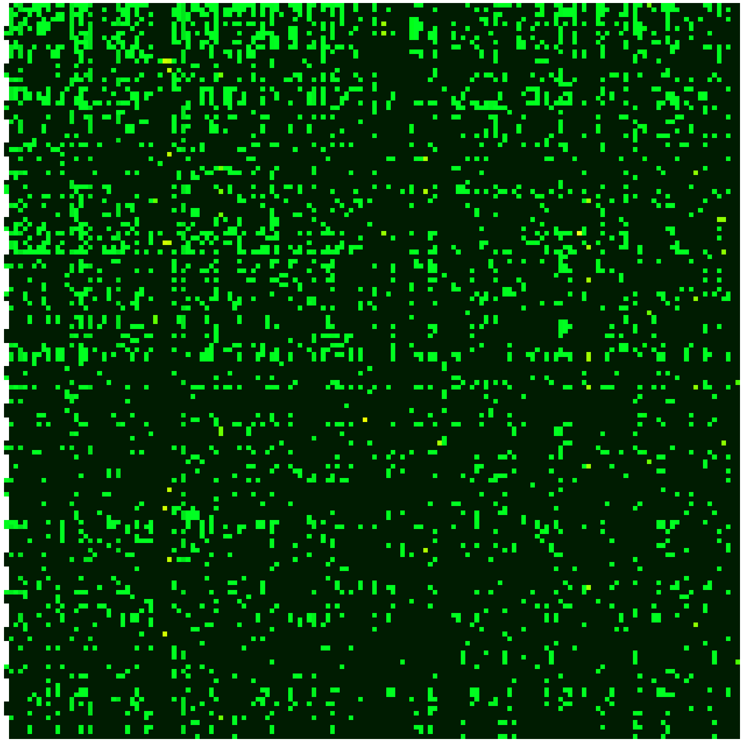
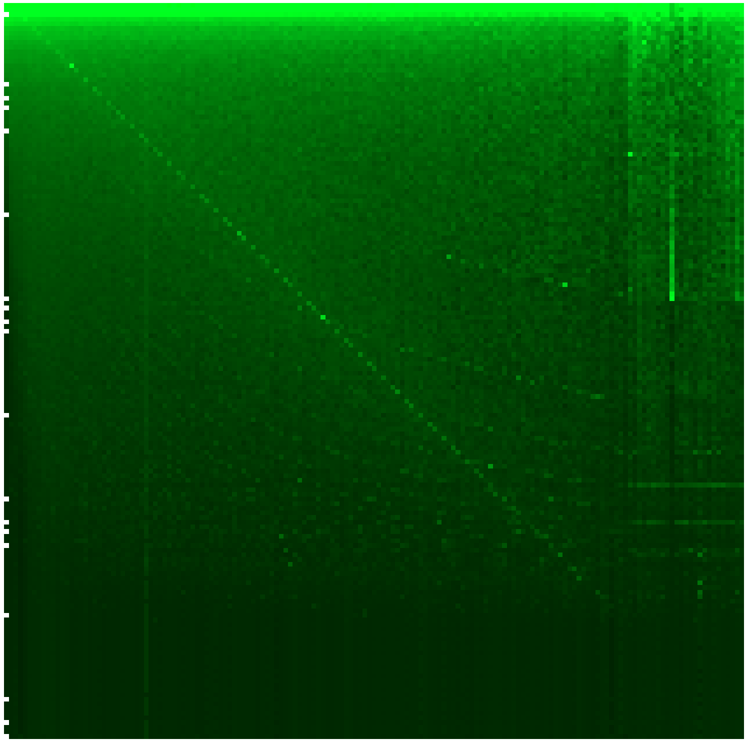
ApliGoogle
Applications of Google matrix to directed networks and Big Data

Défi MASTODONS CNRS
Colloque de restitution 2017

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

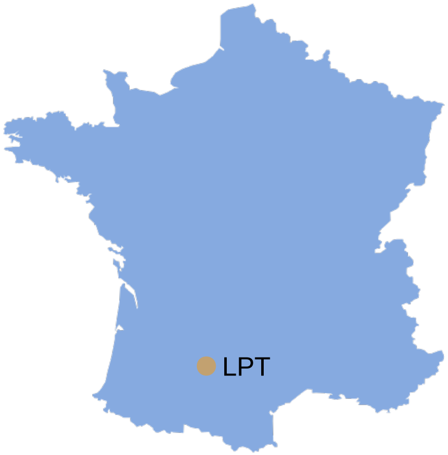
ApliGoogle « Kick-off » meeting, Luchon, May 16th 2016

ApliGoogle partners

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

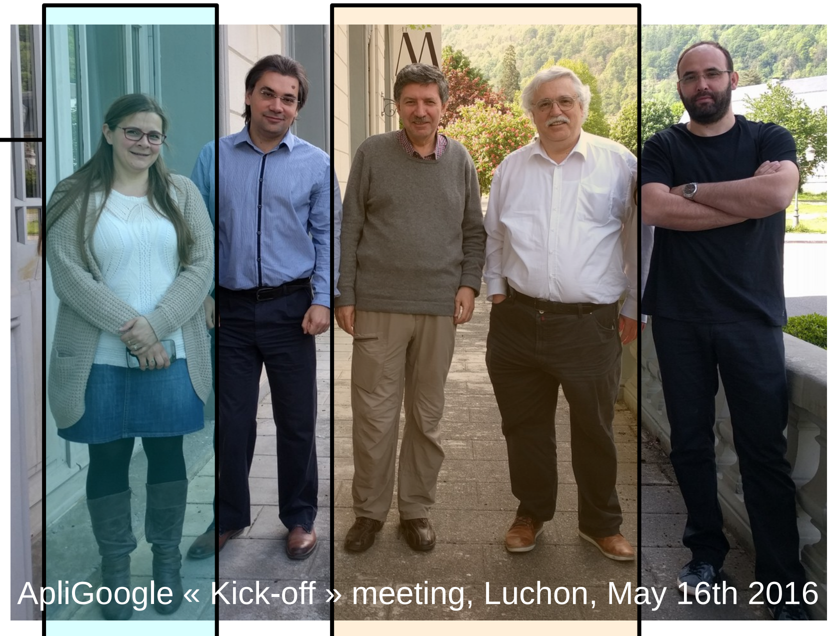
ApliGoogle partners

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

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

Expertises: Wireless networks

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

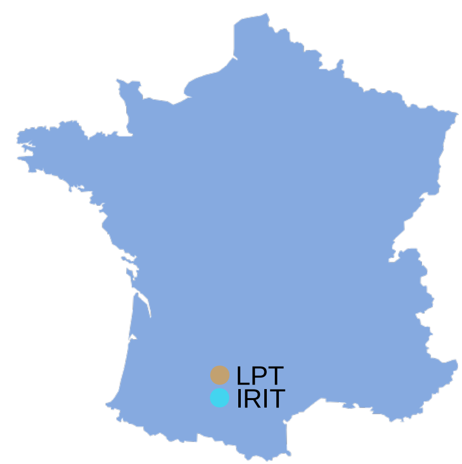


ApliGoogle « Kick-off » meeting, Luchon, May 16th 2016

Laboratoire de Physique Théorique de Toulouse (UMR CNRS 5152)

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

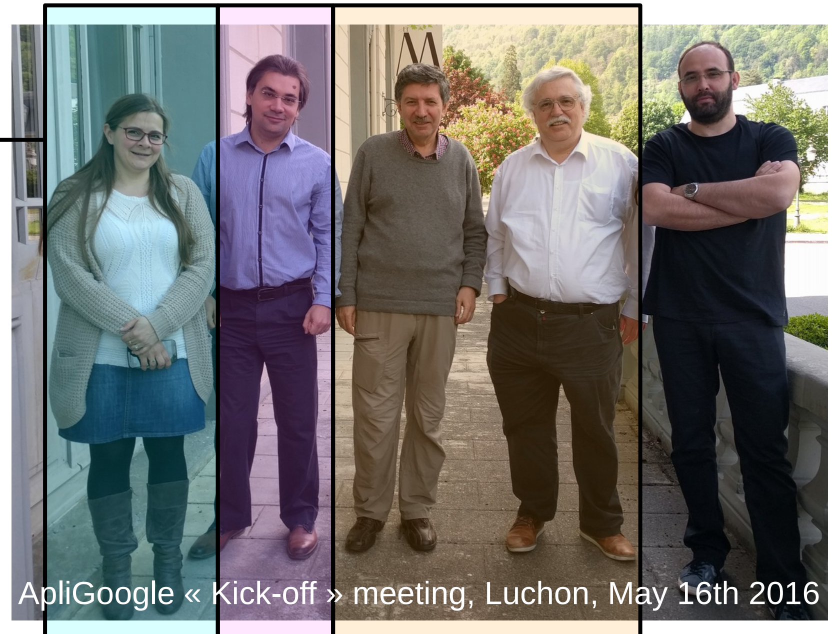
Expertises: Quantum chaos, Random Matrix Theory, Complex directed networks



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**
- S. El Zant (doctorant)
- Dr. T. Peng (post-doctorant)

Expertises: Wireless networks

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

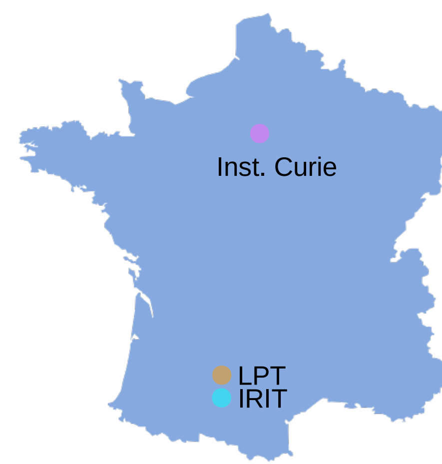
http://www.ihes.fr/~zinovyev/

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

ApliGoogle partners

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

- **MCF K. Jaffrès-Runser**
- S. 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 UTINAM
(UMR CNRS 6213)

- **MCF J. Lages**
- C. Coquidé (doctorant)
- G. Rollin (post-doctorant)

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

Institut UTINAM — UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ — CNRS

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

institut Curie — PSL RESEARCH UNIVERSITY PARIS — Inserm

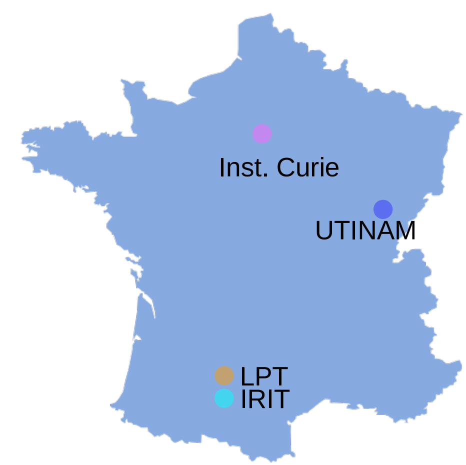
http://www.ihes.fr/~zinovyev/

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

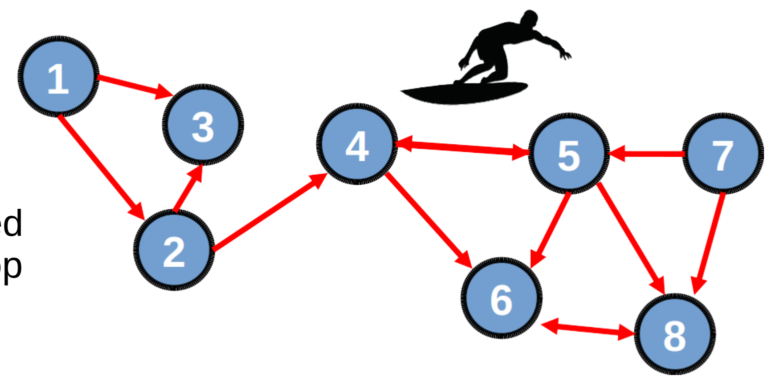


http://www.quantware.ups-tlse.fr/dima

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Main methods

How Google works



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

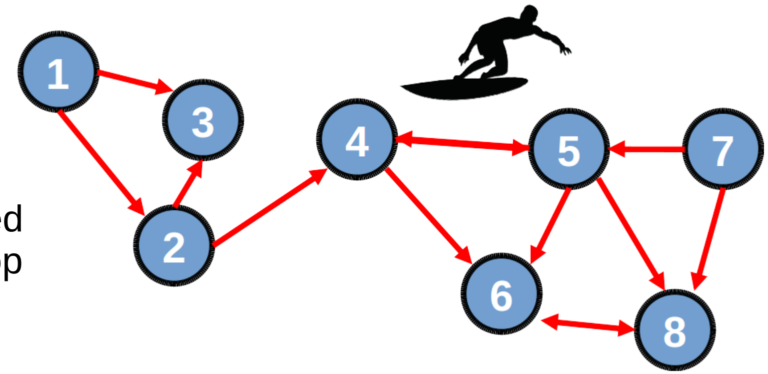
Markovian process : a random surfer probe the structure of a directed network. A each step, the surfer choose randomly an adjacent node to hop and continue its journey.

Adjacency matrix

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

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

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

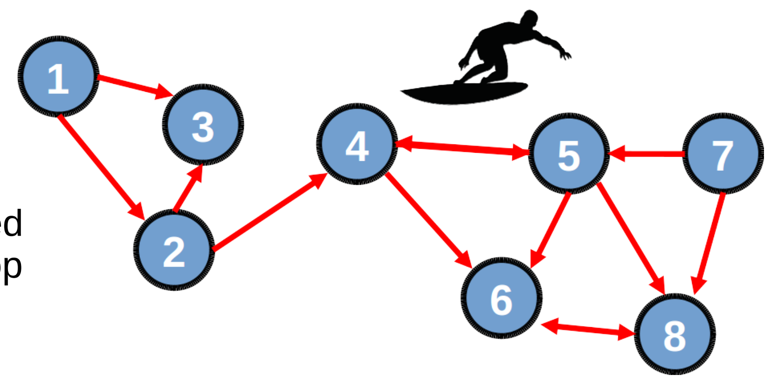
$$A_{ij} = \begin{cases} 1 \text{ si } j \to i \\ 0 \text{ si } j \nrightarrow 1 \end{cases}$$

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

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

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

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

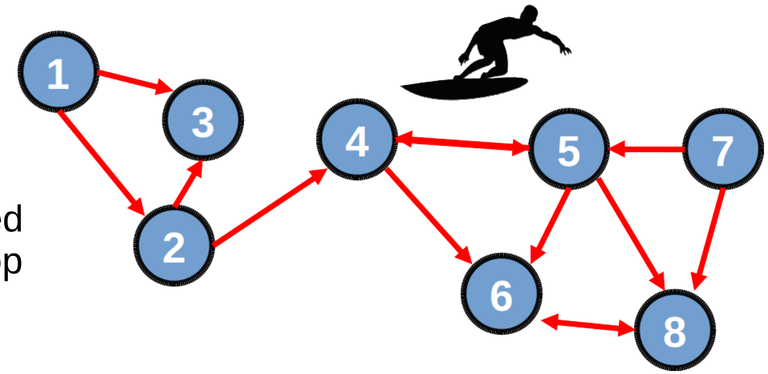
Perron-Frobenius operator

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

$$\alpha = 0.8$$

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

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

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

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

$$\mathbf{P} = \begin{pmatrix} 0.031094525687305 97 \\ 0.0435323361475661 7 \\ 0.0609452708660655 8 \\ 0.0672941236179782 6 \\ 0.0704499859958617 1 \\ \color{red}{0.3518167935609448 9} \\ 0.0310945256873059 7 \\ 0.3437724384369714 3 \end{pmatrix}$$

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

$P(1) = \color{red}{0.35181679356094489}$ 6

$P(2) = 0.34377243843697143$ 8

$P(3) = 0.07044998599586171$ 5

$P(4) = 0.06729412361797826$ 4

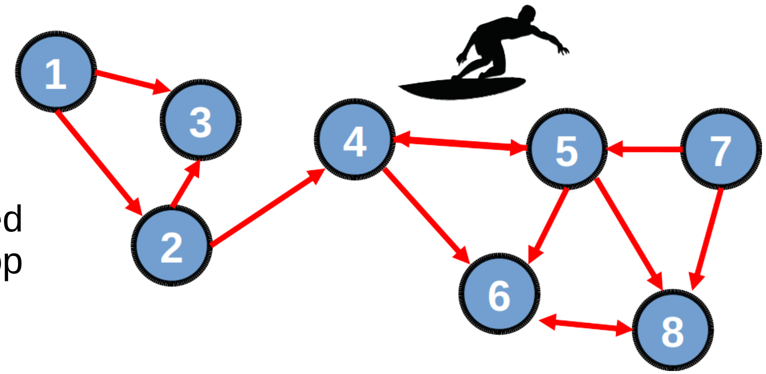
$P(5) = 0.06094527086606558$ 3

$P(6) = 0.04353233614756617$ 2

$P(7) = P(8) = 0.03109452568730597$ 1 7

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

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

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

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

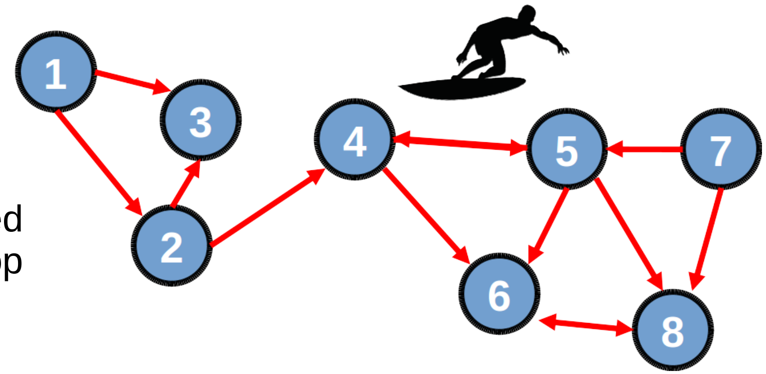
The most important node is the one with the highest probability.
“Recursive definition”: the more a node is pointed by important nodes, the more it is important.

PageRank measures the influence of a node.

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

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

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

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

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

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

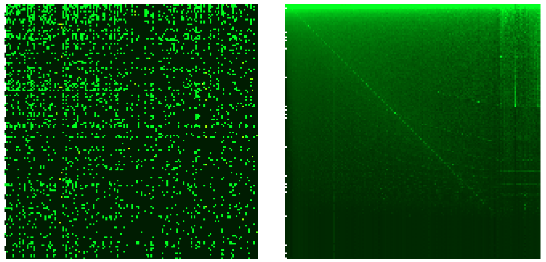
Similar to PageRank of the inverted network. With inverted adjacency matrix elements $A^*_{ij} = A_{ji}$, it is possible to define the stochastic matrix elements $S^*_{ij} \neq S_{ji}$, and the Google matrix elements $G^*_{ij} \neq G_{ji}$ associated to the inverted network (Fogaras ‘03, Chepelianskii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

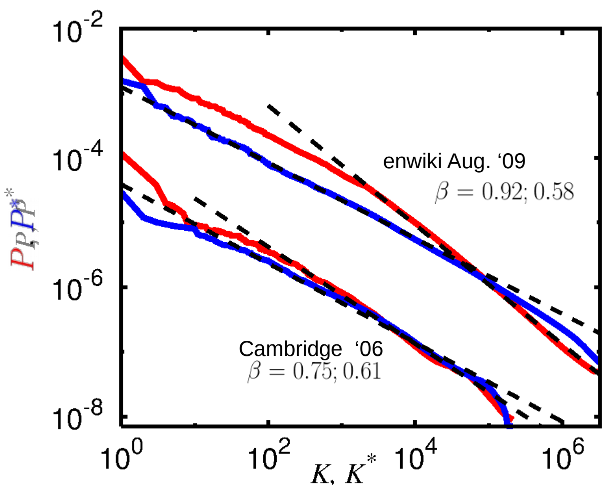
N=3 282 257

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

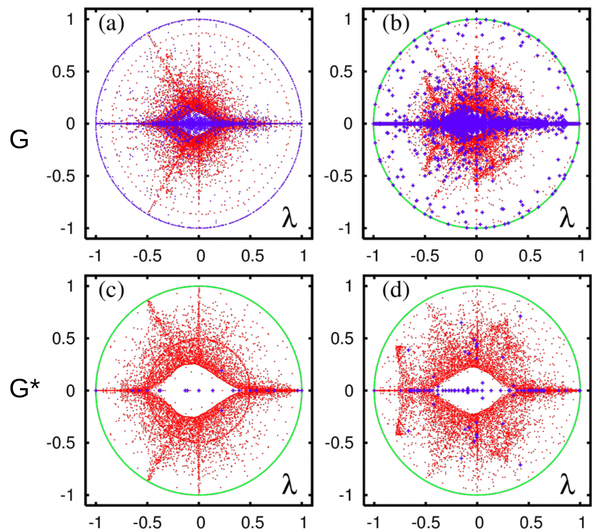
- **"Small world" property:** average distance between two nodes $\sim \log N$
- **Scale-free property:** distribution of the number of ingoing or outgoing links is $\rho(k) \sim k^{-\nu}$

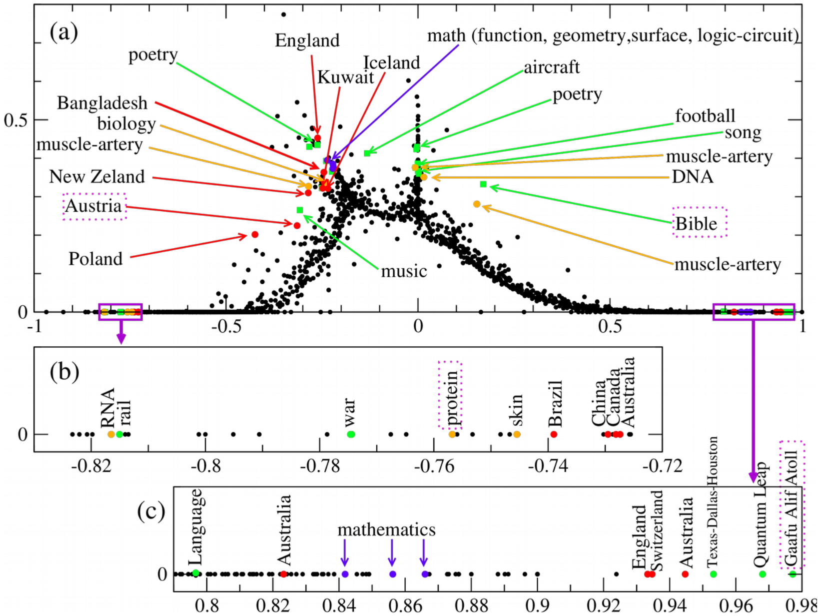
PageRank distribution properties

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



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).

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

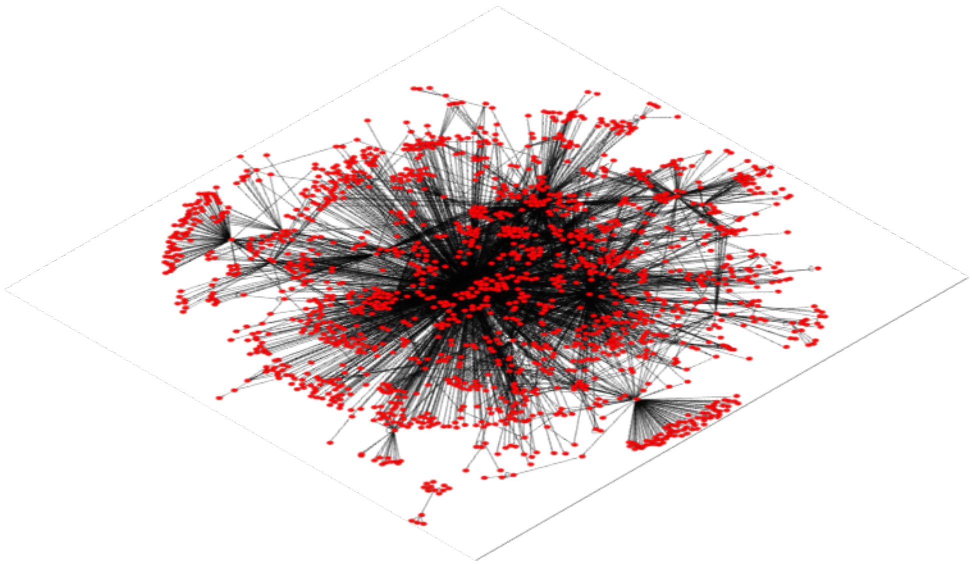
Ermann, Frahm, Shepelyansky (2015), Rev. Mod. Phys. 87, 1261.

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

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Reduced Google matrix

Consider a network with $N \gg 1$ nodes.

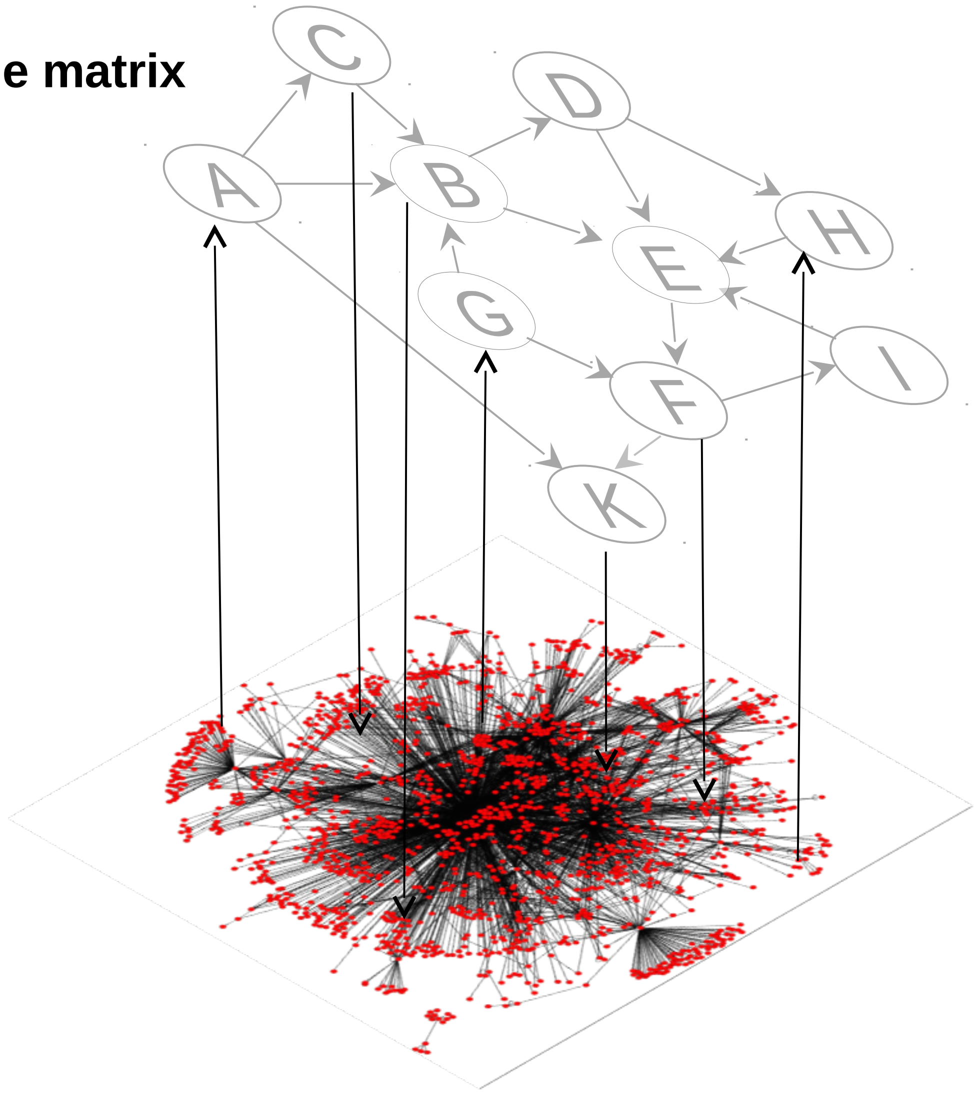


ApliGoogle

Applications of Google matrix to directed networks and Big Data

Reduced Google matrix

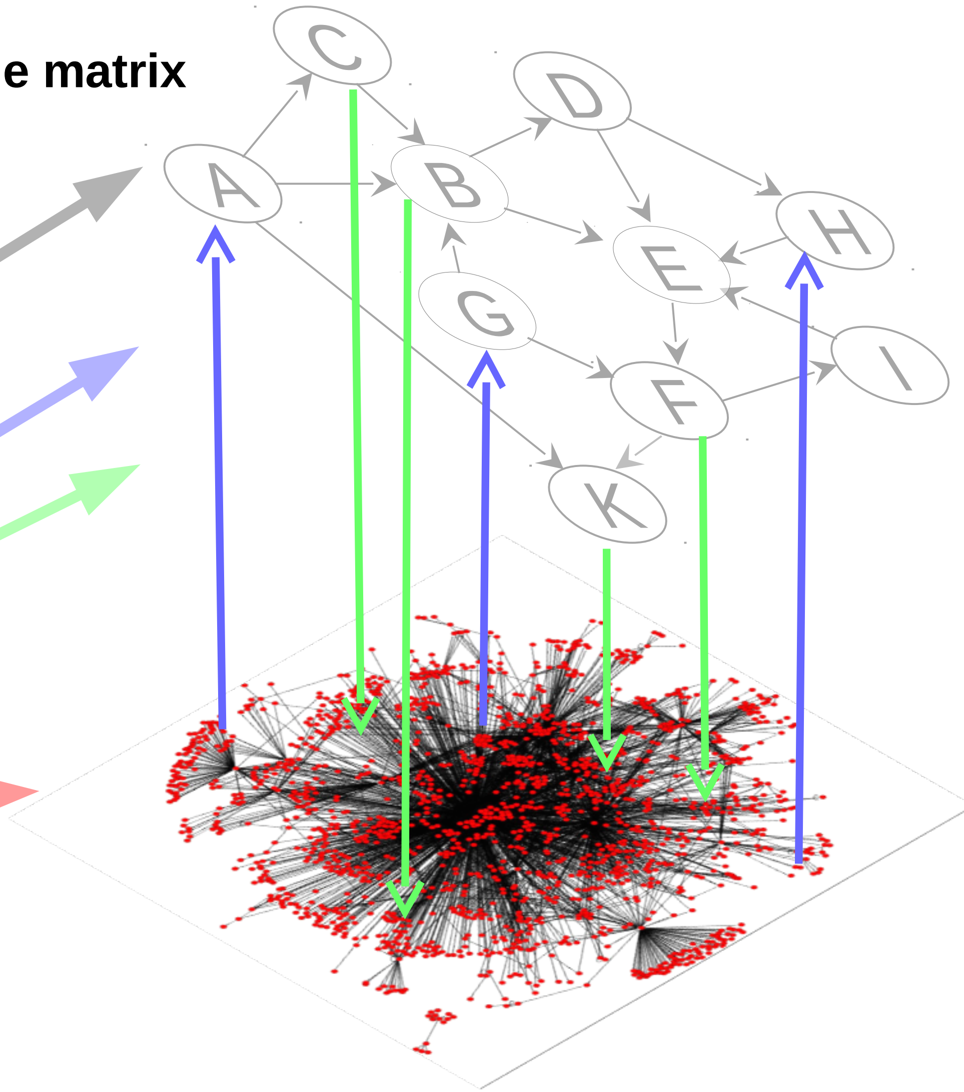
Consider a network with $N \gg 1$ nodes.
Consider a sub-network (a community) of $N_r \ll N$ nodes.



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.
The Google matrix of the size N network and the associated PageRank vector can be written as



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

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix} \qquad \mathbf{GP} = \mathbf{P}$$

Reduced Google matrix

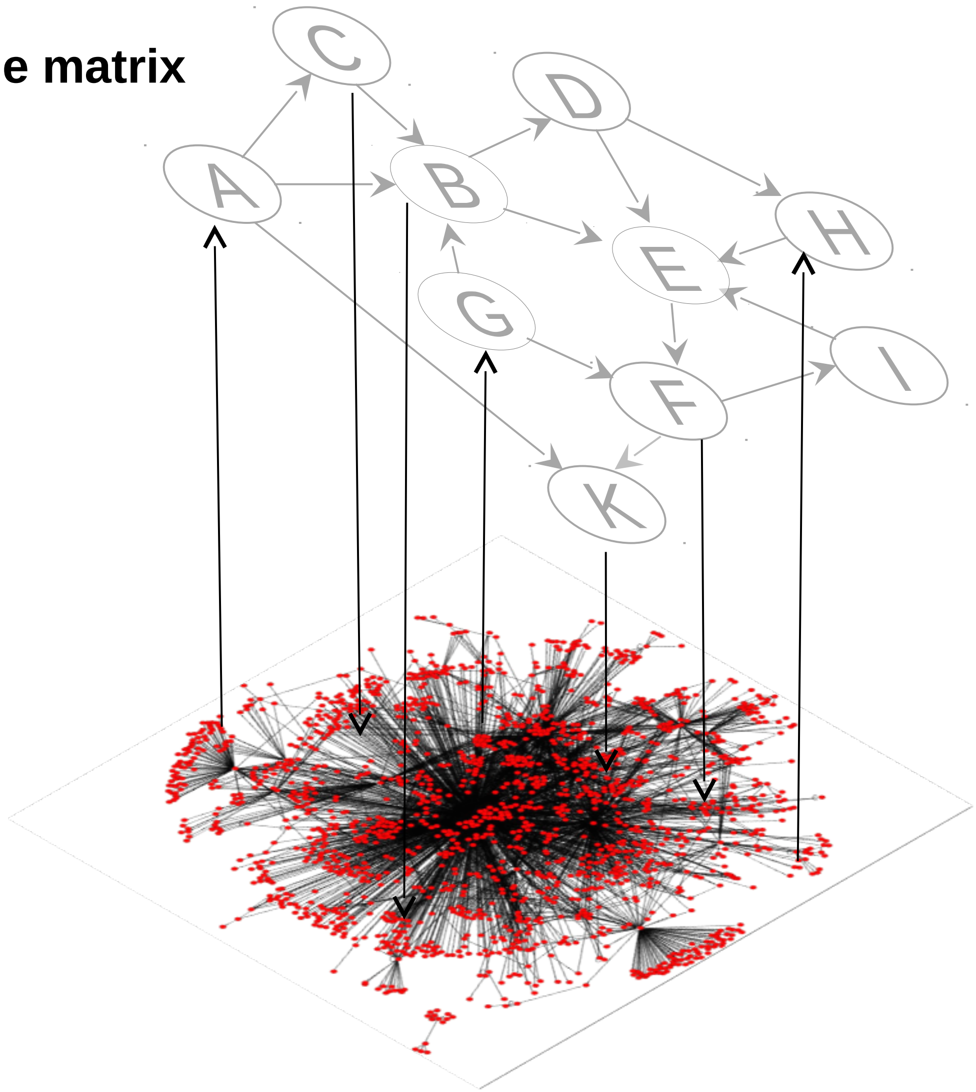
Consider a network with $N \gg 1$ nodes.
Consider a sub-network (a community) of $N_r \ll N$ nodes.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

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

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

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



Reduced Google matrix

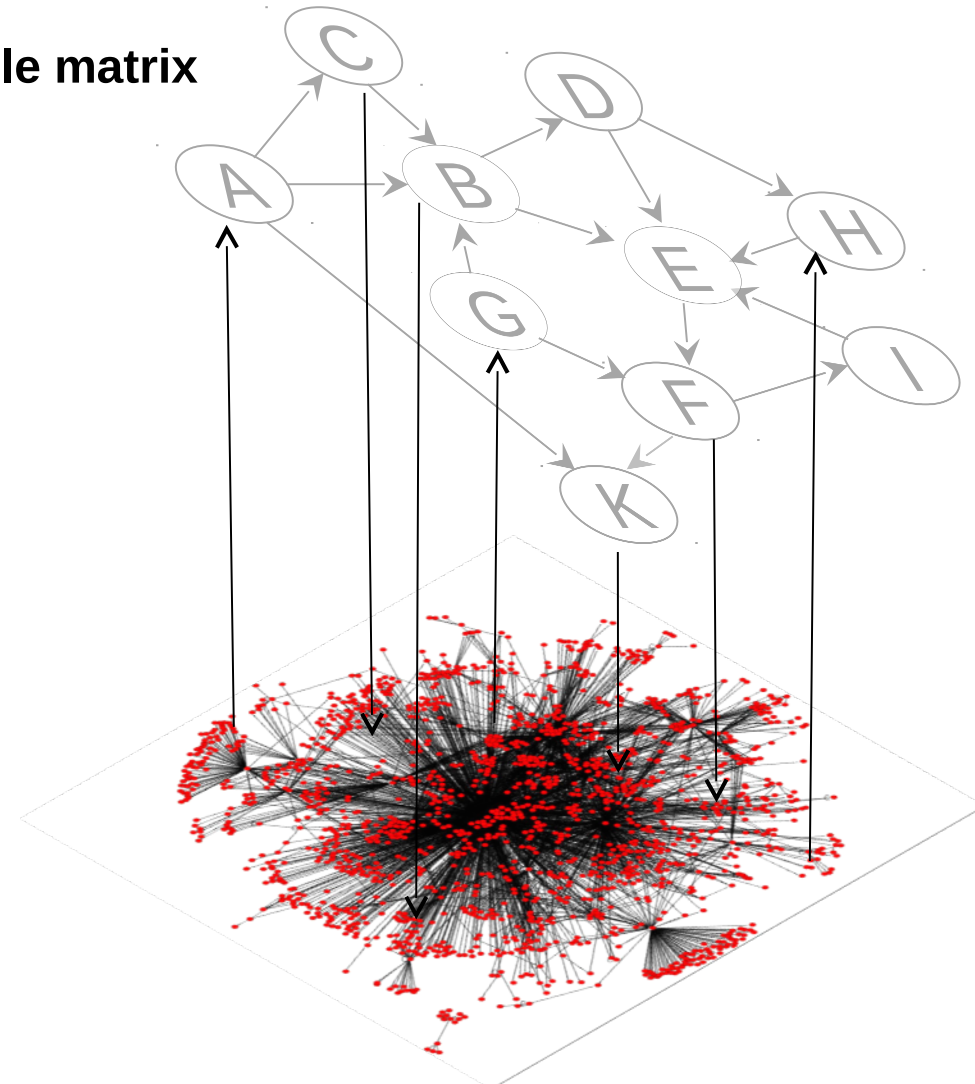
Consider a network with $N \gg 1$ nodes.
Consider a sub-network (a community) of $N_r \ll N$ nodes.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

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

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

Contribution from indirect links (scattering term)

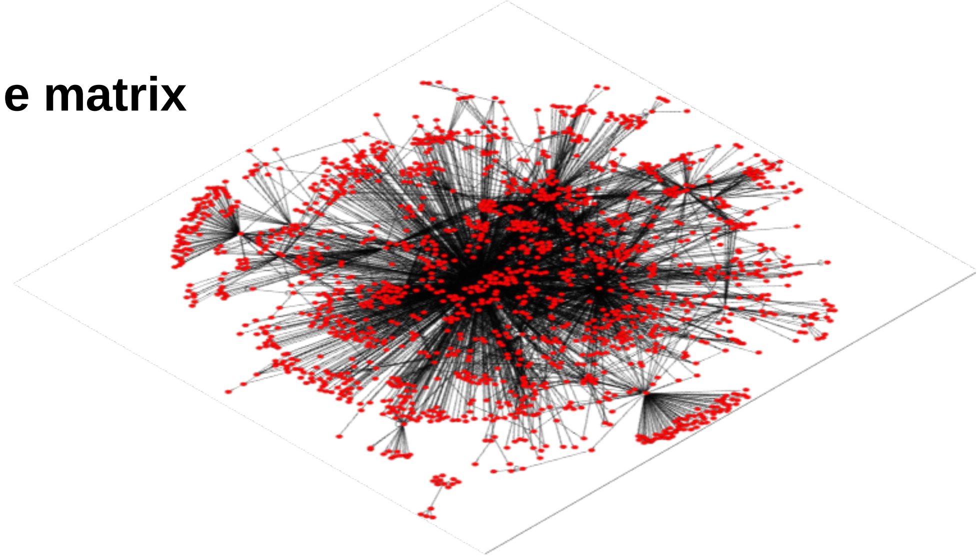
J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, arXiv:1602.02394 [physics.soc-ph]

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

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

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.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

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

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

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

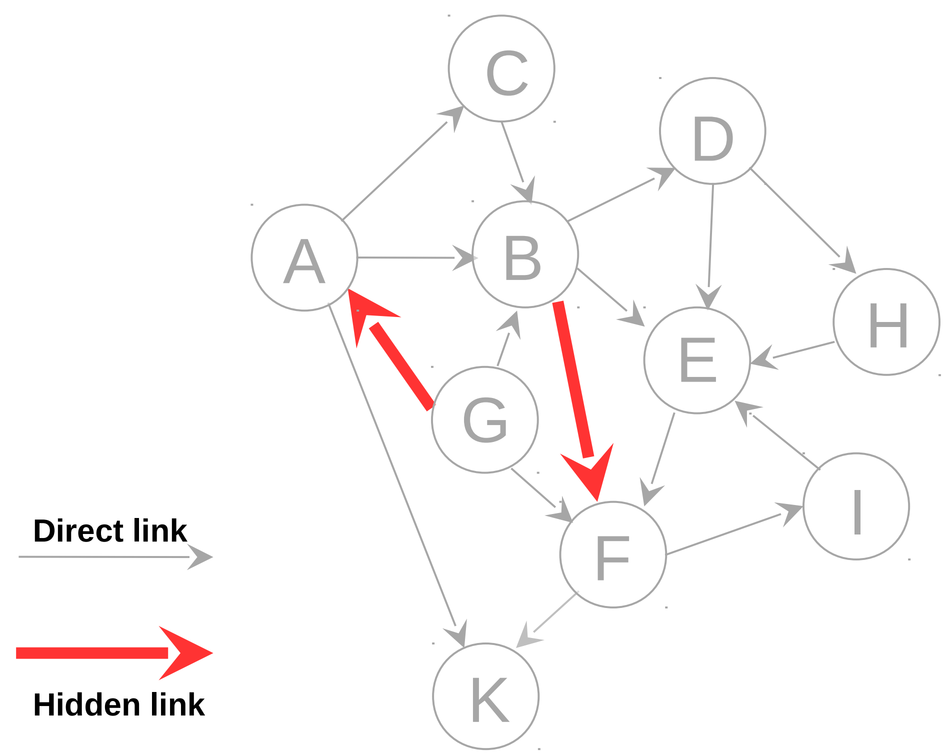
The reduced Google matrix can be written

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

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

$\mathbf{G}_{pr}$: Contribution from « PageRank »

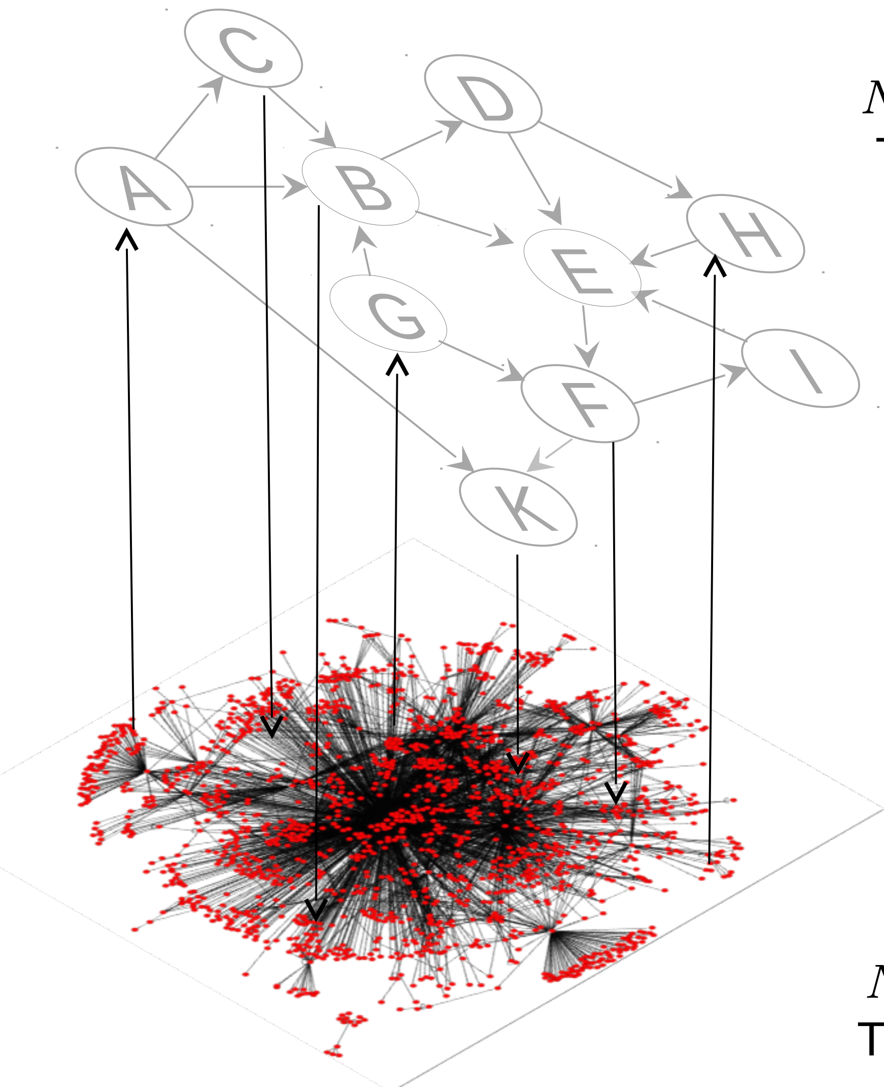
$\mathbf{G}_{qr}$: Contribution from hidden links



ApliGoogle
Applications of Google matrix to directed networks and Big Data

Highlights: major results obtained in 2017

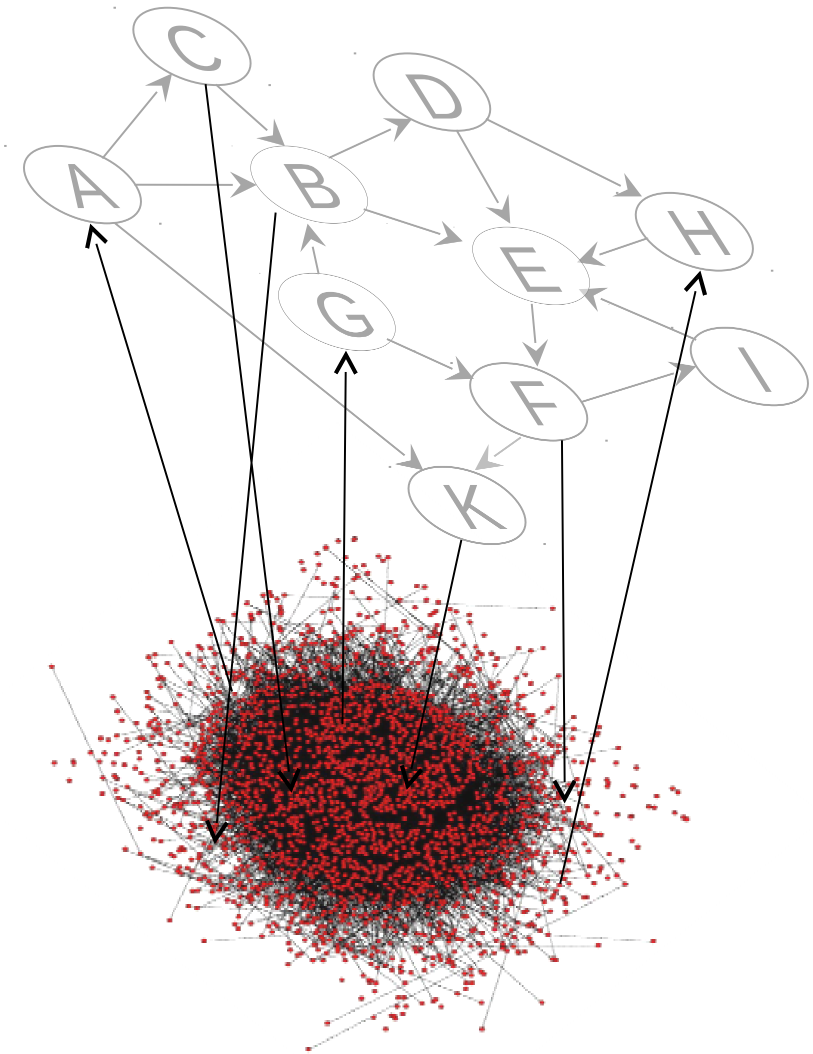
Googlomics
Inferring hidden causal relations between proteins



Nr < 100 proteins
The subnetworks

N = 2432 proteins
The whole network

TRN1 (“normal”)
Normal B-Lymphocytes



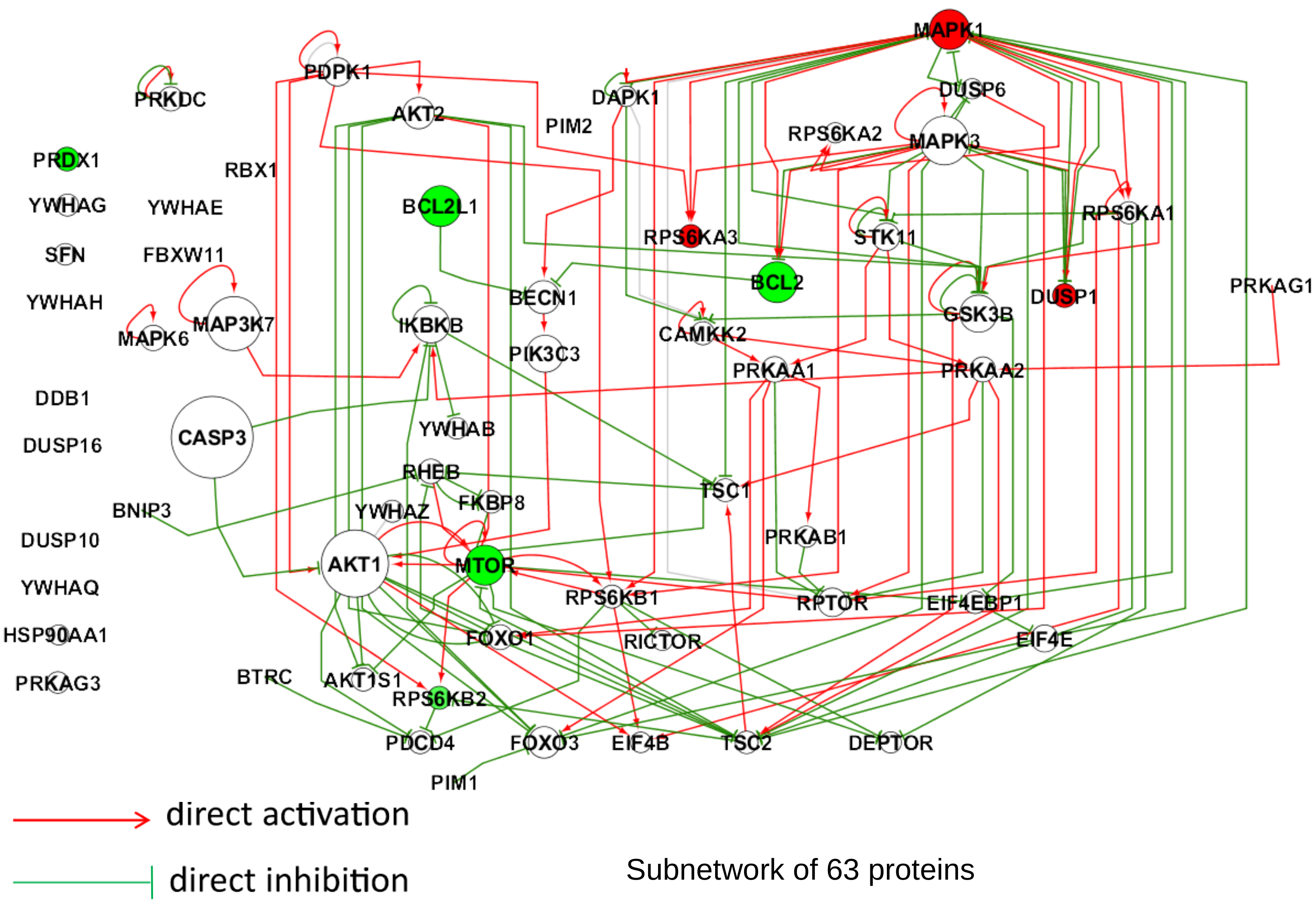
TRN2 (“cancer”)
Leukemia cell line

J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Googlomics
Inferring hidden causal relations between proteins

Inferring indirect (hidden) causal connections between AKT-mTOR pathway members



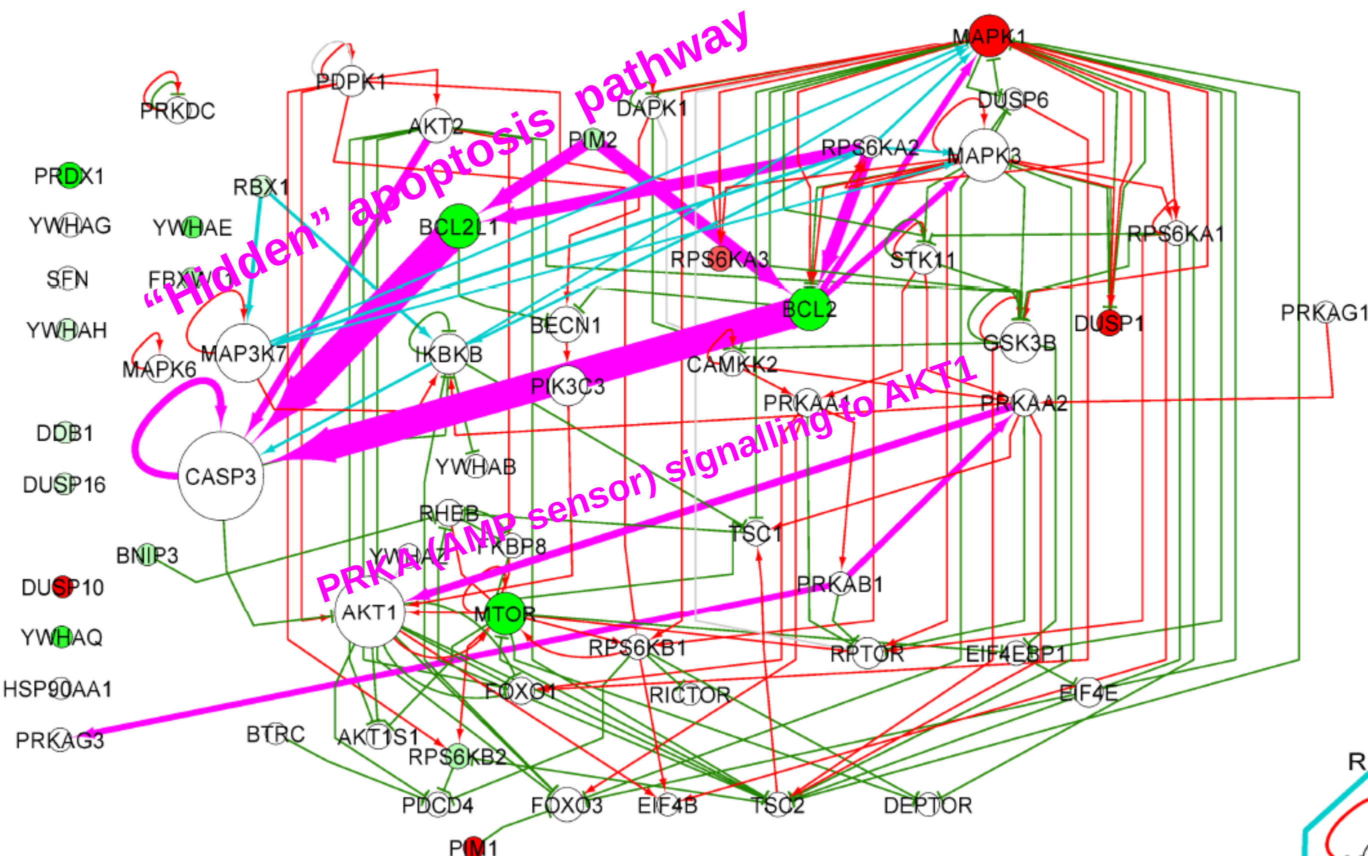
Subnetwork of 63 proteins

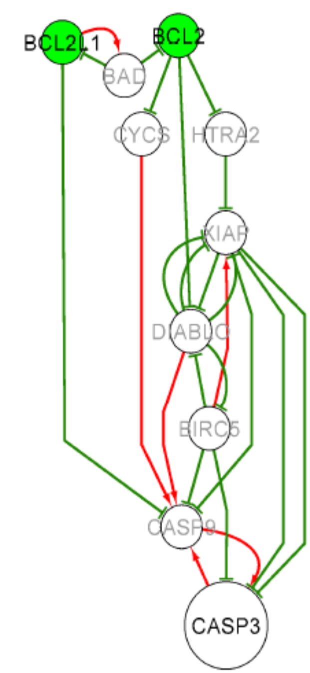
J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Googlomics
Inferring hidden causal relations between proteins

Inferring indirect (hidden) causal connections between AKT-mTOR pathway members





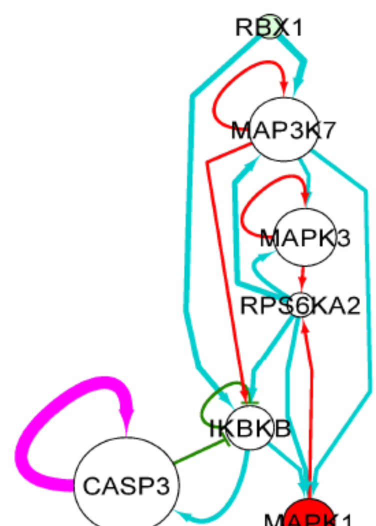
direct activation

direct inhibition

hidden (indirect)

Subnetwork of 63 proteins

hidden emergent oncogenic



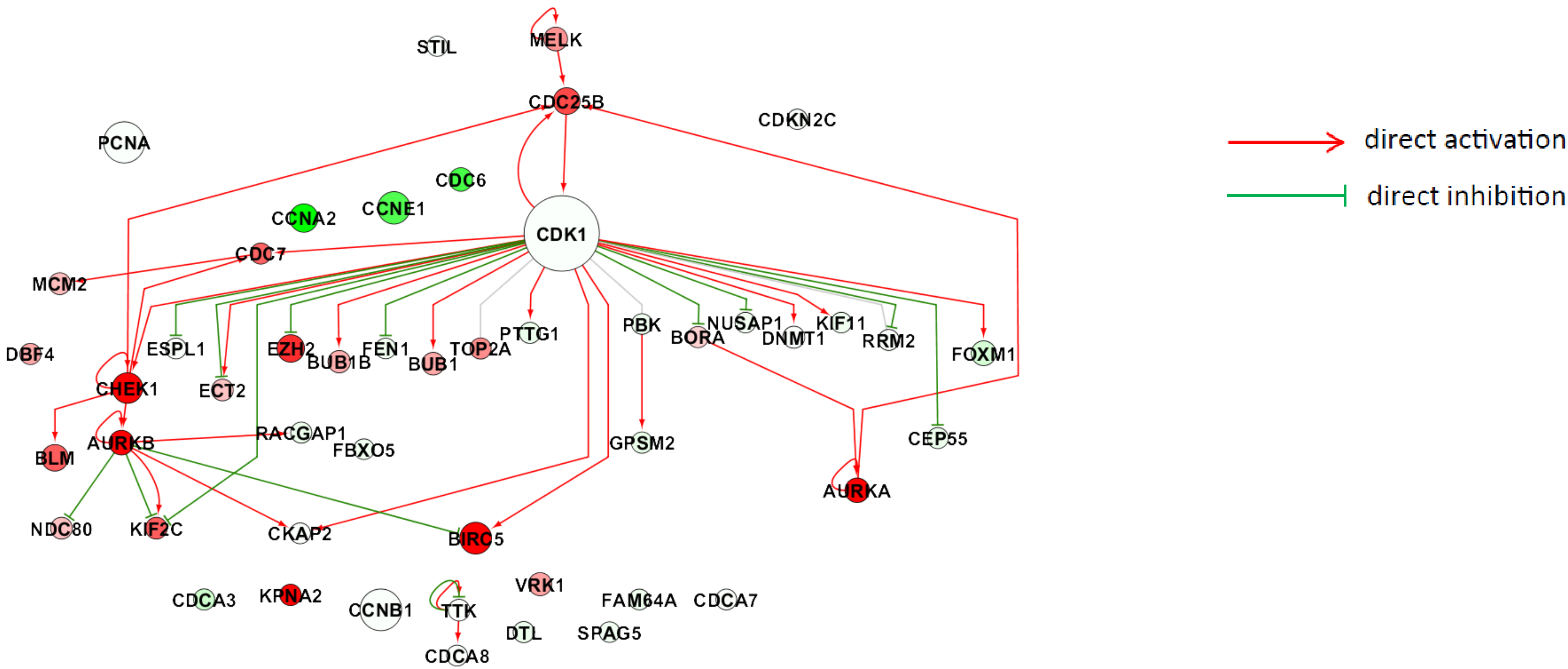
Emergent oncogenic signaling between RBX1 (cell cycle protein degradation proteasome) and MAPK1

J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Googlomics
Inferring hidden causal relations between proteins

Genes of a proliferative signature resulted from pancancer transcriptomic analysis



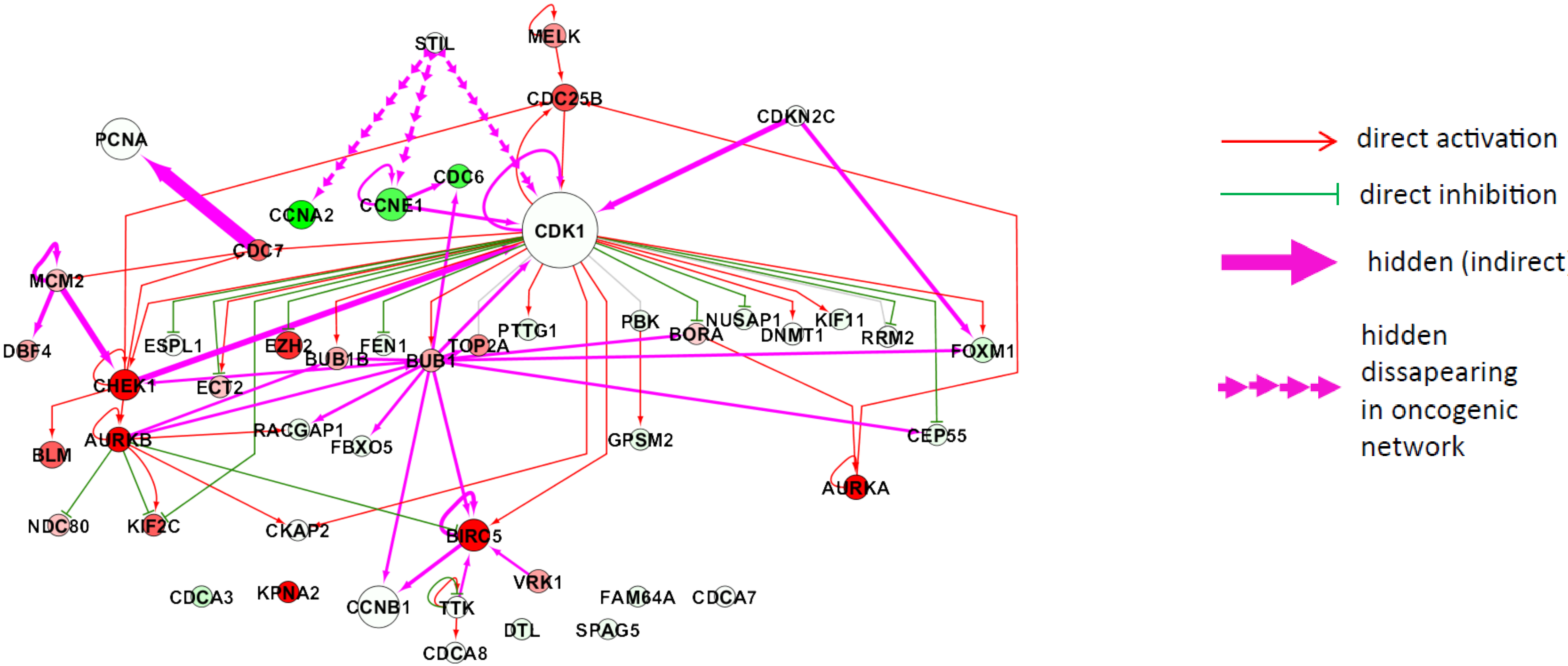
Subnetwork of 49 proteins

J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Googlomics
Inferring hidden causal relations between proteins

Genes of a proliferative signature resulted from pancancer transcriptomic analysis



Subnetwork of 49 proteins

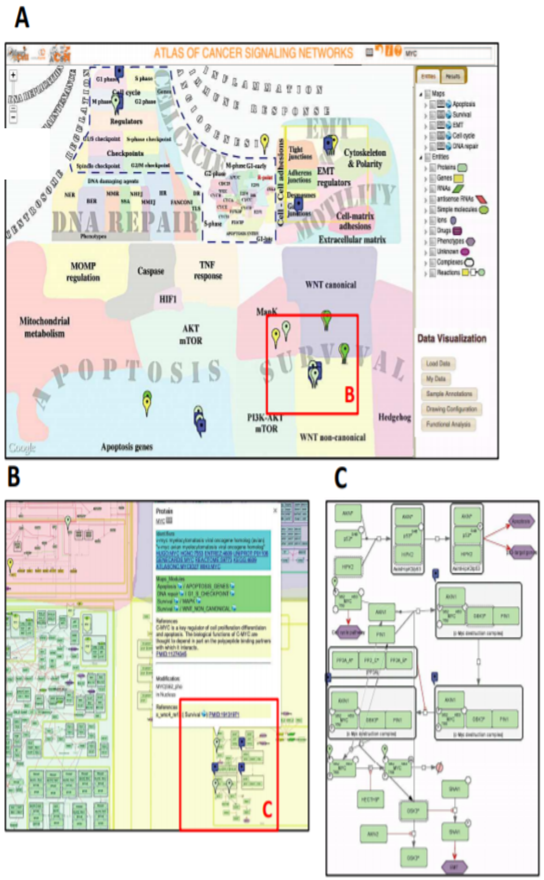
More genes are connected into the network
Emergence of a new “hidden” hub BUB1
Connection to PCNA (DNA replication and DNA repair)
Many cell cycle proteins improves in PageRank (AURK)
Connection between STIL (mitotic spindle checkpoint regulator) and CCNA2, CCNE1

J. Lages, D. Shepelyansky, A. Zinovyev, PLoS ONE 13(1): e0190812 (2018)

ApliGoogle
Applications of Google matrix to directed networks and Big Data

Atlas of Cancer Signaling Network

Atlas of Cancer Signaling Network pathway database and its application in pre-clinical studies (Monraz Gomez et al, http://dx.doi.org/10.1101/234823 (2017), submitted to Briefings in Bioinformatics)

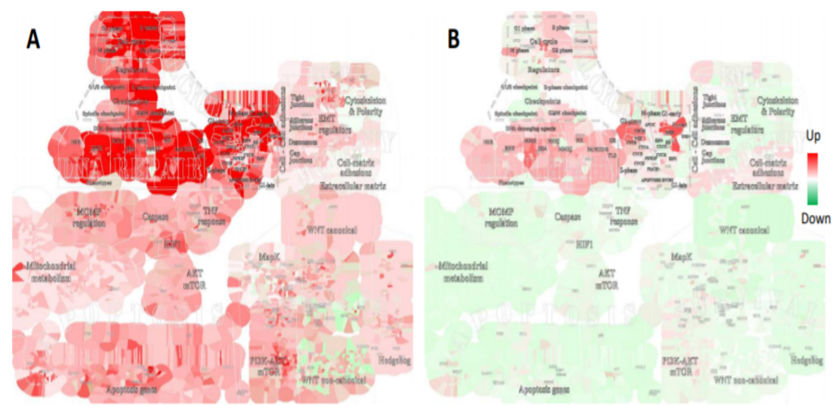


http://acsn.curie.fr

ACSN – pathway database created by Institut Curie, large reaction network describing processes deregulated in cancer

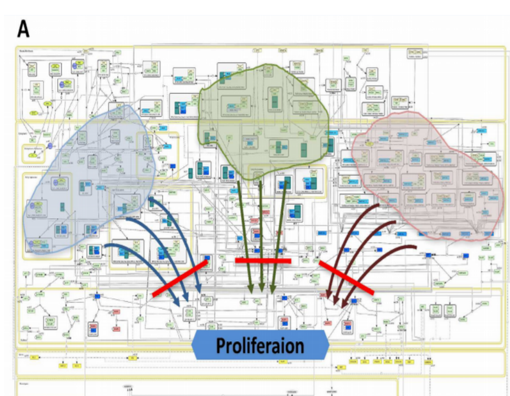
Network propagation methods applied to ACSN allows :

Visualizing molecular omics data for different cancer subtypes (A vs B)



Red –upregulated modules, green - downregulated

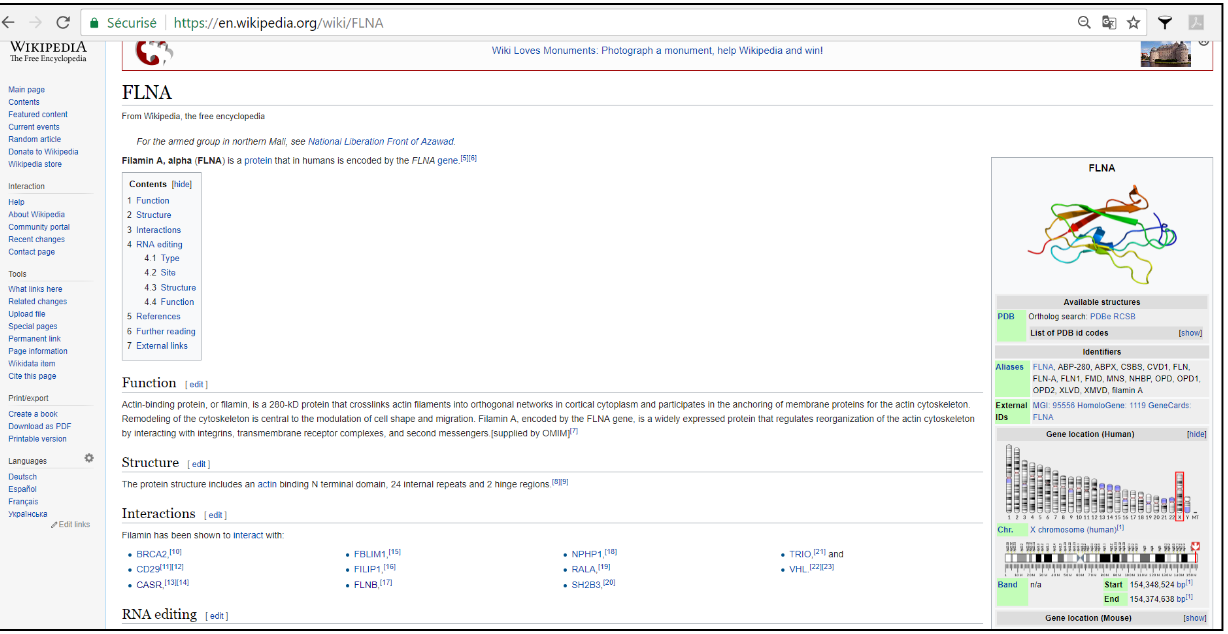
Designing therapeutic interventions to disrupt malignant cancer signaling



ApliGoogle

Applications of Google matrix to directed networks and Big Data

WikiProteins project
Inferring directed network of proteins from Wikipedia using reduced Google Matrix

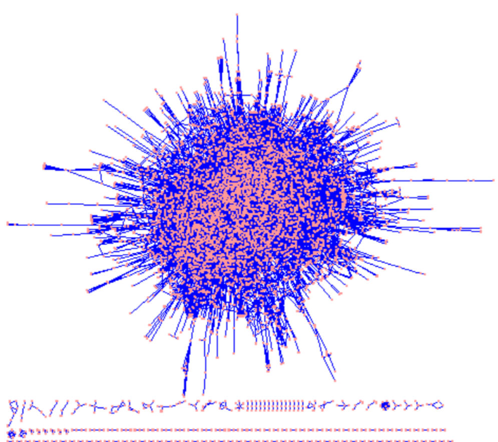


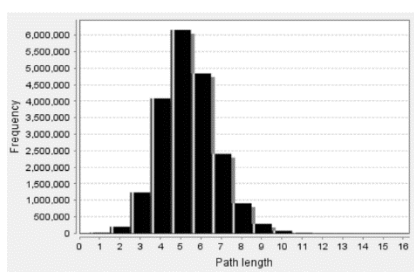
(ongoing work…)

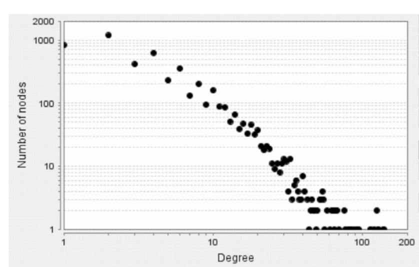
~10000 wiki pages devoted to proteins

5009 proteins with described interactions, 16468 direct connections

The general properties of extracted network is similar to manually curated protein-protein interaction networks







This network is embedded in the global Wikipedia network

The rest of the Wikipedia defines a context of hyperlinks (model of external world)

Reduced Google matrix allows finding “hidden” functional interactions between proteins

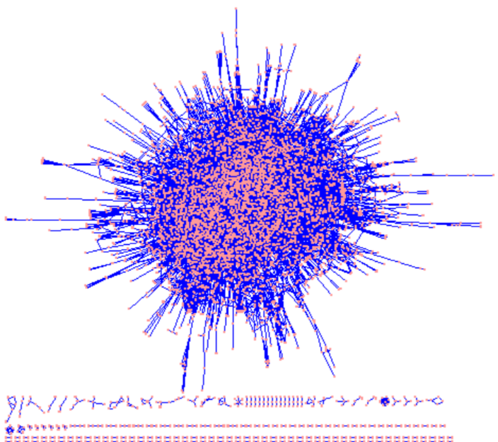
ApliGoogle
Applications of Google matrix to directed networks and Big Data

WikiProteins project
Inferring directed network of proteins from Wikipedia using reduced Google Matrix

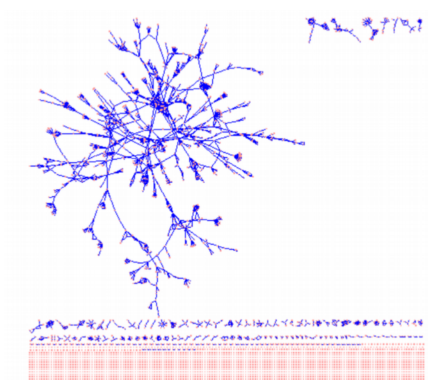
(ongoing work…)

Comparing direct and hidden networks of protein connections

Direct connections

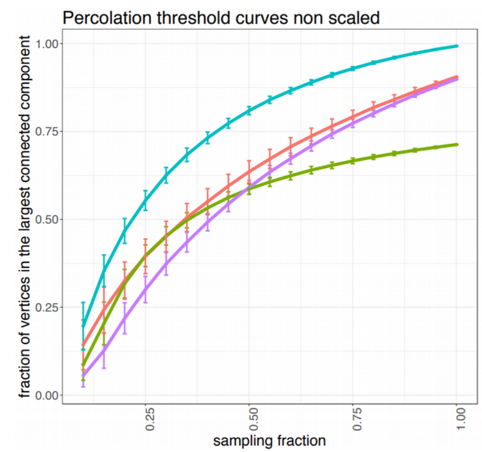


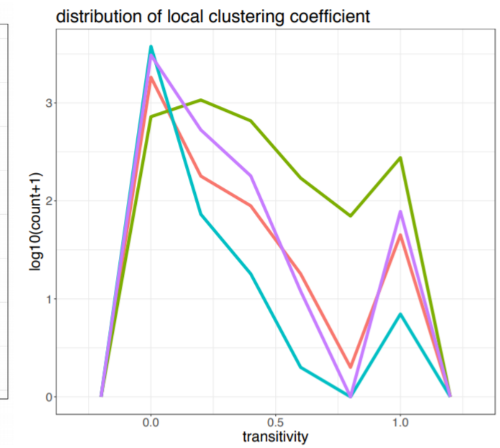
Hidden connections (Gqr>0.005)

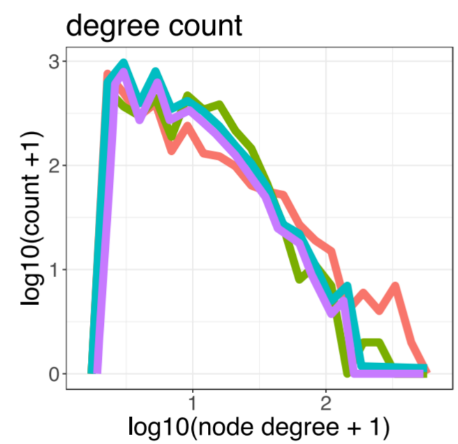


Hidden connections are characterized by existence of tight network communities

Comparative topological analysis







- Manually curated network SIGNOR
- Network of hidden interactions
- Randomized direct WikiProtein network
- Direct WikiProtein network

ApliGoogle
Applications of Google matrix to directed networks and Big Data

WikiProteins project
Inferring directed network of proteins from Wikipedia using reduced Google Matrix

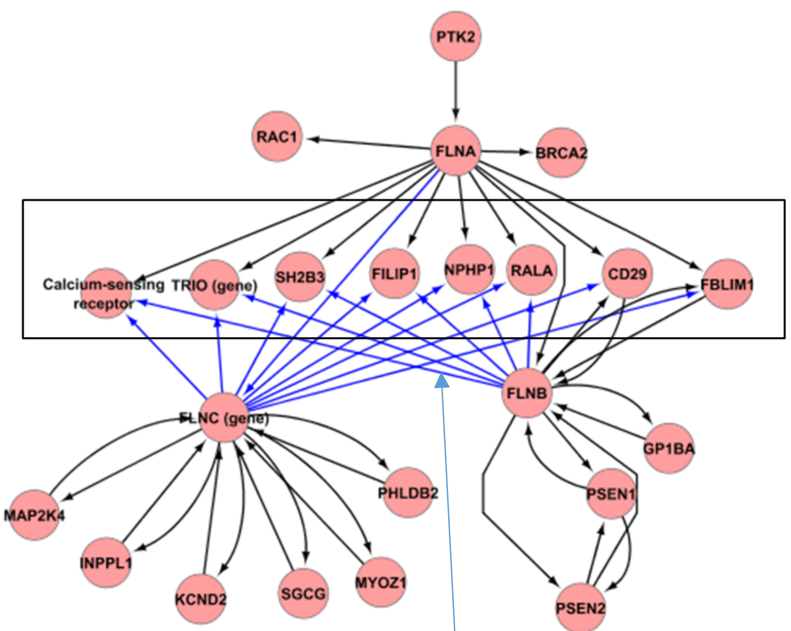
(ongoing work…)

Studying “hidden” communities of proteins in Wikipedia is informative

Example: Discovering common functional partners of filamins

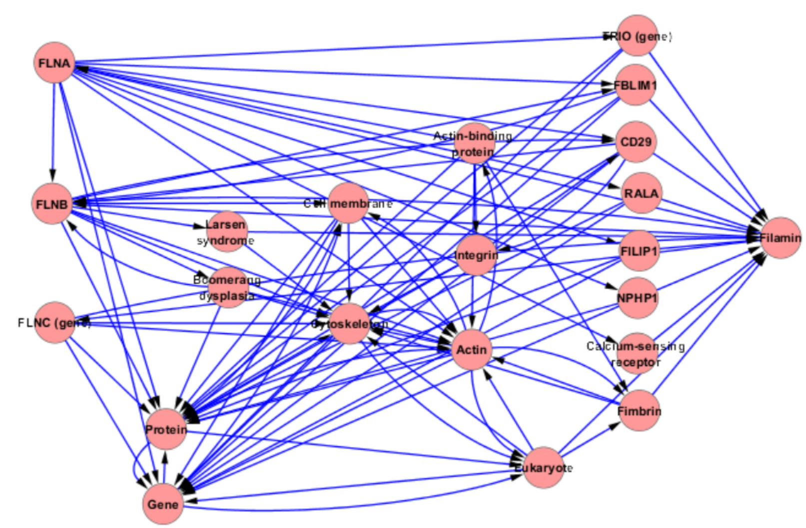
Direct links

Hidden links



“Hidden” connection community

Direct shortest/second shortest paths from filamins FLN* to targets



Direct links

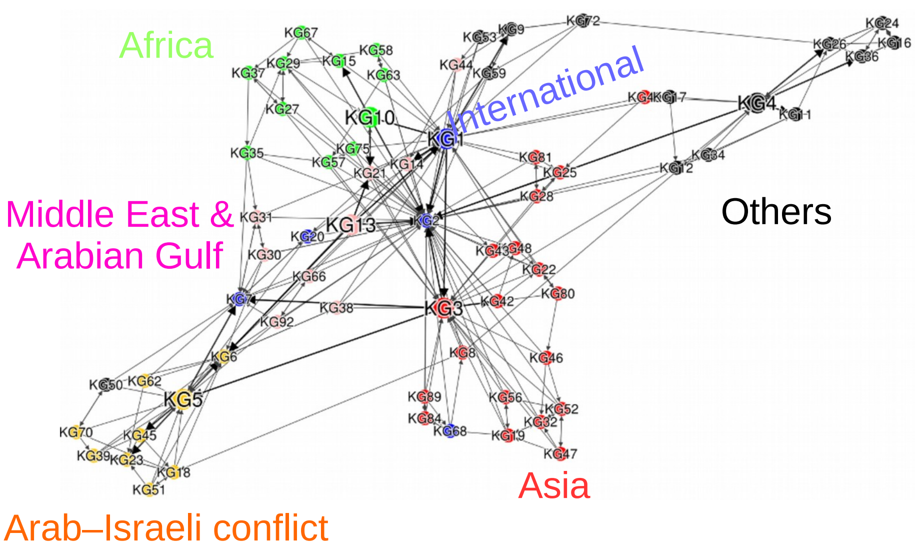
Automatically extracted annotation: Filamins are actin-binding proteins, related to cytoskeleton and involved in Larsen syndrome and Boomerang dysplasia

ApliGoogle
Applications of Google matrix to directed networks and Big Data

World Terror Networks from the reduced Google matrix of Wikipedia

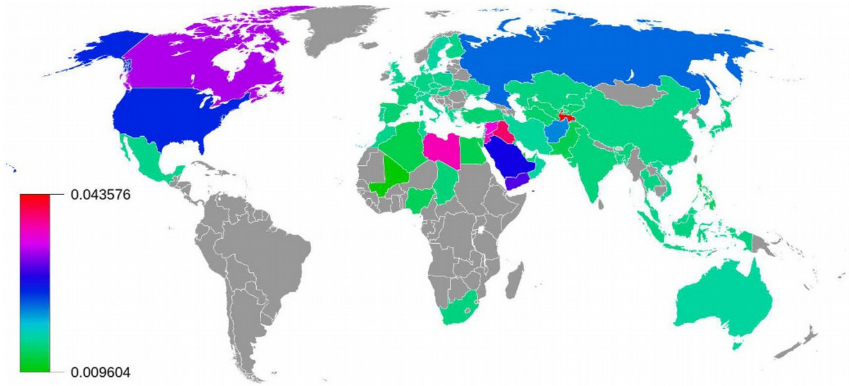
Reduced Google matrix of 95 terrorist group and 64 world countries

Interactions between terrorist groups

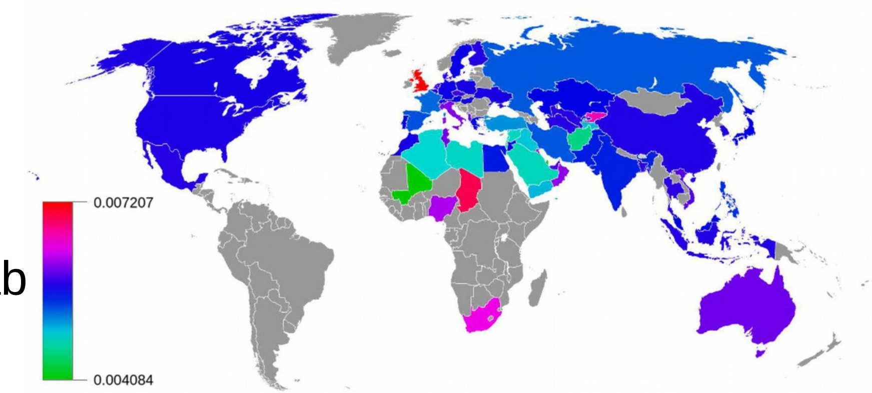


World map of the influence of terrorist groups on countries

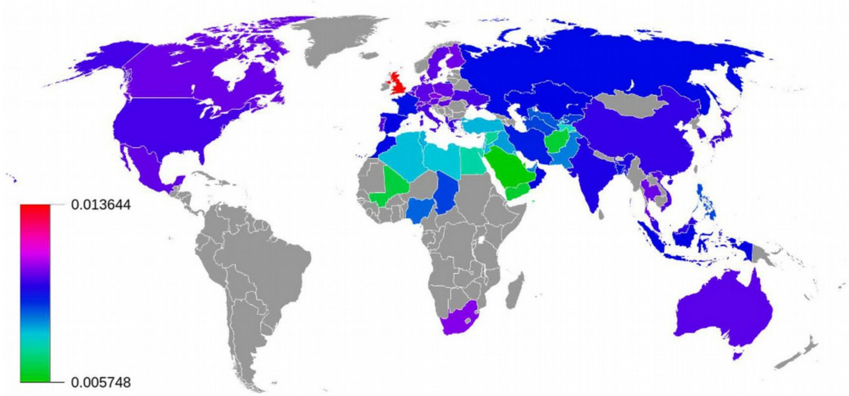
ISIS



Al Shabaab



IRA

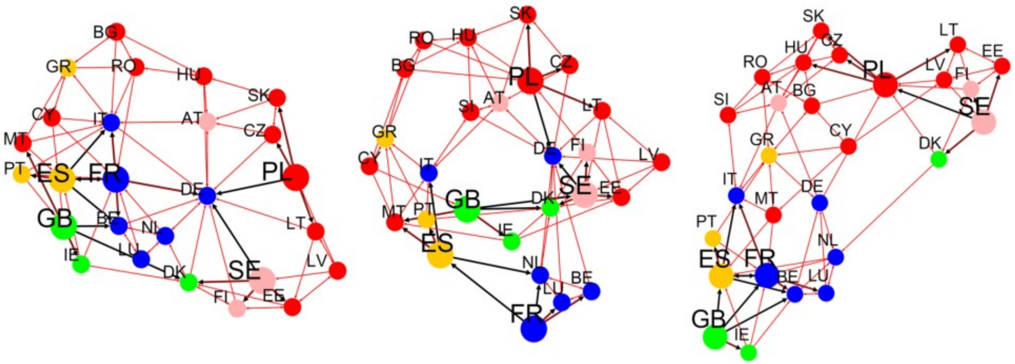


Samer El Zant, Klaus M. Frahm, Katia Jaffres-Runser,
Dima L. Shepelyansky, Eur. Phys. J. B v.91, p,7 (2018)

ApliGoogle

Applications of Google matrix to directed networks and Big Data

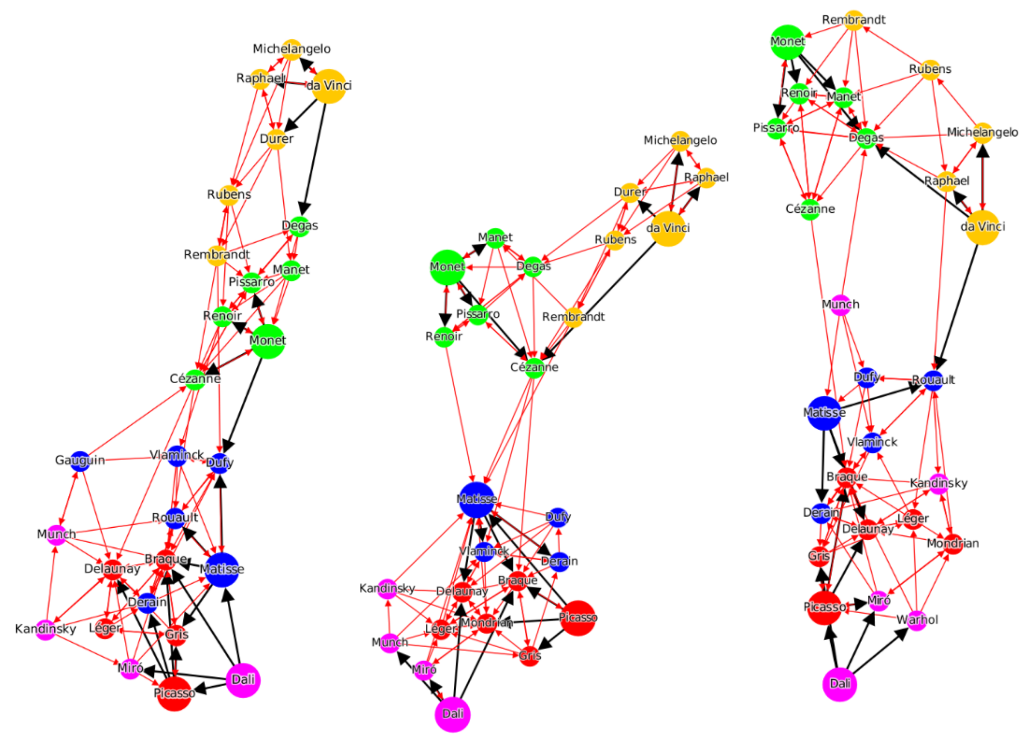
EU countries relationships according to Wikipedia



English edition French edition German edition

S. El Zant, K.Jaffres-Runser, and D.L.Shepelyansky, "Geopolitical interactions from reduced Google matrix analysis of Wikipedia", submitted.

Painters relationships according to Wikipedia



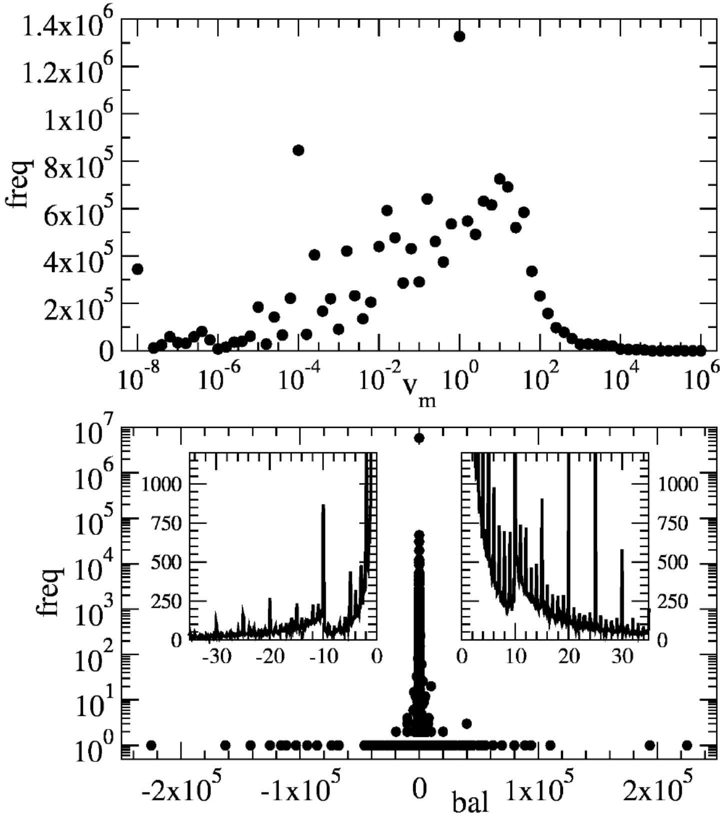
English edition French edition German edition

S. El Zant, K.Jaffres-Runser, K.M.Frahm, and D.L.Shepelyansky, "Reduced Google matrix analysis of Art: Painters", submitted.

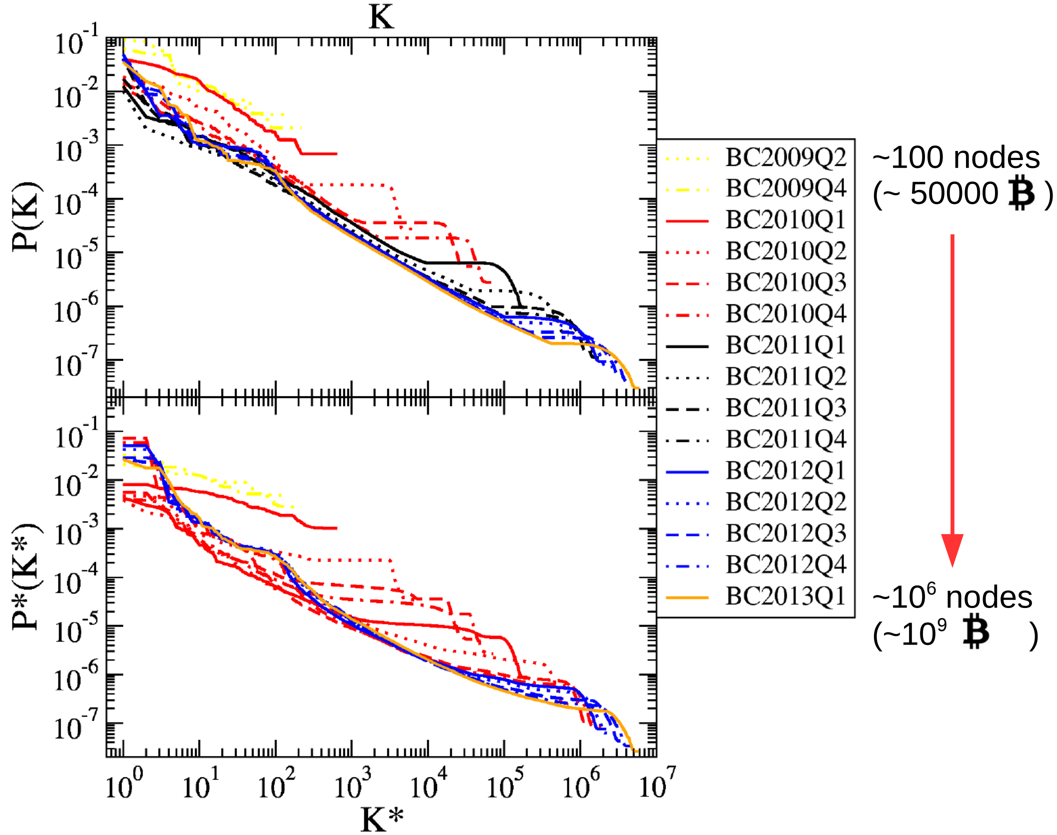
ApliGoogle
Applications of Google matrix to directed networks and Big Data

Google matrix of Bitcoin network

Frequency of Bitcoin transaction volume



Frequency of Bitcoin user balance



PageRank K and CheiRank K* distributions

L.Ermann, K.Frahm and D. Shepelyansky "Google matrix of Bitcoin networks", submitted to Eur. Phys. J. B (2017)

ApliGoogle

Applications of Google matrix to directed networks and Big Data

Planned further works

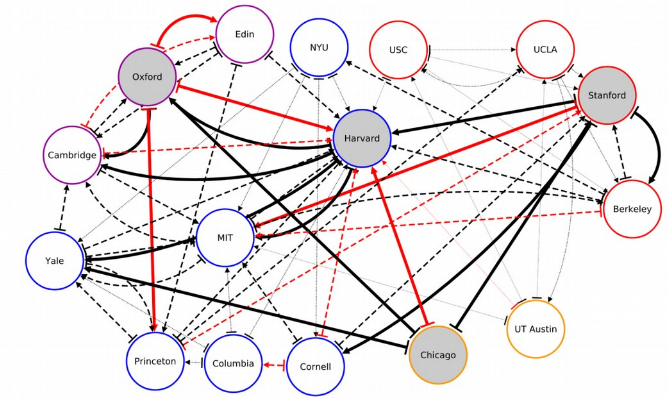
Planned further works

- **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.

- **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).

- **The reduced Google matrix method as a Cytoscape plug-in**
  The reduced Google matrix method will be available as Cytoscape plugin.

- **Googlomics : application of Google matrix to multifunctional network of omics (proteins, genes, …)**
  taking into account the nature of protein-protein interactions.



Hidden relationships between top 20 universities in Wikipedia **[preliminary result]**

Thank You !