EPJ manuscript No.
(will be inserted by the editor)

Time evolution of Wikipedia network ranking

Young-Ho Eom[1], Klaus M. Frahm[1], András Benczúr[2], and Dima L. Shepelyansky[1]

[1] Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, 31062 Toulouse, France

[2] Informatics Laboratory, Institute for Computer Science and Control,
Hungarian Academy of Sciences (MTA SZTAKI), Pf. 63, H-1518 Budapest, Hungary

Received: April 24, 2013

Abstract. We study the time evolution of ranking and spectral properties of the Google matrix of English Wikipedia hyperlink network during years 2003 - 2011. The statistical properties of ranking of Wikipedia articles via PageRank and CheiRank probabilities, as well as the matrix spectrum, are shown to be stabilized for 2007 - 2011. A special emphasis is done on ranking of Wikipedia personalities and universities. We show that PageRank selection is dominated by politicians while 2DRank, which combines PageRank and CheiRank, gives more accent on personalities of arts. The Wikipedia PageRank of universities recovers 80 percents of top universities of Shanghai ranking during the considered time period.

PACS. 89.75.Fb Structures and organization in complex systems – 89.75.Hc Networks and genealogical trees – 89.20.Hh World Wide Web, Internet

1 Introduction

At present Wikipedia [1] became the world largest Encyclopedia with open public access to its contain. A recent review [2] represents a detailed description of publications and scientific research of this modern Library of Babel, which stores an enormous amount of information, approaching the one described by Jorge Luis Borges [3]. The hyperlinks of citations between Wikipedia articles represent a directed network which reminds the structure of the World Wide Web (WWW). Hence, the mathematical tools developed for WWW search engines, based on the Markov chains [4], Perron-Frobenius operators [5] and the PageRank algorithm of the corresponding Google matrix [6,7], give solid mathematical grounds for analysis of information flow on the Wikipedia network. In this work we perform the Google matrix analysis of Wikipedia network of English articles extending the results presented in [8,9],[10,11]. The main new element of this work is the study of time evolution of Wikipedia network during the years 2003 to 2011. We analyze how the ranking of Wikipedia articles and the spectrum of the Google matrix G of Wikipedia are changed during this period.

The directed network of Wikipedia articles is constructed in a usual way: a directed link is formed from an article j to an article i when j quotes i and an element A_{ij} of the adjacency matrix is taken to be unity when there is such a link and zero in absence of link. Then the matrix S_{ij} of Markov transitions is constructed by normalizing elements of each column to unity ($\sum_j S_{ij} = 1$) and replacing columns with only zero elements (*dangling nodes*) by $1/N$, with N being the matrix size. Then the Google matrix of the network takes the form [6,7]:

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

The damping parameter α in the WWW context describes the probability $(1-\alpha)$ to jump to any node for a random surfer. For WWW the Google search engine uses $\alpha \approx 0.85$ [7]. The matrix G belongs to the class of Perron-Frobenius operators [5,7], its largest eigenvalue is $\lambda = 1$ and other eigenvalues have $|\lambda| \le \alpha$. The right eigenvector at $\lambda = 1$, which is called the PageRank, has real nonnegative elements $P(i)$ and gives a probability $P(i)$ to find a random surfer at site i. It is possible to rank all nodes in a decreasing order of PageRank probability $P(K(i))$ so that the PageRank index $K(i)$ counts all N nodes i according their ranking, placing the most popular articles or nodes at the top values $K = 1, 2, 3....$

Due to the gap $1-\alpha \approx 0.15$ between the largest eigenvalue $\lambda = 1$ and other eigenvalues the PageRank algorithm permits an efficient and simple determination of the PageRank by the power iteration method [7]. It is also possible to use the powerful Arnoldi method [12,13],[14] to compute efficiently the eigenspectrum λ_i of the Google matrix:

$$\sum_{k=1}^{N} G_{jk}\psi_i(k) = \lambda_i \psi_i(j) \quad . \tag{2}$$

The Arnoldi method allows to find a several thousands of eigenvalues λ_i with maximal $|\lambda|$ for a matrix size N as large as a few tens of millions [10,11], [14,15]. Usually, at $\alpha = 1$ the largest eigenvalue $\lambda = 1$ is highly degenerate [15] due to many invariant subspaces which define many independent Perron-Frobenius operators providing (at least) one eigenvalue $\lambda = 1$.

In addition to a given directed network A_{ij} it is useful to analyze an inverse network with inverted direction of links with elements of adjacency matrix $A_{ij} \to A_{ji}$. The Google matrix G^* of the inverse network is then constructed via corresponding matrix S^* according to the relations (1) using the same value of α as for the G matrix. This time inversion approach was used in [16,17] but the statistical properties and correlations between direct and inversed ranking were not analyzed there. In [18], on an example of the Linux Kernel network, it was shown thus this approach allows to obtain an additional interesting characterization of information flow on directed networks. Indeed, the right eigenvector of G^* at eigenvalue $\lambda = 1$ gives a probability $P^*(i)$, called CheiRank vector [8]. It determines a complementary rank index $K^*(i)$ of network nodes in a decreasing order of probability $P^*(K^*(i))$ [8, 9],[10,18]. It is known that the PageRank probability is proportional to the number of ingoing links characterizing how popular or known is a given node. In a similar way the CheiRank probability is proportional to the number of outgoing links highlighting the node communicativity (see e.g. [7,19], [20,21],[8,9]). The statistical properties of distribution of indexes $K(i), K^*(i)$ on the PageRank-CheiRank plane are described in [9].

In this work we apply the above mathematical methods to the analysis of time evolution of Wikipedia network ranking using English Wikipedia snapshots dated by December 31 of years 2003, 2005, 2007, 2009, 2011. In addition we use the snapshot of August 2009 (200908) analyzed in [8]. The parameters of networks with the number of articles (nodes) N, number of links N_ℓ and other information are given in Tables 1,2 with the description of notations given in Appendix.

The paper is composed as following: the statistical properties of PageRank and CheiRank are analyzed in Section 2, ranking of Wikipedia personalities and universities are considered in Sections 3, 4 respectively, the properties of spectrum of Google matrix are considered in Section 5, the discussion of the results is presented in Section 6, Appendix Section 7 gives network parameters.

2 CheiRank versus PageRank

The dependencies of PageRank and CheiRank probabilities $P(K)$ and $P^*(K^*)$ on their indexes K, K^* at different years are shown in Fig. 1. The top positions of K are occupied by countries starting from *United States* while at the top positions of K^* we find various listings (e.g. geographical names, prime ministers etc.; in 2011 we have appearance of listings of listings). Indeed, the countries accumulate links from all types of human activities and nature, that make them most popular Wikipedia articles, while listings have the largest number of outgoing links making them the most communicative articles.

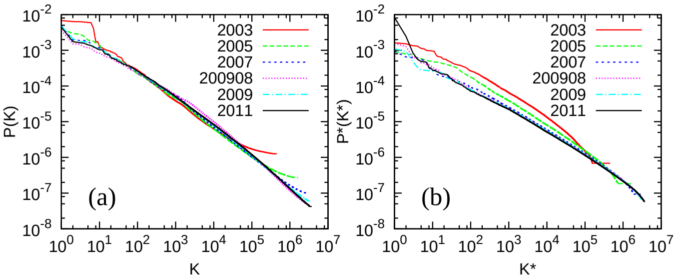

Fig. 1. PageRank probability $P(K)$ (left panel) and CheiRank probability $P^*(K^*)$ (right panel) are shown as a function of the corresponding rank indexes K and K^* for English Wikipedia articles at years 2003, 2005, 2007, 200908, 2009, 2011; here the damping factor is $\alpha = 0.85$.

The data of Fig. 1 show that the global behavior of $P(K)$ remains stable from 2007 to 2011. The probability $P^*(K^*)$ is stable in the time interval 2007 - 2009 while at 2011 we see the appearance of peak at $1 \le K^* < 10$ that is related to introduction of listings of listings which were absent at earlier years. At the same time the behavior of $P^*(K^*)$ in the range $10 \le K^* \le 10^6$ remains stable for 2007 - 2011.

Each article i has its PageRank and CheiRank indexes $K(i)$, $K^*(i)$ so that all articles are distributed on two-dimensional plane of PageRank-CheiRank indexes. Following [8,9] we present the density of articles in the 2D plane (K, K^*) in Fig. 2. The density is computed for 100×100 logarithmically equidistant cells which cover the whole plane (K, K^*) for each year. The density distribution is globally stable for years 2007-2011 even if there are articles which change their location in 2D plane. We see an appearance of a mountain like ridge of probability along a line $\ln K^* \approx \ln K + 4.6$ that indicate the presence of correlation between $P(K(i))$ and $P^*(K^*(i))$. Following [8,9, 18] we characterize the interdependence of PageRank and CheiRank vectors by the correlator

$$\kappa = N \sum_{i=1}^{N} P(K(i)) P^*(K^*(i)) - 1 \ . \quad (3)$$

We find the following values of the correlator at various time slots: $\kappa = 2.837(2003)$, $3.894(2005)$, $4.121(2007)$, $4.084(200908)$, $6.629(2009)$, $5.391(2011)$. During that period the size of the network increased almost by 10 times while κ increased less than 2 times. This confirms the stability of the correlator κ during the time evolution of the Wikipedia network.

In the next two Sections we analyze the time variation of ranking of personalities and universities.

3 Ranking of personalities

To analyze the time evolution of ranking of Wikipedia personalities (persons or humans) we chose the top 100 persons appearing in the ranking list of Wikipedia 200908

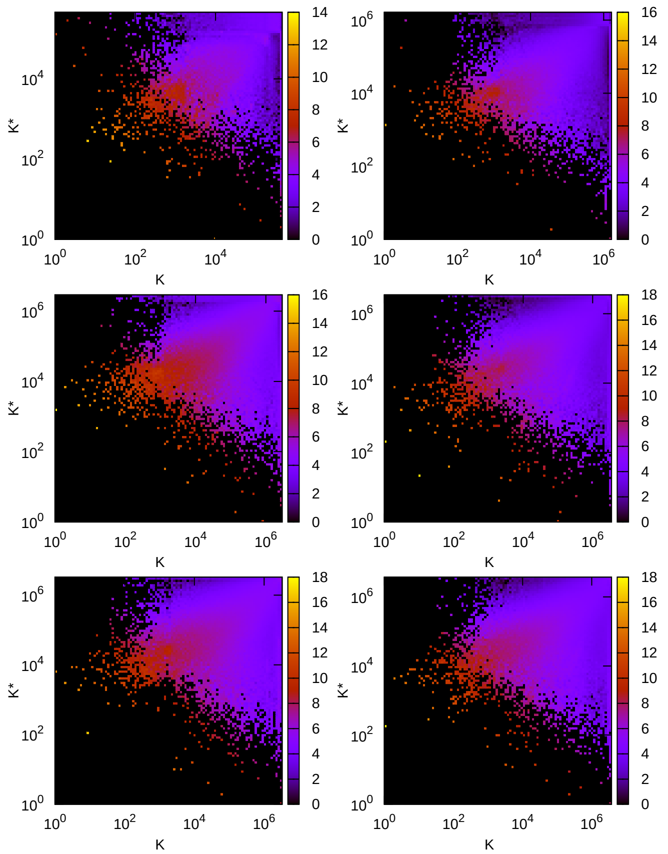

Fig. 2. Density of Wikipedia articles in the CheiRank versus PageRank plane at different years. Color is proportional to logarithm of density changing from minimal nonzero density (dark) to maximal one (white), zero density is shown by black (distribution is computed for 100×100 cells equidistant in logarithmic scale; bar shows color variation of natural logarithm of density); left column panels are for years 2003, 2007, 200908 and right column panels are for 2005, 2009, 2011 (from top to bottom).

given in [8] in order of PageRank, CheiRank and 2DRank. We remind that 2DRank K_2 is obtained by counting nodes in order of their appearance on ribs of squares in (K, K^*) plane with their size growing from $K = 1$ to $K = N$ [8].

The distributions of personalities in PageRank-CheiRank plane is shown at various time slots in Fig. 3. There are visible fluctuations of distribution of nodes for years 2003, 2005 when the Wikipedia size has rapid growth. For other years the distribution of top 100 nodes of PageRank and 2DRank is stable even if individual nodes change their ranking. For top 100 of CheiRank the fluctuations remain strong during all years. Indeed, the number of outgoing links is more easy to be modified by authors writing a given article, while a modification of ingoing links depends on authors of other articles.

In Fig. 3 we also show the distribution of top 100 personalities from Hart's book [22] (the list of names is also available at the web page [8]). This distribution also remains stable in years 2007-2011. It is interesting to note that while top PageRank and 2DRank nodes form a kind

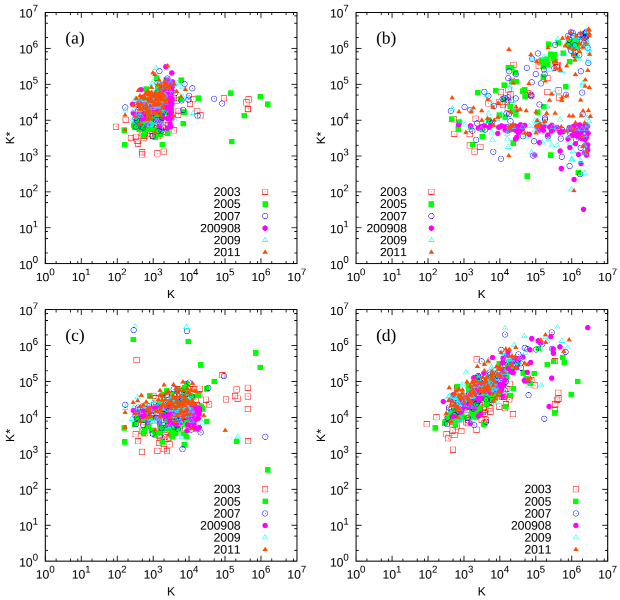

Fig. 3. Change of locations of top-rank persons of Wikipedia in K-K* plane. Each list of top ranks is determined by data of top 100 personalities of time slot 200908 in corresponding rank. Data sets are shown for (a) PageRank, (b) CheiRank, (c) 2DRank, (d) rank from Hart [22].

of droplet in (K, K^*) plane, the distribution of Hart's personalities approximately follows the ridge along the line $\ln K^* \approx \ln K + 4.6$.

The time evolution of top 10 personalities of slot 200908 is shown in Fig. 4 for PageRank and 2DRank. For PageRank the main part of personalities keeps their rank position in time, e.g. G.W.Bush remains at first-second position. B.Obama significantly improves his ranking as a result of president elections. There are strong variations for Elizabeth II which we relate to modification of article name during the considered time interval. We also see a steady improvement of ranking of C.Linnaeus that we attribute to a growth of various botanic descriptions and listings at Wikipedia articles which quote his name. For 2DRank we observe stronger variations of K_2 index with time. Such a politician as R.Nixon has increasing K_2 index with time since the period of his presidency goes in the past. At the same time singers and artists remain at approximately constant level of K_2.

In [8] it was pointed out that the top personalities of PageRank are dominated by politicians while for 2DRank the dominant component of human activity is represented by artists. We analyze the time evolution of the distribution of top 30 personalities over 6 categories of human activity (*politics, arts, science, religion, sport and etc (or others)*). The category *etc* contains only C.Columbus. The results are presented in Fig. 5. They clearly show that the PageRank personalities are dominated by politicians whose percentage increases with time, while the percent of arts decreases. For 2DRank we see that the arts are dominant even if their percentage decreases with time. We also see the appearance of sport which is absent in

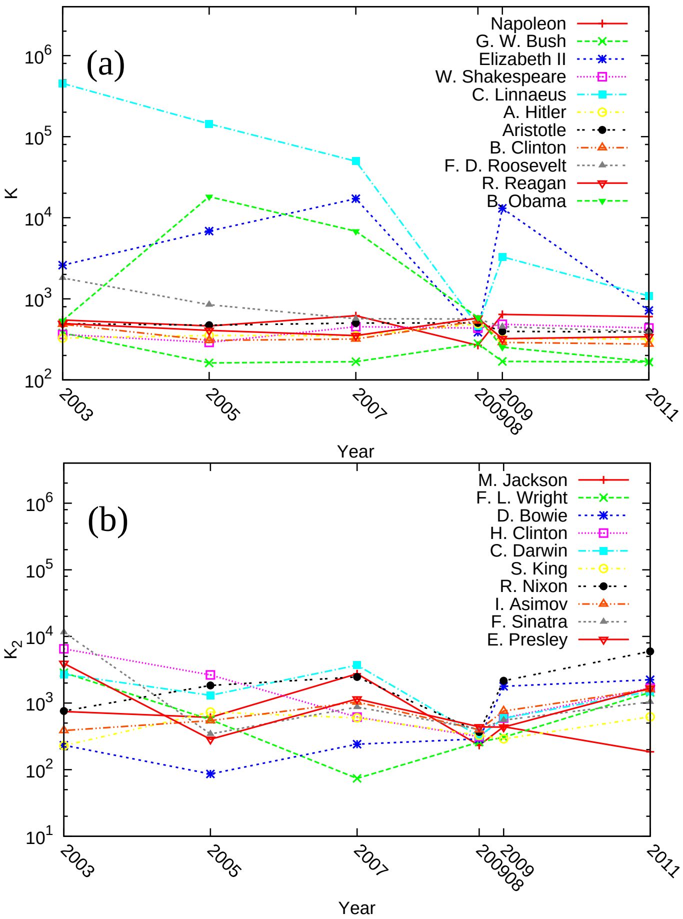

Fig. 4. Time evolution of top 10 personalities of year 200908 in indexes of PageRank K (a) and 2DRank K_2 (b); B.Obama is added in panel (a).

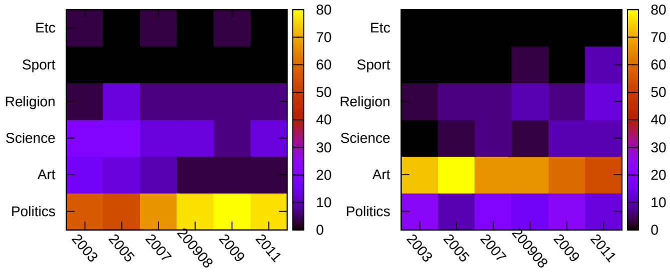

Fig. 5. Left panel: distribution of top 30 PageRank personalities over 6 activity categories at various years of Wikipedia. Right panel: distribution of top 30 2DRank personalities over the same activity categories at same years. Categories are politics, art, science, religion, sport, etc (other). Color shows the number of personalities for each activity expressed in percents.

PageRank. The mechanism of the qualitative ranking differences between two ranks is related to the fact that 2DRank takes into account via CheiRank a contribution of outgoing links. Due to that singers, actors, sportsmen increase their ranking since they are listed in various music albums, movies sport competition results. Due to that the component of arts gets higher positions in 2DRank in contrast to politics dominance in PageRank. Thus the two-dimensional ranking on PageRank-CheiRank plane allows to select qualities of nodes according to their popularity and communicativity.

4 Ranking of universities

The local ranking of top 100 universities is shown in Fig. 6 for years 2003, 2005, 2007 and in Fig. 7 for 2009, 200908, 2011. The local ranking is obtained by selecting top 100 universities appearing in PageRank listing so that they get their university ranking K from 1 to 100. The same procedure is done for CheiRank listing of universities obtaining their local CheiRank index K^* from 1 to 100. Those universities which enter inside 100×100 square on the local index plane (K, K^*) are shown in Figs. 6, 7.

The data show that the top PageRank universities are rather stable in time, e.g. U Harvard is always on the first top position. At the same time the positions in K^* are strongly changing in time. To understand the origin of this variations in CheiRank we consider the case of U Cambridge. Its Wikipedia article in 2003 is rather short but it contains the list of all 31 Colleges with direct links to their corresponding articles. This leads to a high position of U Cambridge with university $K^* = 4$ in 2003 (Fig. 8). However, with time the direct links remain only to about 10 Colleges while the whole number of Colleges are presented by a list of names without links. This leads to a significant increase of index up to $K^* \approx 40$ at Dec 2009. However, at Dec 2011 U Cambridge again improves significantly its CheiRank obtaining $K^* = 2$. The main reason of that is the appearance of section of "Notable alumni and academics" which provides direct links to articles about outstanding scientists studied and/or worked at U Cambridge that leads to second position at $K^* = 2$ among all universities. We note that in 2011 the top CheiRank University is George Mason University with university $K^* = 1$. The main reason of this high ranking is the presence of detailed listings of alumni in politics, media, sport with direct links to articles about corresponding personalities (including former director of CIA). These two examples show that the links, kept with a large number of university alumni, significantly increase CheiRank position of university. We note that artistic and politically oriented universities usually preserve more links with their alumni.

The time evolution of global ranking of top 10 universities of year 200908 for PageRank and 2DRank is shown in Fig. 8. The results show the stability of PageRank order with a clear tendency of top universities (e.g. Harvard) to go with time to higher and higher top positions of K. Thus for U Harvard the global value of K changes from $K \approx 300$ in 2003 to $K \approx 100$ in 2011, while the whole size N of the Wikipedia network increases almost by a factor 10 during this time interval. Since Wikipedia ranks all human knowledge, the stable improvement of PageRank indexes of universities reflects the global growing impor-

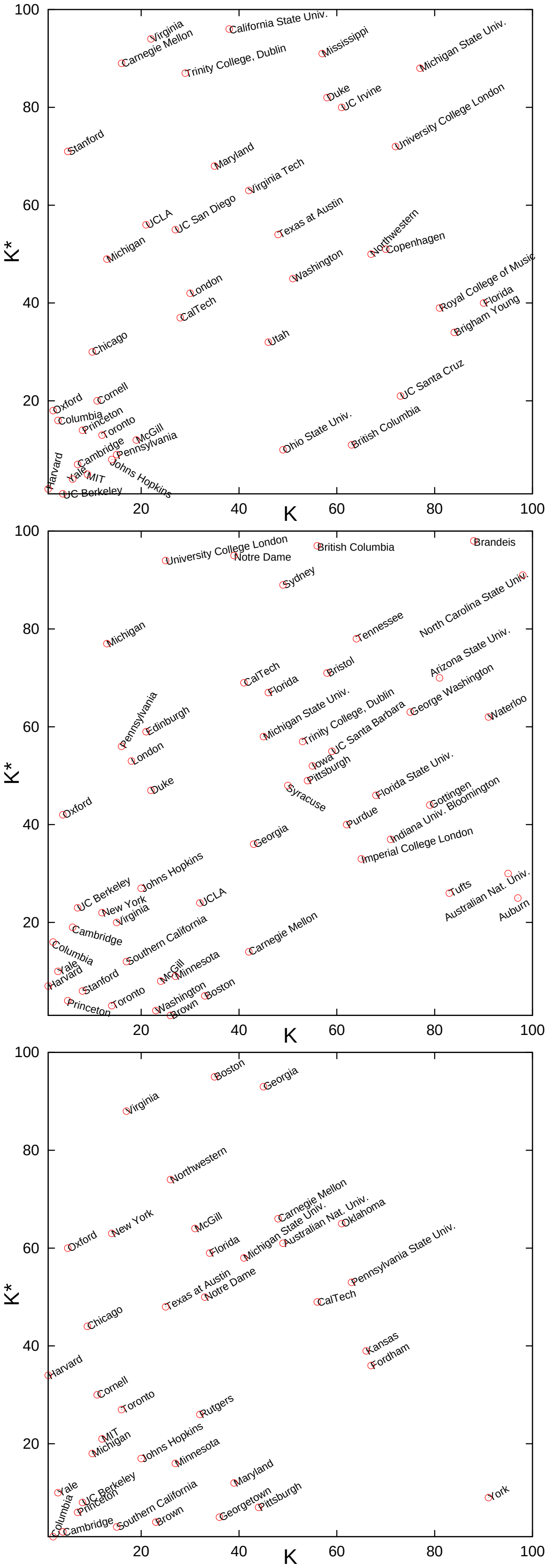

Fig. 6. University of Wikipedia articles in the local CheiRank versus PageRank plane at different years; panels are for years 2003, 2005, 2007 (from top to bottom).

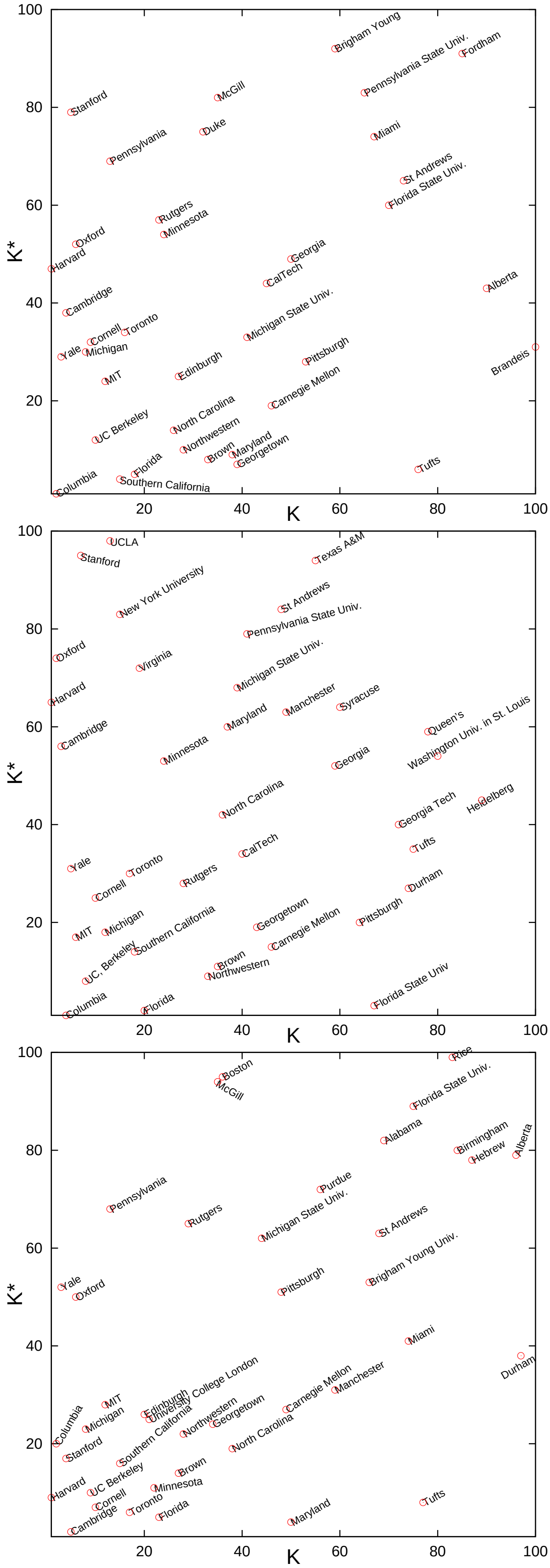

Fig. 7. Same as in Fig. 6 for years 2009, 200908, 2011 (from top to bottom).

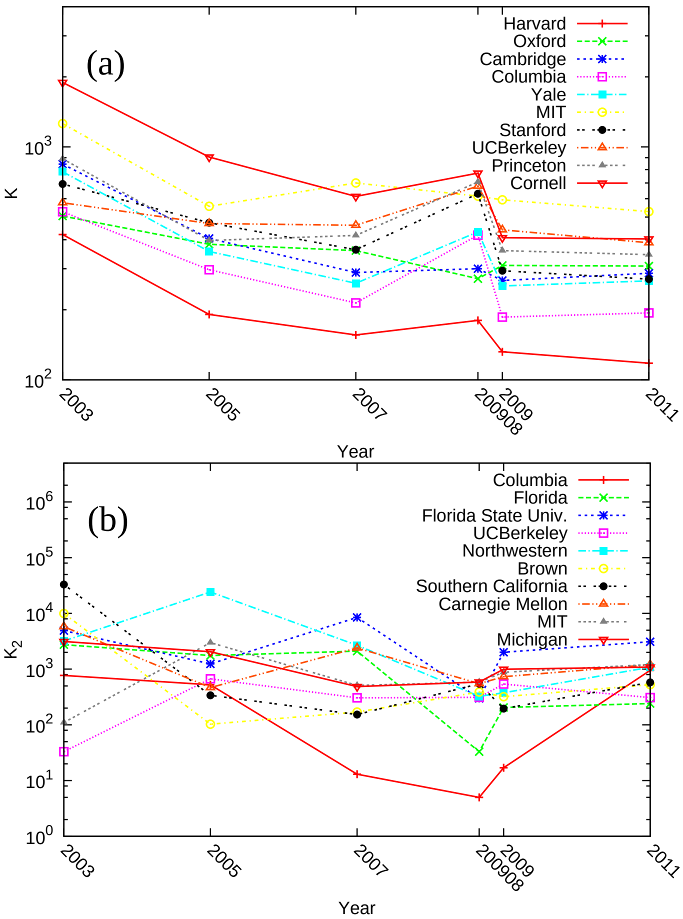

Fig. 8. Time evolution of global ranking of top 10 Universities of year 200908 in indexes of PageRank K (a) and 2DRank K_2 (b).

tance of universities in the world of human activity and knowledge.

The time evolution of the same universities in 2DRank remains stable in time showing certain interchange of their ranking order. We think that an example of U Cambridge considered above explains the main reasons of these fluctuations. In view of 10 times increase of the whole network size during the period 2003 - 2011 the average stability of 2DRank of universities also confirms the significant importance of their place in human activity.

Finally we compare the Wikipedia ranking of universities in their local PageRank index K with those of Shanghai university ranking [23]. In the top 10 of Shanghai university rank the Wikipedia PageRank recovers 9 (2003), 9 (2005), 8 (2007), 7 (2009), 7 (2011). This shows that the Wikipedia ranking of universities gives the results being very close to the real situation. A small decrease of overlap with time can be attributed to earlier launched activity of leading universities on Wikipedia.

5 Google matrix spectrum

Finally we discuss the time evolution of the spectrum of Wikipedia Google matrix taken at $\alpha = 1$. We perform the numerical diagonalization based on the Arnoldi method [12,13] using the additional improvements described in [14,15] with the Arnold dimension $n_A = 6000$. The Google matrix is reduced to the form

$$S = \begin{pmatrix} S_{ss} & S_{sc} \\ 0 & S_{cc} \end{pmatrix} \quad (4)$$

where S_{ss} describes disjoint subspaces V_j of dimension d_j invariant by applications of S; S_{cc} depicts the remaining part of nodes forming the wholly connected *core space*. We note that S_{ss} is by itself composed of many small diagonal blocks for each invariant subspace and hence those eigenvalues can be efficiently obtained by direct ("exact") numerical diagonalization. The total subspace size N_s, the number of independent subspaces N_d, the maximal subspace dimension $d_{\max}$ and the number N_1 of S eigenvalues with $\lambda = 1$ are given in Table 2 (See also Appendix). The spectrum and eigenstates of the core space S_{cc} are determined by the Arnoldi method with Arnoldi dimension n_A giving the eigenvalues λ_i of S_{cc} with largest modulus. Here we restrict ourselves to the statistical analysis of the spectrum λ_i. The analysis of eigenstates ψ_i ($G\psi_i = \lambda_i\psi_i$), which has been done in [11] for the slot 200908, is left for future studies.

The spectrum for all Wikipedia time slots is shown in Fig. 9 for G and in Fig. 10 for G^*. We see that the spectrum remains stable for the period 2007 - 2001 even if there is a small difference of slot 200908 due to a slightly different cleaning link procedure (see Appendix). For the spectrum of G^* in 2007 - 2001 we observe a well pronounced 3-6 arrow star structure. This structure is very similar to those found in random unistochastic matrices of side 3-4 [24] (see Fig.4 therein). This fact has been pointed in [11] for the slot 200908. Now we see that this is a generic phenomenon which remains stable in time. This indicates that there are dominant groups of 3-4 nodes which have structure similar to random unistochastic matrices with strong ties between 3-4 nodes and various random permutations with random hidden complex phases. The spectral arrow star structure is significantly more pronounce for the case of G^* matrix. We attribute this to more significant fluctuations of outgoing links that probably makes sectors of G^* to be more similar to elements of unistochastic matrices. A further detailed analysis will be useful to understand these arrow star structure and its links with various communities inside Wikipedia.

As it is shown in [11] the eigenstates of G and G^* select certain well defined communities of the Wikipedia network. Such an eigenvector detection of the communities provides a new method of communities detection in addition to more standard methods developed in network science and described in [25]. However, the analysis of eigenvectors represents a separate detailed research and in this work we restrict ourselves to PageRank and CheiRank vectors.

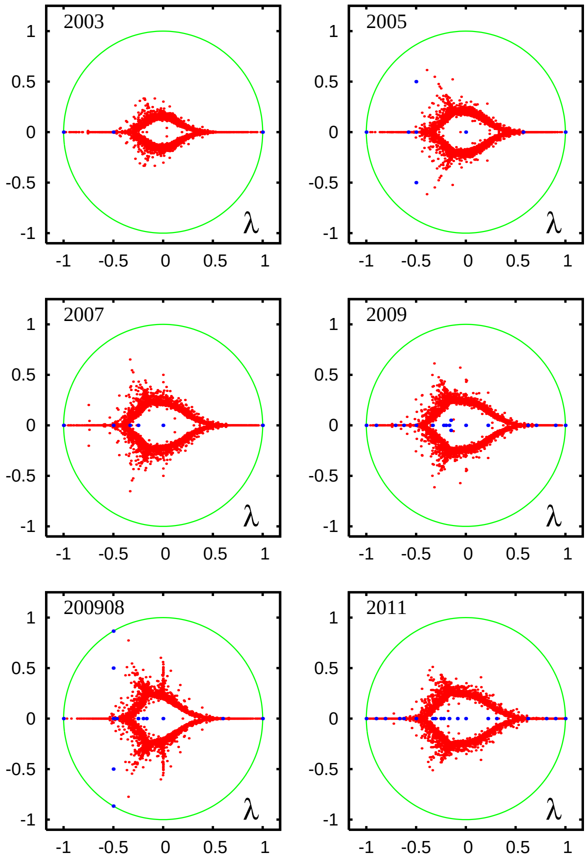

Fig. 9. Spectrum of eigenvalues λ of the Google matrix G of Wikipedia at different years. Red dots are core space eigenvalues, blue dots are subspace eigenvalues and the full green curve shows the unit circle. The core space eigenvalues were calculated by the projected Arnoldi method with Arnoldi dimensions $n_A = 6000$.

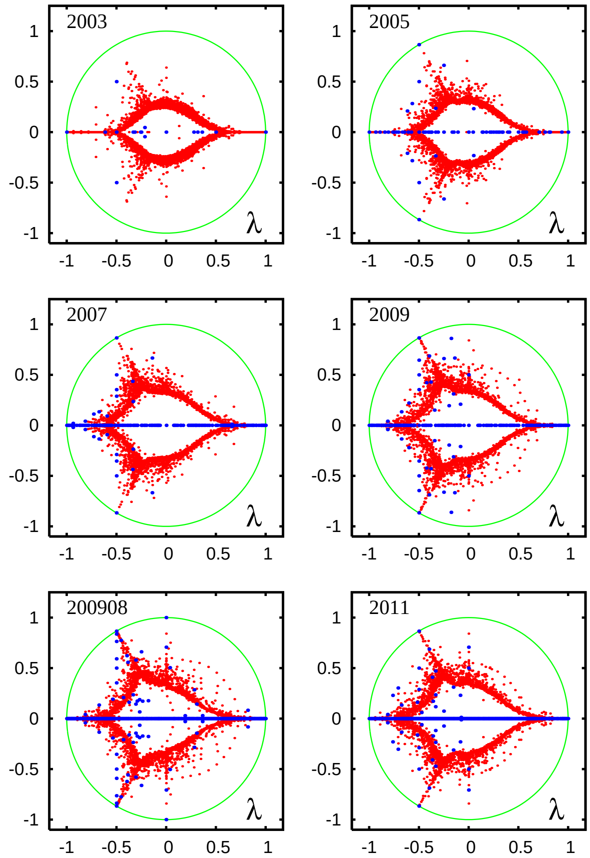

Fig. 10. Same as in Fig. 9 but for the spectrum of matrix G^*.

Finally we note that the fraction of isolated subspaces is very small for G matrix. It is increased approximately by a factor of order 10 for G^* but still it remains very small compared to the networks of UK universities analyzed in [15]. This fact reflects a strong connectivity of network of Wikipedia articles.

6 Discussion

In this work we analyzed the time evolution of ranking of network of English Wikipedia articles. Our study demonstrates the stability of such statistical properties as PageRank and CheiRank probabilities, the article density distribution in PageRank-CheiRank plane during the period 2007 - 2011. The analysis of human activities in different categories shows that PageRank gives main accent to politics while the combined 2DRank gives more importance to arts. We find that with time the number of politicians in the top positions increases. Our analysis of ranking of universities shows that on average the global ranking of top universities goes to higher and higher positions. This clearly marks the growing importance of universities for the whole range of human activities and knowledge. We find that Wikipedia PageRank recovers 70 - 80 % of top 10 universities from Shanghai ranking [23]. This confirms the reliability of Wikipedia ranking.

We also find that the spectral structure of the Wikipedia Google matrix remains stable during the time period 2007 -2011 and show that its arrow star structure reflects certain features of small size unistochastic matrices.

Acknowledgments: Our research presented here is supported in part by the EC FET Open project "New tools and algorithms for directed network analysis" (NADINE *No* 288956). This work was granted access to the HPC resources of CALMIP (Toulouse) under the allocations 2012-P0110, 2013-P0110. We also acknowledge the France-Armenia collaboration grant CNRS/SCS *No* 24943 (IE-017) on "Classical and quantum chaos".

7 Appendix

The tables with all network parameters used in this work are given in the text of the paper. The notations used in the tables are: N is network size, N_ℓ is the number of links, n_A is the Arnoldi dimension used for the Arnoldi method for the core space eigenvalues, N_d is the number of invariant subspaces, $d_{\max}$ gives a maximal subspace dimension, $N_{\text{circ.}}$ notes number of eigenvalues on the unit

	N	N_ℓ	n_A
2003	455436	2033173	6000
2005	1635882	11569195	6000
2007	2902764	34776800	6000
2009	3484341	52846242	6000
200908	3282257	71012307	6000
2011	3721339	66454329	6000

Table 1. Parameters of all Wikipedia networks at different years considered in the paper.

	N_s	N_d	$d_{\max}$	$N_{\text{circ.}}$	N_1
2003	15	7	3	11	7
2003*	940	162	60	265	163
2005	152	97	4	121	97
2005*	5966	1455	1997	2205	1458
2007	261	150	6	209	150
2007*	10234	3557	605	5858	3569
2009	285	121	8	205	121
2009*	11423	4205	134	7646	4221
200908	515	255	11	381	255
200908*	21198	5355	717	8968	5365
2011	323	131	8	222	131
2011*	14500	4637	1323	8591	4673

Table 2. G and G^* eigespectrum parameters for all Wikipedia networks, year marks spectrum of G, year with star marks spectrum of G^*.

circle with $|\lambda_i| = 1$, N_1 notes number of unit eigenvalues with $\lambda_i = 1$. We remark that $N_s \geq N_{\text{circ.}} \geq N_1 \geq N_d$ and $N_s \geq d_{\max}$. The data for G are marked by the corresponding year of the time slot, the data for G^* are marked by the year with a star. Links cleaning procedure eliminates all redirects (nodes with one outgoing link), this procedure is slightly different from the one used for the slot 200908 in [8]. All data sets and high resolution figures are available at the web page [26].

References

1. Wikipedia, *Wikipedia*, `en.wikipedia.org/wiki/Wikipedia`
2. F.A. Nielsen, *Wikipedia research and tools: review and comments*, (2012), available at SSRN: `dx.doi.org/10.2139/ssrn.2129874`
3. J.L. Borges, *The Library of Babel* in *Ficciones* (Grove Press. N.Y. 1962).
4. A.A. Markov, *Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga*, Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya, **15** (1906) 135 (in Russian) [English trans.: *Extension of the limit theorems of probability theory to a sum of variables connected in a chain* reprinted in Appendix B of: R.A. Howard *Dynamic Probabilistic Systems*, volume 1: *Markov models*, Dover Publ. (2007)]
5. M. Brin and G. Stuck, *Introduction to dynamical systems*, Cambridge Univ. Press, Cambridge, UK (2002)
6. S. Brin and L. Page, Computer Networks and ISDN Systems **30**, 107 (1998)
7. A. M. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton (2006)
8. A.O.Zhirov, O.V.Zhirov and D.L.Shepelyansky, Eur. Phys. J. B **77**, 523 (2010); `www.quantware.ups-tlse.fr/QWLIB/2drankwikipedia/`
9. L.Ermann, A.D.Chepelianskii and D.L.Shepelyansky, J. Phys. A: Math. Theor. **45**, 275101 (2012); `www.quantware.ups-tlse.fr/QWLIB/dvvadi/`
10. K.M.Frahm and D.L.Shepelyansky, Eur. Phys. J. B **85**, 355 (2012); `www.quantware.ups-tlse.fr/QWLIB/twittermatrix/`
11. L.Ermann, K.M. Frahm and D.L.Shepelyansky, *Spectral properties of Google matrix of Wikipedia and other networks*, arXiv:1212.1068 [cs.IR] (2012) (Eur. Phys. J. B in press).
12. G.W. Stewart, *Matrix Algorithms Eigensystems*, (SIAM, 2001), Vol. II
13. G.H. Golub and C. Greif, *BIT Num. Math.* **46**, 759 (2006)
14. K.M. Frahm and D.L. Shepelyansky, *Eur. Phys. J. B* **76**, 57 (2010)
15. K.M.Frahm, B.Georgeot and D.L.Shepelyansky, J. Phys, A: Math. Theor. **44**, 465101 (2011)
16. D. Fogaras, *Lect. Notes Comp. Sci.* **2877**, 65 (2003)
17. V. Hrisitidis, H. Hwang and Y. Papakonstantino, *ACM Trans. Database Syst.* **33**, 1 (2008)
18. A. D. Chepelianskii, *Towards physical laws for software architecture*, arXiv:1003.5455[cs.SE] (2010); `www.quantware.ups-tlse.fr/QWLIB/linuxnetwork/`
19. D. Donato, L. Laura, S. Leonardi and S. Millozzi, Eur. Phys. J. B **38**, 239 (2004)
20. G. Pandurangan, P. Raghavan and E. Upfal, Internet Math. **3**, 1 (2005)
21. N. Litvak, W.R.W. Scheinhardt, and Y. Volkovich, Lecture Notes in Computer Science, **4936**, 72 (2008).
22. M.H. Hart, *The 100: ranking of the most influential persons in history*, Citadel Press, N.Y. (1992).
23. `www.shanghairanking.com/`
24. K. Zyczkowski, M. Kus, W. Slomczynski and H.-J. Sommers, J. Phys. A: Math. Gen. **36**, 3425 (2003)
25. S. Fortunato, Phys. Rep. *486*, 75 (2010)
26. `www.quantware.ups-tlse.fr/QWLIB/wikirankevolution/`