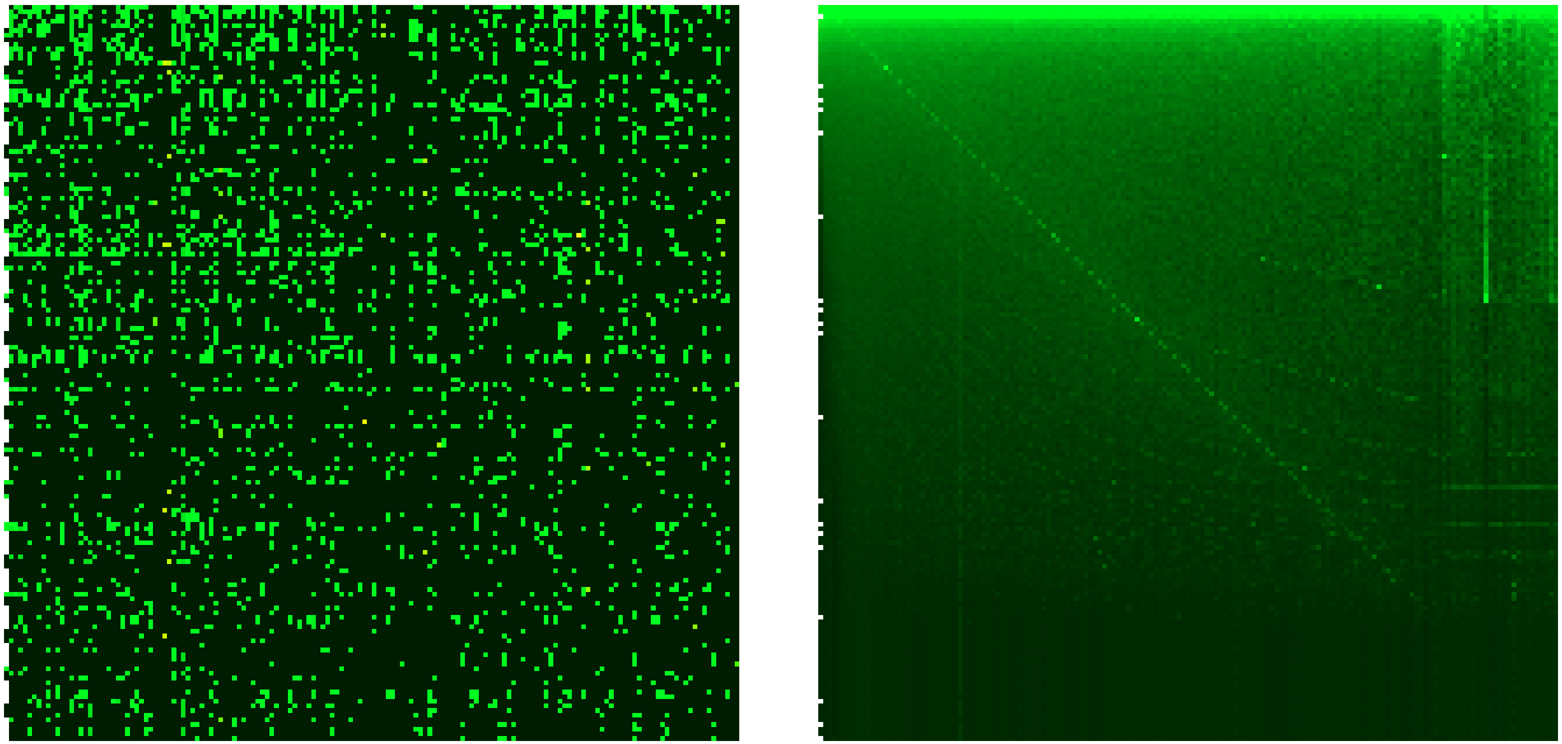
ApliGoogle
Applications of Google matrix to directed networks and Big Data

Colloque de restitution du Défi MASTODONS CNRS 2016



Google matrix G of the English Wikipedia network (Aug. '09) N=3 282 257. Left panel : close up 200x200 first elements.

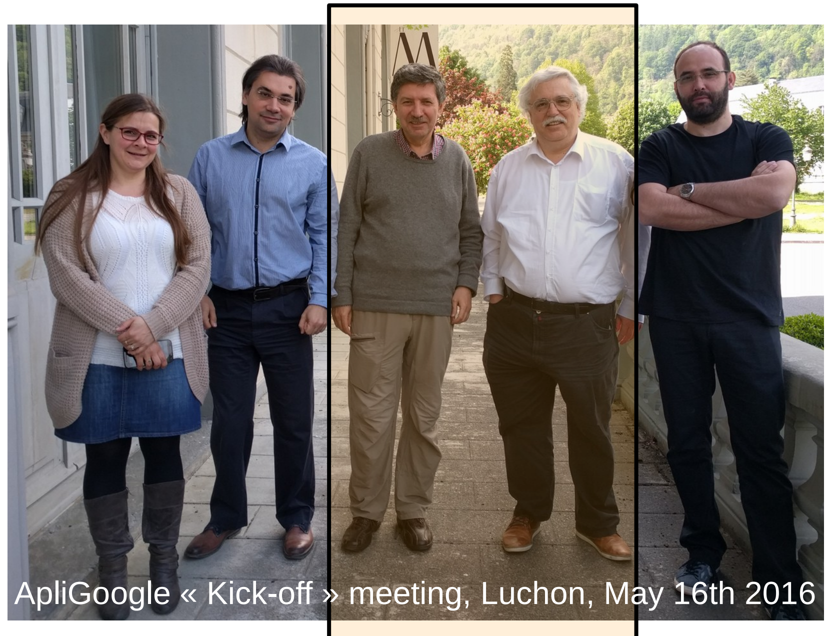
http://www.quantware.ups-tlse.fr/APLIGOOGLE/

ApliGoogle
Applications of Google matrix to directed networks and Big Data

ApliGoogle partners


ApliGoogle « Kick-off » meeting, Luchon, May 16th 2016

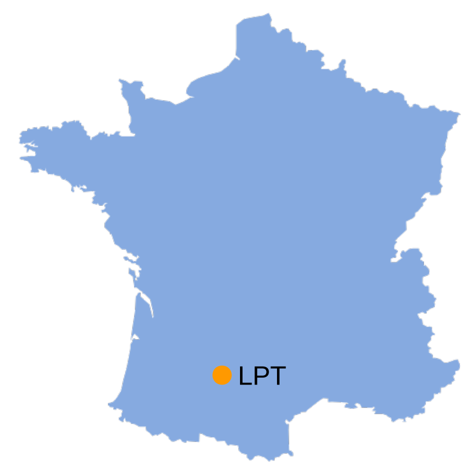
ApliGoogle partners


ApliGoogle « Kick-off » meeting, Luchon, May 16th 2016

Laboratoire de Physique Théorique de Toulouse (UMR CNRS 5152)

- **DR1 D. Shepelyansky (PI)**
- Pr. K. Frahm

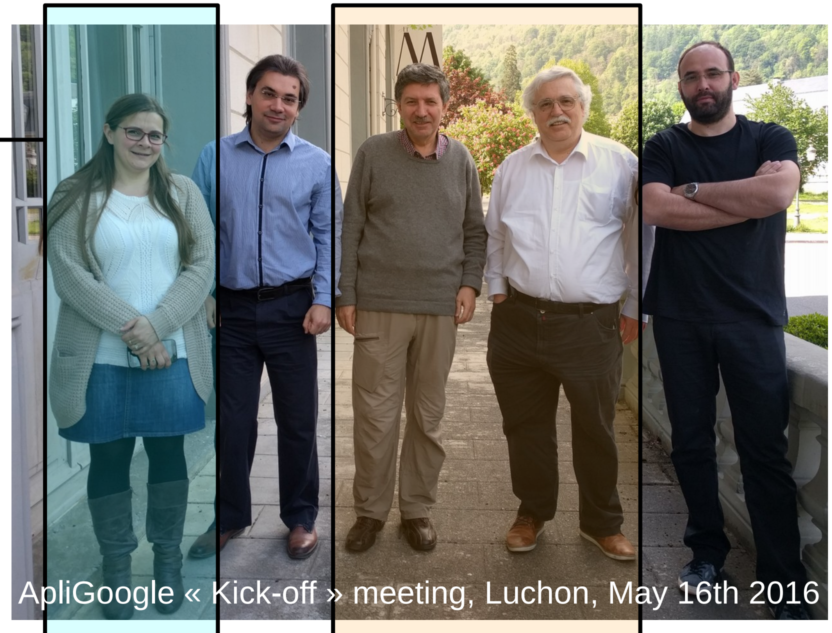
Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks



http://www.quantware.ups-tlse.fr/dima

ApliGoogle
Applications of Google matrix to directed networks and Big Data

ApliGoogle partners



Institut de Recherche en Informatique de Toulouse (UMR CNRS 5152)

- **MCF K. Jaffrès-Runser**
- Samer El Zant (doctorant)
- Dr. T. Peng (post-doctorant)

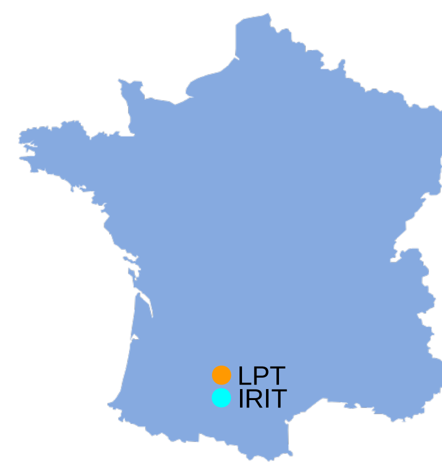
Expertises: Wireless networks

http://www.irit.fr/~Katia.Jaffres

Laboratoire de Physique Théorique de Toulouse (UMR CNRS 5152)

- **DR1 D. Shepelyansky (PI)**
- Pr. K. Frahm

Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks

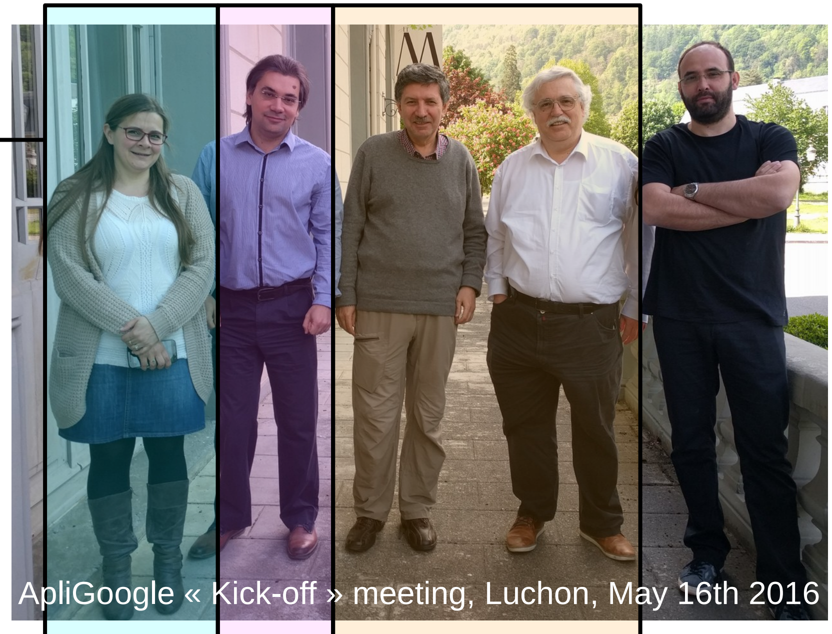


http://www.quantware.ups-tlse.fr/dima

ApliGoogle

Applications of Google matrix to directed networks and Big Data

ApliGoogle partners



Institut de Recherche en Informatique de Toulouse
(UMR CNRS 5152)

- **MCF K. Jaffrès-Runser**
- Samer El Zant (doctorant)
- Dr. T. Peng (post-doctorant)

Expertises: Wireless networks

UNIVERSITÉ TOULOUSE III PAUL SABATIER — cnrs — IRIT CNRS - INPT - UPS - UT1 - UT2J

http://www.irit.fr/~Katia.Jaffres

Institut Curie (INSERM U900)

- **Dr. A. Zinovyev**
- Dr. I. Kuperstein
- Dr. L. Calzone
- U. Czerwinska (doctorante)

Expertises: Computational Biology, Systems Biology of Cancer

institut Curie — PSL RESEARCH UNIVERSITY PARIS — Inserm

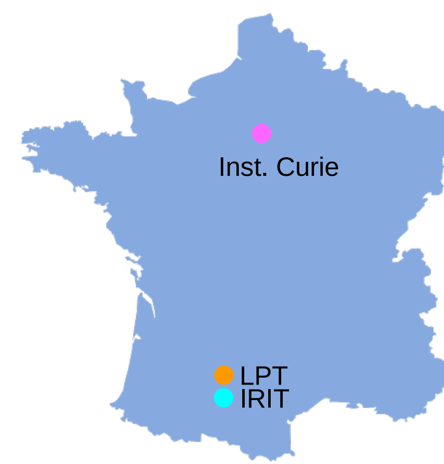
http://www.ihes.fr/~zinovyev/

Laboratoire de Physique Théorique de Toulouse (UMR CNRS 5152)

- **DR1 D. Shepelyansky (PI)**
- Pr. K. Frahm

Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks

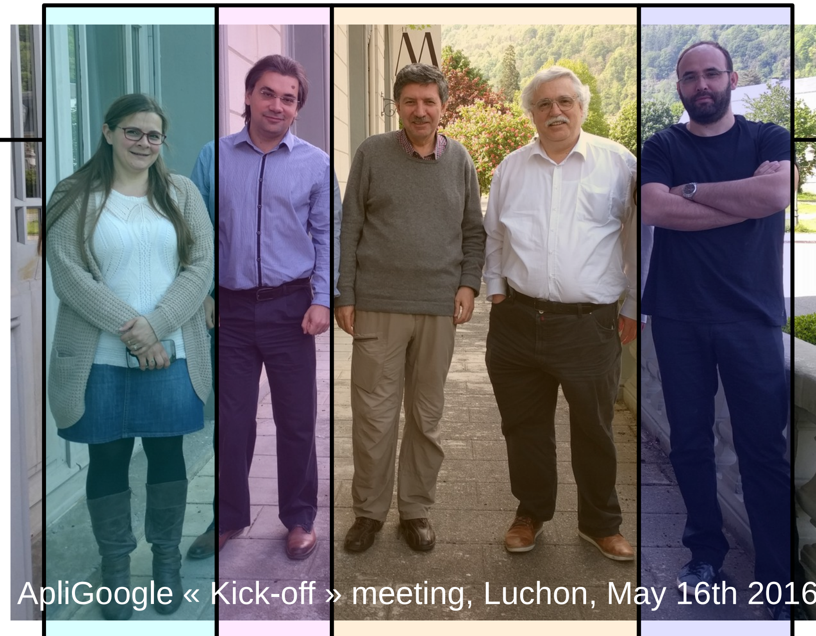
LPT — UNIVERSITÉ TOULOUSE III PAUL SABATIER — cnrs



http://www.quantware.ups-tlse.fr/dima

ApliGoogle
Applications of Google matrix to directed networks and Big Data

ApliGoogle partners



Institut de Recherche en Informatique de Toulouse (UMR CNRS 5152)

- **MCF K. Jaffrès-Runser**
- Samer El Zant (doctorant)
- Dr. T. Peng (post-doctorant)

Expertises: Wireless networks

http://www.irit.fr/~Katia.Jaffres

Institut UTINAM (UMR CNRS 6213)

- **MCF J. Lages**
- S. Diakité (ingénieur)
- F. Gazelle (ingénieur)

Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks

http://perso.utinam.cnrs.fr/~lages

Institut Curie (INSERM U900)

- **Dr. A. Zinovyev**
- Dr. I. Kuperstein
- Dr. L. Calzone
- U. Czerwinska (doctorante)

Expertises: Computational Biology, Systems Biology of Cancer

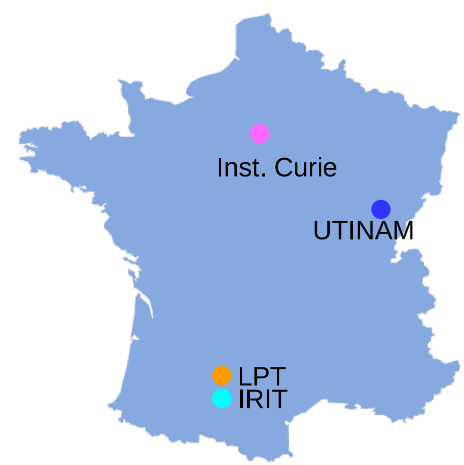
http://www.ihes.fr/~zinovyev/

Laboratoire de Physique Théorique de Toulouse (UMR CNRS 5152)

- **DR1 D. Shepelyansky (PI)**
- Pr. K. Frahm

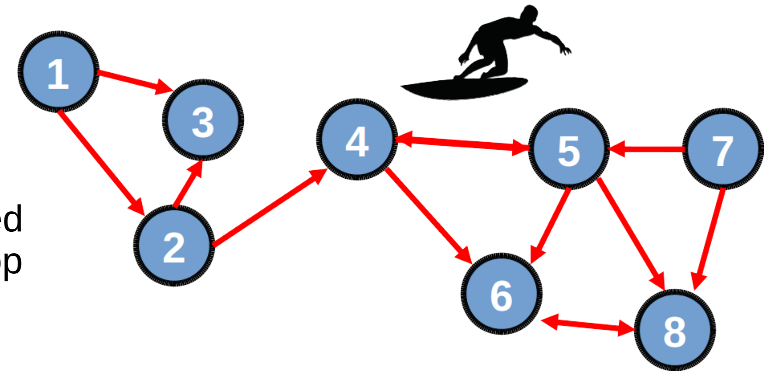
Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks

http://www.quantware.ups-tlse.fr/dima



ApliGoogle
Applications of Google matrix to directed networks and Big Data

How Google works



From Markov (1906) to Brin & Page (1998)

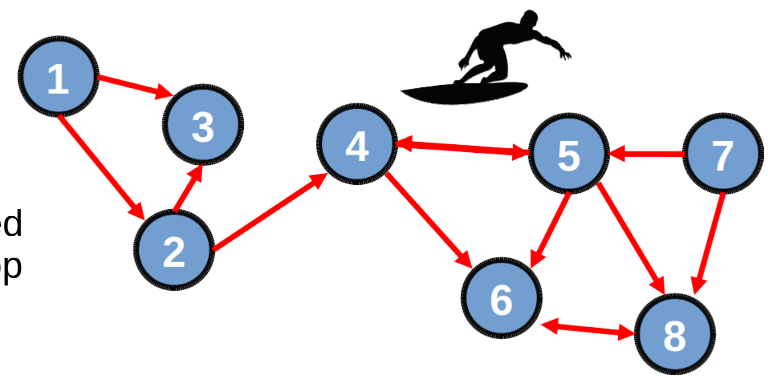
Markovian process : a random surfer probe the structure of a directed network. A each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \rightarrow i \\ 1 & \text{si} \quad j \not\rightarrow i \end{cases}$$

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

How Google works



From Markov (1906) to Brin & Page (1998)

Markovian process : a random surfer probe the structure of a directed network. A each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

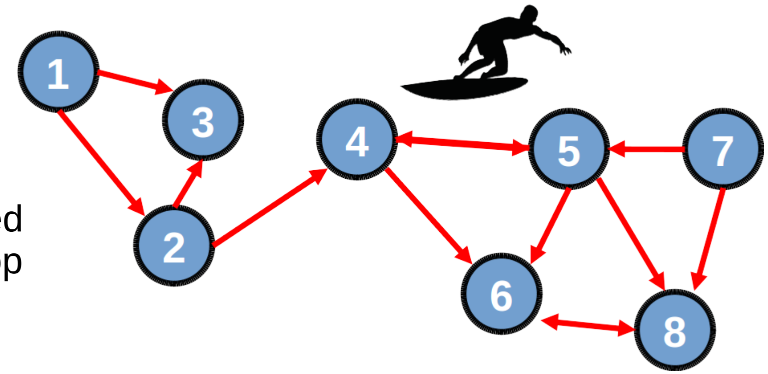
$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \to i \\ 1 & \text{si} \quad j \nrightarrow i \end{cases}$$

Stochastic matrix

$$S_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^{N} A_{kj} & \text{si} \sum_{k=1}^{N} A_{kj} \neq 0 \\ 1/N & \text{sinon} \end{cases}$$

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/8 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/8 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/8 & 1/2 & 1/3 & 0 & 0 & 1 \\ 0 & 0 & 1/8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/8 & 0 & 1/3 & 1 & 1/2 & 0 \end{pmatrix}$$

How Google works



From Markov (1906) to Brin & Page (1998)

Markovian process : a random surfer probe the structure of a directed network. A each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \to i \\ 1 & \text{si} \quad j \nrightarrow i \end{cases}$$

Stochastic matrix

$$S_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^{N} A_{kj} & \text{si} \sum_{k=1}^{N} A_{kj} \neq 0 \\ 1/N & \text{sinon} \end{cases}$$

Google matrix

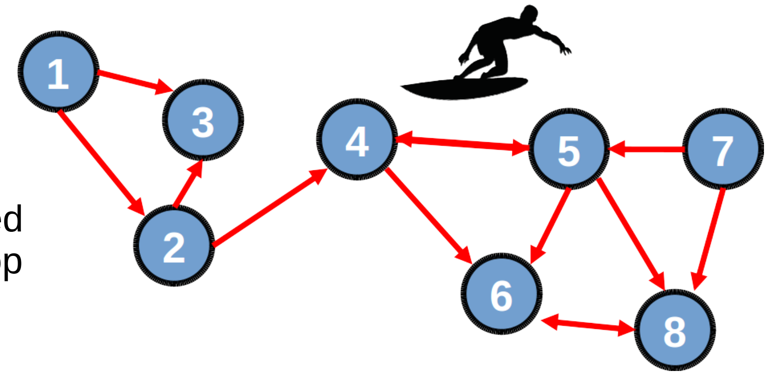
$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N$$
$$\text{avec } 0.5 < \alpha < 1$$

Perron-Frobenius operator

$$\mathbf{G} = \begin{pmatrix} 1/40 & 1/40 & 1/8 & 1/40 & 1/40 & 1/40 & 1/40 & 1/40 \\ 17/40 & 1/40 & 1/8 & 1/40 & 1/40 & 1/40 & 1/40 & 1/40 \\ 17/40 & 17/40 & 1/8 & 1/40 & 1/40 & 1/40 & 1/40 & 1/40 \\ 1/40 & 17/40 & 1/8 & 1/40 & 7/24 & 1/40 & 1/40 & 1/40 \\ 1/40 & 1/40 & 1/8 & 17/40 & 1/40 & 1/40 & 17/40 & 1/40 \\ 1/40 & 1/40 & 1/8 & 17/40 & 7/24 & 1/40 & 1/40 & 33/40 \\ 1/40 & 1/40 & 1/8 & 1/40 & 1/40 & 1/40 & 1/40 & 1/40 \\ 1/40 & 1/40 & 1/8 & 1/40 & 7/24 & 33/40 & 17/40 & 1/40 \end{pmatrix}$$

$\alpha = 0.8$

How Google works



From Markov (1906) to Brin & Page (1998)

Markovian process : a random surfer probe the structure of a directed network. A each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \to i \\ 1 & \text{si} \quad j \not\to i \end{cases}$$

Stochastic matrix

$$S_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^{N} A_{kj} & \text{si} \sum_{k=1}^{N} A_{kj} \neq 0 \\ 1/N & \text{sinon} \end{cases}$$

Google matrix

$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N$$

$$\text{avec } 0.5 < \alpha < 1$$

Perron-Frobenius operator

PageRank vector

$$\mathbf{P} = \lim_{n \to \infty} \mathbf{P}^{(n)} = \lim_{n \to \infty} G^n \mathbf{P}^{(0)}$$

$P_i^{(n)}$ is the probability that random surfer arrives at node i at the nth step.

$\mathbf{P}$ is the $\mathbf{G}$ matrix eigenvector associated with eigenvalue 1

$$\mathbf{P} = \mathbf{GP}$$

$$\mathbf{P} = \begin{pmatrix} 0.031094525687305 97 \\ 0.04353233614756617 \\ 0.06094527086606558 \\ 0.06729412361797826 \\ 0.07044998599586171 \\ \color{red}{0.3518167935609448 9} \\ 0.03109452568730597 \\ 0.3437724384369714 3 \end{pmatrix}$$

Distribution $P(K)$ where K is the rank index:

$P(1) = \color{red}{0.35181679356094489}$

$P(2) = 0.34377243843697143$

$P(3) = 0.07044998599586171$

$P(4) = 0.06729412361797826$

$P(5) = 0.06094527086606558$

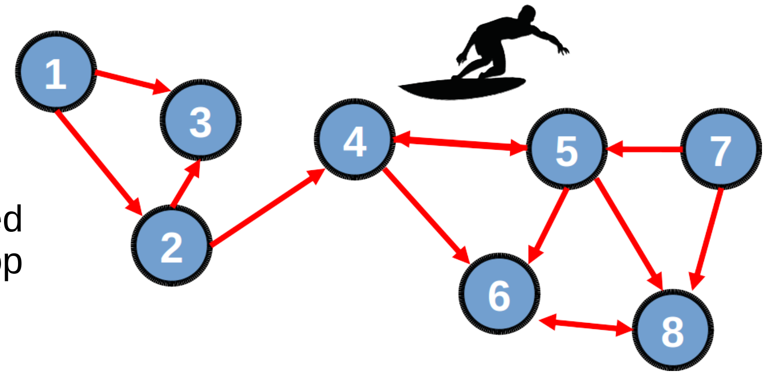
$P(6) = 0.04353233614756617$

$P(7) = P(8) = 0.03109452568730597$

ApliGoogle

Applications of Google matrix to directed networks and Big Data

How Google works



From Markov (1906) to Brin & Page (1998)

Markovian process : a random surfer probe the structure of a directed network. At each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \to i \\ 1 & \text{si} \quad j \not\to i \end{cases}$$

Stochastic matrix

$$S_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^{N} A_{kj} & \text{si} \sum_{k=1}^{N} A_{kj} \neq 0 \\ 1/N & \text{sinon} \end{cases}$$

Google matrix

$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N$$

$$\text{avec } 0.5 < \alpha < 1$$

Perron-Frobenius operator

PageRank vector

$$\mathbf{P} = \lim_{n\to\infty} \mathbf{P}^{(n)} = \lim_{n\to\infty} G^n \mathbf{P}^{(0)}$$

$P_i^{(n)}$ is the probability that random surfer arrives at node i at the nth step.

$\mathbf{P}$ is the $\mathbf{G}$ matrix eigenvector associated with eigenvalue 1

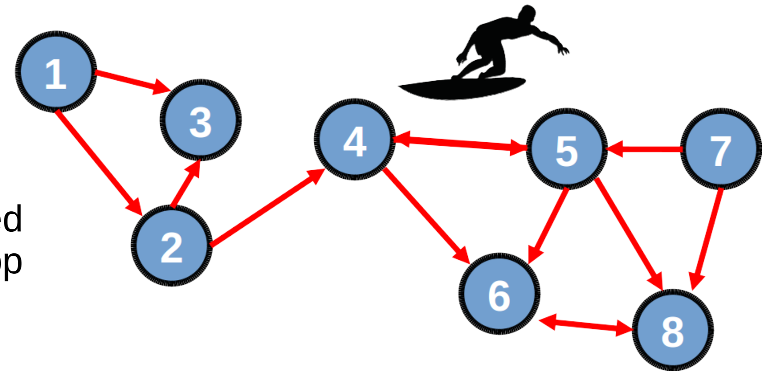
$$\mathbf{P} = \mathbf{G}\mathbf{P}$$

The most important node is the one with the highest probability.
"Recursive definition": the more a node is pointed by important nodes, the more it is important.

PageRank measures the influence of a node.

PageRank is (was?) at the heart of Google search engine (Brin, Page '98).

How Google works



From Markov (1906) to Brin & Page (1998)

Markovian process : a random surfer probe the structure of a directed network. At each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix | **Stochastic matrix** | **Google matrix**

$$A_{ij} = \begin{cases} 0 & \text{si} \quad j \to i \\ 1 & \text{si} \quad j \nrightarrow i \end{cases} \qquad S_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^{N} A_{kj} & \text{si} \sum_{k=1}^{N} A_{kj} \neq 0 \\ 1/N & \text{sinon} \end{cases}$$

$$G_{ij} = \alpha S_{ij} + (1-\alpha)/N$$
$$\text{avec } 0.5 < \alpha < 1$$

Perron-Frobenius operator

PageRank vector

$$\mathbf{P} = \lim_{n\to\infty} \mathbf{P}^{(n)} = \lim_{n\to\infty} G^n \mathbf{P}^{(0)}$$

$P_i^{(n)}$ is the probability that random surfer arrives at node i at the nth step.

$\mathbf{P}$ is the $\mathbf{G}$ matrix eigenvector associated with eigenvalue 1

$$\mathbf{P} = \mathbf{G}\mathbf{P}$$

The most important node is the one with the highest probability.
“Recursive definition”: the more a node is pointed by important nodes, the more it is important.

PageRank measures the influence of a node.

PageRank is (was?) at the heart of Google search engine (Brin, Page ‘98).

CheiRank vector $\mathbf{P}^* = \mathbf{G}^*\mathbf{P}^*$

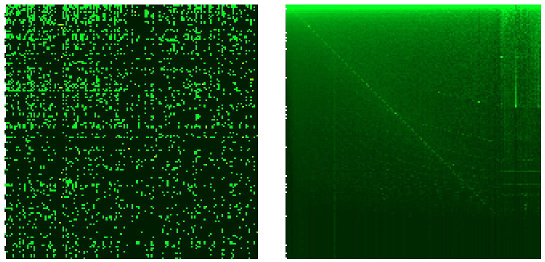
Similar to PageRank of the inverted network. With inverted adjacency matrix elements $A^*_{ij} = A_{ji}$, it is possible to define the stochastic matrix elements $S^*_{ij} \neq S_{ji}$, and the Google matrix elements $G^*_{ij} \neq G_{ji}$ associated to the inverted network (Fogaras ‘03, Chepelianskii ‘10).

“Recursive definition”: the more a node pointed toward important nodes, the more it is important.

CheiRank measures the diffusion/the communication of a node.

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Real directed complex networks



Google matrix G of the English Wikipedia network (Aug. '09)

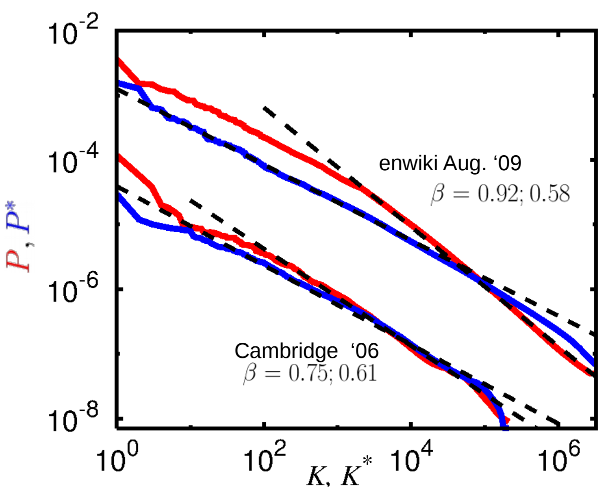
N=3 282 257

Properties of real networks (WWW, social networks, ...)

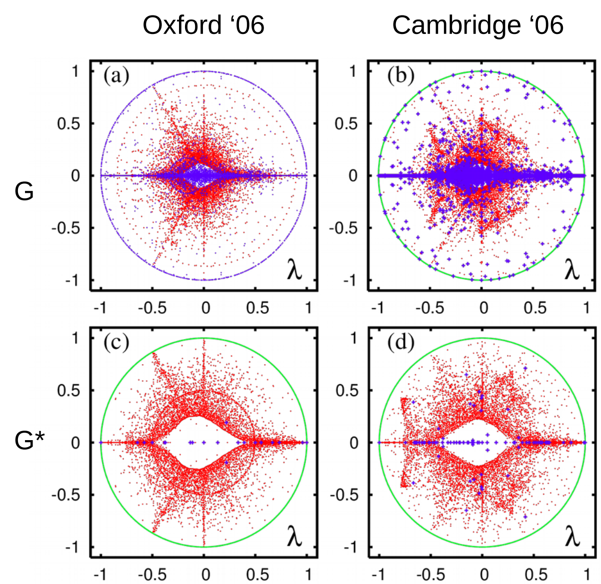
- **"Small world" property:** average distance between two nodes $\sim \log N$
- **Scale-free property:** distribution of the number of ingoing or outgoing links is $\rho(k) \sim k^{-\nu}$

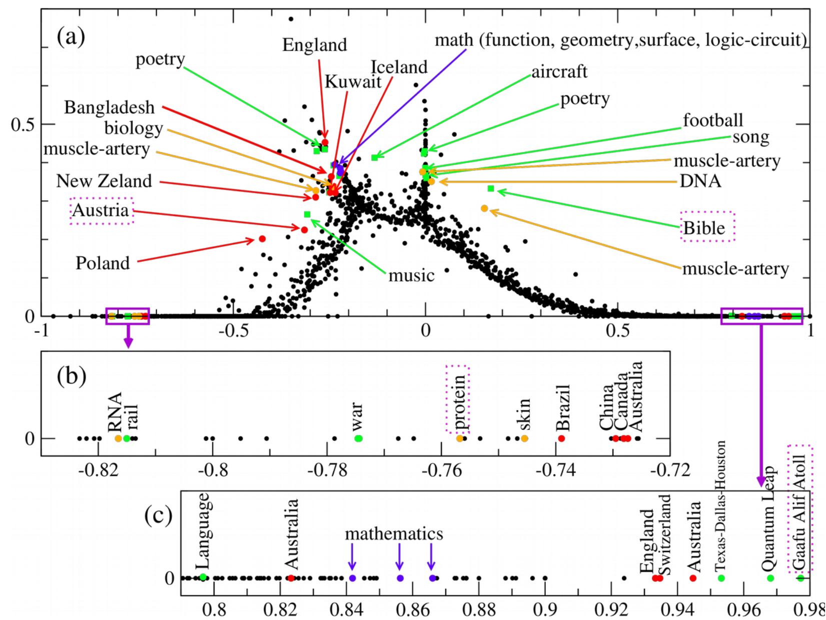
PageRank distribution properties

$$P(K) \sim 1/K^{\beta} \text{ with } \nu_{\text{in}} = 1 + 1/\beta$$
$$P^*(K^*) \sim 1/K^{*\beta} \text{ with } \nu_{\text{out}} = 1 + 1/\beta$$



Google matrix spectrum





Wikipedia english '13 network

Theory of real directed complex networks

Random Matrix Theory introduced by Wigner '67 describes universal spectral properties shared by complex nuclei/atoms/molecules and also mesoscopic and quantum chaos systems (Hermitean and unitary matrices statistical ensembles).

Challenge : A Random Matrix Theory for Markov chains and Google matrix ensembles is still lacking → **We need more examples / more applications**

Figures from :
Ermann, Frahm, Shepelyansky (2016), Scholarpedia, 11(11):30944

Ermann, Frahm, Shepelyansky (2015), Rev. Mod. Phys. 87, 1261.

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Big Data seen as directed networks

Non exhaustive list of applications

Data	Nodes	Links between nodes
WWW	Web pages	Hyperliens
Wikipedia	Wiki articles	Citations intra-wiki
Twitter / social networks	Members	Follow relations
World Trade (from WTO, UN, OECD, ...)	Goods x countries	Economical balance between countries
Omics	Proteins	Inhibition/activation
Linux	Kernel commands	Command succession
DNA	Pattern	Pattern successions
Brain activity	Neurons	Synaptic connections
Go game	Played patterns	Pattern sucessions

For a review, see Ermann, Frahm, Shepelyansky (2015), Rev. Mod. Phys. 87, 1261.

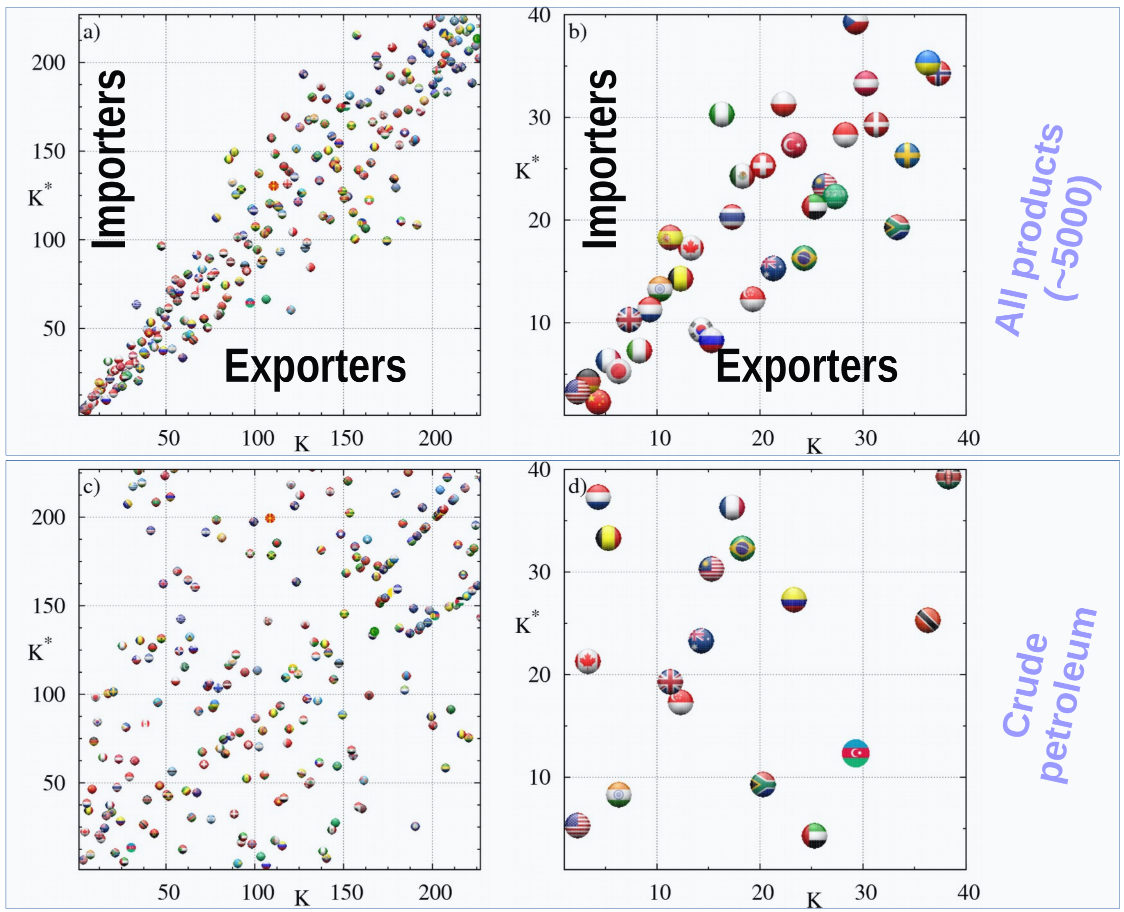
World Trade Google matrix

The analysis of World Trade Google matrix allows to treat every country, poor or rich, in **an equal footing**. Indeed, for a given country, imported quantities from or exported quantities toward another country are normalized by the total imported or exported (different point of view in comparison with usual import-export classification).

PageRank naturally characterizes the capacity of a country to **import**.

CheiRank naturally characterizes the capacity of a country to **export**.

75% overlap between **top 20** countries in K-K* plane and G20 members

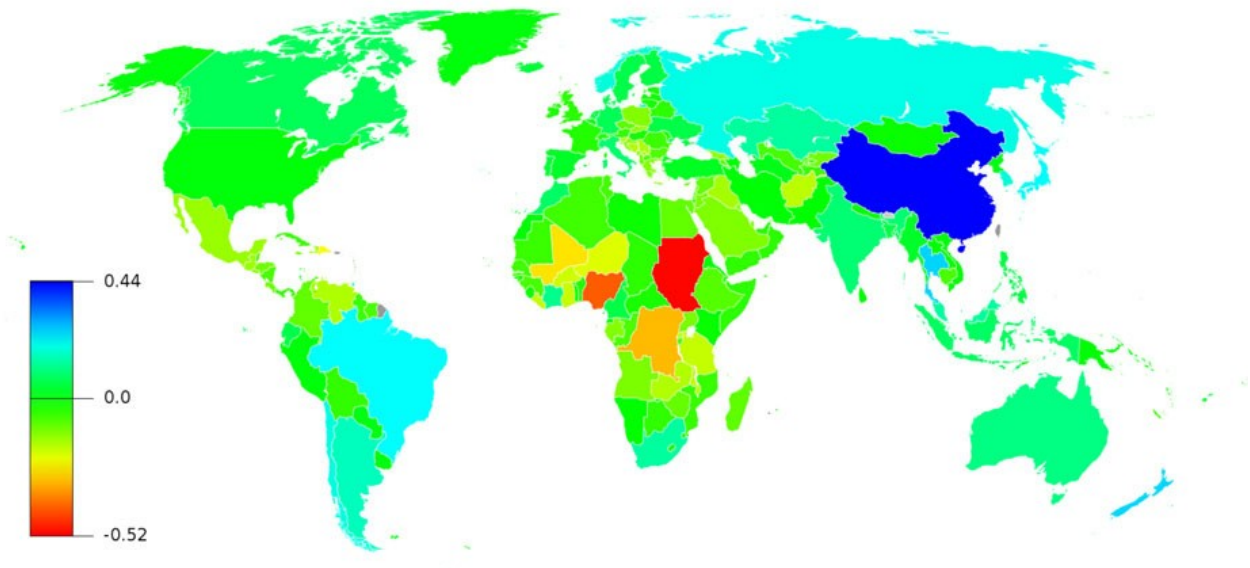


Data from UN-COMTRADE
227 countries
~5000 products

Ermann, Shepelyansky '11

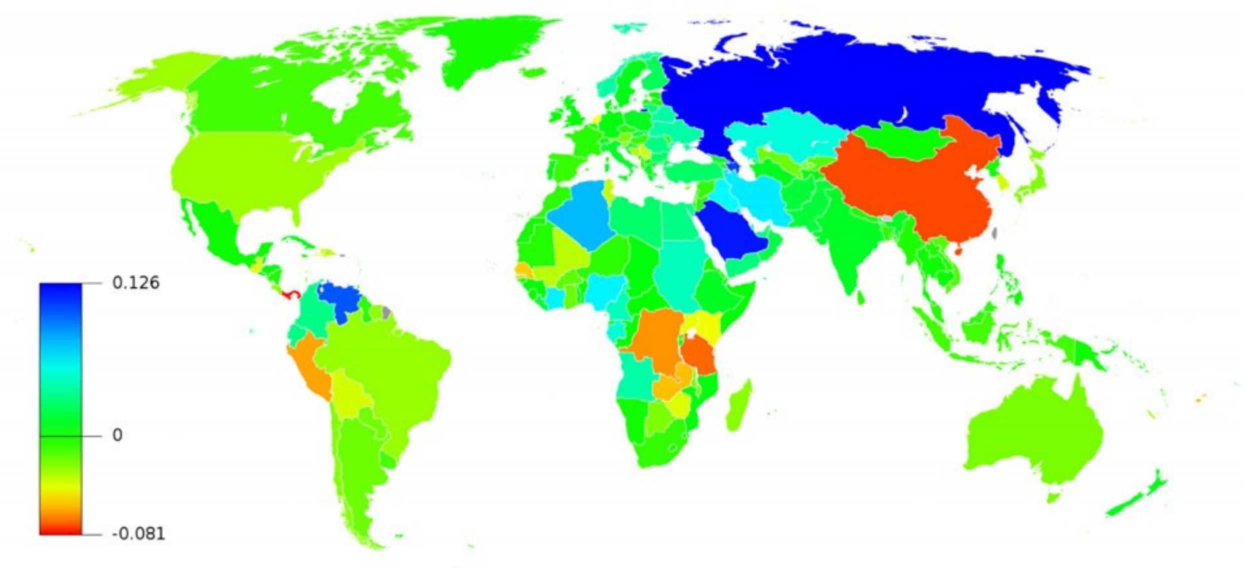
World Trade Google matrix

PageRank-CheiRank balance



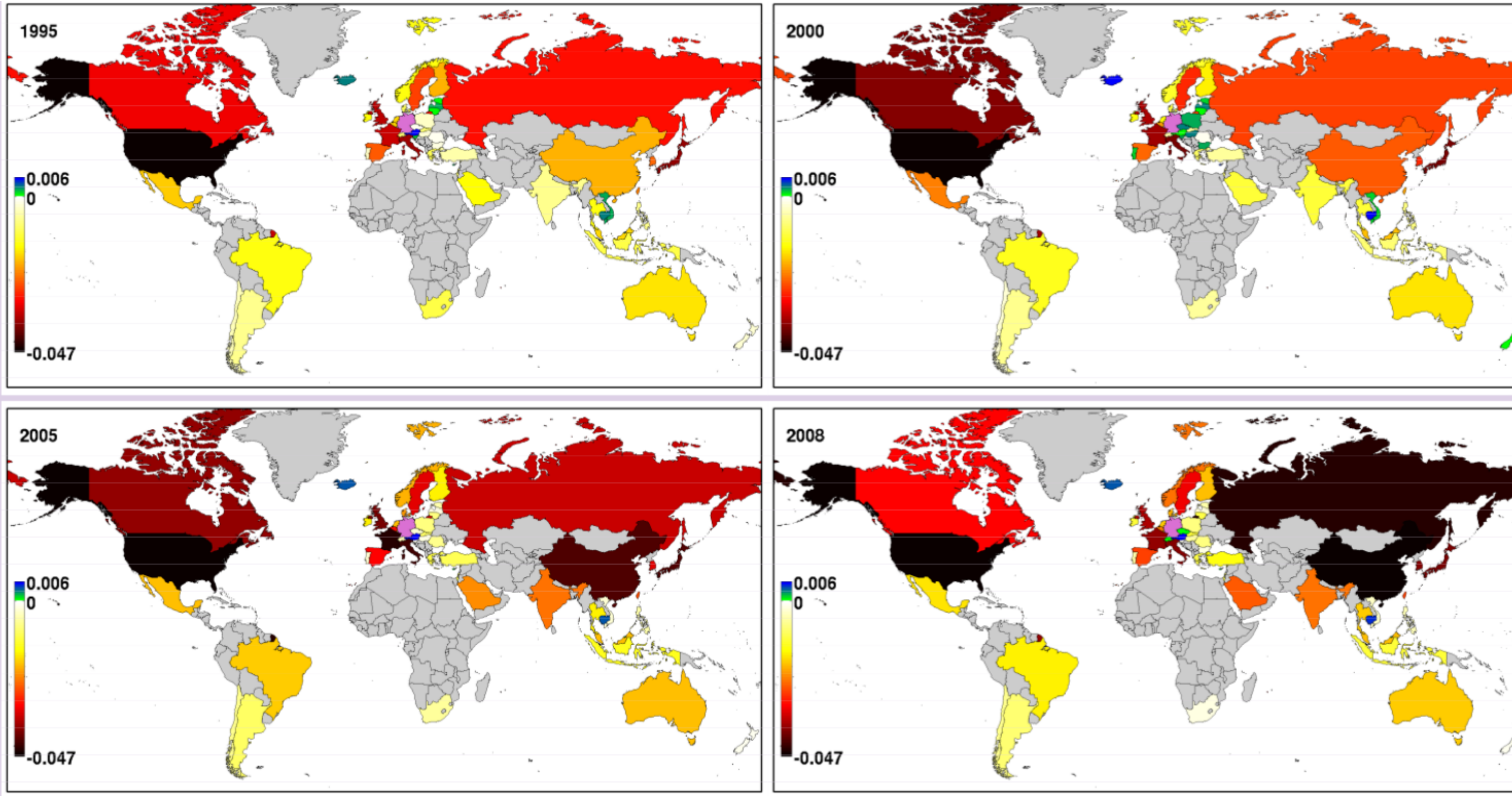
Sensitivity of the PageRank-CheiRank balance to the increase petroleum cost

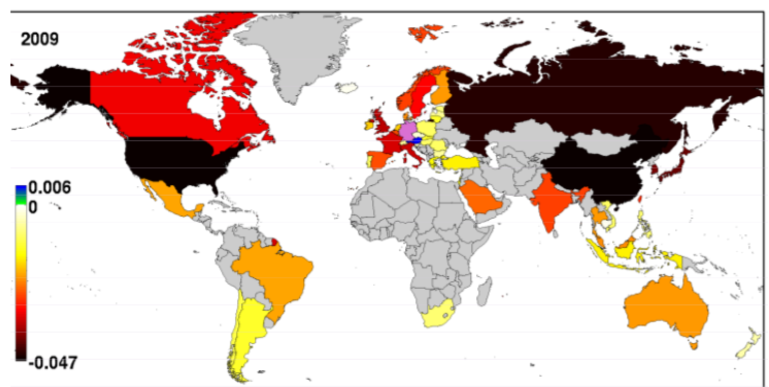
Data from UN '14



World Trade Google matrix

Sensitivity over the years of the PageRank-CheiRank balance to the increase of labor cost in Germany





Study of crisis
Recomendations
Risks

Data from WTO
58 countries
37 activity sectors

Collaboration with H. Escaith from WTO Geneva

Kandiah, Escaith, Shepelyansky ‘15

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Highlight of some results from ApliGoogle project

WRWU: Wikipedia Ranking of World Universities

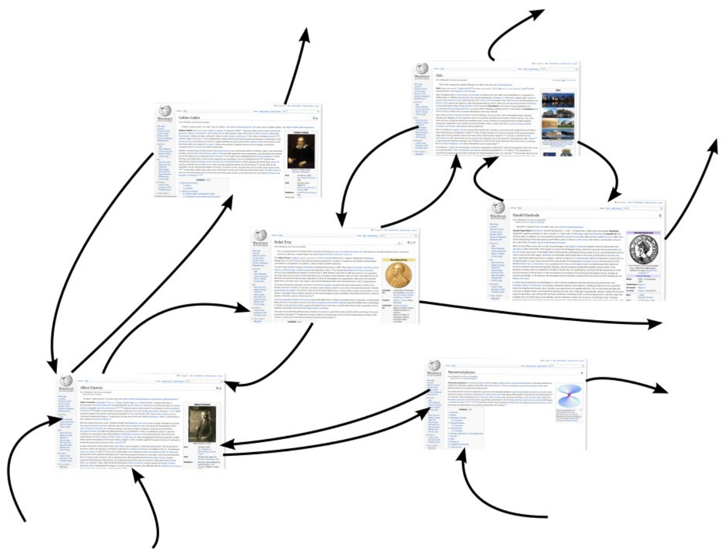
24 Wikipedia language editions
covering 59% of world population
and 68% total Wikipages

Edition	Language	N	Edition	Language	N
EN	English	4 212 493	VI	Vietnamese	594 089
DE	German	1 532 978	FA	Persian	295 696
FR	French	1 352 825	HU	Hungarian	235 212
NL	Dutch	1 144 615	KO	Korean	231 959
IT	Italian	1 017 953	TR	Turkish	206 311
ES	Spanish	974 025	AR	Arabic	203 328
RU	Russian	966 284	MS	Malaysian	180 886
PL	Polish	949 153	DA	Danish	175 228
JA	Japanese	852 087	HE	Hebrew	144 959
SV	Swedish	780 872	HI	Hindi	96 869
PT	Portuguese	758 227	EL	Greek	82 563
ZH	Chinese	663 485	TH	Thai	78 953

About 17M wikipages considered
(March '13)

WIKIPEDIA
The Free Encyclopedia

Each Wikipedia edition is treated as
a complex network



Rankings of World Universities

ACADEMIC RANKING OF WORLD UNIVERSITIES SINCE 2003

THE WORLD UNIVERSITY RANKINGS

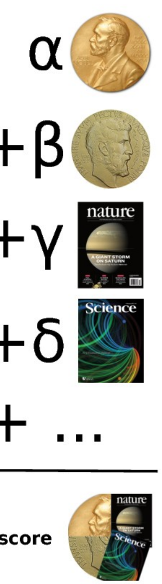
QS WORLD UNIVERSITY RANKINGS

•
•
•

About **20 different** global university rankings are listed in the Wikipedia page "College and university rankings"

All these rankings are composite:

α + β + γ + δ + …



Composite score

Also, universities are preselected

These rankings have an impact on scientific and educational policies of governments

Is there an universal ranking without *a priori* criteria and without cultural bias ?

ApliGoogle
Applications of Google matrix to directed networks and Big Data

WRWU: Wikipedia Ranking of World Universities

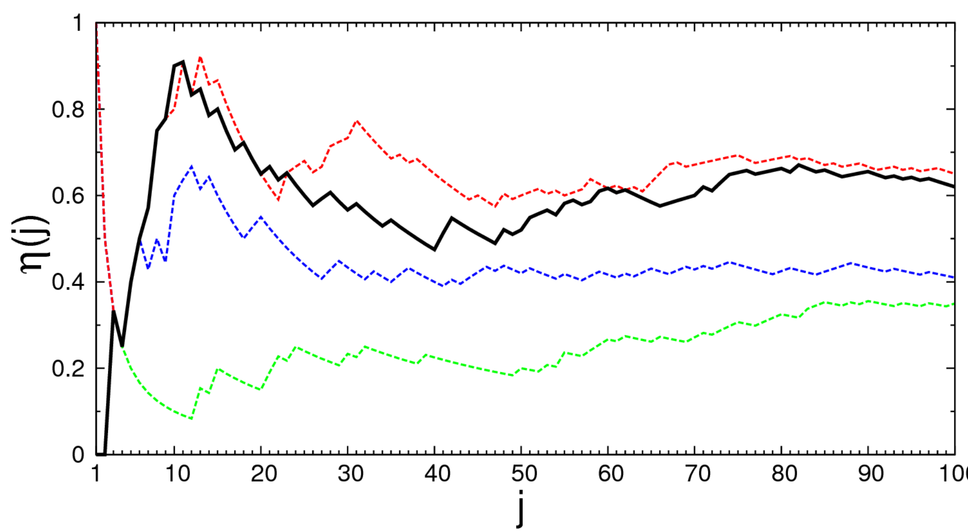
1024 ranked Universities

Wikipedia PageRanking of World Universities WRWU			Academic Ranking of World Universities ARWU ("Shanghai ranking" 2013)	
1st	University of Cambridge		1st	Harvard University (-2)
2nd	University of Oxford		2nd	Stanford University (-6)
3rd	Harvard University		3rd	University of California, Berkeley (-7)
4th	Columbia University		4th	MIT (-2)
5th	Princeton University		5th	University of Cambridge (+4)
6th	MIT		6th	California Institute of Technology (-22)
7th	University of Chicago		7th	Princeton University (+2)
8th	Stanford University		8th	Columbia University (+4)
9th	Yale University		9th	University of Chicago (+2)
10th	University of California, Berkeley		10th	University of Oxford (+8)

90% overlap between top 10s WRWU and ARWU

60% overlap between top 100s WRWU and ARWU

Oxbridge at the top of WRWU followed by **US major universities**



Definitely, as ARWU, WRWU measures academic excellence but not only ...

J. Lages, A. Patt, D. L. Shepelyansky, The European Physical Journal B (2016) 89: 69

ApliGoogle
Applications of Google matrix to directed networks and Big Data

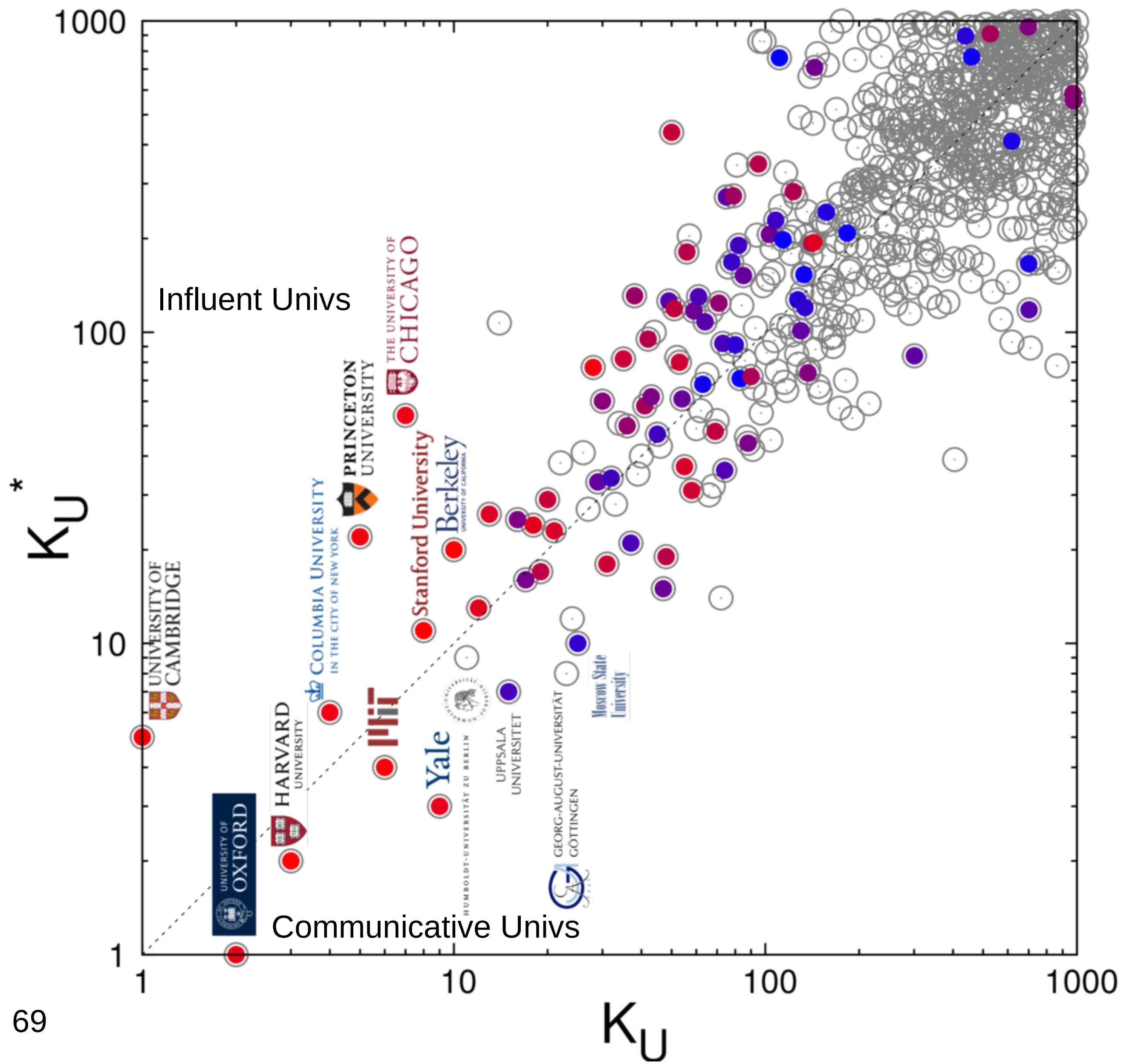
WRWU: Wikipedia Ranking of World Universities

... WRWU measures also historical, societal and regional importance of universities.

WRWU is free from any cultural preferences since :
- it takes into account many cultural point of view as we use encyclopaedic knowledge contained in 24 Wikipedia langage editions (17 millions wiki articles)
- these points of view are treated on equal footing with the same statistical analysis (PageRank, CheiRank, ...)

WRWU can be considered as **complementary** to already existing rankins such as ARWU (Shanghai), THE, ..., but in fact **it encodes already all existing rankings** since Wikipedia contains information on it.

Universal ranking ?



J. Lages, A. Patt, D. L. Shepelyansky
The European Physical Journal B (2016) 89: 69

ApliGoogle
Applications of Google matrix to directed networks and Big Data

WRWU: Wikipedia Ranking of World Universities

Web page of WRWU with detailed data (ranking by country, by foundation century, …)
Is your Univ. well ranked ? Check at : http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/

In the media (100 press articles in 22 countries) : http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/press/Press.html

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Reduced Google matrix

Consider a network with $N \gg 1$ nodes.
Consider a sub-network (a community) of $N_r \ll N$ nodes. Google matrix of the N size network and the associated PageRank vector can be written

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

$$\mathbf{GP} = \mathbf{P}$$

We define the reduced Google matrix $\mathbf{G}_R$ associated to the size N_r community such as

$$\mathbf{G}_R \mathbf{P}_r = \mathbf{P}_r$$

The reduced Google matrix can be written

$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{rs} \left(\mathbf{1} - \mathbf{G}_{ss}\right)^{-1} \mathbf{G}_{sr}$$

Contribution from direct links ($\mathbf{G}_{rr}$)

Contribution from indirect links (scattering term) ($\mathbf{G}_{rs} \left(\mathbf{1} - \mathbf{G}_{ss}\right)^{-1} \mathbf{G}_{sr}$)

$$\left(\mathbf{1} - \mathbf{G}_{ss}\right)^{-1} = \sum_{l=0}^{\infty} \mathbf{G}_{ss}^{l}$$

Very slow convergence since the eigenvalue λ_c of $\mathbf{G}_{ss} \sim \mathbf{G}$ is very close to 1.

K. M. Frahm, and D. L. Shepelyansky,
arXiv:1602.02394 [physics.soc-ph]

Reduced Google matrix

Consider a network with $N \gg 1$ nodes.
Consider a sub-network (a community) of $N_r \ll N$ nodes. Google matrix of the N size network and the associated PageRank vector can be written

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

$$\mathbf{G}\mathbf{P} = \mathbf{P}$$

We define the reduced Google matrix $\mathbf{G}_R$ associated to the size N_r community such as

$$\mathbf{G}_R \mathbf{P}_r = \mathbf{P}_r$$

The reduced Google matrix can be written

$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

Contribution from direct links

Contribution from hidden links

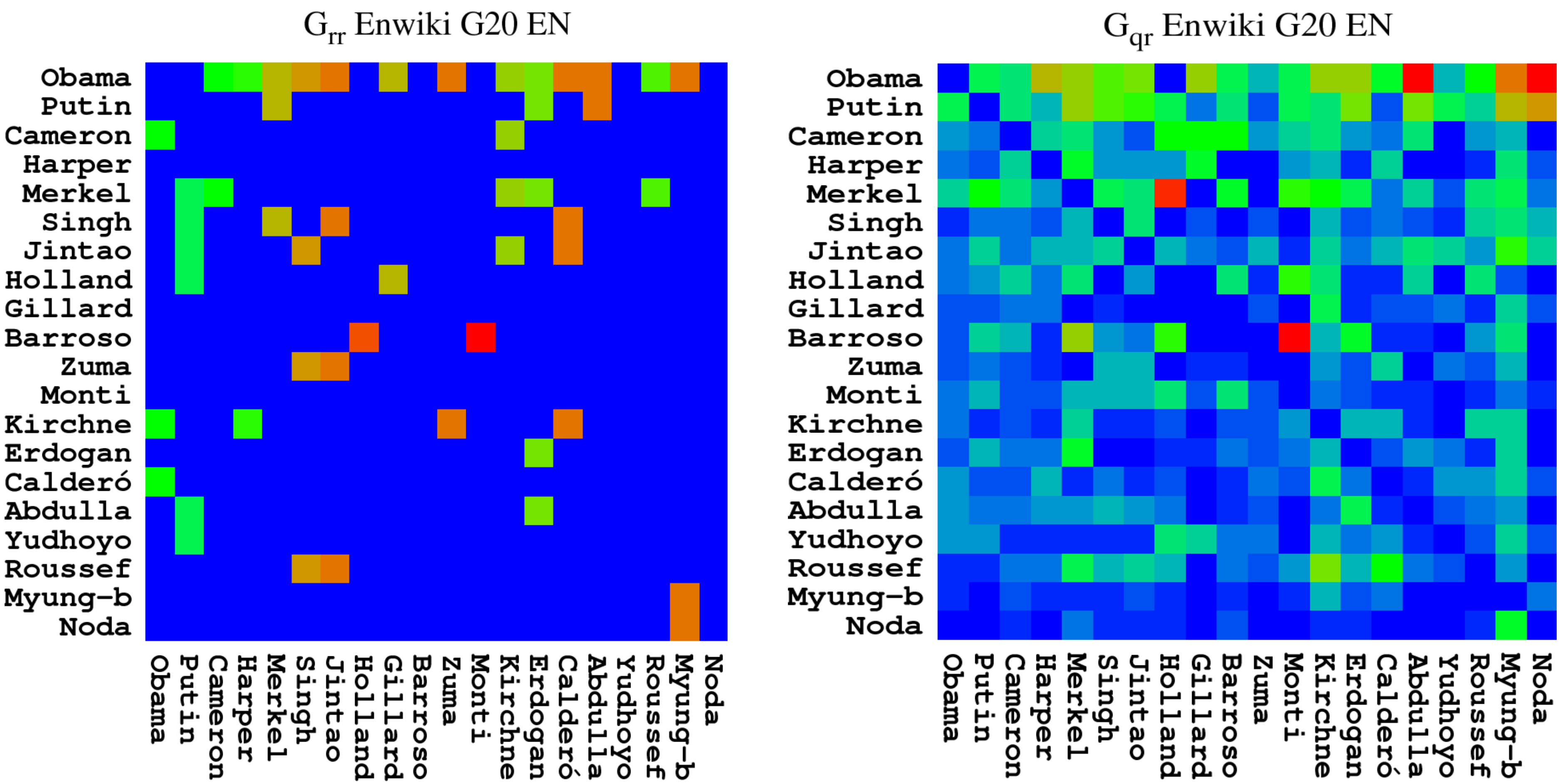
Contribution from « PageRank »

ApliGoogle
Applications of Google matrix to directed networks and Big Data

K. M. Frahm, and D. L. Shepelyansky,
arXiv:1602.02394 [physics.soc-ph]

Wikipedia mining of hidden links between political leaders

2013 Wikipedia edition



$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

direct links **hidden links**

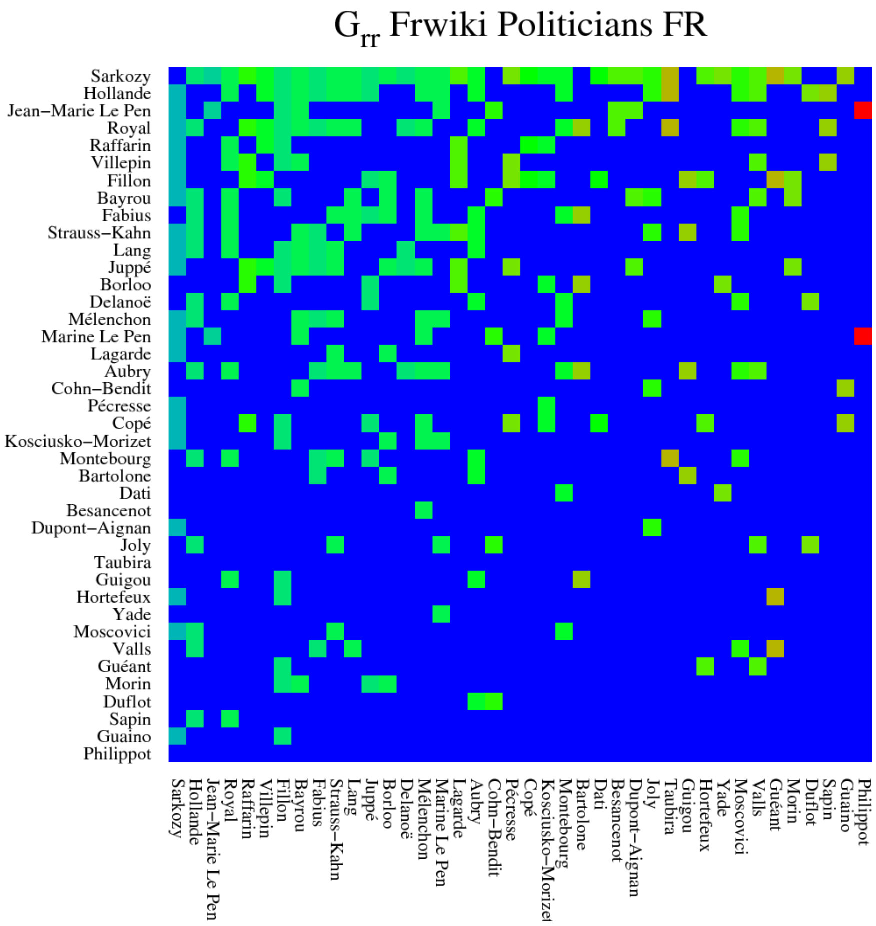
ApliGoogle

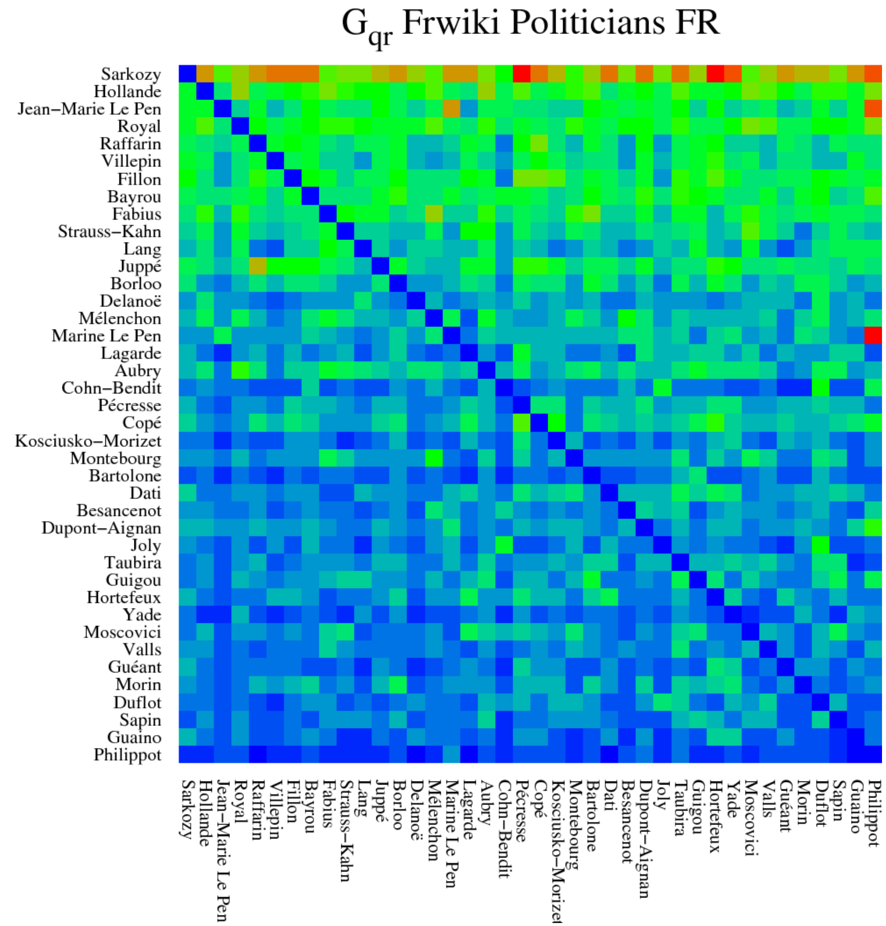
Applications of Google matrix to directed networks and Big Data

Frahm, Jaffrès-Runser, Shepelyansky, Eur. Phys. J. B (2016) 89: 269

Wikipedia mining of hidden links between political leaders

2013 Wikipedia edition





$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

direct links **hidden links**

Frahm, Jaffrès-Runser, Shepelyansky, Eur. Phys. J. B (2016) 89: 269

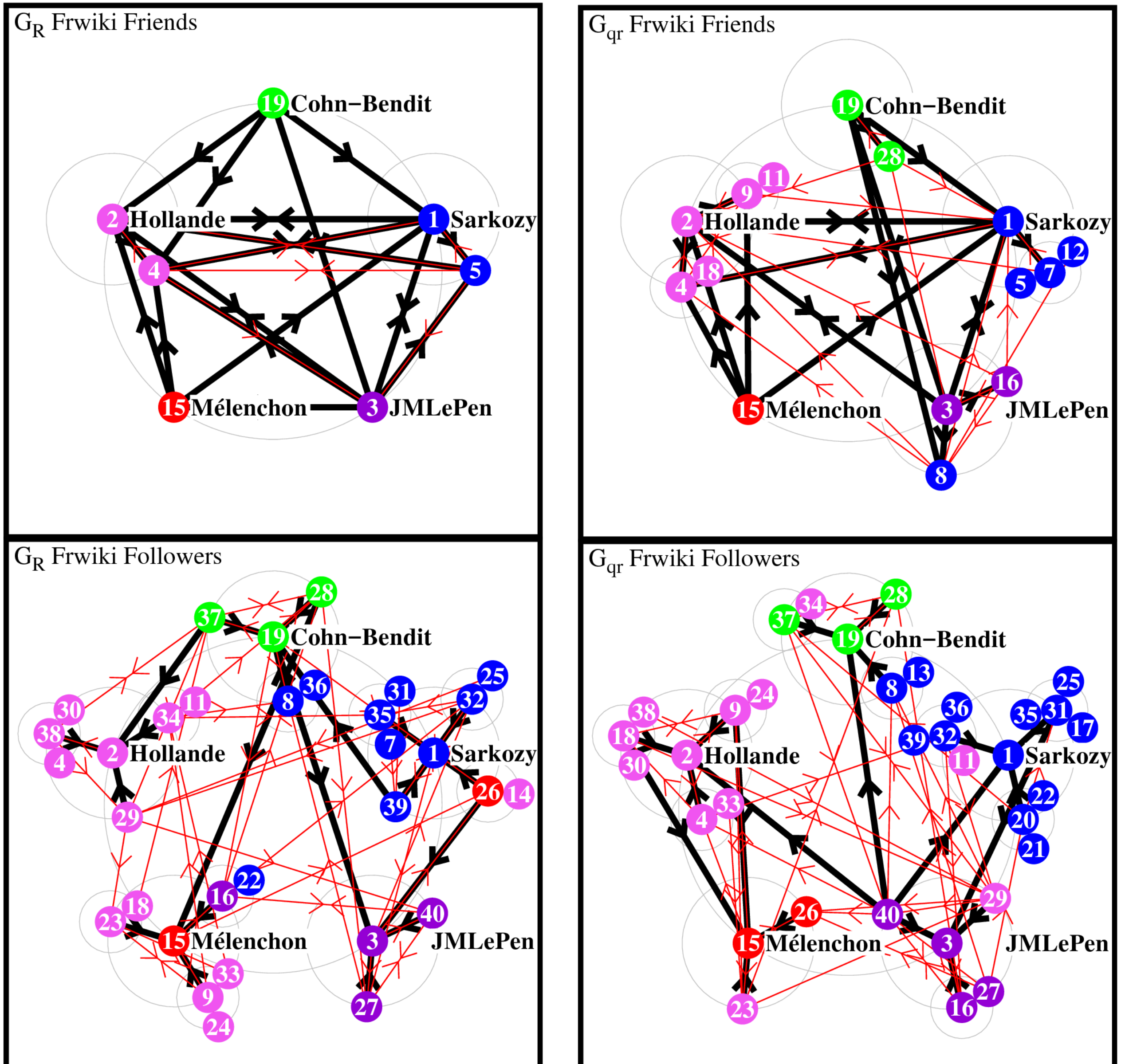
ApliGoogle

Applications of Google matrix to directed networks and Big Data

Wikipedia mining of hidden links between political leaders

Circles of influence

2013 Wikipedia edition



Names (FR)	K
Nicolas Sarkozy CB	1
François Hollande CM	2
Jean-Marie Le Pen CV	3
Ségolène Royal CM	4
Jean-Pierre Raffarin CB	5
Dominique de Villepin CB	6
François Fillon CB	7
François Bayrou CB	8
Laurent Fabius CM	9
Dominique Strauss-Kahn CM	10
Jack Lang CM	11
Alain Juppé CB	12
Jean-Louis Borloo CB	13
Bertrand Delanoë CM	14
Jean-Luc Mélenchon CR	15
Marine Le Pen CV	16
Christine Lagarde CB	17
Martine Aubry CM	18
Daniel Cohn-Bendit CG	19
Valérie Pécresse CB	20
Jean-François Copé CB	21
Nathalie Kosciusko-Morizet CB	22
Arnaud Montebourg CM	23
Claude Bartolone CM	24
Rachida Dati CB	25
Olivier Besancenot CR	26
Nicolas Dupont-Aignan CV	27
Eva Joly CG	28
Christiane Taubira CM	29
Élisabeth Guigou CM	30
Brice Hortefeux CB	31
Rama Yade CB	32
Pierre Moscovici CM	33
Manuel Valls CM	34
Claude Guéant CB	35
Hervé Morin CB	36
Cécile Duflot CG	37
Michel Sapin CM	38
Henri Guaino CB	39
Florian Philippot CV	40

$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

direct links **hidden links**

ApliGoogle

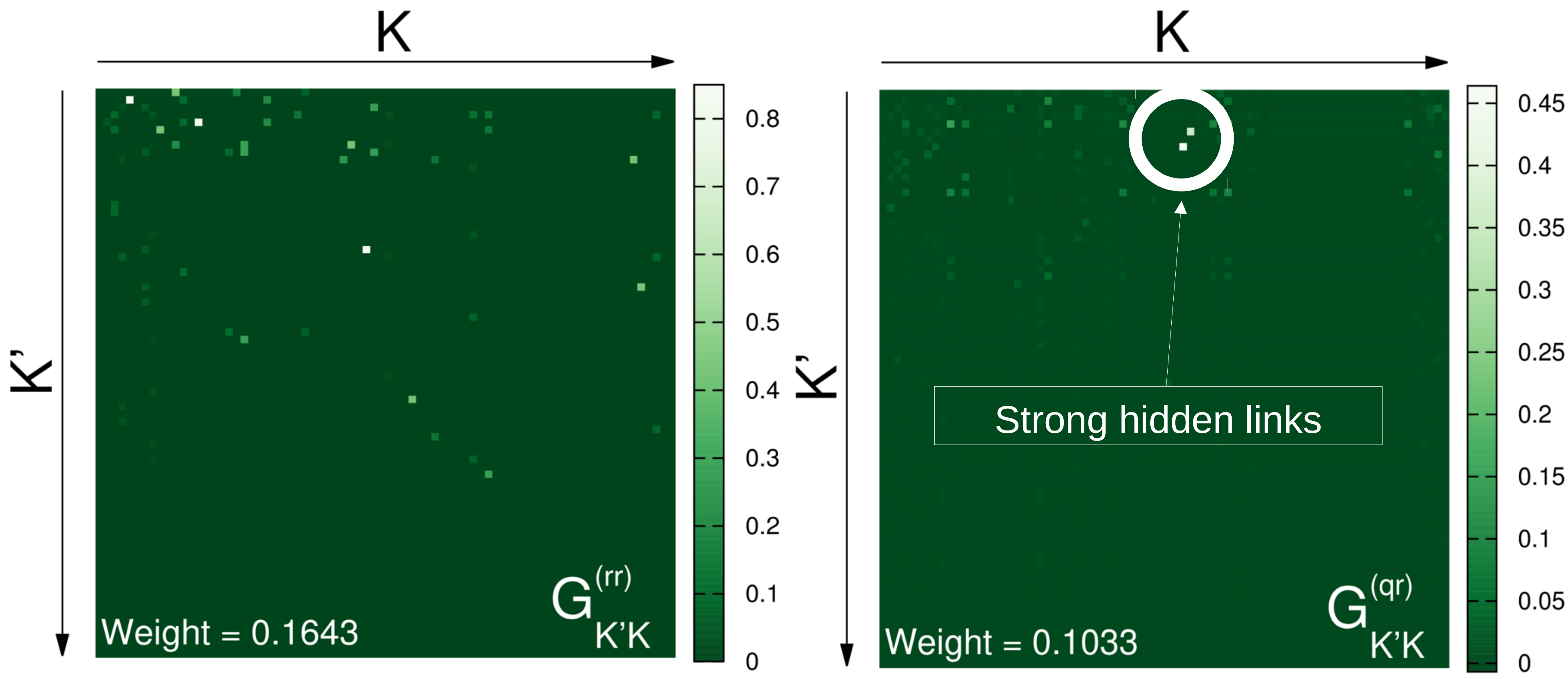
Applications of Google matrix to directed networks and Big Data

Frahm, Jaffrès-Runser, Shepelyansky, Eur. Phys. J. B (2016) 89: 269

Colloque de restitution du Défi MASTODONS CNRS 2016 – jeudi 9 février 2017

Google matrix analysis of causal cancer protein networks

Hidden interactions between proteins
Global network (signaling network SIGNOR) 2432 proteins (nodes)
Example of a sub-network of 76 proteins (direct transcriptional targets of E2F1 protein)



$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

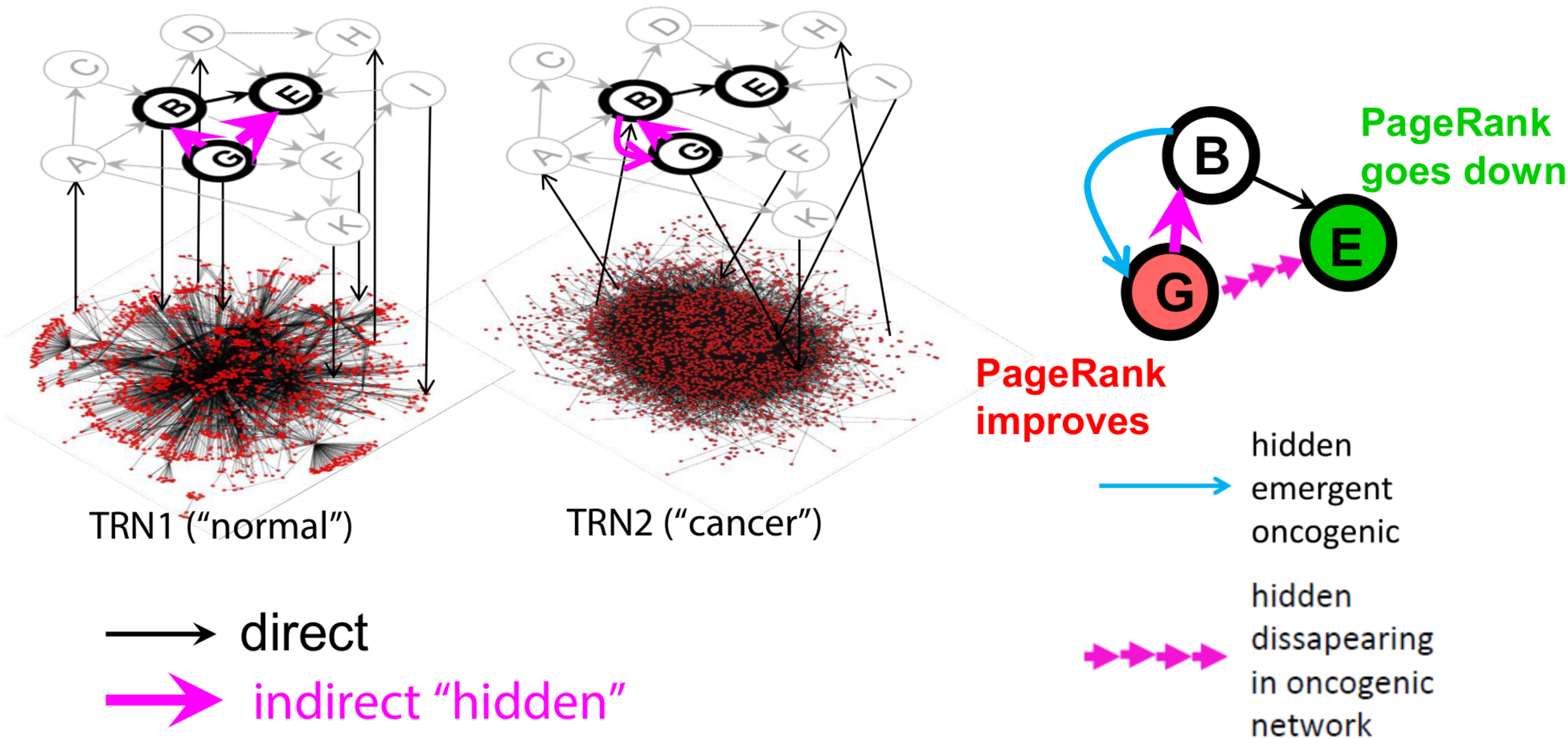
direct links **hidden links**

Lages, Shepelyansky, Zinovyev, bioRxiv 096362, submitted to eLife

ApliGoogle
Applications of Google matrix to directed networks and Big Data

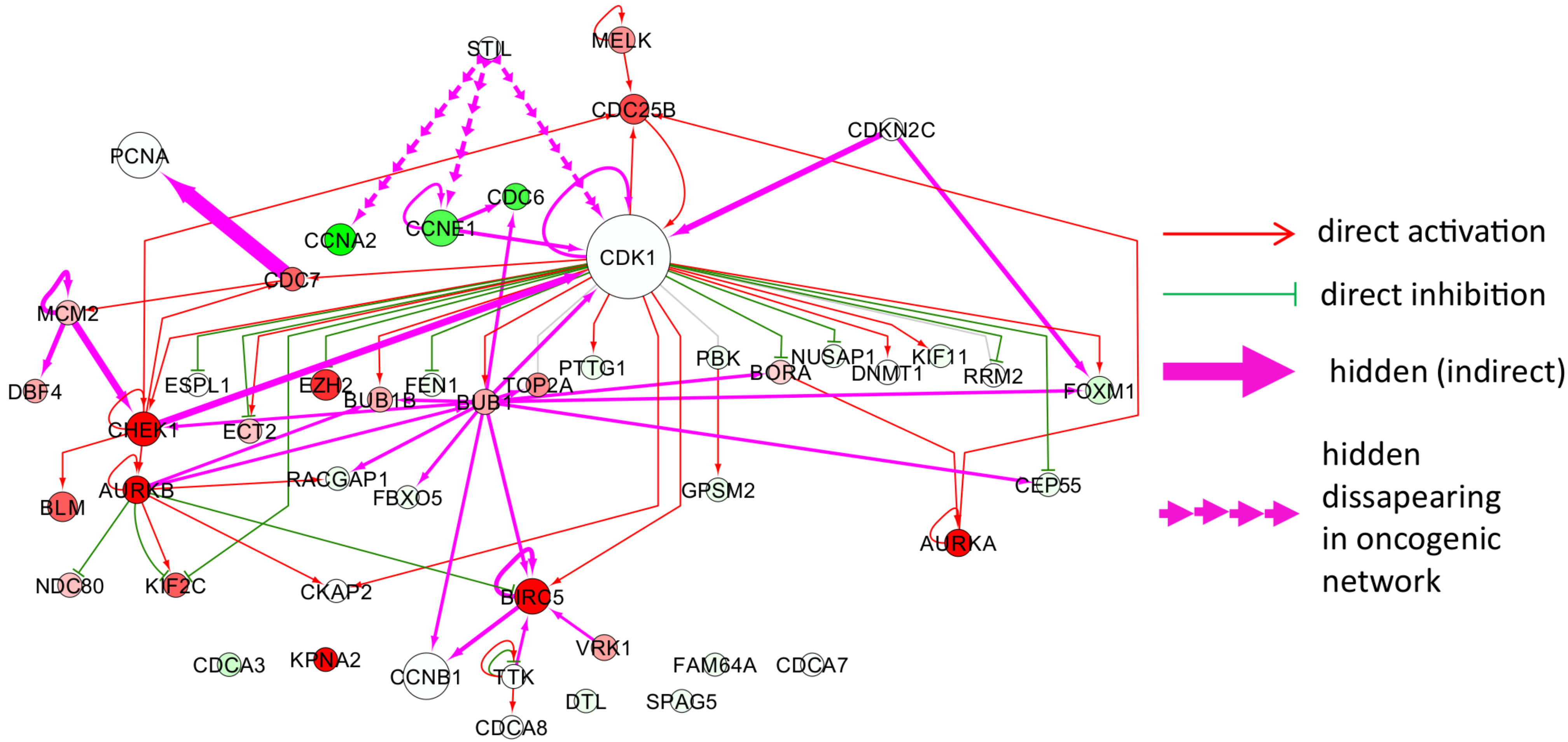
Google matrix analysis of causal cancer protein networks

General scheme



Lages, Shepelyansky, Zinovyev, bioRxiv 096362, submitted to eLife

Google matrix analysis of causal cancer protein networks



Network of proteins shown to be related to cell proliferation by transcriptomic data analysis. The meaning of the node and edge colors is the same as in Figure 6 besides those regulations which disappear in “cancer" network compared to “normal" network (shown by interrupted line arrows).

Lages, Shepelyansky, Zinovyev, bioRxiv 096362, submitted to eLife

Planned further works

Planned further works

- **Global multiproduct world trade network**
  Merging UN, WTO and OECD data to construct a global network of about 2 million nodes.
  Analysis of global world trade along decades.
  Use of reduced Google matrix to find hidden links between activity sectors or products.

- **Bitcoin transactions network**
  Currently about 10 millions transaction per month / 10 billion US dollars volume.

- **Global multilingual Wikipedia network**
  Build a global Wikipedia network (size about 20 million nodes/articles) using the links between articles from different language editions.
  Use of reduced Google matrix to find hidden links between concepts (many possible directions of investigation)

- **Mobile network / Social network**
  With data collected from a bunch of mobile, construct a network of applications/instructions.

- **Googlomics : application of Google matrix to mulfunctionnal network of omics (proteins, genes, ...)**

Bitcoin transactions network density in PageRank-CheiRank plane for all period from creation in 2009 till April 2013 (about 2 million users, the number of transactions is about 20 millions) **[preliminary result]**

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Thank You !