Googlomics:
Reduced Google matrix analysis of directed biological networks

<u>José Lages</u>
jose.lages@utinam.cnrs.fr
Équipe de Physique Théorique, Institut UTINAM, CNRS, Université Bourgogne Franche-Comté, Besançon

Klaus Frahm, **Dima L. Shepelyansky**
Laboratoire de Physique Théorique, CNRS, Université Paul Sabatier, Toulouse

Andrei Zinovyev
Institut Curie, Inserm, PSL Université, Paris

SophI.A summit
Université Nice Sophia Antipolis
November 20 2020

Institut UTINAM

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

Projects

ApliGoogle project (2016-2018) funded by MASTODONS CNRS Mission interdisciplinarité
Partners : LPT, CNRS, UPS, Toulouse / UTINAM, CNRS, UBFC, Besançon / I. Curie, Inserm, PSL, Paris / IRIT, CNRS, UPS, Toulouse

APEX project (2017-2020) funded by the Bourgogne Franche-Comté region council.

GNETWORKS project (2018-2021) funded by ISITE-UBFC (PIA).

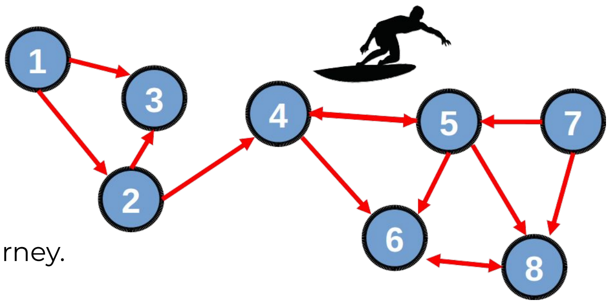
REpTILs project (2020-2023) funded by the Bourgogne Franche-Comté region council.
Partners: UTINAM, CNRS, UBFC, Besançon / IHGT, Inserm, UBFC, Besançon / PEPITE, UFC

Projects devoted to the physical analysis of complex networks and the application of Google matrix based analysis to complex systems.

How Google search engine works

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

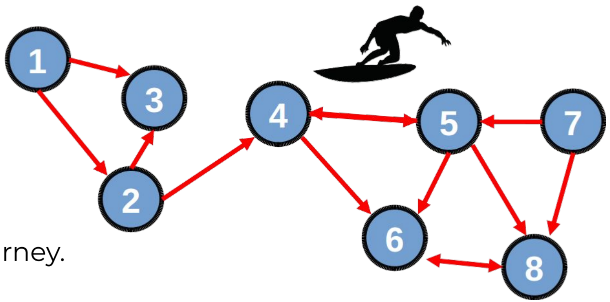
Markovian process : a random surfer probe the structure of a directed network.
A each step, the random surfer jumps randomly on an adjacent node and continue its journey.



Adjacency matrix

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

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



How Google search engine works

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

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

Adjacency matrix

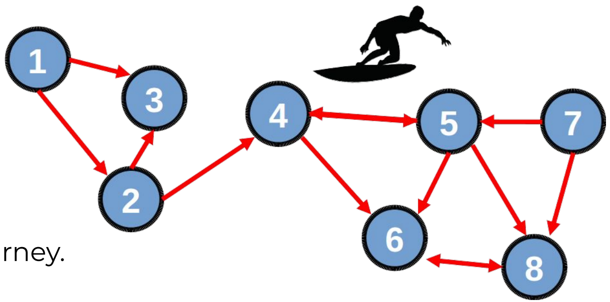
$$A_{ij} = \begin{cases} 1 \text{ si } j \rightarrow i \\ 0 \text{ si } j \nrightarrow 1 \end{cases}$$

Stochastic matrix

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

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

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



How Google search engine works

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

Markovian process : a random surfer probe the structure of a directed network.
A each step, the random surfer jumps randomly on an adjacent node 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

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

Google matrix

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

with $0.5 < \alpha < 1$

Perron-Frobenius operator

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

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

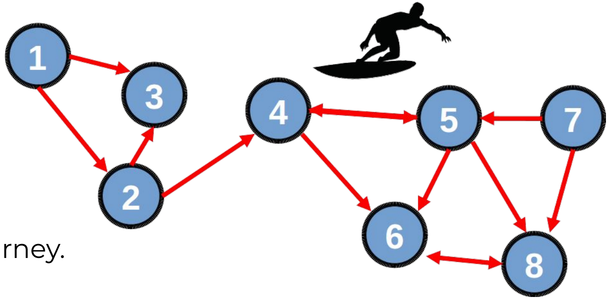
$$\mathbf{G} = \begin{pmatrix} 1/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 17/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 17/40&17/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 1/40&17/40&1/8&1/40&7/24&1/40&1/40&1/40 \\ 1/40&1/40&1/8&17/40&1/40&1/40&17/40&1/40 \\ 1/40&1/40&1/8&17/40&7/24&1/40&1/40&33/40 \\ 1/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40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

$$\alpha = 0.8$$

How Google search engine works

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

Markovian process : a random surfer probe the structure of a directed network.
A each step, the random surfer jumps randomly on an adjacent node 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

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

Google matrix

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

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

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

Steady-state

$$\mathbf{P} = \begin{pmatrix} 0.031094525687305 97 \\ 0.04353233614756617 \\ 0.06094527086606558 \\ 0.06729412361797826 \\ 0.07044998599586171 \\ 0.3518167935609448 9 \\ 0.03109452568730597 \\ 0.34377243843697143 \end{pmatrix}$$

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

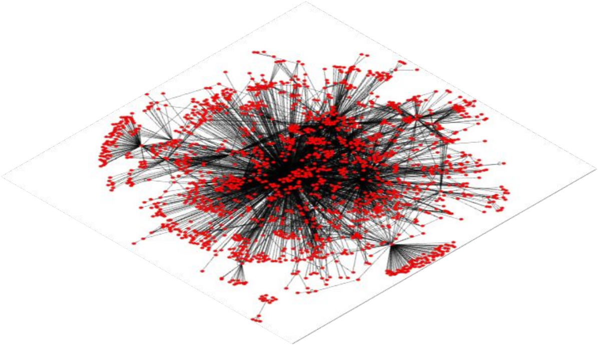
$P(1) = 0.35181679356094489$ 6
$P(2) = 0.34377243843697143$ 8
$P(3) = 0.07044998599586171$ 5
$P(4) = 0.06729412361797826$ 4
$P(5) = 0.06094527086606558$ 3
$P(6) = 0.04353233614756617$ 2
$P(7) = P(8) = 0.03109452568730597$ 1 7

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

The reduced Google matrix

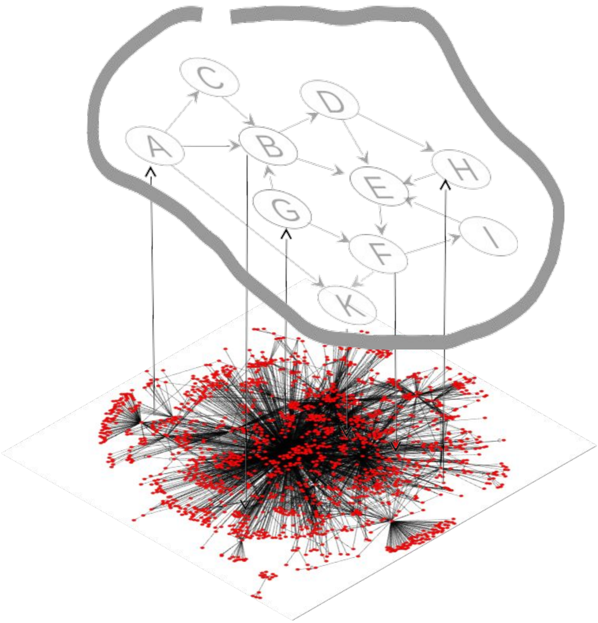
Let us consider a very large network with $N \gg 1$.



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

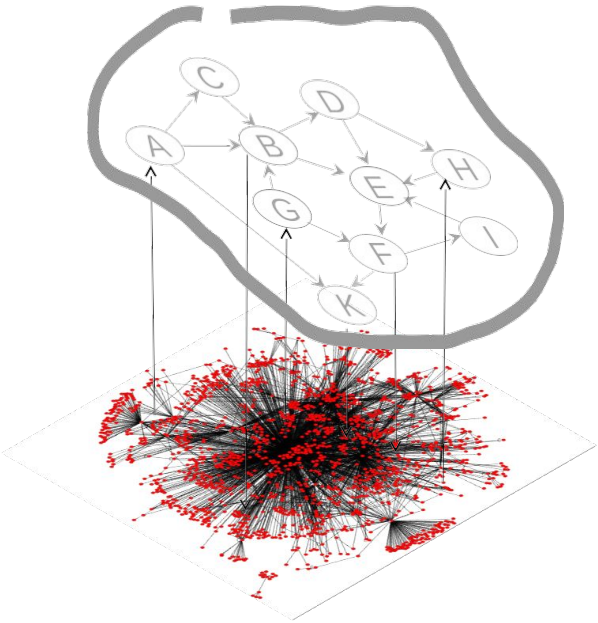
Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

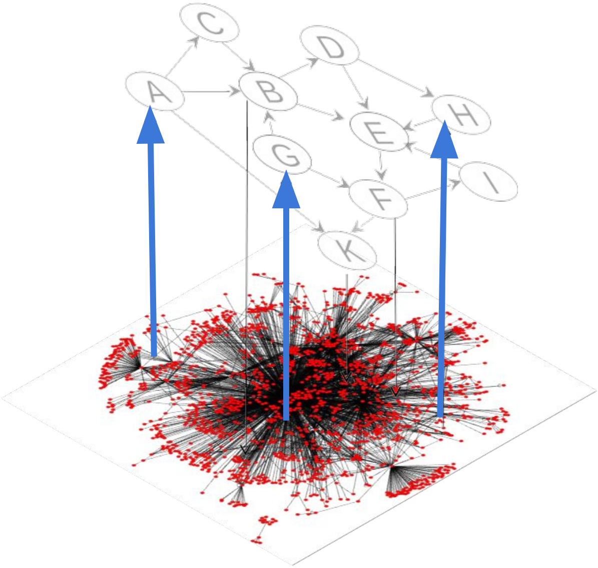


$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

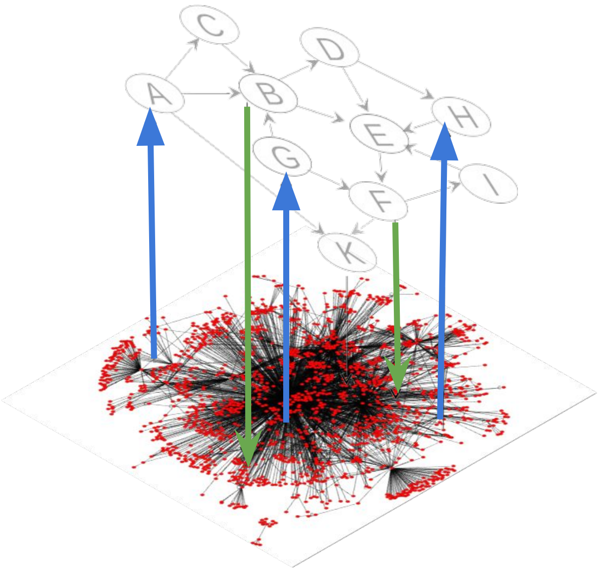


$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as



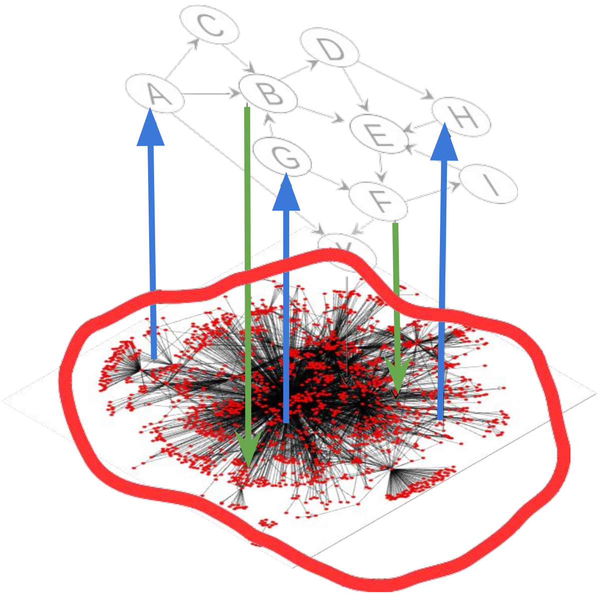
$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_{r} \\ \mathbf{P}_{s} \end{pmatrix}$$

J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

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



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

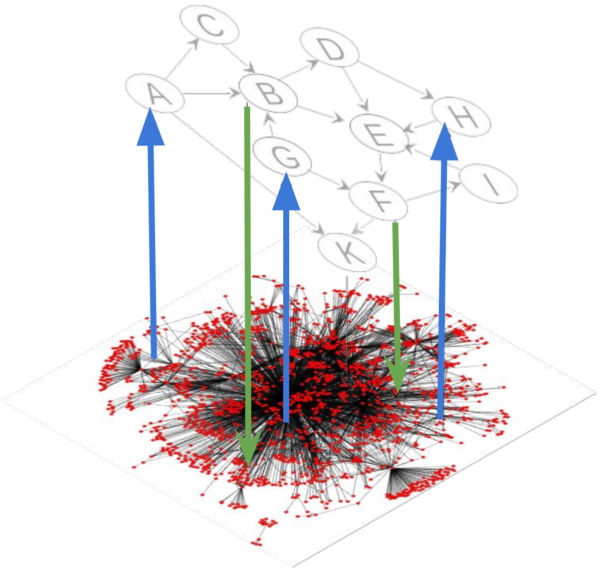
$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

For the global matrix, we have

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

We define the reduced Google matrix G_R associated to the N_r-size subset of interest such as

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



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

For the global matrix, we have

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

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

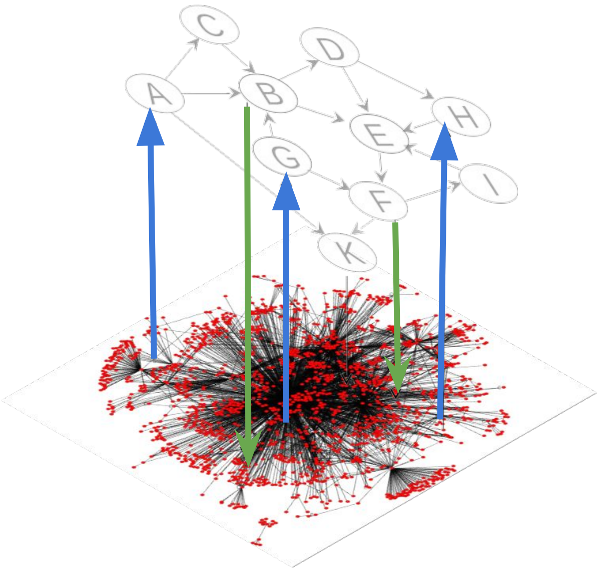
$$\mathbf{G}_R \mathbf{P}_r = \mathbf{P}_r$$

The reduced Google matrix can be written as

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

Contributions from **direct links**

Contributions from **indirect links** (*scattering terms*)



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

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

J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

For the global matrix, we have

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

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

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

The reduced Google matrix can be written as

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

Contributions from **direct links**

Contributions from **indirect links** (*scattering terms*)

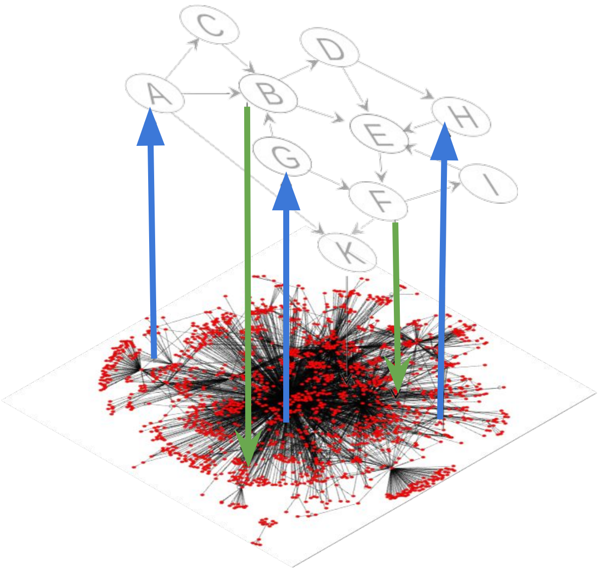
Projection onto the subspace associated to the leading eigenvalue $\lambda \approx 1$ with

$$(1 - \mathbf{G}_{ss})^{-1} = (1-\lambda)^{-1} \pi + \pi_c \sum_{l=0}^{\infty} (\pi_c \mathbf{G}_{ss} \pi_c)^l$$

$$\pi \mathbf{G}_{ss} \pi = \lambda \pi$$

$$\pi_c = \mathbf{1} - \pi$$

Contribution from the **"PageRank"** since $\pi \approx (\mathbf{P}\mathbf{P}\ldots\mathbf{P})$



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The Google matrix of the size N network and the associated PageRank vector can be written as

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

For the global matrix, we have

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

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

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

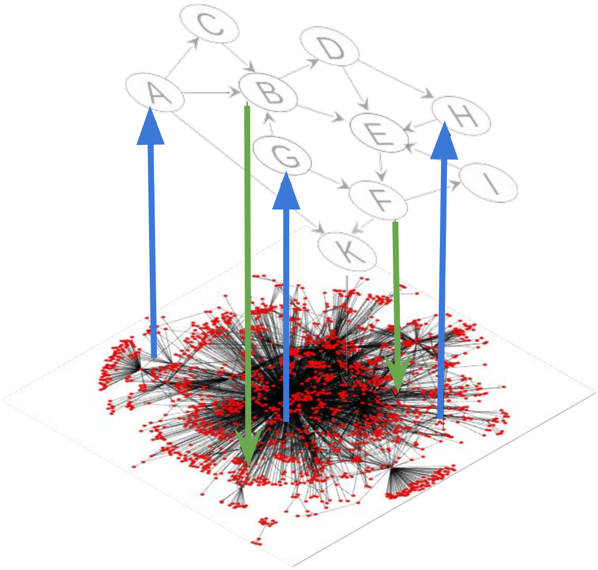
The reduced Google matrix can be written as

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

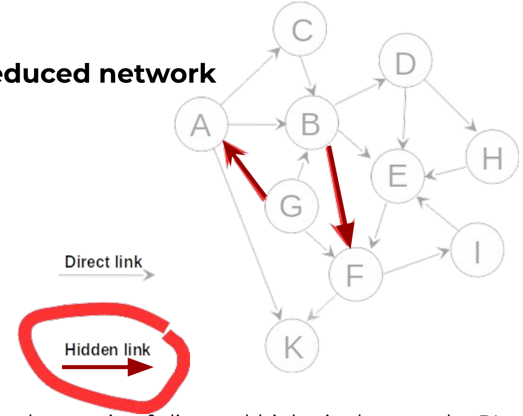
Contributions from **direct links**

Contributions from the **PageRank**

Contributions from **hidden links**



The reduced network



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

The reduced Google matrix

Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The reduced Google matrix can be written as

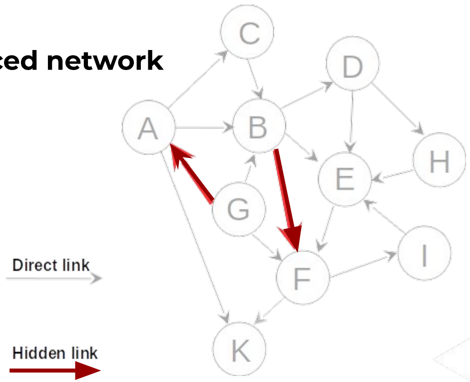
$$\mathbf{G}_R = \mathbf{G}_{rr} + \mathbf{G}_{pr} + \mathbf{G}_{qr}$$

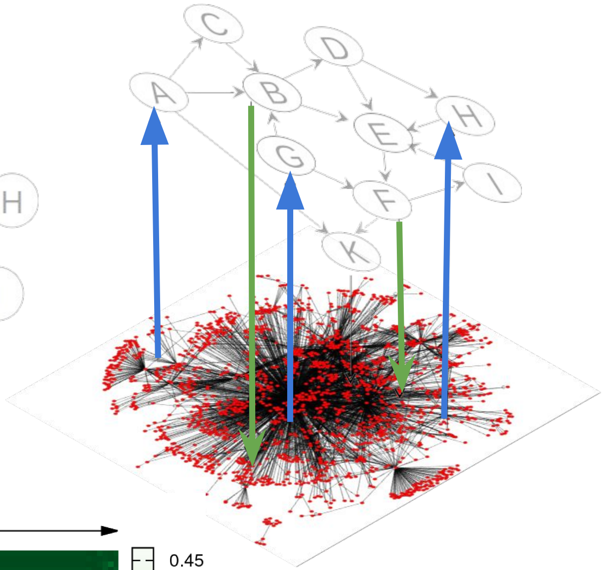
Contributions from **direct links**

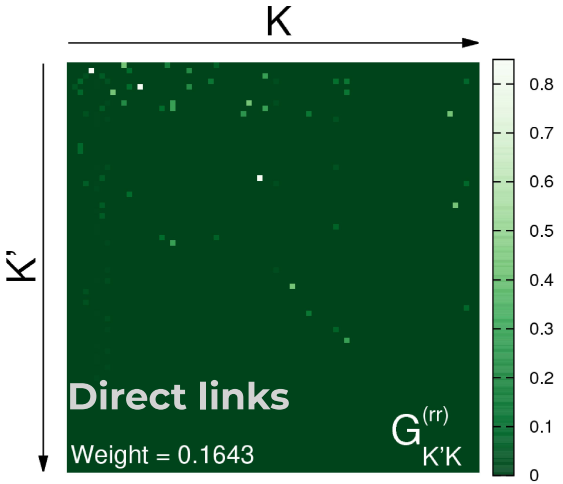
Contributions from **hidden links**

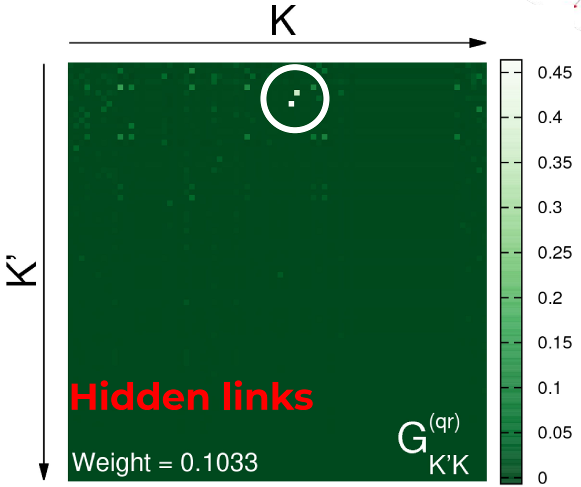
Contributions from the **PageRank**

The reduced network





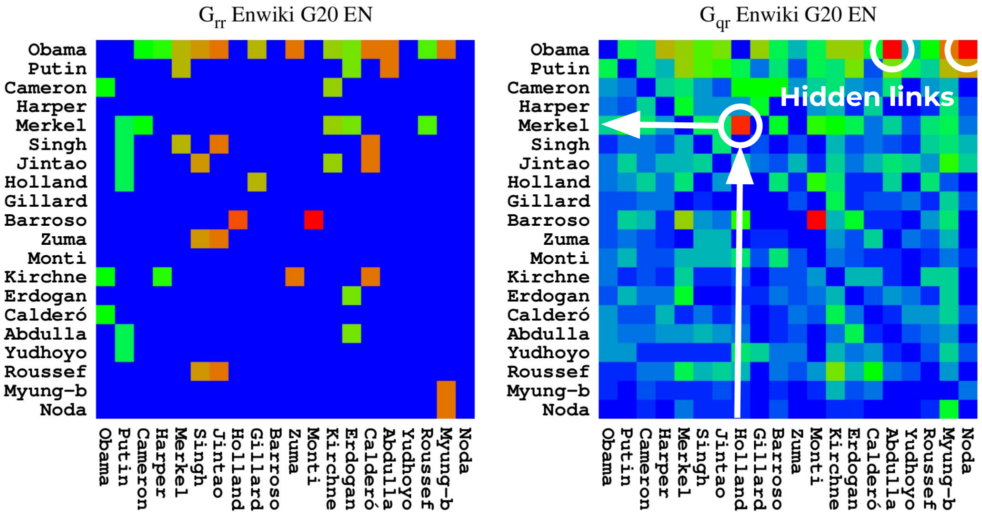




J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

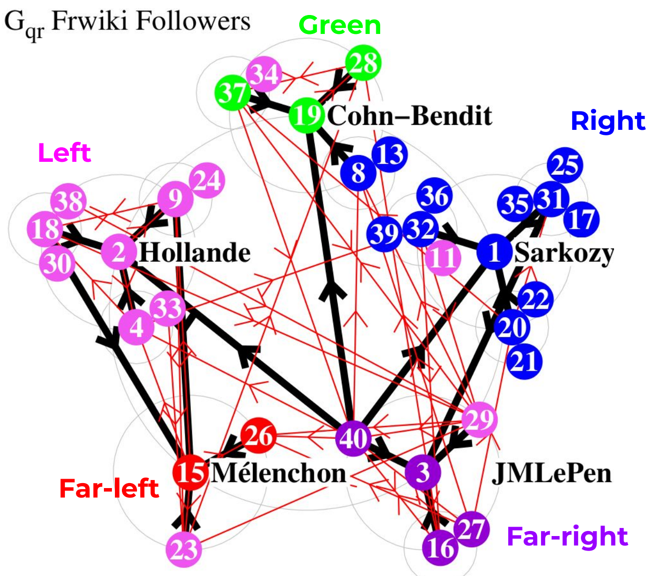
Proof of concept with Wikipedia as a complex network

Hidden links between political leaders



Analysis of hidden links between 2012 **G20 leaders** from the English edition Wikipedia (extracted in 2013)

El Zant, S., Frahm, K.M., Jaffrès-Runser, K. et al. *Analysis of world terror networks from the reduced Google matrix of Wikipedia*. Eur. Phys. J. B 91, 7 (2018)



Analysis of hidden links between 2013 **French politics** from the French edition Wikipedia (extracted in 2013)

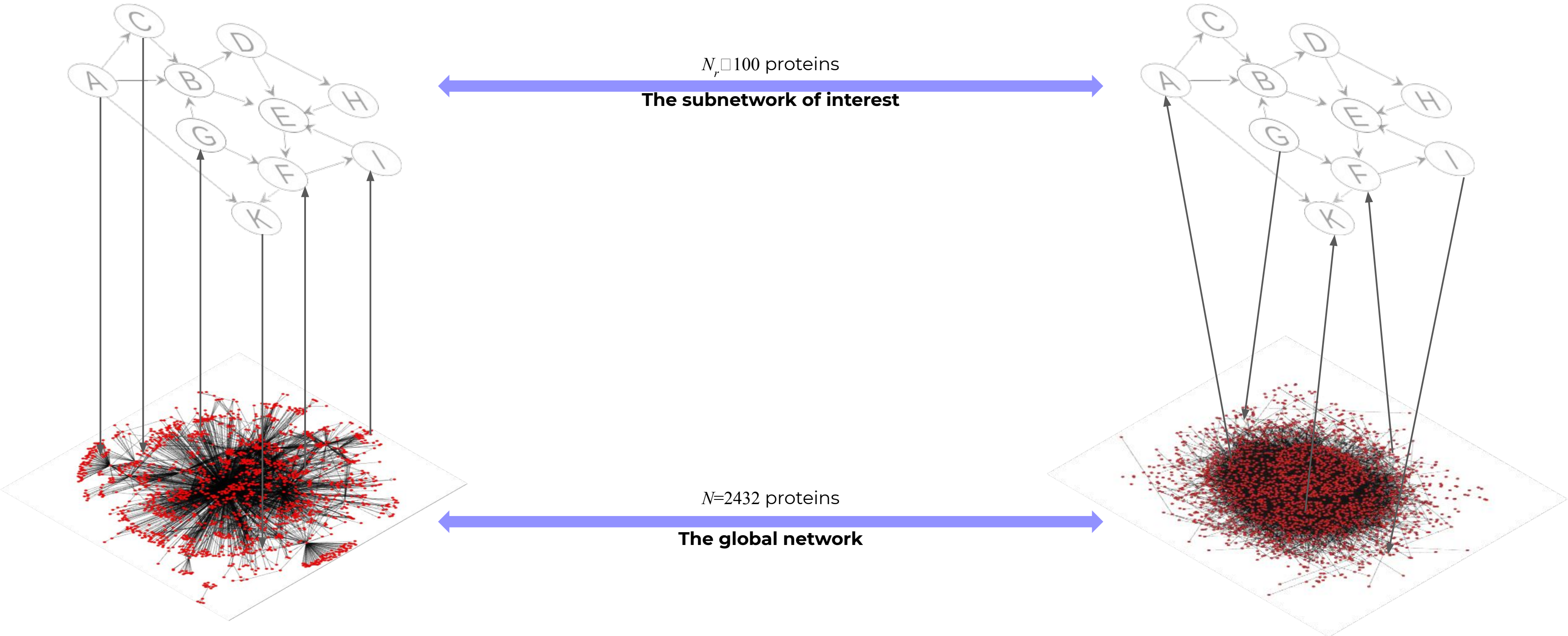
We retrieve knowledge about known political acquaintances (not trivially stated in Wikipedia).

The reduced Google matrix approach was also used for the **network analysis** of:

terrorist groups, **pharmaceutical groups,** **infectious diseases,** (within Wikipedia)

bitcoin transactions, **the world trade, …** (within corresp. economical networks)

Googlomics : Inferring hidden causal relations between proteins



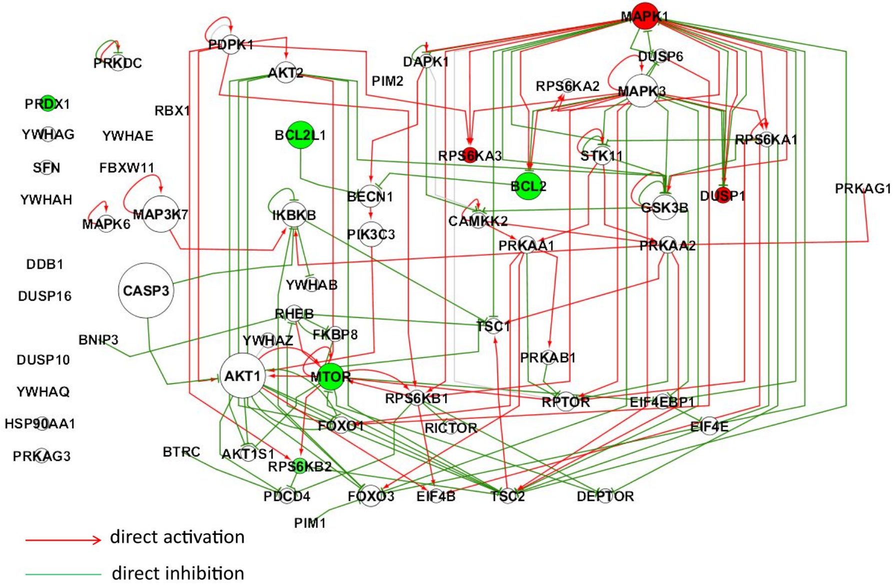
TRN1 ("normal")
Normal B-Lymphocytes

TRN2 ("cancer")
Leukemia cell line

J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)

Googlomics : Inferring hidden causal relations between proteins

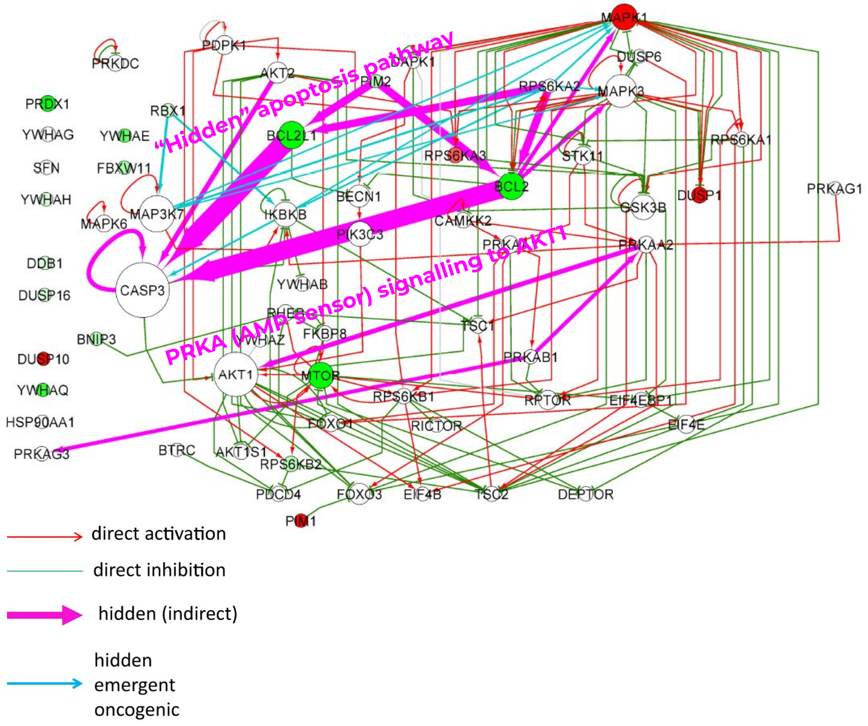
Inferring indirect (hidden) causal connections between **AKT-mTOR pathway members** (subnetwork of 63 proteins)

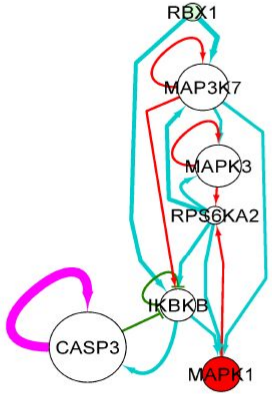


J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)

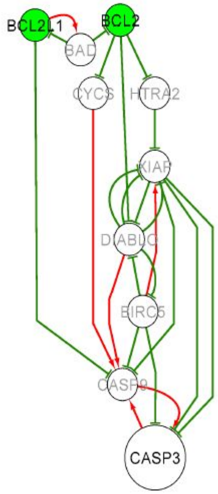
Googlomics : Inferring hidden causal relations between proteins

Inferring indirect (hidden) causal connections between **AKT-mTOR pathway members** (subnetwork of 63 proteins)





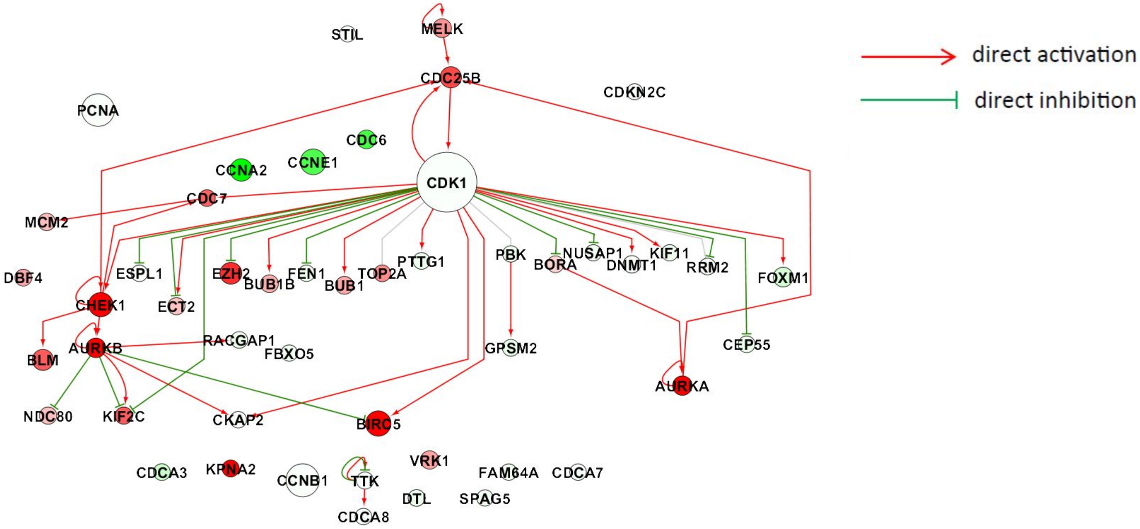
Emergent oncogenic signaling between **RBX1** (cell cycle protein degradation proteasome) and **MAPK1**.



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)

Googlomics : Inferring hidden causal relations between proteins

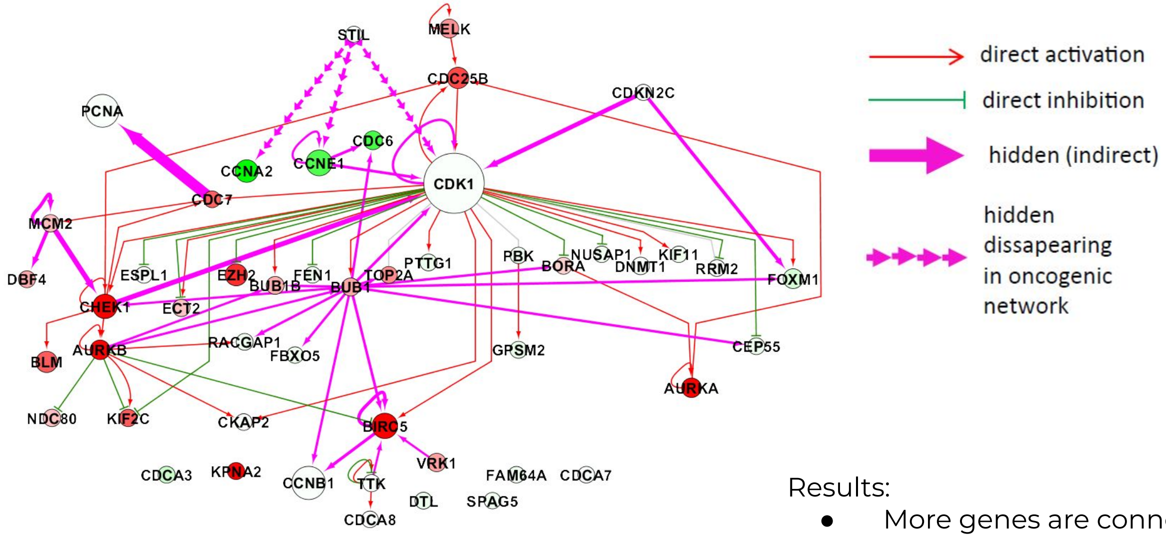
Genes of a proliferative signature resulted from pancancer transcriptomic analysis (subnetwork of 49 proteins)



J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)

Googlomics : Inferring hidden causal relations between proteins

Genes of a proliferative signature resulted from pancancer transcriptomic analysis (subnetwork of 49 proteins)



Results:

- 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, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)

Take home messages

- **Reduced Google Matrix: analytical approach for inferring hidden indirect connections within a set of nodes embedded in a very large network**
- **In the case of the proteome, hidden signaling pathways can be detected**
- **Structural changes in transcriptional network lead to implicit rewiring of pathways in cancer**
  - **Emergence of oncogenic pathways**
  - **Disappearance of hidden indirect connections in oncogenic networks**
- **Upstream from an AI treatment, the reduced Google matrix can considerably reduce the size of very large networks**

Thank you for your attention !!

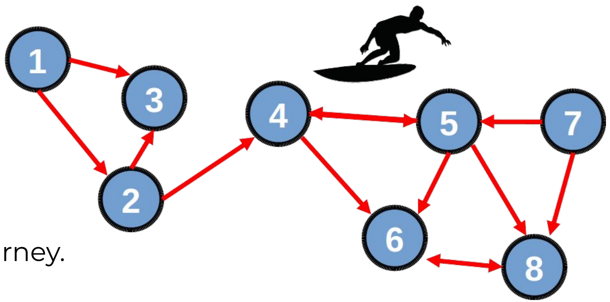
Main references:

J. Lages, D. Shepelyansky, A. Zinovyev,
Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks,
PLoS ONE 13(1): e0190812 (2018)

K. M. Frahm, and D. L. Shepelyansky,
Reduced Google Matrix,
arXiv:1602.02394

How Google search engine works

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



Markovian process : a random surfer probe the structure of a directed network.
A each step, the random surfer jumps randomly on an adjacent node 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

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

Google matrix

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

with $0.5 < \alpha < 1$

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

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

Steady-state

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

CheiRank vector $\mathbf{P}^* = \mathbf{G}^*\mathbf{P}^*$
Similar to the PageRank vector for the network with inverted links. 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 points toward important nodes, the more it is important.

The CheiRank measures the diffusion/the communication of a node.