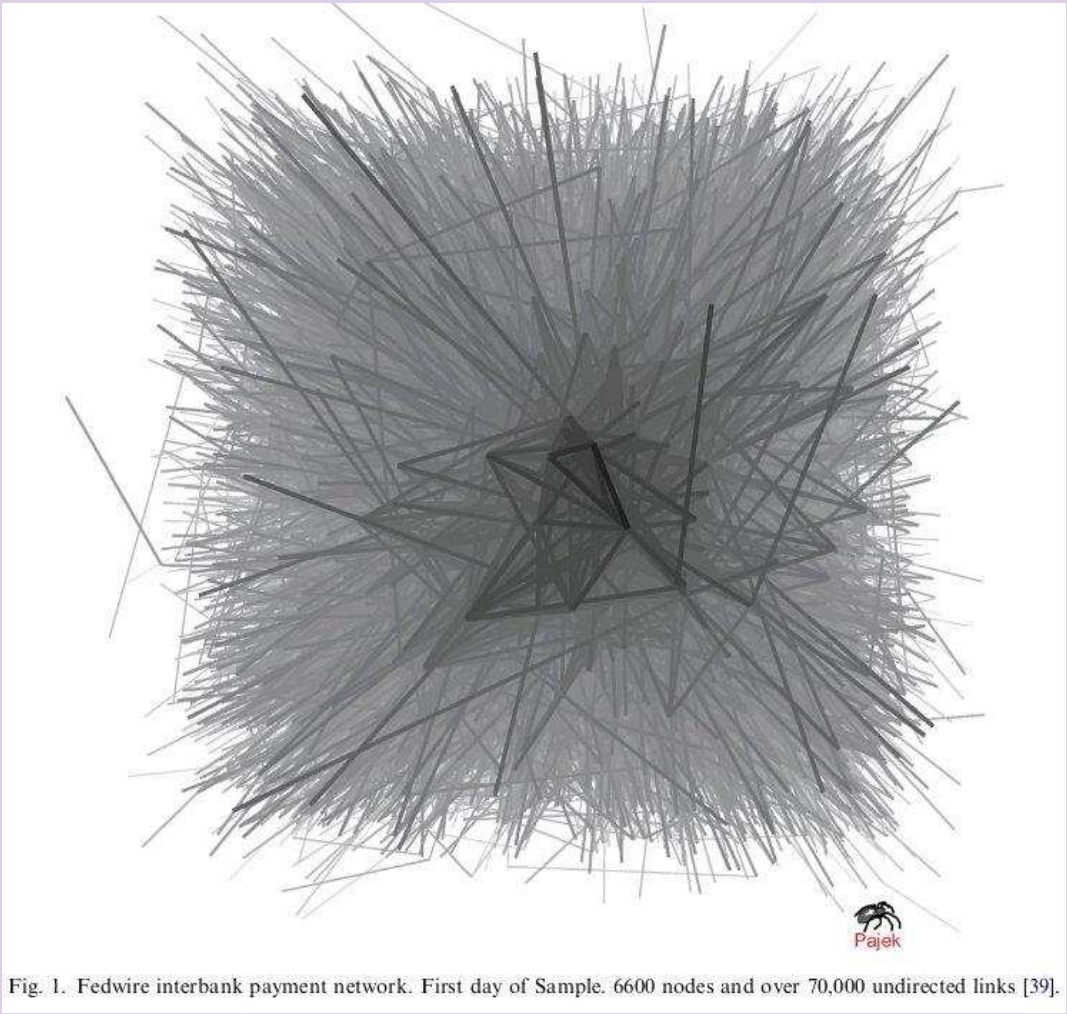


Fig. 1. Fedwire interbank payment network. First day of Sample. 6600 nodes and over 70,000 undirected links [39].

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

Google Matrix Applications

practically to everything .... ECT* Workshop 23-27 July 2012, Trento
=> www.quantware.ups-tlse.fr/complexnetworks2012/

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

(“PETRO primo CATHARINA secunda”, St. Petersburg 1782)

References:

1. S.Brin and L.Page, *The anatomy of a large-scale hypertextual Web search engine*, Comp. Networks ISDN Systems **30**, 107 (1998)

2. 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)].

3. 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)

4. Wikipedia articles *PageRank*, *CheiRank*, *Google matrix* (2008-2011)

5. D.Fogaras, *Where to start browsing the web?*, Lect. Notes Computer Sci. **2877**, 65 (2003)

6. V.Hrisitidis, H.Hwang and Y.Papakonstantinou, *Authority-based keyword search in databases*, ACM Trans. Database Syst. **33**, 1 (2008)

7. A.D.Chepelianksii, *Towards physical laws for software architecture* arXiv:1003.5455[cs.SE] (2010)

8. A.O.Zhirov, O.V.Zhirov and D.L.Shepelyansky, *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77**, 523 (2010)

9. 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):

10. K.M.Frahm and D.L.Shepelyansky, *Ulam method for the Chirikov standard map*, Eur. Phys. J. B **76**, 57 (2010)
11. L.Ermann and D.L.Shepelyansky, *Ulam method and fractal Weyl law for Perron-Frobenius operators*, Eur. Phys. J. B **75**, 299 (2010)
12. L.Ermann, A.D.Chepelianskii and D.L.Shepelyansky, *Fractal Weyl law for Linux Kernel Architecture*, Eur. Phys. J. B **79**, 115 (2011)
13. L.Ermann and D.L.Shepelyansky, *Google matrix of the world trade network*, arxiv:1103.5027 (2011)
14. L.Ermann, A.D.Chepelianskii and D.L.Shepelyansky, *Towards two-dimensional search engines*, arxiv:1106.6215[cs.IR] (2011)
15. K.M.Frahm, B.Georgeot and D.L.Shepelyansky, *Universal emergence of PageRank*, arxiv:1105.1062[cs.IR] (2011)
16. L.Ermann and D.L.Shepelyansky, *Ecological analysis of world trade*, arXiv:1201.3584[q-fin.GN] (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).