Reduced Google Matrix

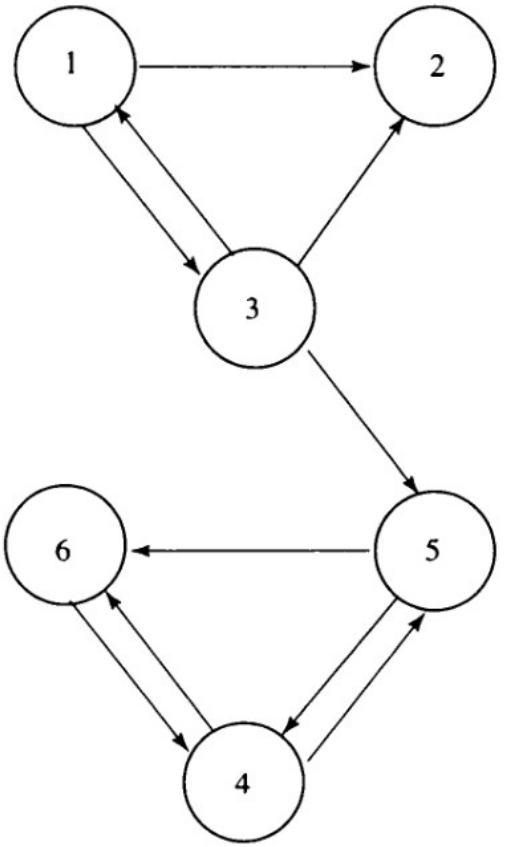
Samer EL ZANT

University of Toulouse
INP-IRIT

17 May 2016

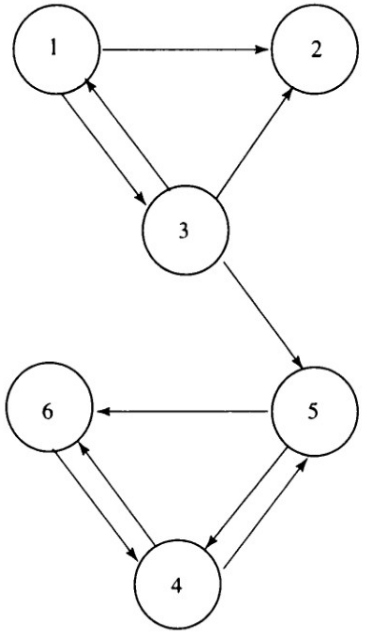
Introduction

- A network is composed of many nodes.
- Each node has his own score.
- This score is calculated by taking into consideration many factors (popularity-content-area-cookies...).
- One of these factors is based on the link analysis which is the popularity or importance score.
- A node is important if it is pointed by other important pages.



Page Rank

- Page Rank method is used to calculate the rank vector or the importance of each node.



$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|} \qquad r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|}$$

	Iteration 0	Rank 0	Iteration 1	Rank 1	Iteration 2	Rank 2
$r(P_1)$	1/6	1	1/18	5	1/36	5
$r(P_2)$	1/6	1	5/36	3	1/18	4
$r(P_3)$	1/6	1	1/12	4	1/36	5
$r(P_4)$	1/6	1	1/4	1	17/72	1
$r(P_5)$	1/6	1	5/36	3	11/72	3
$r(P_6)$	1/6	1	1/6	2	14/72	2

- B_{pi}:set of pages pointing into P_i.
- $|P_j|$:number of out-links from page P_j.
- $r_{k+1}(P_i)$:page-rank of page P_i at iteration $k+1$.

Google Matrix

- Using matrices we compute a page rank vector at each iteration $\pi^{(k+1)T} = \pi^{(k)T}\mathbf{H}$

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	1/2	1/2	0	0	0
P_2	0	0	0	0	0	0
P_3	1/3	1/3	0	0	1/3	0
P_4	0	0	0	0	1/2	1/2
P_5	0	0	0	1/2	0	1/2
P_6	0	0	0	1	0	0

- H matrix
- $H_{ij}=1/|P_i|$ if there is a link going from I to j and zero otherwise.
- sub-stochastic

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	1/2	1/2	0	0	0
P_2	1/6	1/6	1/6	1/6	1/6	1/6
P_3	1/3	1/3	0	0	1/3	0
P_4	0	0	0	0	1/2	1/2
P_5	0	0	0	1/2	0	1/2
P_6	0	0	0	1	0	0

- S matrix
- Prevent rank sink $\pi^{13} = [0,0,0,2/3,1/3,1/5]$
- stochastic

$$\mathbf{S} = \mathbf{H} + \mathbf{a}(1/n\,\mathbf{e}^T)$$

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	1/60	7/15	7/15	1/60	1/60	1/60
P_2	1/6	1/6	1/6	1/6	1/6	1/6
P_3	19/60	19/60	1/60	1/60	19/60	1/60
P_4	1/60	1/60	1/60	1/60	7/15	7/15
P_5	1/60	1/60	1/60	7/15	1/60	7/15
P_6	1/60	1/60	1/60	11/12	1/60	1/60

- G matrix
- Google Matrix
- To guarantee convergence.
- By using α , we take into consideration the probability that a surfer follows hyper-link or a new node.

$$\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)1/n\,\mathbf{e}\mathbf{e}^T$$

$$\pi^{(k+1)T} = \pi^{(k)T}\mathbf{G}$$

Reduced Google Matrix

- Describes the properties and interactions of a certain subsets of selected nodes belonging to a much larger network.
- Example: top 40 French politicians in French wikipedia.
  Number of nodes of frwiki = N = 1,352,825.
  number of selected node = N_r = 40.

Name	Order/ index	Name	Order/ Index	Name	Order/ Index	Name	Order/ Index
NicolasSarkozy	1	JackLang	11	JeanFrancoisCope	21	BriceHortefeux	31
FrancoisHollande	2	AlainJuppe	12	NathalieKosciuskoMorizet	22	RamaYade	32
JeanMarieLePen	3	JeanLouisBorloo	13	ArnaudMontebourg	23	PierreMoscovici	33
SegoleneRoyal	4	BertrandDelanoe	14	ClaudeBartolone	24	ManuelValls	34
JeanPierreRaffarin	5	JeanLucMelenchon	15	RachidaDati	25	ClaudeGueant	35
DominiquedeVillepin	6	MarineLePen	16	OlivierBesancenot	26	HerveMorin	36
FrancoisFillon	7	ChristineLagarde	17	NicolasDupontAignan	27	CecileDuflot	37
FrancoisBayrou	8	MartineAubry	18	EvaJoly	28	MichelSapin	38
LaurentFabius	9	DanielCohnBendit	19	ChristianeTaubira	29	HenriGuaino	39
DominiqueStraussKahn	10	ValeriePecresse	20	ElisabethGuigou	30	FlorianPhilippot	40

Reduced Google Matrix Dima and Klaus algorithm

- N_r is the number of nodes in the reduced network.
- N is the number of nodes in the whole network.
- N_s=N-N_r
- We can write $G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix}$
- $G\,P = P$
- The page rank vector $P = \begin{pmatrix} P_r \\ P_s \end{pmatrix}$
- As a result we obtain: $G_{\mathrm{R}} P_r = P_r \quad , \quad G_{\mathrm{R}} = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1} G_{sr}$
- G_R is the reduced Google matrix of size N_r*N_r

Reduced Google Matrix
Dima and Klaus algorithm

- To avoid the problem of slow convergence in G_R a slight modification can be done.

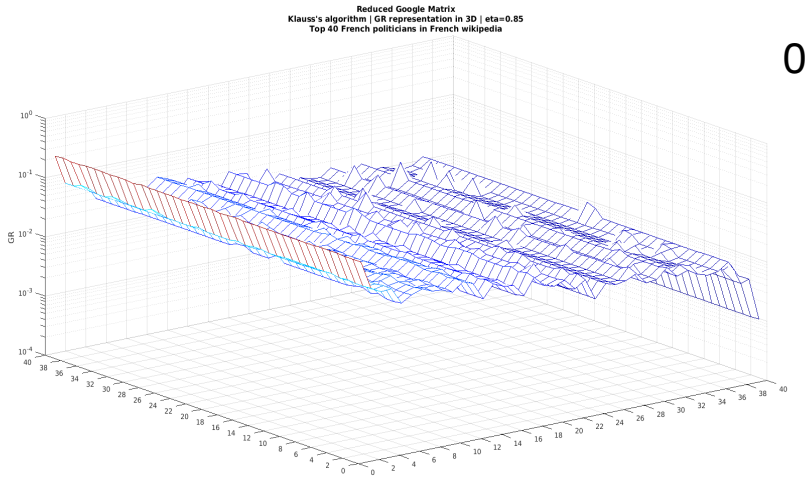
$$G_{\rm mod} = \begin{pmatrix} \mathbf{1} & (1-\eta)U_{rs} \\ 0 & \eta\mathbf{1} \end{pmatrix} \times \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix}$$

- η is an additional damping factor (0.5<η<1).
- $U_{rs} = (1/N_r)E_rE_s^T$
- E^T=(1,1,1,.....,1,1,1) of size N.

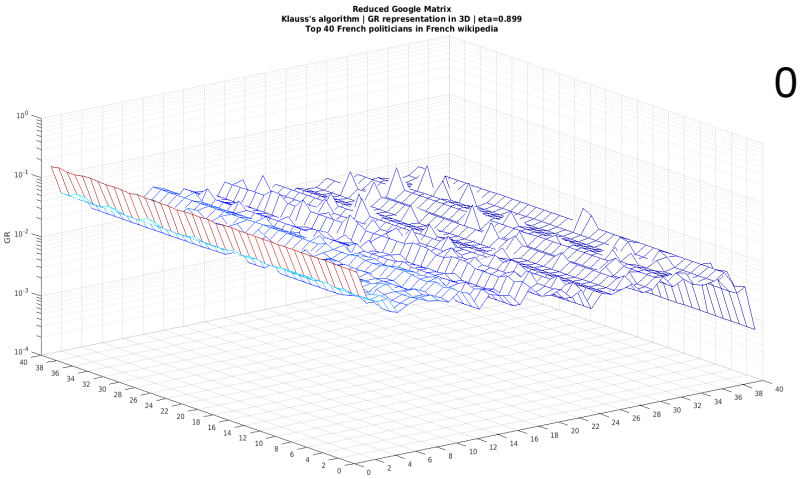
$$G_{\rm Rmod} = G_{rr} + (1-\eta)U_{rs}G_{sr} + \eta[G_{rs} + (1-\eta)U_{rs}G_{ss}](\mathbf{1} - \eta G_{ss})^{-1}G_{sr}$$

$$(\mathbf{1} - G_{ss})^{-1} = \sum_{l=0}^{\infty} G_{ss}^{l}$$

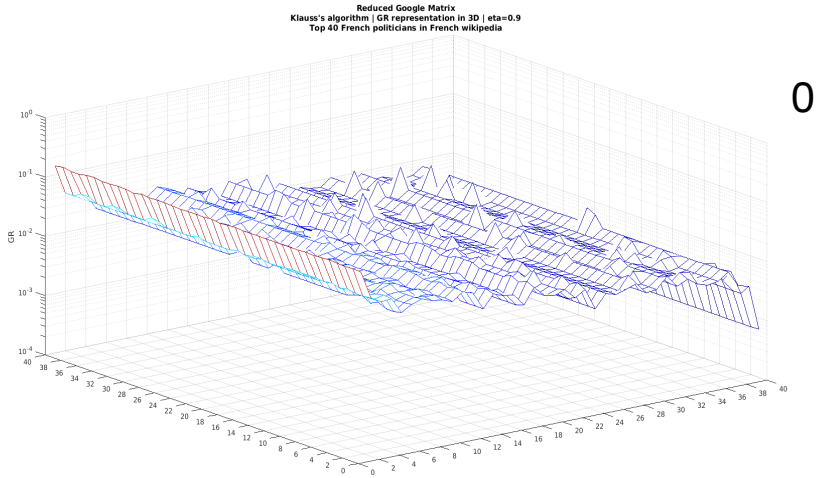
Reduced Google Matrix
Dima and Klaus algorithm - Results



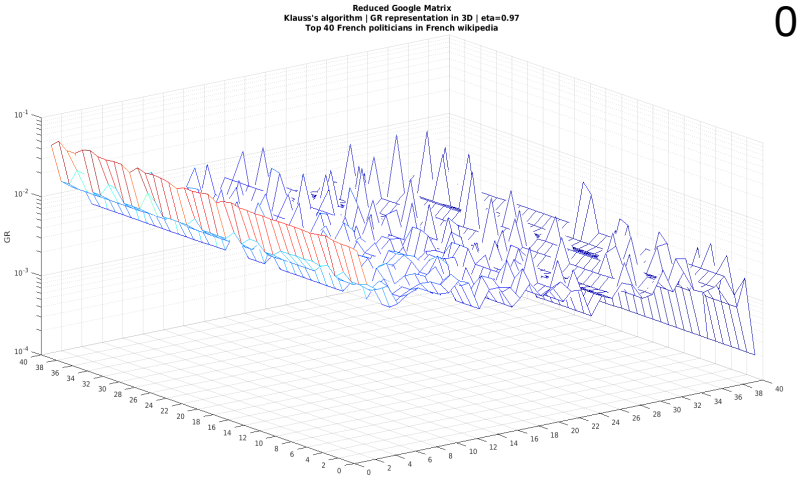
0.85



0.899

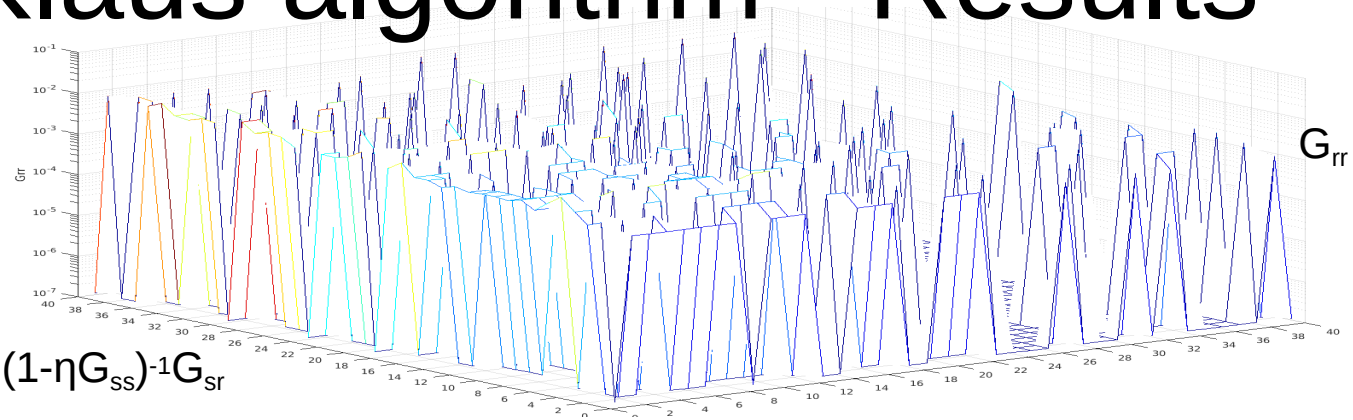


0.9



0.97

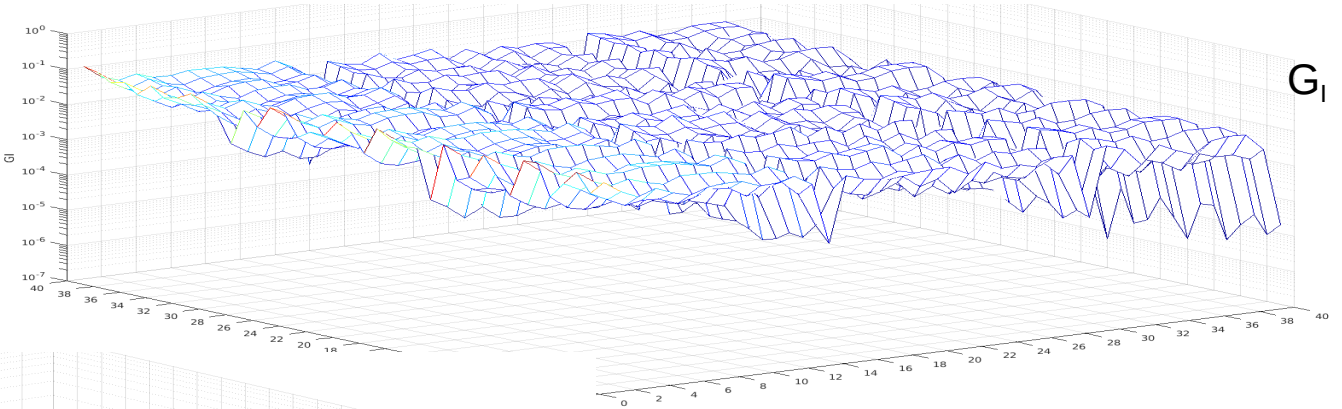
Reduced Google Matrix
Dima and Klaus algorithm - Results



G_{rr}

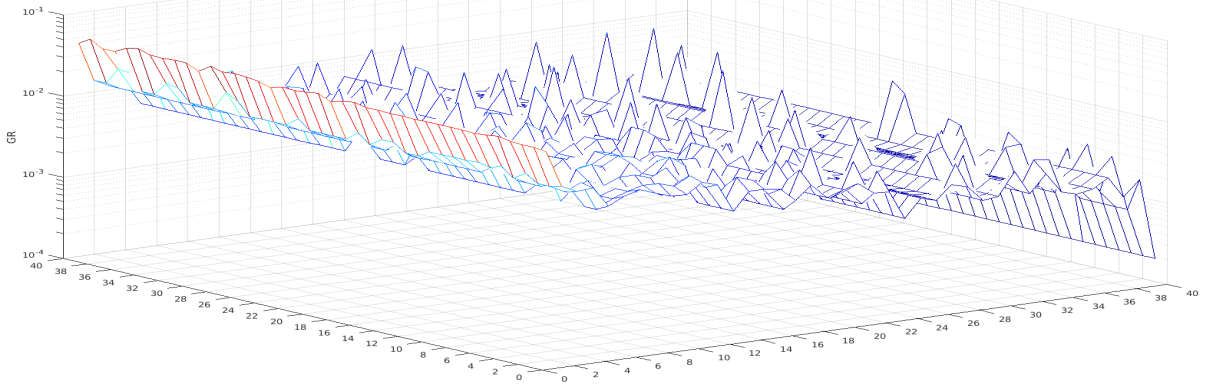
- $G_{Rmod}=G_{rr}+G_I$
- $G_I=(1-\eta)U_{rs}G_{sr}+\eta[G_{rs}+(1-\eta)U_{rs}G_{ss}](1-\eta G_{ss})^{-1}G_{sr}$

$$G_{\mathrm{Rmod}} = G_{rr} + (1-\eta)U_{rs}G_{sr} + \eta[G_{rs} + (1-\eta)U_{rs}G_{ss}](\mathbf{1} - \eta G_{ss})^{-1}G_{sr}$$



G_I

G_{Rmod}



Reduced Google Matrix
Dima and Klaus algorithm - Results

	Name	frwiki_norm	pg_order	Eta_0.85	Eta_85	Eta_0.9	Eta_9	Eta_0.899	Eta_899	Eta_0.97	Eta_97	fr_85	fr_9	fr_899	fr_97	85_97	9_97
8	NicolasSarkozy	0.193771494	1	0.1929990298	1	0.1917553425	1	0.1917918248	1	0.1838686995	1	0	0	0	0	0	0
6	FrancoisHollande	0.0663077616	2	0.0639258053	2	0.0639188775	2	0.0639190617	2	0.063906395	2	0	0	0	0	0	0
8	JeanMarieLePen	0.0555352675	3	0.0508677407	4	0.0505696593	4	0.0505784552	4	0.0485939435	4	1	1	1	1	0	0
3	SegoleneRoyal	0.0549899728	4	0.0530477781	3	0.0532011062	3	0.0531966035	3	0.0541834108	3	1	1	1	1	0	0
7	JeanPierreRaffarin	0.0431100728	5	0.0276338479	13	0.0276440214	14	0.0276437251	14	0.0277033643	14	8	9	9	9	1	0
8	DominiquedeVillepin	0.0421307556	6	0.0354592703	7	0.0354700139	7	0.0354697064	7	0.0355241196	9	1	1	1	3	2	2
2	FrancoisFillon	0.0418714731	7	0.041064781	5	0.0411440093	5	0.0411416911	5	0.0416348587	5	2	2	2	2	0	0
8	FrancoisBayrou	0.037068743	8	0.0348237858	8	0.0349333024	8	0.0349300998	8	0.0356115109	8	0	0	0	0	0	0
5	LaurentFabius	0.0352170941	9	0.0344733076	9	0.0344942584	10	0.0344935971	10	0.0347050148	10	0	1	1	1	1	0
9	DominiqueStraussKahn	0.0347213746	10	0.0285822134	12	0.0288540928	11	0.0288461033	11	0.0306013503	11	2	1	1	1	1	0
8	JackLang	0.0331928182	11	0.0391170232	6	0.0390972257	6	0.0390978088	6	0.0389695952	6	5	5	5	5	0	0
7	AlainJuppe	0.03311447	12	0.034402622	10	0.0346046018	9	0.0345986869	9	0.0358676025	7	2	3	3	5	3	2
5	JeanLouisBorloo	0.0254923511	13	0.0177592751	19	0.017805316	18	0.0178039574	18	0.0181095857	18	6	5	5	5	1	0
9	BertrandDelanoe	0.0237575599	14	0.0288258571	11	0.0287734542	12	0.0287749688	12	0.0284789401	13	3	2	2	1	2	1
7	JeanLucMelenchon	0.0229707059	15	0.0239460576	15	0.0240683172	15	0.0240647267	15	0.0248504999	15	0	0	0	0	0	0
8	MarineLePen	0.0191072018	16	0.0220121036	16	0.0220160519	16	0.0220159656	16	0.0219922175	16	0	0	0	0	0	0
7	ChristineLagarde	0.0182841153	17	0.0134271798	23	0.0134616946	23	0.0134607015	23	0.0136483819	23	6	6	6	6	0	0
9	MartineAubry	0.0174884028	18	0.0275979108	14	0.0277626474	13	0.0277577692	13	0.0288839637	12	4	5	5	6	2	1
7	DanielCohnBendit	0.0173995352	19	0.0177912531	18	0.0176627235	19	0.01766648	19	0.0168694676	20	1	0	0	1	2	1
2	ValeriePecresse	0.0153641952	20	0.0151191045	21	0.0150381409	21	0.0150405347	21	0.0144939269	22	1	1	1	2	1	1
8	JeanFrancoisCope	0.0141486119	21	0.0185537971	17	0.0187504537	17	0.0187446977	17	0.0199734654	17	4	4	4	4	0	0
2	NathalieKosciuskoMorizet	0.0118838329	22	0.0117446485	26	0.011822461	26	0.0118201929	26	0.0122916426	25	4	4	4	3	1	1
5	ArnaudMontebourg	0.0116355277	23	0.0168873578	20	0.0170445528	20	0.0170399252	20	0.0180691501	19	3	3	3	4	1	1
8	ClaudeBartolone	0.0114906493	24	0.0074262016	37	0.007485888	37	0.0074841217	37	0.0078917415	36	13	13	13	12	1	1
0	RachidaDati	0.0111766309	25	0.0120475246	25	0.0119478434	25	0.0119507705	25	0.0113112315	26	0	0	0	1	1	1
7	OlivierBesancenot	0.0111050871	26	0.0105794617	27	0.0104749625	27	0.0104780254	27	0.0098165562	31	1	1	1	5	4	4
3	NicolasDupontAignan	0.0098456554	27	0.0104533848	29	0.0104522053	28	0.0104522571	28	0.0104166279	28	2	1	1	1	1	0
9	EvaJoly	0.0096459119	28	0.0141354919	22	0.0141981043	22	0.014196269	22	0.0145929345	21	6	6	6	7	1	1
3	ChristianeTaubira	0.0091570795	29	0.008792094	33	0.0086754519	34	0.0086788752	34	0.0079332504	35	4	5	5	6	2	1
4	ElisabethGuigou	0.0089108003	30	0.0096810899	31	0.0097517209	31	0.0097496297	31	0.0102324093	29	1	1	1	1	2	2
5	BriceHortefeux	0.0081200595	31	0.0128026652	24	0.0128081676	24	0.0128080261	24	0.0128107662	24	7	7	7	7	0	0
9	RamaYade	0.0081034805	32	0.0089036607	32	0.0088188064	32	0.0088212977	32	0.0082768785	34	0	0	0	2	2	2
2	PierreMoscovici	0.0076839294	33	0.0097899723	30	0.0098968324	30	0.0098937031	30	0.0105660055	27	3	3	3	6	3	3
0	ManuelValls	0.007296955	34	0.0086842276	34	0.0087936638	33	0.0087904501	33	0.0094940056	32	0	1	1	2	2	1
8	ClaudeGueant	0.0070754558	35	0.0075199995	36	0.0075254797	36	0.0075253108	36	0.0075728661	37	1	1	1	2	1	1
9	HerveMorin	0.0069324402	36	0.0081825842	35	0.0082349328	35	0.0082333859	35	0.0085848926	33	1	1	1	3	2	2
0	CecileDuflot	0.0062199551	37	0.0104721241	28	0.0104330145	29	0.0104341664	29	0.0101787471	30	9	8	8	7	2	1
5	MichelSapin	0.0047755043	38	0.005866071	38	0.0059314698	38	0.0059295487	38	0.0063505553	38	0	0	0	0	0	0
0	HenriGuaino	0.0035622722	39	0.0037293212	39	0.0038182859	39	0.0038156947	39	0.0043519423	39	0	0	0	0	0	0
1	FlorianPhilippot	0.0003348118	40	0.0008723755	40	0.0008608463	40	0.0008611847	40	0.0007874833	40	0	0	0	0	0	0
												102	102	102	122	42	30
												2.55	2.55	2.55	3.05	1.05	0.75

Reduced Google Matrix
Leo's algorithm

- Leo's proposal defined a reduced google matrix in dimension r+1 as follows

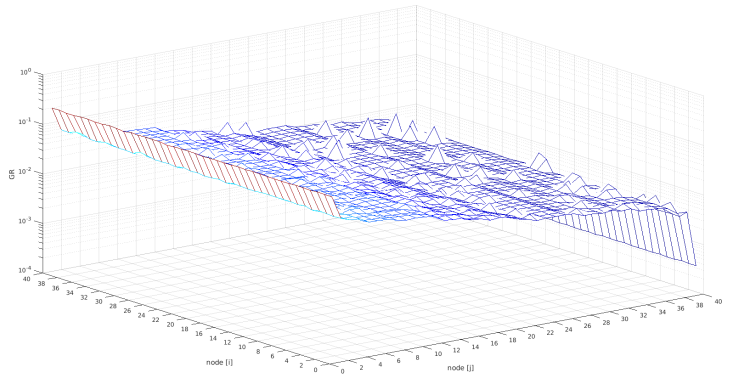
$$\tilde{G}\tilde{P} = \begin{pmatrix} \tilde{G}_{rr} & \tilde{G}_{r1} \\ \tilde{G}_{1r} & \tilde{G}_{11} \end{pmatrix} \begin{pmatrix} \tilde{P}_r \\ \tilde{P}_1 \end{pmatrix} = \begin{pmatrix} \tilde{P}_r \\ \tilde{P}_1 \end{pmatrix}$$

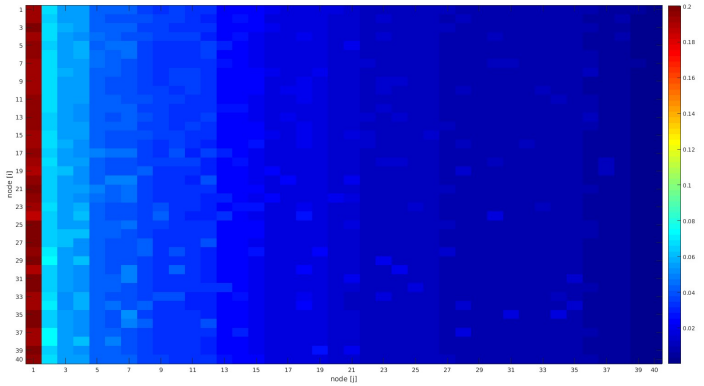
$$\tilde{G}_{1r} = 1 - \sum_{rows} G_{rr} = \sum_{rows} G_{sr} \qquad \tilde{P}_r = P_r \qquad \tilde{G}_{rr} = G_{rr} \qquad \tilde{P}_1 = \sum P_s = 1 - \sum P_r$$

$$\tilde{G}_{r1} = \frac{1}{\tilde{P}_1} G_{rs} P_s$$

$$\boxed{G_R^E = G_{rr} + \frac{1}{1 - \tilde{G}_{11}} \tilde{G}_{r1} \tilde{G}_{1r}}$$

Reduced Google Matrix
Leo's algorithm - Results





Order\Index	Name	Order\Index	Name
1	NicolasSarkozy	21	JeanFrancoisCope
2	FrancoisHollande	22	NathalieKosciuskoMorizet
3	JeanMarieLePen	23	ArnaudMontebourg
4	SegoleneRoyal	24	ClaudeBartolone
5	JeanPierreRaffarin	25	RachidaDati
6	DominiquedeVillepin	26	OlivierBesancenot
7	FrancoisFillon	27	NicolasDupontAignan
8	FrancoisBayrou	28	EvaJoly
9	LaurentFabius	29	ChristianeTaubira
10	DominiqueStraussKahn	30	ElisabethGuigou
11	JackLang	31	BriceHortefeux
12	AlainJuppe	32	RamaYade
13	JeanLouisBorloo	33	PierreMoscovici
14	BertrandDelanoe	34	ManuelValls
15	JeanLucMelenchon	35	ClaudeGueant
16	MarineLePen	36	HerveMorin
17	ChristineLagarde	37	CecileDuflot
18	MartineAubry	38	MichelSapin
19	DanielCohnBendit	39	HenriGuaino
20	ValeriePecresse	40	FlorianPhilippot

- It has the same order of pagerank vector for whole network.
- The problem with this proposal is that the pagerank of whole network has full influence on the results.

Conclusion

- Klaus's algorithm of reduced Google matrix analyzes effective interactions between a subset of selected nodes belonging to a much larger network taking into consideration the indirect links via whole network.

Thank you