

Analyse des interactions géopolitiques par la matrice de Google réduite

S. El Zant¹ K. M. Frahm² K. Jaffrès-Runser¹ et D. L. Shepelyansky²

¹*Institut de Recherche en Informatique de Toulouse, Université de Toulouse, INPT*

²*Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, CNRS, UPS*

La géopolitique étudie les relations entre des puissances politiques par rapport à leur situation géographique. Les interactions entre les pays du monde ont été largement étudiées à différentes échelles, en analysant par exemple des informations issues des échanges économiques, des faits historiques ou l'évolution des réglementations internationales. Cette étude propose d'exploiter les connaissances encyclopédiques de Wikipédia pour extraire les interactions entre un ensemble de pays choisis. Wikipédia enregistre ses connaissances sous la forme d'un réseau de pages Internet reliées entre elles par des hyperliens. La structure de ce réseau est porteuse d'informations très précieuses car elle représente les inter-dépendances entre les sujets étudiés. Wikipedia est édité en plusieurs langues, la plus volumineuse étant la langue anglaise. L'édition anglaise comporte plusieurs millions de pages et une étude minutieuse des interactions entre certaines pages passe par la définition d'outils mathématiques efficaces et qui capturent les inter-dépendances qui existent dans un nombre limité de pages. Dans cet article, nous utilisons la méthode d'analyse de la matrice réduite de Google récemment proposée pour extraire les interactions entre 40 pays au niveau mondial. Cette étude est menée pour 5 éditions de pages Wikipédia. Cette approche permet d'extraire du réseau, en plus des liens directs, des relations indirectes (ou cachées) entre les 40 pays sélectionnés. Ces relations indirectes sont particulièrement intéressantes car elles ne sont pas influencées par la mesure de PageRank et proposent ainsi un axe d'étude supplémentaire pertinent.

Mots-clefs : Google Matrix, Reduced Google Matrix, Geopolitics, Wikipedia

1 Introduction

The major finding of this paper is to show that meaningful worldwide interactions can be extracted from the free online encyclopaedia Wikipedia [1] for a given set of countries. All information gathered in this collaborative knowledge base can be leveraged to provide a valuable picture of countries relationships. Unique to Wikipedia is that articles make citations to each other, providing a direct relationship between webpages and topics. Wikipedia generates a larger directed network of article titles with a rather clear meaning. For these reasons, it is interesting to apply algorithms developed for search engines of World Wide Web (WWW) like PageRank algorithm [2] to analyze the ranking properties and relations between Wikipedia articles. In this paper, we are interested in capturing the interactions of the 40 countries represented in Table 1 using the networks extracted from five Wikipedia language editions. However, the interactions between these 40 countries that represent the nodes of the network, should be determined taking into account indirect links between the 40 webpages via all other nodes of the network. In this paper, reduced Google matrix analysis proposed in [5] is applied to a set of 40 countries on networks representing five different Wikipedia editions: English 'EnWiki', Arabic 'ArWiki', Russian 'RuWiki', French 'FrWiki' and German 'DeWiki' editions. Our study shows the impact of the cultural bias when comparing direct and hidden networks of friends (or followers) among different language editions. This paper introduces first the main elements of reduced Google matrix theory in Section 2. Next, Section 3 highlights the main results of reduced Google matrix analysis on the selected data. From this analysis, we are able to capture direct and hidden interactions between countries that lead us to the identification of friend or follower relationships between countries.

Country	EnWiki		ArWiki		RuWiki		Country	EnWiki		ArWiki		RuWiki	
	K	K*	K	K*	K	K*		K	K*	K	K*	K	K*
United States (US)	1	9	1	5	2	27	Norway (NO)	21	32	35	16	27	11
France (FR)	2	19	3	31	3	14	Romania (RO)	22	10	19	6	32	36
United Kingdom (UK)	3	25	6	20	7	3	Turkey (TR)	23	7	15	13	21	38
Germany (DE)	4	33	8	14	4	24	South Africa (ZA)	24	24	29	39	35	20
Canada (CA)	5	26	13	19	12	26	Belgium (BE)	25	18	27	37	29	30
India (IN)	6	23	9	25	13	8	Austria (AT)	26	39	28	28	14	28
Australia (AU)	7	35	16	22	18	12	Greece (GR)	27	21	10	36	25	25
Italy (IT)	8	15	5	1	6	32	Argentina (AR)	28	1	32	29	33	23
Japan (JP)	9	4	11	9	11	7	Philippines (PH)	29	17	36	21	39	33
China (CN)	10	8	12	17	9	21	Portugal (PT)	30	36	24	12	17	9
Russia (RU)	11	6	7	2	1	2	Pakistan (PK)	31	5	25	35	37	29
Spain (ES)	12	30	4	8	8	15	Denmark (DK)	32	16	33	10	31	19
Poland (PL)	13	12	26	32	10	17	Israel (IL)	33	20	17	18	28	6
The Netherlands (NL)	14	37	18	33	15	31	Finland (FI)	34	14	37	4	26	16
Iran (IR)	15	2	14	15	30	22	Egypt (EG)	35	31	2	3	24	39
Brazil (BR)	16	3	21	26	20	1	Indonesia (ID)	36	13	31	11	34	10
Sweden (SE)	17	22	22	7	19	5	Hungary (HU)	37	11	40	40	23	40
New Zealand (NZ)	18	28	34	24	36	4	Taiwan (TW)	38	27	39	27	40	34
Mexico (MX)	19	40	23	38	22	37	South Korea (KR)	39	34	38	30	38	35
Switzerland (CH)	20	38	20	34	16	18	Ukraine (UA)	40	29	30	23	5	13

Tab. 1: List of names of 40 selected countries with PageRank K , CheiRank K^* for EnWiki, ArWiki and RuWiki.

2 Reduced Google matrix theory

It is convenient to describe the network of N Wikipedia articles by the Google matrix G constructed from the adjacency matrix A_{ij} with elements 1 if article (node) j points to article (node) i and zero otherwise. In this case, elements of the Google matrix take the standard form as shown in Eq. 1 [2, 3].

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N, (1) \quad G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix}, (2)$$

$$P = \begin{pmatrix} P_r \\ P_s \end{pmatrix}, (3) \quad G_R P_r = P_r, \quad G_R = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1} G_{sr}, (4) \quad (\mathbf{1} - G_{ss})^{-1} = \sum_{l=0}^{\infty} G_{ss}^l, (5)$$

In Eq. 1, S is the matrix of Markov transitions with elements $S_{ij} = A_{ij}/k_{out}(j)$, $k_{out}(j) = \sum_{i=1}^N A_{ij} \neq 0$ being the node j out-degree (number of outgoing links) and with $S_{ij} = 1/N$ if j has no outgoing links (dangling node). Let G be a typical Google matrix of Perron-Frobenius type for a network with N nodes such that $G_{ij} \geq 0$. We consider a sub-network with $N_r < N$ nodes, called “reduced network”. In this case we can write G in a block form as in Eq. 2. where the index “ r ” refers to the nodes of the reduced network and “ s ” to the other $N_s = N - N_r$ nodes. We denote the PageRank vector of the full network as in Eq. 3, which satisfies the equation $GP = P$. Reduced Google matrix G_R of size $N_r \times N_r$ is defined in Eq. 4. It represents a concise version of G , still of Perron-Frobenius form. The matrix elements of G_R are non-negative since the matrix inverse in Eq. (4) can be expanded as shown in Eq. 5. The integer l represents the order of indirect links used to connect indirectly two nodes of the reduced network through G_{ss} [6].

By solving Eq.(4)-(5) using the projector method as shown in [6], G_R can be presented as a sum of three components: $G_R = G_{rr} + G_{pr} + G_{qr}$, with the 1st component G_{rr} given by direct matrix elements of G among the selected N_r nodes. G_{pr} is the projector component that captures mainly the effect of the first eigenvector of G (i.e. Pagerank vector). The 3rd component G_{qr} is of particular interest in this study as it characterizes the impact of indirect or hidden links going through the N_s scattering nodes (i.e. the rest of the network). In other words, G_{qr} captures the impact of the $N - N_r$ nodes on G_R . It can be further decomposed into the sum of its non-diagonal terms G_{qmd} and of its diagonal terms G_{qrd} .

3 Matrices of world countries

Our study focuses on the networks extracted from 5 different Wikipedia editions[†] collected in [4]: EnWiki, ArWiki, RuWiki, DeWiki and FrWiki that contain 4.212, 0.203, 0.966, 1.533 and 1.353 millions of article

[†] Data collected mid February 2013

Formatting a submission for AlgoTel

Top country	G_R Wiki friends present in			G_R Wiki friends followers present in		
	all 5 editions	4 out of 5 editions	3 out of 5 editions	all 5 editions	4 out of 5 editions	3 out of 5 editions
US	FR	DE	UK	MX - CA	IL	UK
BR	US - FR	DE		AR - PT		MX
FR	US	DE		BE	ES	IT
RU	US - FR	DE		UA - FI		PL
TR	US - FR	DE		GR - IR		
JP	US - FR	DE		KR - TW - CN	PH	
IN	US - FR		DE	PK - ID		ZA

Top country	G_{qr} Wiki friends present in			G_{qr} Wiki followers present in		
	all 5 editions	4 out of 5 editions	3 out of 5 editions	all 5 editions	4 out of 5 editions	3 out of 5 editions
US	CA	UK - MX	FR		UK	CA - MX
BR	PT - AR	ES - MX		AR	PT	UK
FR	IT - DE		US - BE		UK - BE - ES	CH
RU	UA	DE	CN - PL - US	FI	UA - PL	IR
TR	GR	IR - RU		GR - IR	EG - RO	
JP		CN - TW - US - KR		KR - CN - TW		
IN	PK - CN	ID	IR	PK - ID		CN

Tab. 2: Cross-editions direct friends and followers extracted from G_R matrix (top table) and from G_{qrd} (bottom table) for the top countries of each area. For each top country, we list the direct friends (followers) present in the direct friends list given by all 5 Wiki editions, the ones present in 4 editions out of 5 and the ones present in 3 editions out of five.

respectively. The countries listed in Table 1 have been selected from EnWiki network by keeping the 40 countries with largest PageRank probability. They are ordered by a local PageRank index K varying between 1 and 40. The most influential countries are found to be at the top values $K = 1, 2, \dots$. We determine the local CheiRank \ddagger index K^* of the selected countries using the CheiRank vector of the global network. At the top of K^* we have the most communicative countries. Table 1 lists K and K^* for all countries and all five Wikipedia editions. Countries that belong to the same region or having a common piece of history may probably exhibit stronger interactions. It is convenient to plot all nodes in the (K, K^*) plane to highlight the countries that are the most influential ($K = 1, 2, \dots$) and the most communicative ($K^* = 1, 2, \dots$) at the same time. Table 1 ranks all 40 countries in the (K, K^*) plane for EnWiki, ArWiki and RuWiki editions. It is interesting to look at the set of non-dominated countries. A country is non-dominated if there is no country that is beating it for both K and K^* . Cultural bias is obvious here as for EnWiki, this set is composed of {US, JP, IR, AR}, for ArWiki of {US, EG, IT} and for RuWiki of only Russia.

We are now extracting from G_{qrd} and G_R the main relationships 7 leading countries experience with the 33 other ones. For both matrices, we extract the top 4 *Friends* (resp. *Followers*) for each leading country as following. For a given country j , its top 4 Friends (resp. Followers) are given by the 4 best values of the elements of column j (resp. of line j) in G_{qrd} or G_R . Top 4 friends and followers have been extracted from G_{qrd} and G_R for 7 leading countries selected as: US for English speaking countries, Brazil for South America, France for Europe, Japan for North-East Asia, India for South-East Asia, Russia for the Soviet block and Turkey for the Middle-East. To pick them inside each group, we have chosen the country whose worst PageRank order over all 5 Wikipedia editions is the highest. A summary of relevant results is given in Table 2 to extract the cross-culture interactions. Looking at G_R friends, top friends of leading countries are strongly related to the top PageRank countries (we have a predominance of US, France and Germany). Similarly, cross-edition followers of US are Mexico and Canada, and followers of Japan are China, Korea and Taiwan. On the opposite, the indirect interactions of G_{qrd} are not as much influenced by PageRank. As shown Canada is always identified as a hidden friend of USA while it was never the case in G_R . Similarly, Ukraine is always tagged as a hidden friend of Russia; Italy and Spain as friends of France. G_{qrd} seems to emphasize more fine-grained regional interactions.

4 Networks of 40 countries

This study concentrates on the same 7 leading countries. Top 4 friends and top 4 followers of these leading countries are extracted from G_R and G_{qrd} to plot the graphs of Figure 1. The black thick arrows identify the

\ddagger CheiRank is calculated like PageRank but for the transposed version of A

top 4 friends and top 4 followers interactions. Red arrows represent the friends of friends (respectively the followers of followers) interactions that are computed recursively until no new edge is added to the graph.

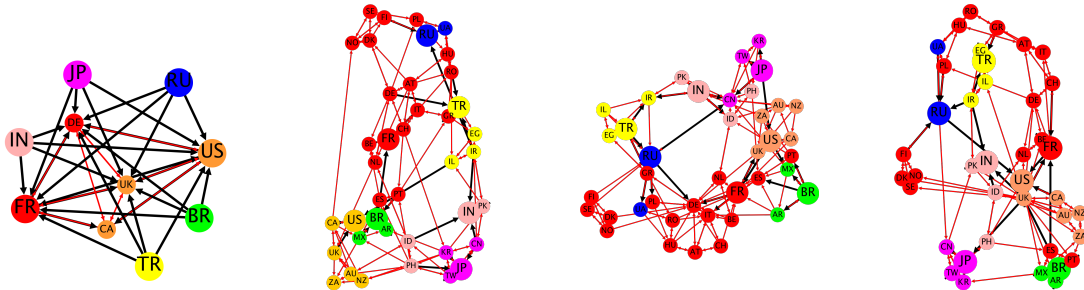


Fig. 1: Network structure for EnWiki of friends (first) and followers (second) induced by the 7 top countries of each geographical area (US, FR, IN, JP, BR, TR, RU) in G_R . While, third and fourth images show hidden relationships from G_{qmd} . The top country node points (is pointed by) with a bold black arrow to its top 4 friends (followers). Red arrows represent the friends of friends (resp. followers) interactions computed until no new edges are added to the graph.

The networks of friends obtained from G_R never expands to the full set of nodes. They only concentrate on around 10 countries. This results from G_R begin dominated by the projector component (i.e. influenced by PageRank). Looking at the follower graphs, they show that North-East Asian, Middle-Eastern and South American create a cluster of nodes densely interconnected. European countries englobe Russia and Ukraine as these countries are linked to EU countries that were part of the former Soviet Union zone of influence. The networks of followers end up almost spanning the full set of 40 countries. The networks of friends obtained from G_{qmd} don't concentrate to a limited set of countries, they end up spanning the full set of countries. The hidden friend links show that the interactions between the geographical groups are coherent.

5 Conclusion

This work offers a new perspective for future geopolitics studies. It is possible to extract from multi-cultural Wikipedia networks a global understanding of the interactions between countries, continental or regional scale. Reduced Google matrix theory has been shown to capture hidden interactions among countries.

Acknowledgements

This work was supported by APR 2015 call of University of Toulouse and by Région Occitanie (project GOMOBILE), MASTODONS-2016 CNRS project APLIGOOGLLE,

References

- [1] <http://www.wikipedia.org> (accessed August 2016)
- [2] S. Brin and L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems **30**, 107 (1998)
- [3] A.M. Langville and C.D. Meyer, *Google's PageRank and beyond: the science of search engine rankings*, Princeton University Press, Princeton (2006)
- [4] Y.-H.Eom, P.Aragon, D.Laniado, A.Kaltenbrunner, S.Vigna and D.L.Shepelyansky, *Interactions of cultures and top people of Wikipedia from ranking of 24 language editions*, PLoS ONE **10(3)** (2015)
- [5] K.M. Frahm and D.L. Shepelyansky, *Reduced Google matrix*, arXiv:1602.02394[physics.soc] (2016)
- [6] K.M. Frahm, K. Jaffrès-Runser and D.L. Shepelyansky, *Wikipedia mining of hidden links between political leaders*, arXiv:1609.01948[physics.soc] (2016)