Statistical properties of the MetaCore network of protein-protein interactions

Ekaterina Kotelnikova
Clarivate Analytics, Barcelona, Spain

José Lages
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

French Regional Conference on Complex Systems
FRCCS 2021
Dijon, France

Projects

ApliGoogle
Applications of Google matrix to directed networks and Big Data

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
Analyse Physique des résEaux compleXes

REGION
BOURGOGNE
FRANCHE
COMTE

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

GNETWORKS
Google Matrix Analysis of Real Complex Networks

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

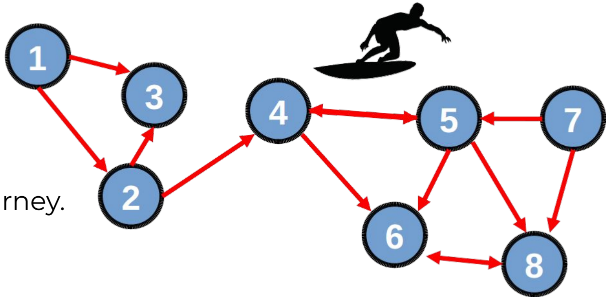
INVESTIR L'AVENIR
UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

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.

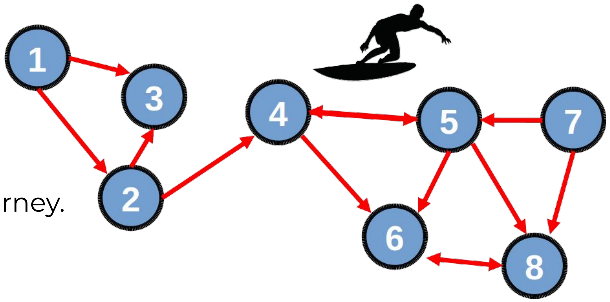
Adjacency matrix

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

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

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

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

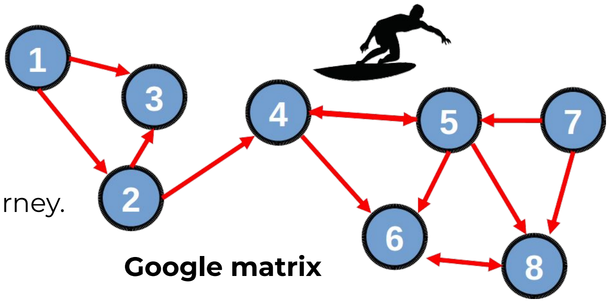
$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

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

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

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

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

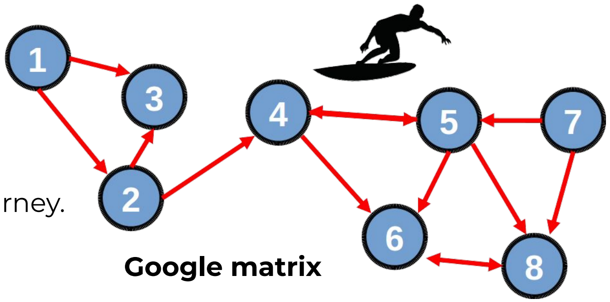
$$\mathbf{G} = \begin{pmatrix} 1/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 17/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 17/40&17/40&1/8&1/40&1/40&1/40&1/40&1/40 \\ 1/40&17/40&1/8&1/40&7/24&1/40&1/40&1/40 \\ 1/40&1/40&1/8&17/40&1/40&1/40&17/40&1/40 \\ 1/40&1/40&1/8&17/40&7/24&1/40&1/40&33/40 \\ 1/40&1/40&1/8&1/40&1/40&1/40&1/40&1/40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

$$\alpha = 0.8$$

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

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

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

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

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

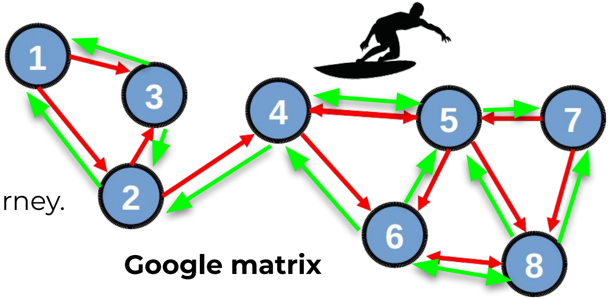
$P(1) = 0.35181679356094489$ ⑥
$P(2) = 0.34377243843697143$ ⑧
$P(3) = 0.07044998599586171$ ⑤
$P(4) = 0.06729412361797826$ ④
$P(5) = 0.06094527086606558$ ③
$P(6) = 0.04353233614756617$ ②
$P(7) = P(8) = 0.03109452568730597$ ① ⑦

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).

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

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

$P_i^{(n)}$ 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

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

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.

Statistical properties of the multi-omic MetaCore network

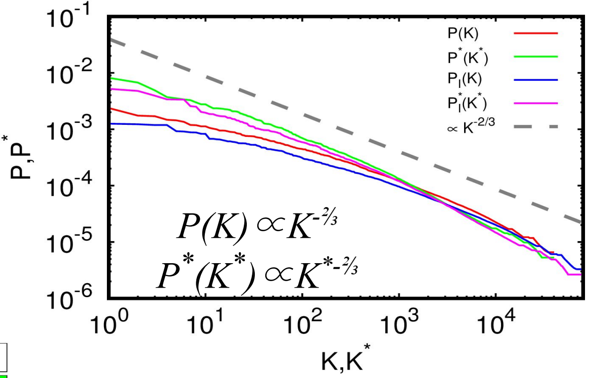
Top 40 PageRank
“Process outcomes”

K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol $K^+ \rightarrow$ extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

Top 40 CheiRank
“Process triggers”

K^*	$P^*(K^*)$ (10^{-2})	k^*	$M(K^*)$	Name	Class	Localization
1	1.1464	1	0.006332	c-Myc	Transcription factor	Nucleus
2	0.8172		0.035667	eIF2C2 (Argonaute-2)	Generic enzyme	Cytoplasm
3	0.6722		-0.174071	IGF2BP3	Generic binding protein	Cytoplasm
4	0.4890		0.680968	Ubiquitin	Generic binding protein	Cytoplasm
5	0.3719		0.110759	SOX9	Transcription factor	Nucleus
6	0.3529	2	-0.028308	p53	Transcription factor	Nucleus
7	0.3373		0.228978	c-Fos	Transcription factor	Nucleus
8	0.3276		0	CUX1 (p110)	Transcription factor	Nucleus
9	0.2989		-0.057557	SP1	Transcription factor	Nucleus
10	0.2770	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.2769	4	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
12	0.2534		-0.010911	eIF2C1 (Argonaute-1)	Generic binding protein	Cytoplasm
13	0.2354		0.062046	Androgen receptor	Transcription factor	Nucleus
14	0.2350		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
15	0.2330		-0.075622	BRD4	Generic binding protein	Nucleus
16	0.2308		0.153950	Oct-3/4	Transcription factor	Nucleus
17	0.2259		-0.001577	PUM2	Generic binding protein	Cytoplasm
18	0.2239		0.188479	EZH2	Generic enzyme	Nucleus
19	0.2193		0.208146	p300	Generic enzyme	Nucleus
20	0.2072		-0.407833	TUG1	RNA	Cytoplasm
21	0.2072		-0.118501	E2F1	Transcription factor	Nucleus
22	0.2062		0	ASCC2	Generic binding protein	Nucleus
23	0.2005		0	LIMR	Generic receptor	Plasma membrane
24	0.1903		0.148471	BRG1	Generic enzyme	Nucleus
25	0.1871	5	0.250910	STAT3	Transcription factor	Nucleus
26	0.1811		0.381258	RBM24	Generic binding protein	Cytoplasm
27	0.1789		0.746981	SUMO-1	Generic binding protein	Nucleus
28	0.1728		0.140357	c-IAP2	Generic binding protein	Cytoplasm
29	0.1699		0.038221	HIF1A	Transcription factor	Nucleus
30	0.1677		0	Zn^{2+} cytosol	Inorganic ion	Cytosol
31	0.1623		-0.013644	CDK9	Protein kinase	Cytoplasm
32	0.1587		-0.223816	MeCP2	Generic binding protein	Nucleus
33	0.1533		-0.053592	ELAVL1 (HuR)	Generic binding protein	Nucleus
34	0.1497		0.120649	HDAC1	Generic enzyme	Nucleus
35	0.1473		-0.034082	BRD7	Generic binding protein	Nucleus
36	0.1452		0.131956	CREB1	Transcription factor	Nucleus
37	0.1449		0	Zn^{2+} nucleus	Inorganic ion	Nucleus
38	0.1423		0.096830	SUMO-2	Generic binding protein	Cytoplasm
39	0.1400		-0.051730	BRD2	Protein kinase	Cytoplasm
40	0.1343		0.228824	C/EBPbeta	Transcription factor	Nucleus

Statistical properties of the multi-omic MetaCore network



The PageRank $P(K)$ and CheiRank $P^*(K^*)$ probabilities have similar distributions $P(K)\propto K^{-2/3}$ and $P(K^*)\propto K^{*-2/3}$. On average, **equal importance of ingoing and outgoing links** in the MetaCore network.

Top 40 PageRank
“Process outcomes”

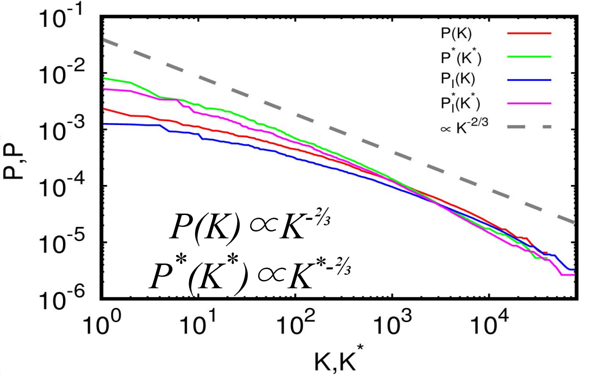
K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol K^+ → extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

Top 40 CheiRank
“Process triggers”

K^*	$P^*(K^*)$ (10^{-2})	k^*	$M(K^*)$	Name	Class	Localization
1	1.1464	1	0.006332	c-Myc	Transcription factor	Nucleus
2	0.8172		0.035667	eIF2C2 (Argonaute-2)	Generic enzyme	Cytoplasm
3	0.6722		-0.174071	IGF2BP3	Generic binding protein	Cytoplasm
4	0.4890		0.680968	Ubiquitin	Generic binding protein	Cytoplasm
5	0.3719		0.110759	SOX9	Transcription factor	Nucleus
6	0.3529	2	-0.028308	p53	Transcription factor	Nucleus
7	0.3373		0.228978	c-Fos	Transcription factor	Nucleus
8	0.3276		0	CUX1 (p110)	Transcription factor	Nucleus
9	0.2989		-0.057557	SP1	Transcription factor	Nucleus
10	0.2770	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.2769	4	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
12	0.2534		-0.010911	eIF2C1 (Argonaute-1)	Generic binding protein	Cytoplasm
13	0.2354		0.062046	Androgen receptor	Transcription factor	Nucleus
14	0.2350		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
15	0.2330		-0.075622	BRD4	Generic binding protein	Nucleus
16	0.2308		0.153950	Oct-3/4	Transcription factor	Nucleus
17	0.2259		-0.001577	PUM2	Generic binding protein	Cytoplasm
18	0.2239		0.188479	EZH2	Generic enzyme	Nucleus
19	0.2193		0.208146	p300	Generic enzyme	Nucleus
20	0.2072		-0.407833	TUG1	RNA	Cytoplasm
21	0.2072		-0.118501	E2F1	Transcription factor	Nucleus
22	0.2062		0	ASCC2	Generic binding protein	Nucleus
23	0.2005		0	LIMR	Generic receptor	Plasma membrane
24	0.1903		0.148471	BRG1	Generic enzyme	Nucleus
25	0.1871	5	0.250910	STAT3	Transcription factor	Nucleus
26	0.1811		0.381258	RBM24	Generic binding protein	Cytoplasm
27	0.1789		0.746981	SUMO-1	Generic binding protein	Nucleus
28	0.1728		0.140357	c-IAP2	Generic binding protein	Cytoplasm
29	0.1699		0.038221	HIF1A	Transcription factor	Nucleus
30	0.1677		0	Zn^{2+} cytosol	Inorganic ion	Cytosol
31	0.1623		-0.013644	CDK9	Protein kinase	Cytoplasm
32	0.1587		-0.223816	MeCP2	Generic binding protein	Nucleus
33	0.1533		-0.053592	ELAVL1 (HuR)	Generic binding protein	Nucleus
34	0.1497		0.120649	HDAC1	Generic enzyme	Nucleus
35	0.1473		-0.034082	BRD7	Generic binding protein	Nucleus
36	0.1452		0.131956	CREB1	Transcription factor	Nucleus
37	0.1449		0	Zn^{2+} nucleus	Inorganic ion	Nucleus
38	0.1423		0.096830	SUMO-2	Generic binding protein	Cytoplasm
39	0.1400		-0.051730	BRD2	Protein kinase	Cytoplasm
40	0.1343		0.228824	C/EBPbeta	Transcription factor	Nucleus

Statistical properties of the multi-omic MetaCore network

The PageRank $P(K)$ and CheiRank $P^*(K^*)$ probabilities have similar distributions $P(K) \propto K^{-2/3}$ and $P(K^*) \propto K^{*-2/3}$. On average, **equal importance of ingoing and outgoing links** in the MetaCore network.

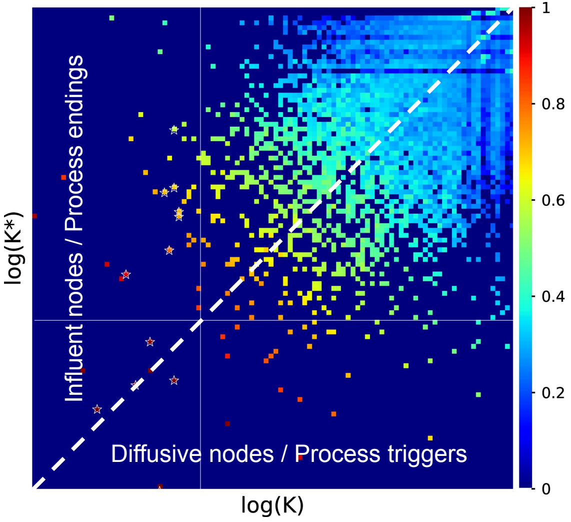


Top 40 PageRank
“Process outcomes”

K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol $K^+ \rightarrow$ extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

Top 40 CheiRank
“Process triggers”

K^*	$P^*(K^*)$ (10^{-2})	k^*	$M(K^*)$	Name	Class	Localization
1	1.1464	1	0.006332	c-Myc	Transcription factor	Nucleus
2	0.8172		0.035667	eIF2C2 (Argonaute-2)	Generic enzyme	Cytoplasm
3	0.6722		-0.174071	IGF2BP3	Generic binding protein	Cytoplasm
4	0.4890		0.680968	Ubiquitin	Generic binding protein	Cytoplasm
5	0.3719		0.110759	SOX9	Transcription factor	Nucleus
6	0.3529	2	-0.028308	p53	Transcription factor	Nucleus
7	0.3373		0.228978	c-Fos	Transcription factor	Nucleus
8	0.3276		0	CUX1 (p110)	Transcription factor	Nucleus
9	0.2989		-0.057557	SP1	Transcription factor	Nucleus
10	0.2770	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.2769	4	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
12	0.2534		-0.010911	eIF2C1 (Argonaute-1)	Generic binding protein	Cytoplasm
13	0.2354		0.062046	Androgen receptor	Transcription factor	Nucleus
14	0.2350		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
15	0.2330		-0.075622	BRD4	Generic binding protein	Nucleus
16	0.2308		0.153950	Oct-3/4	Transcription factor	Nucleus
17	0.2259		-0.001577	PUM2	Generic binding protein	Cytoplasm
18	0.2239		0.188479	EZH2	Generic enzyme	Nucleus
19	0.2193		0.208146	p300	Generic enzyme	Nucleus
20	0.2072		-0.407833	TUG1	RNA	Cytoplasm
21	0.2072		-0.118501	E2F1	Transcription factor	Nucleus
22	0.2062		0	ASCC2	Generic binding protein	Nucleus
23	0.2005		0	LIMR	Generic receptor	Plasma membrane
24	0.1903		0.148471	BRG1	Generic enzyme	Nucleus
25	0.1871	5	0.250910	STAT3	Transcription factor	Nucleus
26	0.1811		0.381258	RBM24	Generic binding protein	Cytoplasm
27	0.1789		0.746981	SUMO-1	Generic binding protein	Nucleus
28	0.1728		0.140357	c-IAP2	Generic binding protein	Cytoplasm
29	0.1699		0.038221	HIF1A	Transcription factor	Nucleus
30	0.1677		0	Zn^{2+} cytosol	Inorganic ion	Cytosol
31	0.1623		-0.013644	CDK9	Protein kinase	Cytoplasm
32	0.1587		-0.223816	MeCP2	Generic binding protein	Nucleus
33	0.1533		-0.053592	ELAVL1 (HuR)	Generic binding protein	Nucleus
34	0.1497		0.120649	HDAC1	Generic enzyme	Nucleus
35	0.1473		-0.034082	BRD7	Generic binding protein	Nucleus
36	0.1452		0.131956	CREB1	Transcription factor	Nucleus
37	0.1449		0	Zn^{2+} nucleus	Inorganic ion	Nucleus
38	0.1423		0.096830	SUMO-2	Generic binding protein	Cytoplasm
39	0.1400		-0.051730	BRD2	Protein kinase	Cytoplasm
40	0.1343		0.228824	C/EBPbeta	Transcription factor	Nucleus



Ising Metacore network: bi-functionality of the interactions

We take into account the nature of the interaction:

- either,
  j **activates** i
- or,
  j **inhibits** i.

Nodes are doubled, the node i is now i+ and i-.
The matrix element G_{ij}, associated to the link $j \to i$, is replaced by one of the following 2x2 matrices:

$$\sigma_+ = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$ if j **activates** i,

$$\sigma_- = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$ if j **inhibits** i.

$$\sigma_0 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

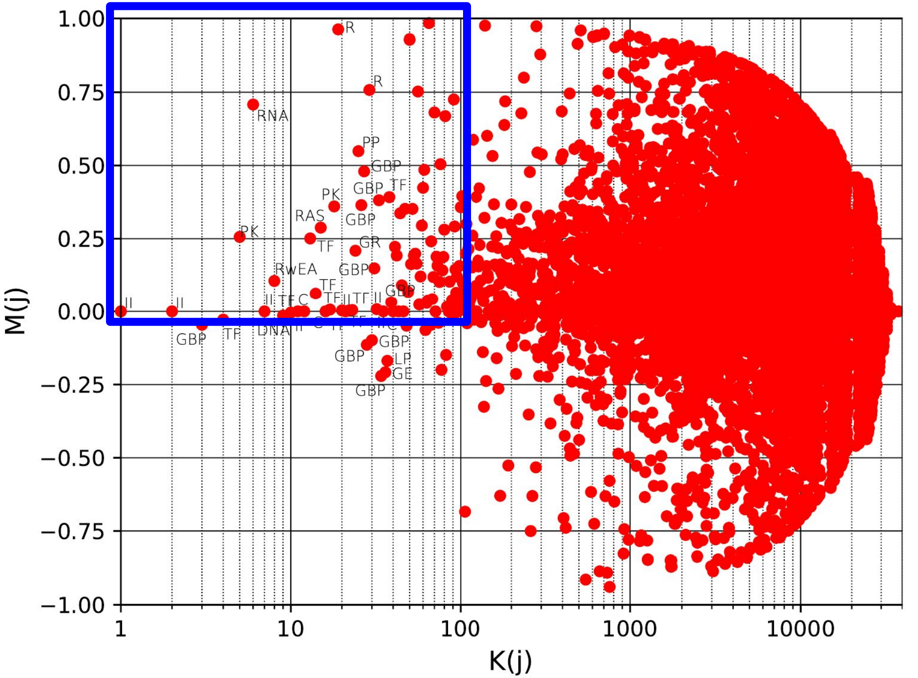
if interaction is **neutral** or **unknown**.

Consequently,
$P_+(i)$ is the probability that node i is activated,
and
$P_-(i)$ is the probability that node i is inhibited.

PageRank Magnetization

$$M(j) = \frac{P_+(j) - P_-(j)}{P_+(j) + P_-(j)}$$

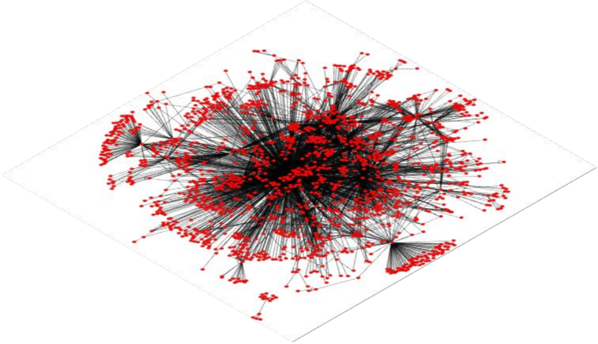
$$-1 \leq M(j) \leq 1$$



Most of the omics in **top 100 PageRank (K≤100) are more activated** (M>0) than inhibited.

The reduced Google matrix

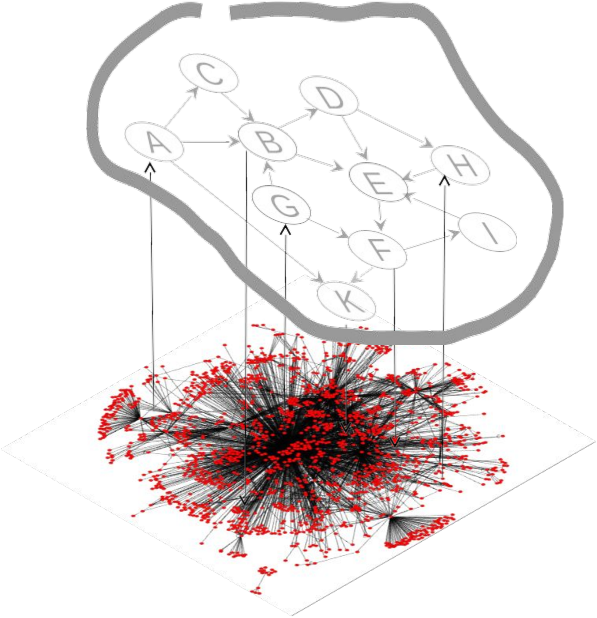
Let us consider a very large network with $N \gg 1$.



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

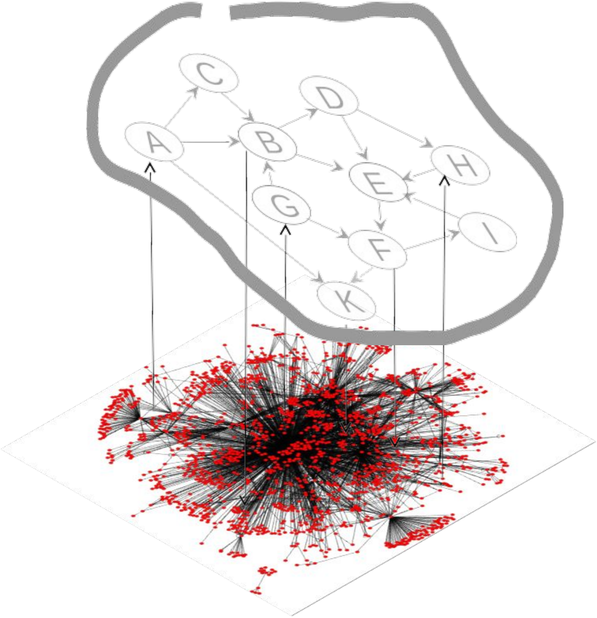
Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.



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

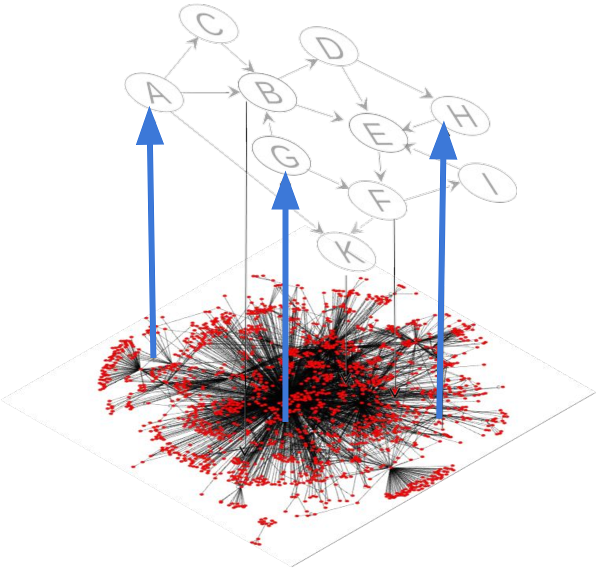


$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

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



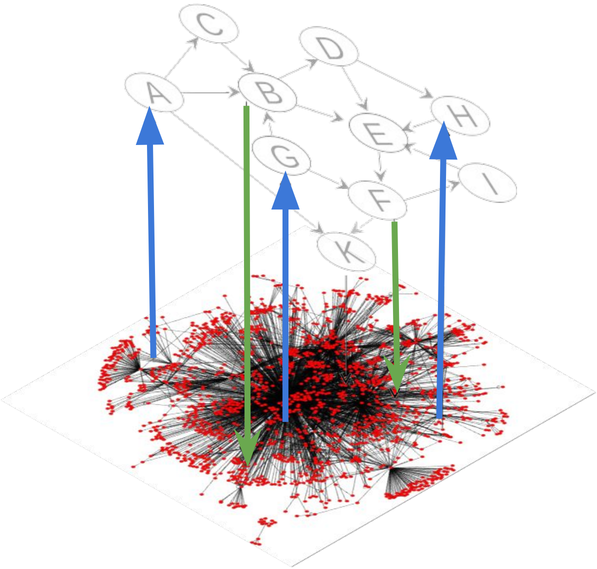
$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

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

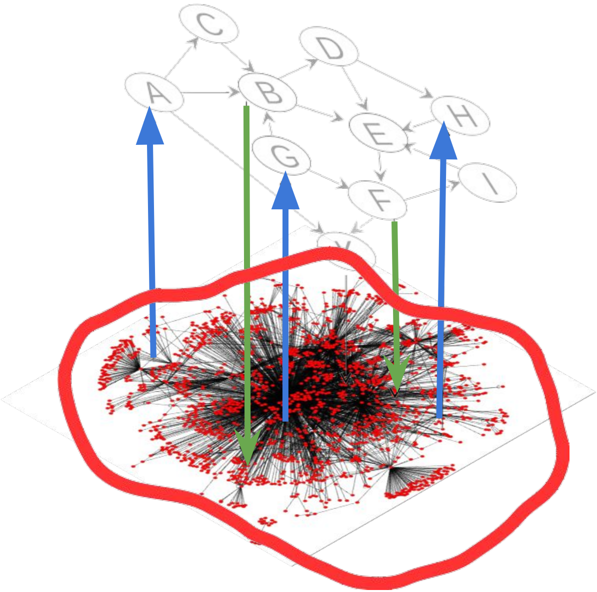
$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$



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

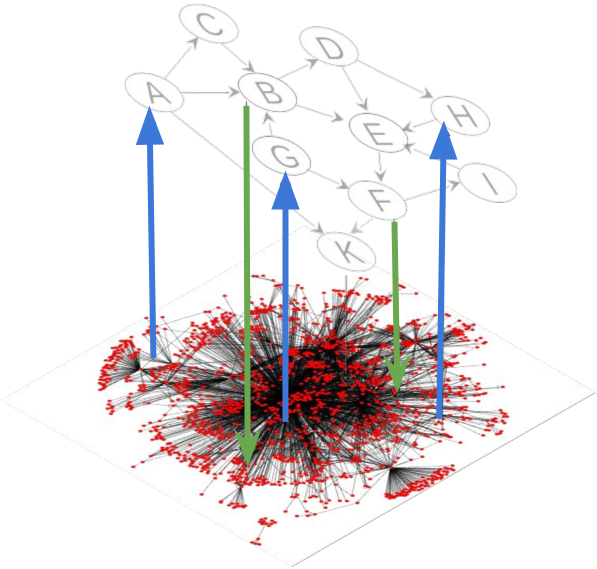
$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_{rr} & \mathbf{G}_{rs} \\ \mathbf{G}_{sr} & \mathbf{G}_{ss} \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} \mathbf{P}_r \\ \mathbf{P}_s \end{pmatrix}$$

For the global matrix, we have

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

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

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



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

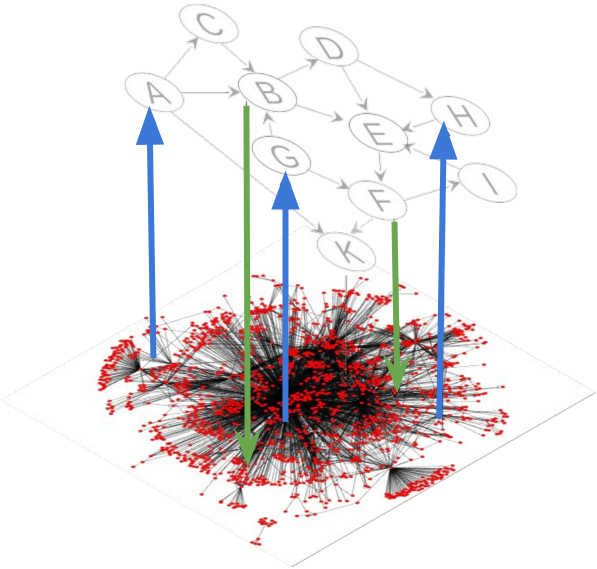
$$\mathbf{G}_R = \boxed{\mathbf{G}_{rr}} + \boxed{\mathbf{G}_{rs}\left(\mathbf{1} - \mathbf{G}_{ss}\right)^{-1}\mathbf{G}_{sr}}$$

Contributions
from **direct links**

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

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

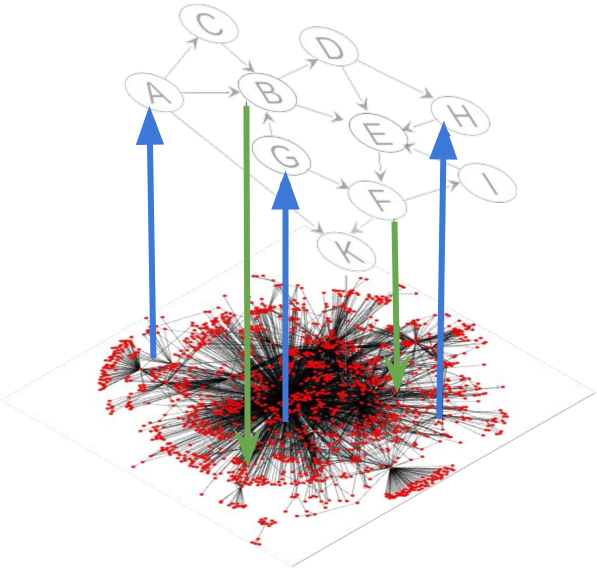
$$\left(\mathbf{1} - \mathbf{G}_{ss}\right)^{-1} = \sum_{l=0}^{\infty} \mathbf{G}_{ss}^{l}$$



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

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

Contributions
from **direct links**

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

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

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

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

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

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

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

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

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

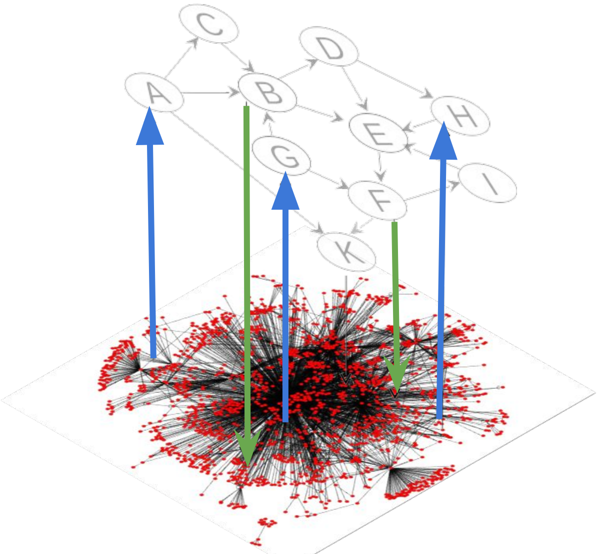
The reduced Google matrix can be written as

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

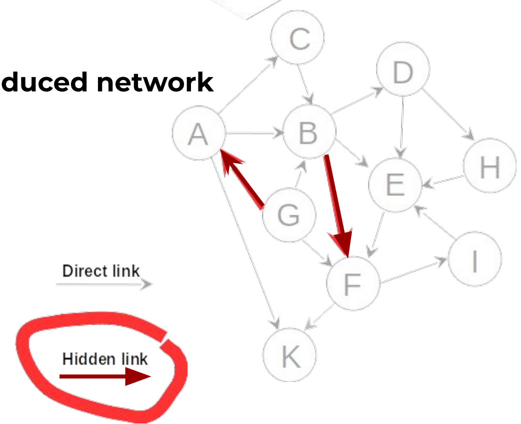
Contributions from **direct links**

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]

The reduced Google matrix

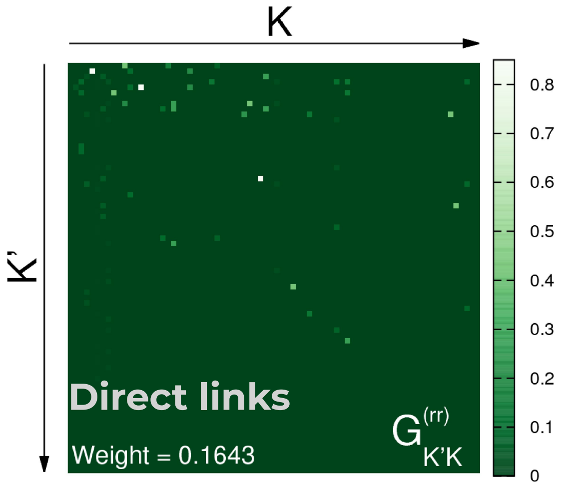
Let us consider a very large network with $N \gg 1$.
Consider a sub-network of $N_r \ll N$ nodes of interest.
The reduced Google matrix can be written as

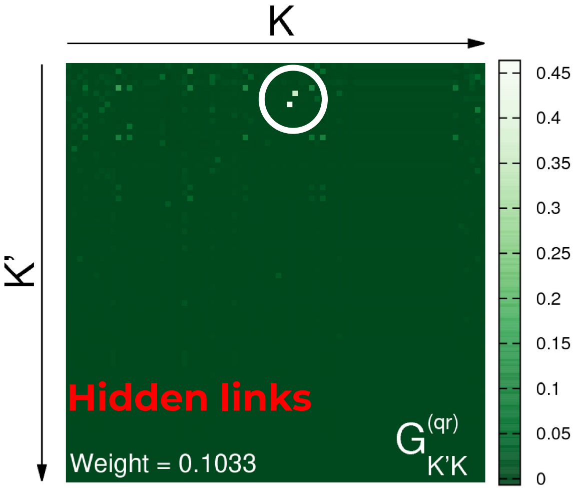
$$\mathbf{G}_R = \boxed{\mathbf{G}_{rr}} + \boxed{\mathbf{G}_{pr}} + \boxed{\mathbf{G}_{qr}}$$

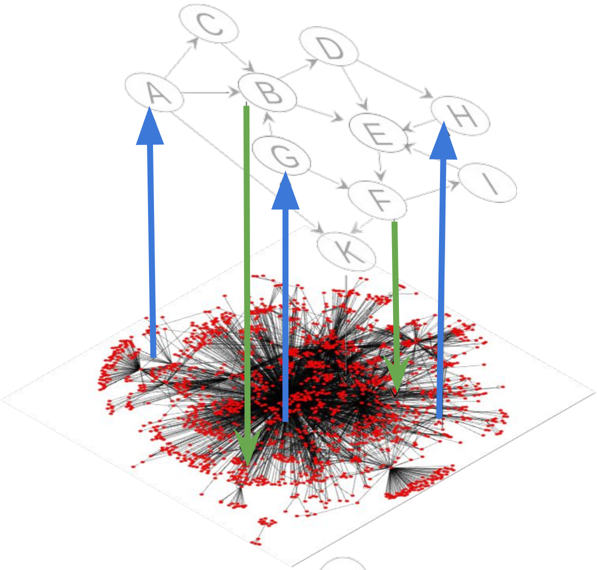
Contributions from **direct links**

Contributions from the **PageRank**

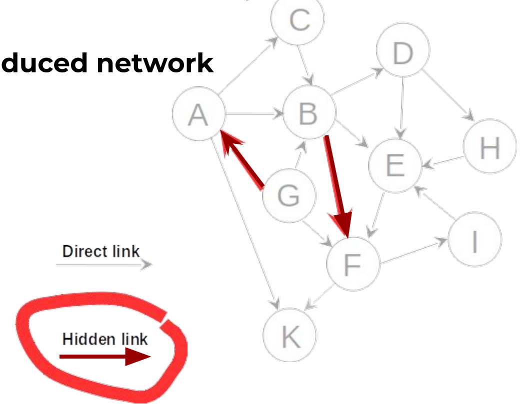
Contributions from **hidden links**







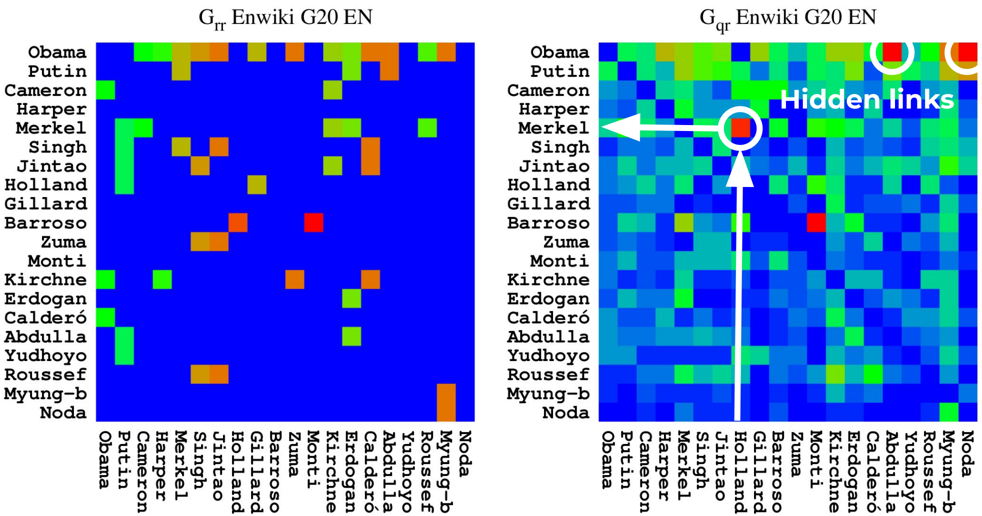
The reduced network



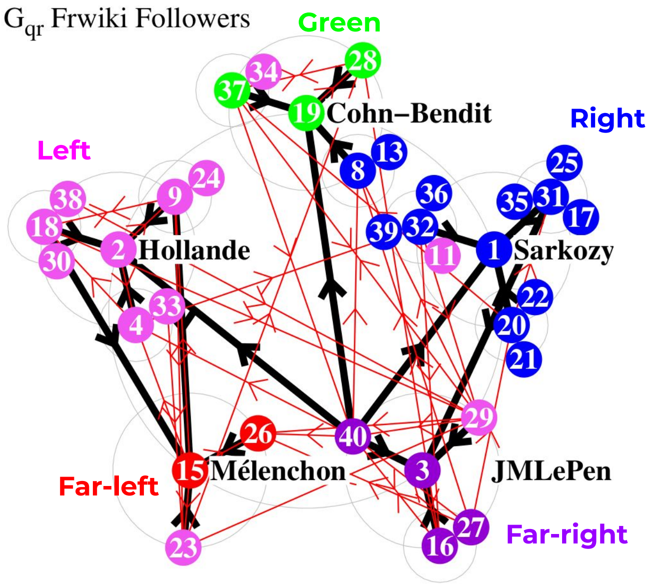
J. Lages, D. Shepelyansky, A. Zinovyev, *Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks*, PLoS ONE 13(1): e0190812 (2018)
K. M. Frahm, and D. L. Shepelyansky, *Reduced Google Matrix*, arXiv:1602.02394 [physics.soc-ph]

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)



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

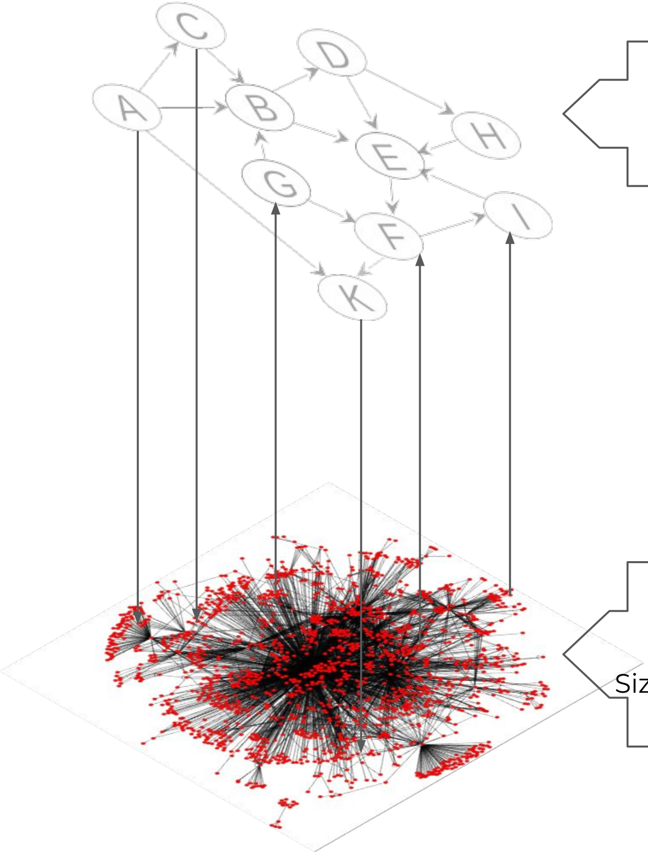
We retrieve knowledge about known political acquaintances (not trivially stated in Wikipedia).

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 analysis of the MetaCore multi-omic network



The subnetwork of interest

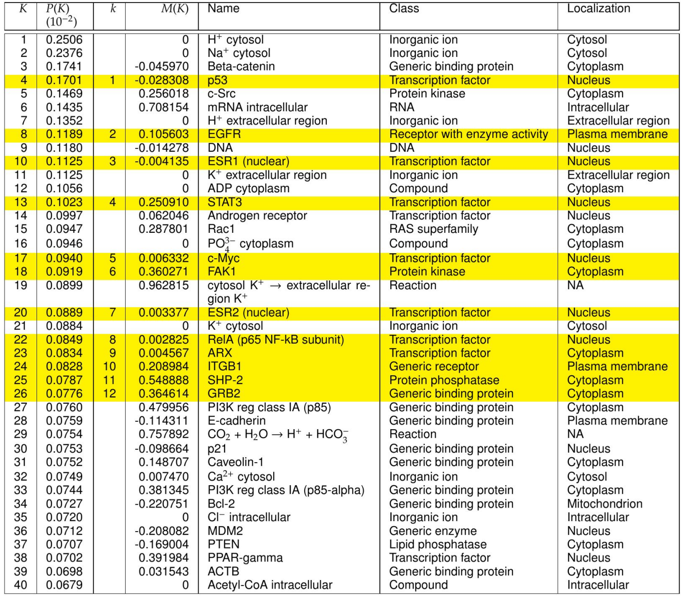
Size of $N_r \sim \mathbf{100}$ proteins

The MetaCore network
(multi omic database)

Size of $N \approx \mathbf{40\ 000}$ proteins, complexes, chemical elements, …

Googlomics analysis of the MetaCore multi-omic network

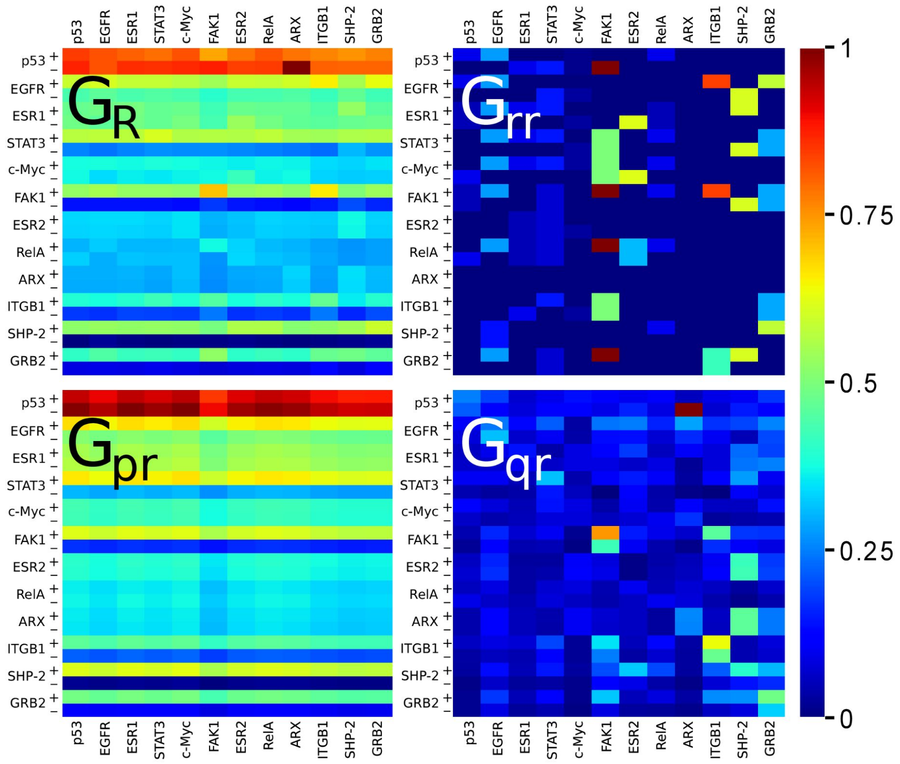
Top 40 PageRank
“Process outcomes”



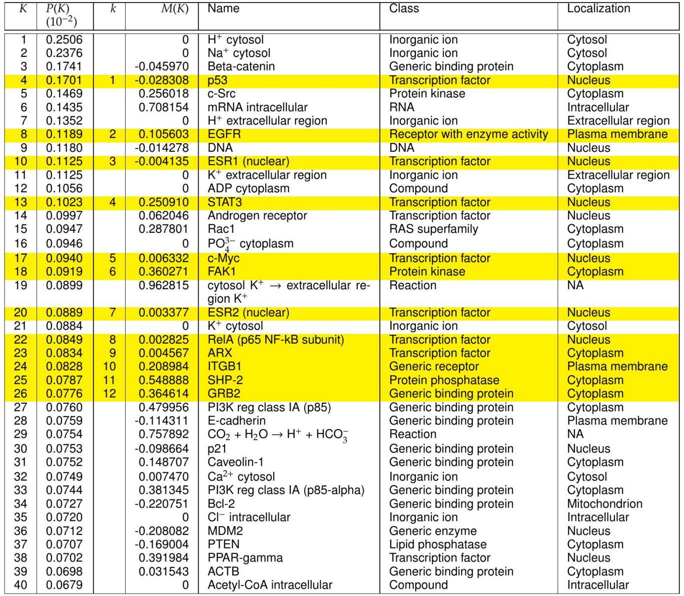
K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol $K^+ \rightarrow$ extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

We select **12 proteins of interest** (in yellow) among the top 40 PageRank

Googlomics analysis of the MetaCore multi-omic network



Top 40 PageRank
“Process outcomes”

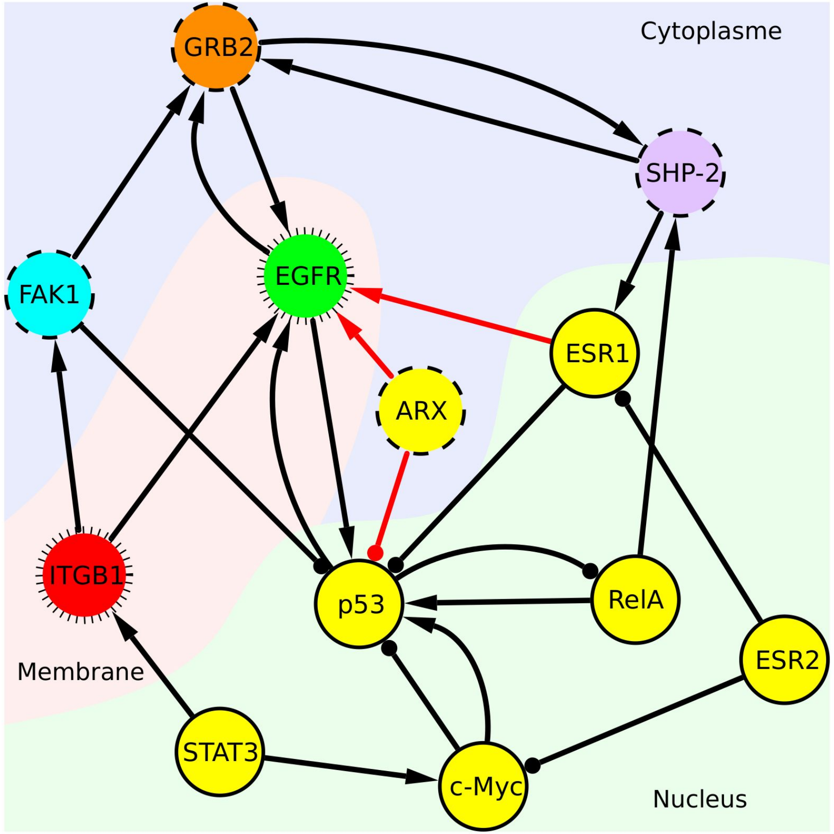


K	$P(K)$ (10^{-2})	k	$M(K)$	Name	Class	Localization
1	0.2506		0	H^+ cytosol	Inorganic ion	Cytosol
2	0.2376		0	Na^+ cytosol	Inorganic ion	Cytosol
3	0.1741		-0.045970	Beta-catenin	Generic binding protein	Cytoplasm
4	0.1701	1	-0.028308	p53	Transcription factor	Nucleus
5	0.1469		0.256018	c-Src	Protein kinase	Cytoplasm
6	0.1435		0.708154	mRNA intracellular	RNA	Intracellular
7	0.1352		0	H^+ extracellular region	Inorganic ion	Extracellular region
8	0.1189	2	0.105603	EGFR	Receptor with enzyme activity	Plasma membrane
9	0.1180		-0.014278	DNA	DNA	Nucleus
10	0.1125	3	-0.004135	ESR1 (nuclear)	Transcription factor	Nucleus
11	0.1125		0	K^+ extracellular region	Inorganic ion	Extracellular region
12	0.1056		0	ADP cytoplasm	Compound	Cytoplasm
13	0.1023	4	0.250910	STAT3	Transcription factor	Nucleus
14	0.0997		0.062046	Androgen receptor	Transcription factor	Nucleus
15	0.0947		0.287801	Rac1	RAS superfamily	Cytoplasm
16	0.0946		0	PO_4^{3-} cytoplasm	Compound	Cytoplasm
17	0.0940	5	0.006332	c-Myc	Transcription factor	Nucleus
18	0.0919	6	0.360271	FAK1	Protein kinase	Cytoplasm
19	0.0899		0.962815	cytosol $K^+ \rightarrow$ extracellular region K^+	Reaction	NA
20	0.0889	7	0.003377	ESR2 (nuclear)	Transcription factor	Nucleus
21	0.0884		0	K^+ cytosol	Inorganic ion	Cytosol
22	0.0849	8	0.002825	RelA (p65 NF-kB subunit)	Transcription factor	Nucleus
23	0.0834	9	0.004567	ARX	Transcription factor	Cytoplasm
24	0.0828	10	0.208984	ITGB1	Generic receptor	Plasma membrane
25	0.0787	11	0.548888	SHP-2	Protein phosphatase	Cytoplasm
26	0.0776	12	0.364614	GRB2	Generic binding protein	Cytoplasm
27	0.0760		0.479956	PI3K reg class IA (p85)	Generic binding protein	Cytoplasm
28	0.0759		-0.114311	E-cadherin	Generic binding protein	Plasma membrane
29	0.0754		0.757892	$CO_2 + H_2O \rightarrow H^+ + HCO_3^-$	Reaction	NA
30	0.0753		-0.098664	p21	Generic binding protein	Nucleus
31	0.0752		0.148707	Caveolin-1	Generic binding protein	Cytoplasm
32	0.0749		0.007470	Ca^{2+} cytosol	Inorganic ion	Cytosol
33	0.0744		0.381345	PI3K reg class IA (p85-alpha)	Generic binding protein	Cytoplasm
34	0.0727		-0.220751	Bcl-2	Generic binding protein	Mitochondrion
35	0.0720		0	Cl^- intracellular	Inorganic ion	Intracellular
36	0.0712		-0.208082	MDM2	Generic enzyme	Nucleus
37	0.0707		-0.169004	PTEN	Lipid phosphatase	Cytoplasm
38	0.0702		0.391984	PPAR-gamma	Transcription factor	Nucleus
39	0.0698		0.031543	ACTB	Generic binding protein	Cytoplasm
40	0.0679		0	Acetyl-CoA intracellular	Compound	Intracellular

We select **12 proteins of interest** (in yellow) among the top 40 PageRank

Googlomics analysis of the MetaCore multi-omic network

Reduced network
for 12 proteins of
interest



Types of links

Direct activation links
→

Hidden activation links
→

Direct inhibition links
—●

Hidden inhibition links
—●

Type of proteins

🟡 Transcription factors

● Protein kinases

🔴 Generic receptors

🟢 Receptors
with enzyme activity

🟠 General
binding proteins

🟣 Protein phosphatases

Take home messages

- We have presented a **detailed description of the statistical properties of the multi-omic interactions MetaCore network** obtained with extensive **Google matrix analysis**.
- Proteins and molecules which are at the **top PageRank are at “outcomes” of multiple cascades of causal interactions**.
- Proteins and molecules which are at the **top CheiRank are “triggers” of multiple cascades of causal interactions**.
- The **reduced Google Matrix** is an analytical approach for **inferring hidden indirect connections** within a set of nodes embedded in a very large network.
- **In the case of the MetaCore database, hidden signaling pathways can be detected**.
- **The reduced Google matrix analysis allows to determine the effective interactions between proteins of interest taking into account all the indirect pathways between these proteins through the global MetaCore network**.
- **The reduced Ising Google matrix** algorithm, based on Ising spin description, allows to take into account the **bi-functional nature of the protein-protein interactions (activation or inhibition)** and to determine the **average action type (or magnetization) of each protein**.
- **Further studies will address biological effects inferred from the reduced Google matrix analysis of the MetaCore network**.

Thank you for your attention !!