Eur. Phys. J. B (2019) 92: 3
https://doi.org/10.1140/epjb/e2018-90532-7

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

World influence and interactions of universities from Wikipedia networks

Célestin Coquidé[1], José Lages[1], and Dima L. Shepelyansky[2,a]

[1] Institut UTINAM, Observatoire des Sciences de l'Univers THETA, CNRS, Université de Bourgogne Franche-Comté, Besançon, France
[2] Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

Received 5 September 2018 / Received in final form 14 November 2018
Published online 9 January 2019

Abstract. We present Wikipedia Ranking of World Universities (WRWU) based on analysis of networks of 24 Wikipedia editions collected in May 2017. With PageRank and CheiRank algorithms we determine ranking of universities averaged over cultural views of these editions. The comparison with the Shanghai ranking gives overlap of 60% for top 100 universities showing that WRWU gives more significance to their historical development. We show that the new reduced Google matrix algorithm allows to determine interactions between leading universities on a scale of ten centuries. This approach also determines the influence of specific universities on world countries. We also compare different cultural views of Wikipedia editions on significance and influence of universities.

1 Introduction

The importance of universities for progress of humanity is broadly recognized worldwide. Thus in 2017 UNESCO emphasizes the role of universities and higher education institutes in fostering sustainable development and empowering learners [1]. The efficiency of university education gained a high political importance in many world countries. Various tools have been developed to measure quantitatively this efficiency among which the ranking of universities gained significant importance as reviewed in [2]. Thus the Academic Ranking of World Universities (ARWU), compiled by Shanghai Jiao Tong University since 2003 (Shanghai ranking) [3], generated a significant political impact on evaluation of higher education efficiency in many countries [2]. For example, the ARWU stimulated the emergence of LABEX, IDEX projects in France [4] and the Russian Academic Excellence Project [5] with allocation of significant financial supports. In addition to ARWU other international ranking systems of universities appeared (see e.g. [6–8]). Various strong and weak features of ranking methodology are reviewed in [9–11]. Of course, the ranking systems are based on different specific criteria with different cultural preferences of rather larger groups realizing these rankings. Already, the presence of many ranking systems indicates the presence of bias in each of above ranking systems.

Another purely mathematical and statistical approach to ranking of world universities has been developed in [12–14] on the basis of Google matrix analysis of Wikipedia networks. For each Wikipedia language edition, a network is composed by all Wikipedia articles with directed links between them generated by mutual quotations of a given article to other articles. In [12], the analysis was performed only for English Wikipedia (ENWIKI) of year 2009, other years for ENWIKI were considered in [13], while in [14] this analysis was done with 24 language Wikipedia editions of 2013 that allowed to reduce significantly cultural bias (these 24 networks had been collected and analyzed for historical figures in [15] in the frame of EC FET Open project NADINE [16]). The Google matrix analysis [12–14] is based on the PageRank algorithm [17] which detailed description is given in [18]. Some additional characteristics have also been used for the description of network nodes (Wikipedia articles), like CheiRank and 2DRank, as described in [19,20]. Thus, the Wikipedia Ranking of World Universities (WRWU2013) from 24 Wikipedia networks was introduced in [14] and it was shown that its top 100 universities have 62% overlap with ARWU. In addition, WRWU2013 attracted a significant interest worldwide (see [21] and various press highlights listed at [22]). Other research groups also start to apply Wikipedia ranking in Wikiometrics [23]. We also note the growing interest to scientific analysis of several language editions of Wikipedia [24].

In this work, we extend the WRWU studies started in [14]. The new elements are: we use 24 Wikipedia editions collected in May 2017 [25] and also we apply the recently invented reduced Google matrix (REGOMAX) method [26]. This new method allows to determine effective interactions between a selected relatively small subset of network nodes taking into account all pathways

[a] e-mail: dima@irsamc.ups-tlse.fr

Table 1. Wikipedia directed networks of 2017 from 24 considered language editions; here N is the number of articles. Wikipedia data were collected in May 2017.

Edition	Language	N	Edition	Language	N
EN	English	5416537	ZH	Chinese	939625
SV	Swedish	3786455	FA	Persian	539926
DE	German	2057898	AR	Arabic	519714
NL	Dutch	1900222	HU	Hungarian	409297
FR	French	1866546	KO	Korean	380086
RU	Russian	1391225	TR	Turkish	291873
IT	Italian	1353276	MS	Malaysian	289234
ES	Spanish	1287834	DA	Danish	225523
PL	Polish	1219733	HE	Hebrew	205411
VI	Vietnamese	1155932	EL	Greek	130429
JP	Japanese	1058950	HI	Hindi	121503
PT	Portuguese	967162	TH	Thai	116495

between them via the global huge network with millions of nodes. The efficiency of the REGOMAX method has been demonstrated on examples of analysis of interactions of political leaders [27], terror networks [28] and protein–protein interactions in cancer networks [29]. Here, using the REGOMAX method we obtain effective interactions between a group of selected universities and determine their influence on world countries. The new ranking of universities from Wikipedia 2017 editions is compared with those of 2013.

The paper is composed as follows: Section 2 gives description of network datasets; Section 3 describes Google matrix construction and PageRank, CheiRank algorithms with overview of the reduced Google matrix approach; Section 4 presents results on global ranking of universities from 24 editions and their distribution over world countries; Section 5 provides REGOMAX results for English edition determining the world influence of specific universities; the interactions between top 20 universities are analyzed in Section 6 comparing views of English, French, German and Russian editions; in Section 7, we obtained the reduced Google matrix of top 100 universities averaged over 24 editions and analyze the interactions between universities on the scale of 10 centuries and all continents; discussion of the results is given in Section 8. All detailed ranking results of WRWU2017 are available at [30] and arXiv version of this work [31], which contains Supplementary Information with many ranking lists and additional figures being too heavy to be included here.

2 Datasets

We use the datasets of 24 Wikipedia editions extracted in May 2017 [25] (see also [30]). The size of each network is given in Table 1. Compared to 2013 discussed in [15] there is a significant size increase for each edition, especially for Swedish (SV) where a part of articles is now computer generated. The number of links is given at [30]. On average there are about 20 links per node. Self-citation links are not considered (references on the article inside the same article are eliminated).

3 Description of algorithms and methods

3.1 Google matrix, PageRank and CheiRank algorithms

The mathematical grounds of this study are based on Markov chain theory and, in particular, on the Google matrix analysis initially introduced in 1998 by Google's co-founders, Brin and Page [17], for hypertext analysis of the World Wide Web. Let us consider the network of the N articles of a given Wikipedia edition. The network adjacency matrix element A_{ij} is equal to 1 if article j quotes article i and equal to 0 otherwise. The Google matrix element $G_{ij} = \alpha S_{ij} + (1-\alpha)/N$ gives a transition probability that a random reader jumps from article j to article i. The stochastic matrix element S_{ij} is $S_{ij} = A_{ij}/\sum_{i=1}^{N} A_{ij}$ if article j quotes at least one other article, otherwise $S_{ij} = 1/N$. The second term in G proportional to $(1-\alpha)$, where $0.5 < \alpha < 1$ is the damping factor, allows to a random reader to escape from isolated sets of articles. More details can be found in [18]. Here, we use the value $\alpha = 0.85$ typical for WWW studies [18]. The Google matrix G, constructed as described above, belongs to the class of Perron–Frobenius operators [18]. The eigenvector $\mathbf{P}$ with the largest eigenvalue $\lambda = 1$ is the solution of equation $G\mathbf{P} = \mathbf{P}$. This PageRank vector $\mathbf{P}$ has positive or zero components and describes the steady-state probability distribution of the Markov process encoded in the Google matrix G. Assuming an infinite random process, the vector component P_i is proportional to the number of times a random reader reaches an article i. It is convenient to sort the vector components $P_1, \ldots, P_N$ in descending order: the article associated to the highest (lowest) vector component has the top (last) rank index $K = 1$ ($K = N$). The PageRank algorithm measures the relative influence of articles. Recursively, more an article is quoted by influent articles, more high is its probability.

As proposed in [32] we also consider the same network of articles but with inverted links, i.e., article j points toward article i if article j is quoted by article i. This inverted network is defined by the adjacency matrix elements, $A^*_{ij} = A_{ji}$, which can be used to build successively the corresponding stochastic matrix elements, S^*_{ij}, and the corresponding Google matrix elements, G^*_{ij}. The CheiRank vector $\mathbf{P}^*$ is then defined such as $G^*\mathbf{P}^* = \mathbf{P}^*$ and the CheiRank is constructed similarly to the PageRank [12,19,32]. The CheiRank algorithm measures the relative communicative ability of the articles. Recursively, the more an article quotes very communicative articles, the more it is communicative.

The properties of the Google matrix spectrum and eigenstates and their various applications are discussed in detail in [18–20].

3.2 The reduced Google matrix

The concept of reduced Google matrix (REGOMAX) was introduced in [26] and tested with Wikipedia networks in [27,28] and protein–protein networks [29]. The method is based on the construction of a Google matrix for a relatively small subset of nodes embedded into a much larger network taking into account all indirect interactions between subset nodes via the remaining huge part of the network.

Let us consider a small subset $\mathcal{S}_\mathrm{r}$ of $n_\mathrm{r} \ll N$ articles, and the complementary subset $\mathcal{S}_\mathrm{s}$ of the $n_\mathrm{s} = N - n_\mathrm{r} \simeq N$ remaining articles. For convenience, the Google matrix can be rewritten as

$$G = \begin{pmatrix} G_\mathrm{rr} & G_\mathrm{rs} \\ G_\mathrm{sr} & G_\mathrm{ss} \end{pmatrix} \tag{1}$$

where the submatrix G_rr, of size $n_\mathrm{r} \times n_\mathrm{r}$, encodes the transitions between articles of the subset $\mathcal{S}_\mathrm{r}$, the submatrix G_ss, of size $n_\mathrm{s} \times n_\mathrm{s}$, encodes the transitions between articles of the subset $\mathcal{S}_\mathrm{s}$, the submatrix G_rs, of size $n_\mathrm{r} \times n_\mathrm{s}$, encodes the transitions from articles of the subset $\mathcal{S}_\mathrm{s}$ toward articles of the subset $\mathcal{S}_\mathrm{r}$, the submatrix G_sr, of size $n_\mathrm{s} \times n_\mathrm{r}$, encodes the transitions from articles of the subset $\mathcal{S}_\mathrm{r}$ toward articles of the subset $\mathcal{S}_\mathrm{s}$. Since $G\mathbf{P} = \mathbf{P}$, the PageRank vector can be rewritten as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_\mathrm{r} \\ \mathbf{P}_\mathrm{s} \end{pmatrix} \tag{2}$$

where the vector $\mathbf{P}_\mathrm{r}$ ($\mathbf{P}_\mathrm{s}$) of size n_r (n_s) contains the PageRank vector components associated to articles of the $\mathcal{S}_\mathrm{r}$ ($\mathcal{S}_\mathrm{s}$) subset. The reduced Google matrix G_R associated to articles of the subset $\mathcal{S}_\mathrm{r}$ is the $n_\mathrm{r} \times n_\mathrm{r}$ matrix defined implicitly by the following relation $G_\mathrm{R}\mathbf{P}_\mathrm{r} = \mathbf{P}_\mathrm{r}$. After some algebra, the reduced Google matrix can be written as [26,27,29]

$$G_\mathrm{R} = G_\mathrm{rr} + G_\mathrm{ind} \text{ where } G_\mathrm{ind} = G_\mathrm{rs}\left(1_\mathrm{s} - G_\mathrm{ss}\right)^{-1} G_\mathrm{sr}. \tag{3}$$

Here 1_s is the $n_\mathrm{s} \times n_\mathrm{s}$ identity matrix. The reduced Google matrix G_R is composed by the G_R-submatrix G_rr which encodes the direct links (direct quotations) between the n_r articles of the $\mathcal{S}_\mathrm{r}$ subset and by an additional scattering term G_ind which quantifies the indirect links between articles. If there is no direct link from article $j \in \mathcal{S}_\mathrm{r}$ to article $i \in \mathcal{S}_\mathrm{r}$, i.e., $A_{ij} = 0$, then the corresponding G_rr element will be minimum ($G_{\mathrm{rr}_{ij}} \sim 1/N \sim 10^{-7}$ for the May 2017 English Wikipedia network). Conversely, the corresponding G_ind element can be very high highlighting the fact that two articles can be strongly indirectly linked through successive direct links between articles of the $\mathcal{S}_\mathrm{s}$ subset (e.g., $j \in \mathcal{S}_\mathrm{r} \to k_1 \in \mathcal{S}_\mathrm{s} \to k_2 \in \mathcal{S}_\mathrm{s} \to \cdots \to k_n \in \mathcal{S}_\mathrm{s} \to i \in \mathcal{S}_\mathrm{r}$). The PageRank vector of G_R has the same components of n_r nodes as in the global matrix G (up to a constant normalization factor). The reduced Google matrix G_R, which conserves the global Google matrix PageRank hierarchy between the n_r articles of the $\mathcal{S}_\mathrm{r}$ subset, encodes direct links and effective indirect links between articles. The direct calculation of G_ind converges very slowly since the matrix $\left(1_\mathrm{s} - G_\mathrm{ss}\right)^{-1}$ is almost singular, indeed as $n_\mathrm{r} \ll n_\mathrm{s}$, $G_\mathrm{ss} \sim G$, the leading eigenvalue of G_ss is $\lambda_c \sim 1$. Let us associate to the eigenvalue λ_c the right eigenvector $\mathbf{\Psi}_\mathrm{R}$ and the left eigenvector $\mathbf{\Psi}_L$ such as $G_\mathrm{ss}\mathbf{\Psi}_\mathrm{R} = \lambda_c\mathbf{\Psi}_\mathrm{R}$ and $\mathbf{\Psi}_L^T\mathbf{\Psi}_\mathrm{R} = 1$. To speed up calculations, we follow the same procedure as in [26,27,29], splitting the term $\left(1_\mathrm{s} - G_\mathrm{ss}\right)^{-1}$ in a term $\mathbf{\Psi}_\mathrm{R}\mathbf{\Psi}_L^T\left(1 - \lambda_c\right)^{-1}$ which is a projection onto the subspace associated to λ_c and a term $\left(1_\mathrm{s} - \mathbf{\Psi}_\mathrm{R}\mathbf{\Psi}_L^T\right)\left(1_\mathrm{s} - G_\mathrm{ss}\right)^{-1}$ which is a projection onto the complementary subspace. This procedure enables us to rewrite the reduced Google matrix as

$$G_\mathrm{R} = G_\mathrm{rr} + G_\mathrm{pr} + G_\mathrm{qr} \tag{4}$$

where $G_\mathrm{pr} = G_\mathrm{rs}\mathbf{\Psi}_\mathrm{R}\mathbf{\Psi}_L^T G_\mathrm{sr}\left(1 - \lambda_c\right)^{-1}$ encodes essentially already known information concerning the PageRank (since $\mathbf{\Psi}_\mathrm{R} \sim \mathbf{P}$) and $G_\mathrm{qr} = G_\mathrm{rs}(1_\mathrm{s} - \mathbf{\Psi}_\mathrm{R}\mathbf{\Psi}_L^T)\left(1_\mathrm{s} - G_\mathrm{ss}\right)^{-1} G_\mathrm{sr}$ encodes hidden interactions between articles which appear due to indirect links via the global network [26,27,29]. In the following, we perform analysis of the three components present in (4), we also consider the matrix component G_qrnd obtained from G_qr by taking out the diagonal terms since the self-citations are not very interesting.

4 Wikipedia Ranking of World Universities from 24 Wikipedia editions of 2017

Once the articles of 24 considered editions are ranked using PageRank and CheiRank algorithms, we extract for each edition the top 100 articles devoted to institutions of higher education and research. We also consider 2DRank which is a combination of PageRank and CheiRank (see [12,14]; 2DRank results are available at [30]). As in [14,15], from these 24 top 100 listings we obtain the following cumulative score for a given university U

$$\Theta_{U,A} = \sum_E \left(101 - R_{U,E,A}\right) \tag{5}$$

where A denotes the algorithm used for the ranking (PageRank or CheiRank or 2DRank), E the Wikipedia edition, $R_{U,E,A}$ the rank of university U in the top 100 universities obtained using algorithm A from edition E of Wikipedia. If a university U' is absent from the top 100 universities obtained from an edition E' with an algorithm A' then we

Table 2. List of the first 10 universities of the 2017 Wikipedia Ranking of World Universities using PageRank algorithm. For a given university, the score Θ_{PR} is defined by (5), N_a is the number of appearances in the top 100 lists of 24 Wikipedia editions, CC is the country code, LC is the language code, and FC is the foundation century.

Rank	Θ_{PR}	N_a	University	CC	LC	FC
1st	2281	24	University of Oxford	UK	EN	11
2nd	2278	24	University of Cambridge	UK	EN	13
3rd	2277	24	Harvard University	US	EN	17
4th	2099	24	Columbia University	US	EN	18
5th	1959	23	Yale University	US	EN	18
6th	1917	24	University of Chicago	US	EN	19
7th	1858	23	Princeton University	US	EN	18
8th	1825	21	Stanford University	US	EN	19
9th	1804	21	Massachusetts Institute of Technology	US	EN	19
10th	1693	20	University of California, Berkeley	US	EN	19

Table 3. List of the first 10 universities of ARWU2017 [3]. The last columns show the difference between ARWU2017 rank and WRWU2017 rank.

Rank	ARWU17	WRWU17
1st	Harvard University	−2
2nd	Stanford University	−6
3rd	University of Cambridge	+1
4th	Massachusetts Institute of Technology	−5
5th	University of California, Berkeley	−5
6th	Princeton University	−1
7th	University of Oxford	+6
8th	Columbia University	+4
9th	California Institute of Technology	−13
10th	University of Chicago	+4

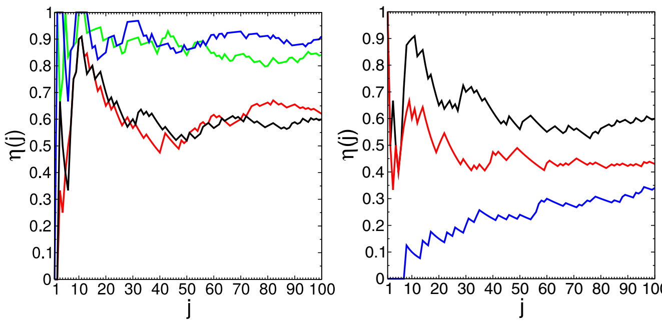

Fig. 1. Left panel: overlap $\eta(j) = j_c/j$ of ARWU2017 with WRWU2017 as a function of the rank j of WRWU2017. Here, j_c gives the number of common universities among the first j universities of the two rankings. The color curves show the overlap of ARWU2017 with WRWU2017 (black curve), of ARWU2013 with WRWU2013 (red curve), of ARWU2013 with ARWU2017 (green curve), and of WRWU2013 with WRWU2017 (blue curve). Right panel: overlap of ARWU2017 with ENWRWU2017 (black curve), of ARWU2017 with FRWRWU2017 (red curve) and of ARWU2017 with DEWRWU2017 (blue curve). The horizontal axis label j is the rank of ARWU2017.

artificially set $R_{U',E',A'} = 101$. We use ISO 3166-1 alpha-2 country codes [33] (all the used country codes are available at [30]).

The top 10 universities from WRWU2017 with PageRank algorithm and from ARWU2017 are given in Tables 2 and 3, respectively. The top 3 places of WRWU are occupied by Oxford, Cambridge and Harvard while for ARWU it is Harvard, Stanford and Cambridge. The universities with significantly lower positions in WRWU (compared to ARWU) are MIT, Berkeley and Caltech, while Oxford significantly improves its position at WRWU going to the first place from 7th at ARWU.

The overlap between different rankings is presented in Figure 1. For the top 100 universities we have 60% overlap between WRWU PageRank list and ARWU list in 2017. For 2013 this overlap was slightly higher at 62%. The overlap between WRWU2017 list and WRWU2013 list is 91% and between ARWU2017 list and ARWU2013 list is 84%. WRWU appears to be stable, top 10s in 2013 (Tab. 2) and 2017 (Tab. 3 in [14]) contain the same universities but with some changes in places: Oxbridge keeps the two first places but Oxford supersedes Cambridge at the first place; Yale (9 → 5) and Chicago (7 → 6) improve their ranking whereas Princeton (5 → 7) and MIT (6 → 9) recede; Harvard (3 → 3), Columbia (4 → 4), Stanford (8 → 8) and Berkeley (10 → 10) keep their positions. Between 2013 [22] and 2017 [30], only 9 universities went out from top 100 (IPSA/Karolinska Institutet/Rockefeller/Rutgers/Tsinghua/Amsterdam/Hamburg/Strasbourg/Wrocław) and 9 new universities enter in the top 100 (Seoul National University/TU Munich/UC, San Diego/Boulder/Freiburg/Kiel/Marburg/Salamanca/Sydney).

The above numbers and comparisons show that WRWU approach gives a reliable ranking which remains relatively close to ARWU at different years. At the same time WRWU has about 40% of different universities compared to ARWU. The origin of this difference is based on variety of cultural views well present in 24 editions. Thus, the right panel of Figure 1 shows a spectacular difference between EN, FR and DE editions: for top 10 universities the German edition has only about 10% overlap with ARWU while EN and FR have about 50%. For top 100 this difference still remains significant being approximately 34% for DE, 43% for FR and 60% for EN. Thus, the case of German edition demonstrates rather different cultural view on importance of their universities. As discussed in [14] there are clear historical grounds for this

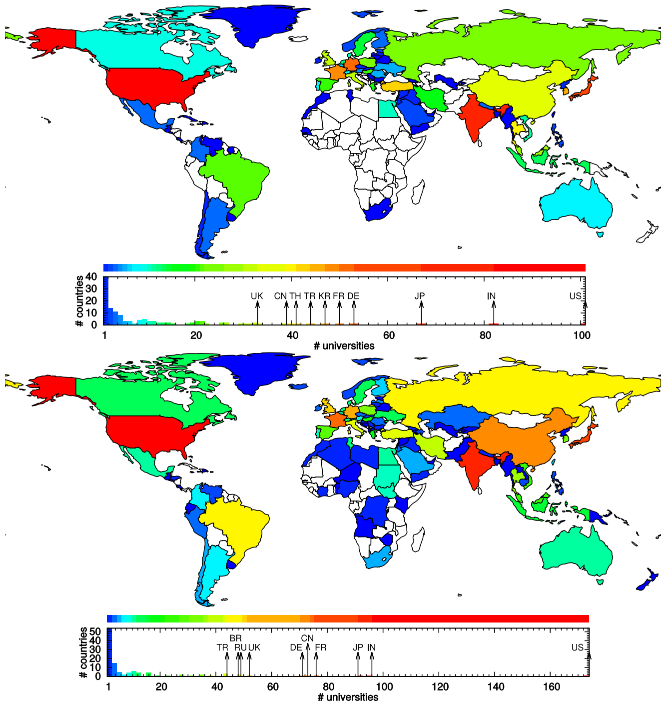

Fig. 2. Geographical distribution of the universities entering the 2017 Wikipedia Ranking of World Universities using PageRank (top panel) and CheiRank algorithms (bottom panel). The total number of universities is 1011 (1464) for WRWU using PageRank (CheiRank) algorithm. US universities are the most numerous: 101 (174) universities for WRWU using PageRank (CheiRank) algorithm. Countries with white color have no universities in the top 100 edition lists. Here and below the color categories are obtained using the Jenks natural breaks classification method [34].

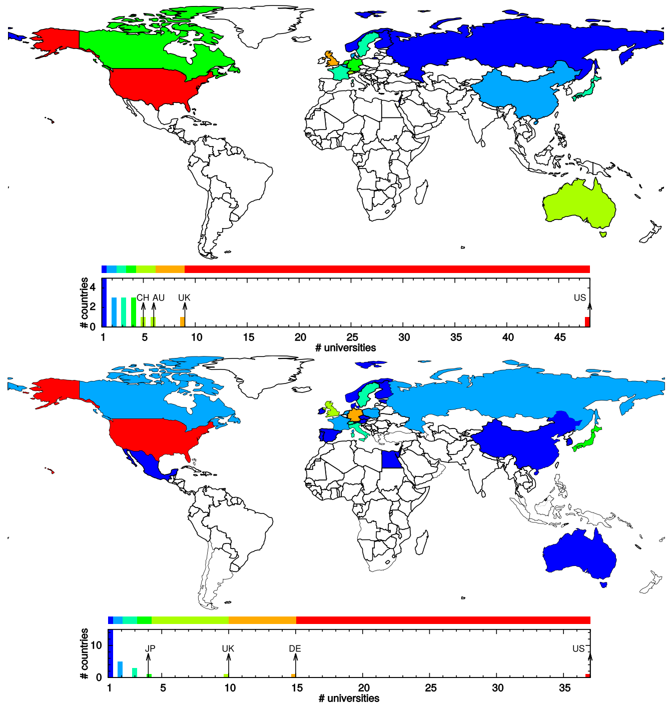

Fig. 3. Geographical distribution of the first 100 universities from ARWU2017 (top panel) and WRWU2017 from PageRank (bottom panel). US universities are the most numerous, 48 universities for ARWU2017 and 37 universities for WRWU2017.

difference related to the world dominance of German and Italian universities before 19th century as it is clearly seen in Figure 10 in [14].

In total for all 24 editions, we find 1011 and 1464 different universities with PageRank and CheiRank algorithms, respectively. Their geographical distribution over the country world map is shown in Figure 2. The largest number of top universities is in US but we see a significant numbers also for India, Japan, Germany and France for PageRank which characterizes the university influence. The communicativity is highlighted by CheiRank with top countries being US, India, Japan, France and China. Of course, Hindi, Japan and Chinese editions give certain preference to their own universities but in global the high positions of these universities and countries reflect significant efforts in higher education performed by these countries.

The geographical distributions of top 100 universities of ARWU2017 and WRWU2017 are presented in Figure 3. For ARWU the top countries are US, UK, Australia and China while for WRWU we find US, Germany, UK and Japan. It is clear that ARWU gives too high significance to Anglo-Saxon and Chinese universities while WRWU provides more balanced historical view taking into account a significant role played e.g. by German universities.

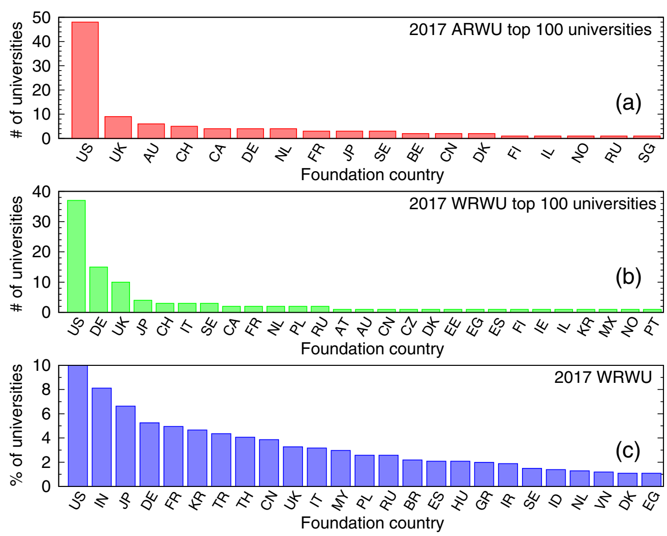

Fig. 4. Distribution over countries of (a) 2017 ARWU top 100 universities and of (b) 2017 WRWU top 100 universities. Panel (c) gives the percentage per country of universities among the 1011 universities listed in 2017 WRWU, countries with less than 10% are not shown. Countries with equal number of universities are sorted by alphabetic order.

A more detailed view on the universities distribution over countries for ARWU and WRWU is shown in Figure 4. The ranking of WRWU universities inside each country is given in [30].

Table 4. List of the PageRank top 20 universities of English edition WRWU2017. The color code corresponds to the regional location of universities: red for US west coast, orange for US central region, blue for US east coast and violet for UK.

Rank	University	Rank	University
1st	**Harvard University**	11th	**University of Michigan**
2nd	**University of Oxford**	12th	**Cornell University**
3rd	**University of Cambridge**	13th	**University of California, Los Angeles**
4th	**Columbia University**	14th	**University of Pennsylvania**
5th	**Yale University**	15th	**New York University**
6th	**Stanford University**	16th	**University of Texas at Austin**
7th	**Massachusetts Institute of Technology**	17th	**University of Florida**
8th	**University of California, Berkeley**	18th	**University of Edinburgh**
9th	**Princeton University**	19th	**University of Wisconsin-Madison**
10th	**University of Chicago**	20th	**University of Southern California**

5 Influence of world universities on countries from English Wikipedia edition

In this section, we use the REGOMAX approach to analyze the influence of universities on world countries. With this aim the reduced Google matrix is constructed for the subset of articles devoted to the PageRank top 20 universities of ENWRWU (see Tab. 4) and articles devoted to the 85 countries to which belong universities appearing in WRWU (see Tab. 5). Thus, the total size of the reduced Google matrix is $n_r = 105$, to be compared to the global ENWIKI network size which is about 5.4 million articles.

The images of the corresponding reduced Google matrix $G_{\rm R}$ and its components $G_{\rm rr}$, $G_{\rm pr}$ and $G_{\rm qr}$ are shown in Figure 5. As discussed above and in [27] the $G_{\rm pr}$ component is rather close to the matrix composed by identical columns of PageRank vector of n_r nodes, the direct links are presented by the component $G_{\rm rr}$ and indirect links by $G_{\rm qr}$ (and related $G_{\rm qrnd}$). The weights of these three components (sum of elements of all columns divided by matrix size n_r) are respectively $W_{\rm R} = 1$, $W_{\rm pr} = 0.948273$, $W_{\rm rr} = 0.0144137$ and $W_{\rm qr} = 0.0373132$. The weights of the components $G_{\rm rr}$ and $G_{\rm qr}$ are small, compared to those of $G_{\rm pr}$, but these two components provide new important information on interactions between nodes. The weight of indirect links is larger than the direct ones $W_{\rm qr} > W_{\rm rr}$.

The knowledge of all matrix elements of $G_{\rm R}$ allows us to determine the influence or sensitivity of a given university u on a given country c. To measure the sensitivity we change the matrix element $G_{\rm R}(u \to c)$ by a factor $(1+\delta)$ with $\delta \ll 1$, we renormalize to 1 the sum of the column associated to university u, and we compute the logarithmic derivative of PageRank probability $P(c)$ associated to country c: $D(u \to c, c) = d \ln P(c)/d\delta$ (diagonal sensitivity). It is also possible to consider the nondiagonal sensitivity $D(u \to c, c') = d \ln P(c')/d\delta$ when the variation is done for the link from u to c and the derivative of PageRank probability is computed for another country c'. This approach was already used in [28,35] showing its efficiency.

The world maps of university influence on countries, expressed by the diagonal sensitivity $D(u \to c, c)$ for 4 selected universities, are shown in Figure 6.

For Harvard the most sensitive country is South Africa (ZA) due to the well known scandal linked to Harvard investments in apartheid ZA pointed on the Harvard wikipage. Puerto Rico also appears on this wikipage in relation with oldest universities in the America. However, next influenced countries are Georgia (GE), Israel (IL) and Ireland (IR) which are not present on the wikipage which appearance we attribute to indirect links.

For Chicago the most influenced countries are Singapore (SG), Puerto Rico (PR) and India (IN). The first two countries SG and PR are not present on the wikipage showing that indirect links play an important role for them. India has a direct link related to the following facts: Chicago campus opened in India and a faculty member was erstwhile governor of India central bank.

For Stanford the top countries are Spain (ES), present on the wikipage, PR and ZA appearing due to indirect links.

For Oxford the top three countries are Jordan (JO), appearing of wikipage since Abdullah II of JO has been educated at Oxford; Iraq (IQ), also appearing on wikipage since T.E. Lawrence educated at Oxford played a major role in establishing and administering the modern state of IQ; IR which is not present on wikipage but has close links with UK.

Examples of nondiagonal sensitivity are shown in Figure 7 for the link variation Harvard University to US and University of Oxford to UK. For the link Harvard→US the most influenced countries are Suriname (SR), People's Republic of Korea (KP) and Puerto Rico (PR). For the link variation Oxford→UK the most influenced countries are Qatar (QA), Ireland (IR) and Singapore (SG). By definition the nondiagonal sensitivity contains indirect effects and it is not so easy to find the pathways of links which are responsible for this dominant influence. These examples show the strength of reduced Google matrix approach.

Finally we discuss an example of Rice University. It has rank 74 in ARWU2017 being at position 37 inside USA, but according to WRWU2017 its PageRank position is only 357 and 56 inside USA. This shows that the Wikipedia article of Rice University is not sufficiently developed and its visibility via Wikipedia is rather low and can be improved by more skillful organization of its wikipage. We note that Wikipedia visibility of a university plays rather

Table 5. List of the countries associated to the universities appearing in WRWU2017. Countries are ranked from 2017 Wikipedia English edition using PageRank algorithm.

Rank	University	CC	Rank	University	CC
1	United States	US	44	Chile	CL
2	France	FR	45	Republic of Ireland	IE
3	Germany	DE	46	Singapore	SG
4	United Kingdom	UK	47	Serbia	RS
5	Iran	IR	48	Vietnam	VN
6	India	IN	49	Nepal	NP
7	Canada	CA	50	Estonia	EE
8	Australia	AU	51	Iraq	IQ
9	China	CN	52	Bangladesh	BD
10	Italy	IT	53	Syria	SY
11	Japan	JP	54	Myanmar	MM
12	Russia	RU	55	Slovakia	SK
13	Brazil	BR	56	Venezuela	VE
14	Spain	ES	57	Morocco	MA
15	Netherlands	NL	58	Cuba	CU
16	Poland	PL	59	Puerto Rico	PR
17	Sweden	SE	60	Saudi Arabia	SA
18	Mexico	MX	61	Lithuania	LT
19	Turkey	TR	62	Lebanon	LB
20	Romania	RO	63	Cyprus	CY
21	South Africa	ZA	64	Latvia	LV
22	Norway	NO	65	Belarus	BY
23	Switzerland	CH	66	United Arab Emirates	AE
24	Philippines	PH	67	Uruguay	UY
25	Austria	AT	68	North Korea	KP
26	Belgium	BE	69	Yemen	YE
27	Argentina	AR	70	Costa Rica	CR
28	Indonesia	ID	71	Tunisia	TN
29	Greece	GR	72	Jordan	JO
30	Denmark	DK	73	Guatemala	GT
31	South Korea	KR	74	Greenland	GL
32	Israel	IL	75	Dominican Republic	DO
33	Hungary	HU	76	Uzbekistan	UZ
34	Finland	FI	77	Kuwait	KW
35	Egypt	EG	78	Qatar	QA
36	Portugal	PT	79	Senegal	SN
37	Taiwan	TW	80	El Salvador	SV
38	Ukraine	UA	81	Suriname	SR
39	Czech Republic	CZ	82	Faroe Islands	FO
40	Malaysia	MY	83	Brunei	BN
41	Thailand	TH	84	Palestine	PS
42	Colombia	CO	85	Georgia	GE
43	Bulgaria	BG			

important role since very often it is on top lines of Google search and many language editions spread the wikipage content worldwide free of charge. Also the WRWU position 357 is only due to appearance of Rice in top 100 universities of FA, KO and TH language editions probably due to wikipage creation and information given by Rice alumni speaking these languages. This indicates an importance of links between university and its alumni.

The world influence of Rice University, expressed via its diagonal sensitivity, is shown in Figure 8. The most influenced countries are Kuwait (KW), Mexico (MX) and Puerto Rico (PR). These countries are not present on the wikipage of Rice University and their appearance is related to indirect links.

6 Reduced networks of world universities

In this section, we analyze the interactions between top 20 universities obtained from the REGOMAX approach. We consider the different cultural views from EN, FR, DE and RU editions and make a comparative analysis. For EN edition, we also consider the links of selected universities with countries.

6.1 Top 20 universities in English Wikipedia edition

Top 20 PageRank universities of ENWRWU2017 are given in Table 4. They are either US or UK universities. We define 4 regional groups with their PageRank leaders:

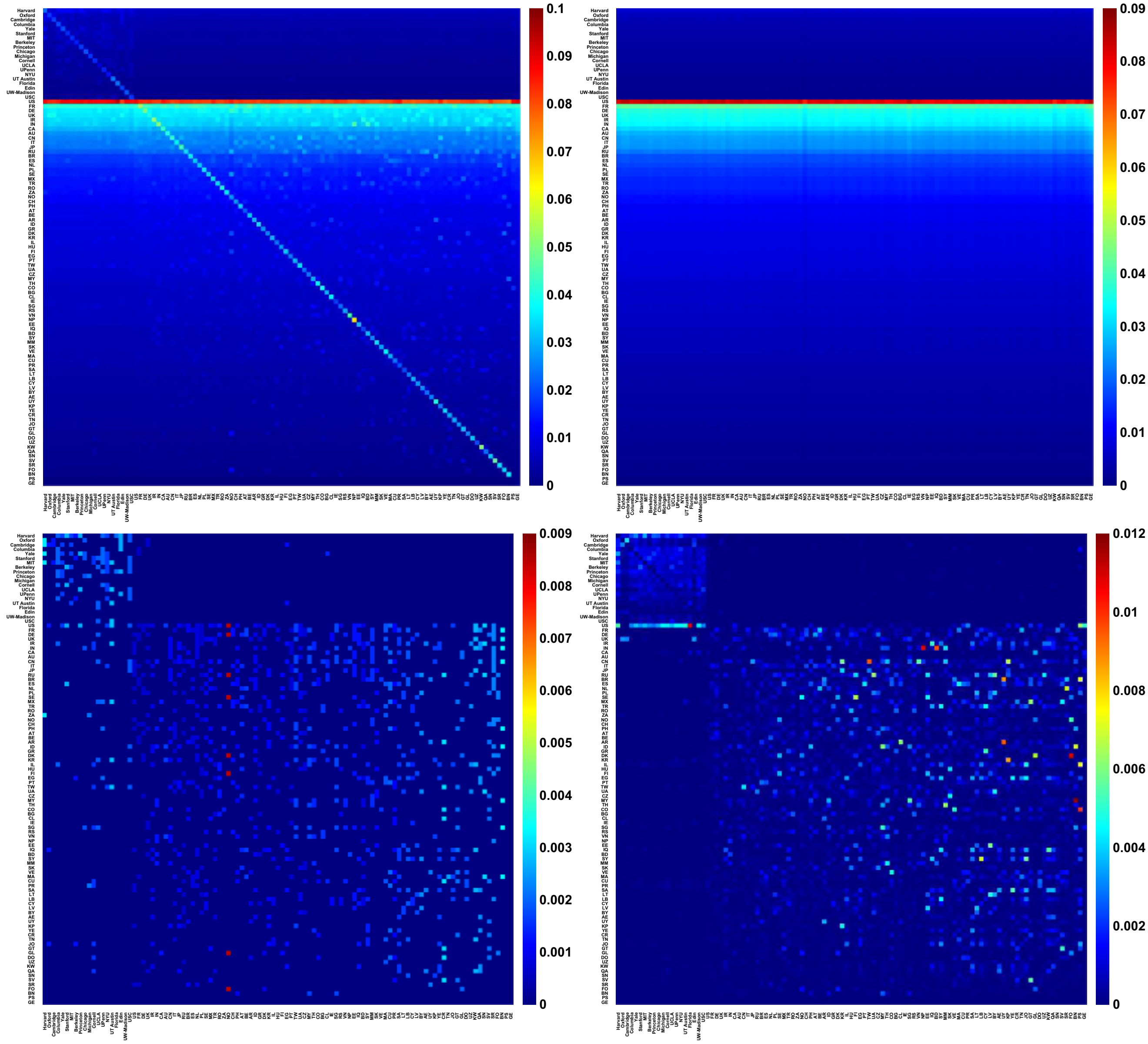

Fig. 5. Reduced Google matrix $G_{\rm R}$ for the first 20 universities ranked in ENWRWU (Tab. 4) and the 85 countries to which universities from WRWU belong (Tab. 5). The full reduced Google matrix $G_{\rm R}$ is presented in top left panel, $G_{\rm pr}$ in top right panel, $G_{\rm rr}$ in bottom left panel and $G_{\rm qrnd}$ in bottom right panel. The weights of $G_{\rm R}$ matrix components are $W_{\rm R} = 1$, $W_{\rm pr} = 0.948273$, $W_{\rm rr} = 0.0144137$ and $W_{\rm qr} = 0.0373132$. The node order presents first 20 universities in order of Table 4 and then 85 countries in order of Table 5.

Stanford University for US west coast, University of Chicago for US central region, Harvard University for US east coast and University of Oxford for UK; each group is marked by color in Table 4.

The components of reduced Google matrix describing direct links $G_{\rm rr}$ and indirect links $G_{\rm qrnd}$ are shown in Figure 9 (only nondiagonal links are shown for $G_{\rm qr}$). The weight of indirect nondiagonal links is about 50% percent stronger than the weight of direct links. This shows the importance of indirect interactions between top 20 universities.

To characterize the importance of direct and indirect links, we consider the sum of two matrix components given by $G_{\rm rr} + G_{\rm qrnd}$. From this matrix we construct the network of friends of regional leaders shown in Figure 10 (the drawings of networks have been produced using Cytoscape [36]). First we place the regional leader on a circle (1st level of possible friendship). From a regional leader we look at the four biggest outgoing links in $G_{\rm rr} + G_{\rm qrnd}$, these four links define the four best friends of the regional leader. If these friends are not present in the network of friends, i.e., they are not themselves regional leaders, then they are placed on the circle around the regional leader (2nd level of friendship). If several regional leaders share the same friend, by preference, the friend is placed in the circle around the leader of its region. Then, from each

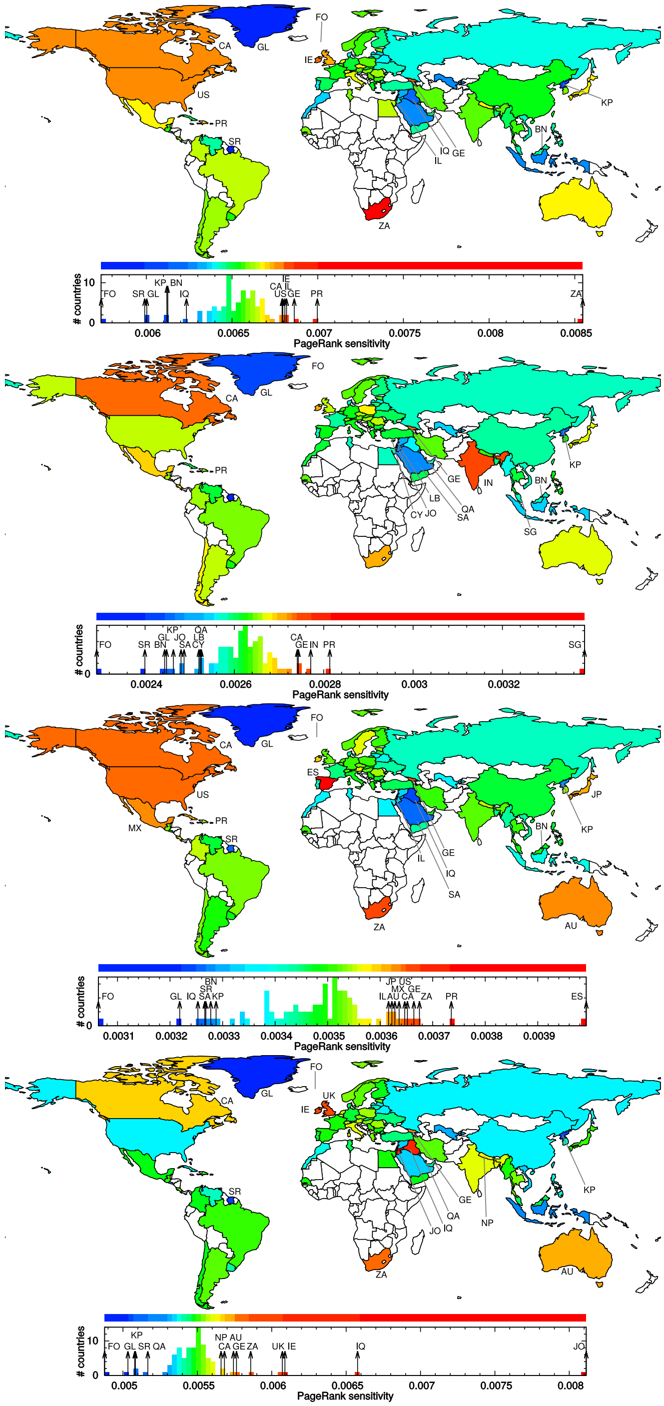

Fig. 6. World map of diagonal sensitivity of world countries $D(u \to c, c)$ to the change of the reduced Google matrix link university $u \to$ country c. The cases of 4 universities are shown from top to bottom: Harvard University, University of Chicago, Stanford University and University of Oxford.

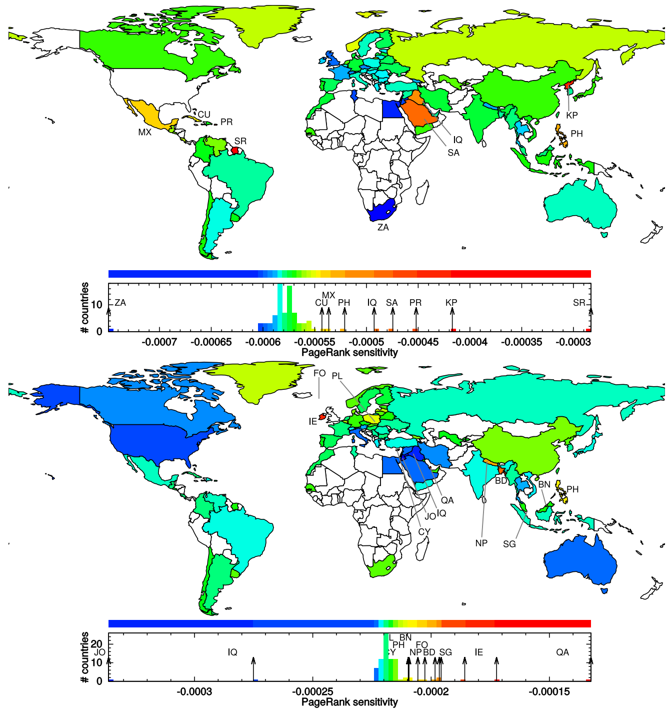

Fig. 7. World map of nondiagonal sensitivity $D(u \to c, c')$ of country c' to the change of the reduced Google matrix link Harvard $\to$US (top panel) and Oxford $\to$UK (bottom panel). Red (blue) color corresponds to the greatest (lowest) absolute value. Diagonal values $D(u \to c, c' = c)$ are not shown.

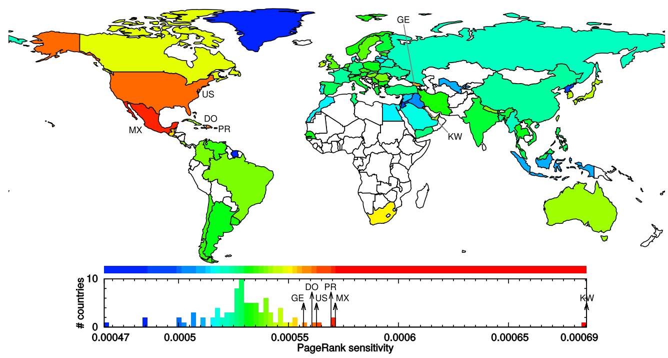

Fig. 8. Same as in Figure 6 with diagonal sensitivity D("Rice University" $\to c, c)$ to the change of the reduced Google matrix link Rice University $\to$ country c.

friends of regional leaders, we define in the same way four new friends. Each new friend is either already placed in the friendship network or not. If the new friend is not present, it is placed in the circle around the corresponding friend of regional leader (3rd level of friendship). If a new friend is shared by several friends of regional leaders, the new friend is placed by preference on the circle around the friend of regional leader belonging to its region. In the same manner, we then define the 4th level of friendship and so on. The procedure continues until no new friends can be placed on the friendship network (because already placed on it). For the 2017 ENWRWU top 20 the procedure stops after four levels of friendship. A red arrow represents a pure hidden link, i.e., a link from university u to university u' with a null adjacency matrix entry, $A_{u'u} = 0$, or otherwise stated, with a minimal value in $G_{\rm rr}$, $(G_{\rm rr})_{u'u} = (1-\alpha)/N$.

This network presentation of friends in Figure 10 shows that the close friends are mainly located in same region;

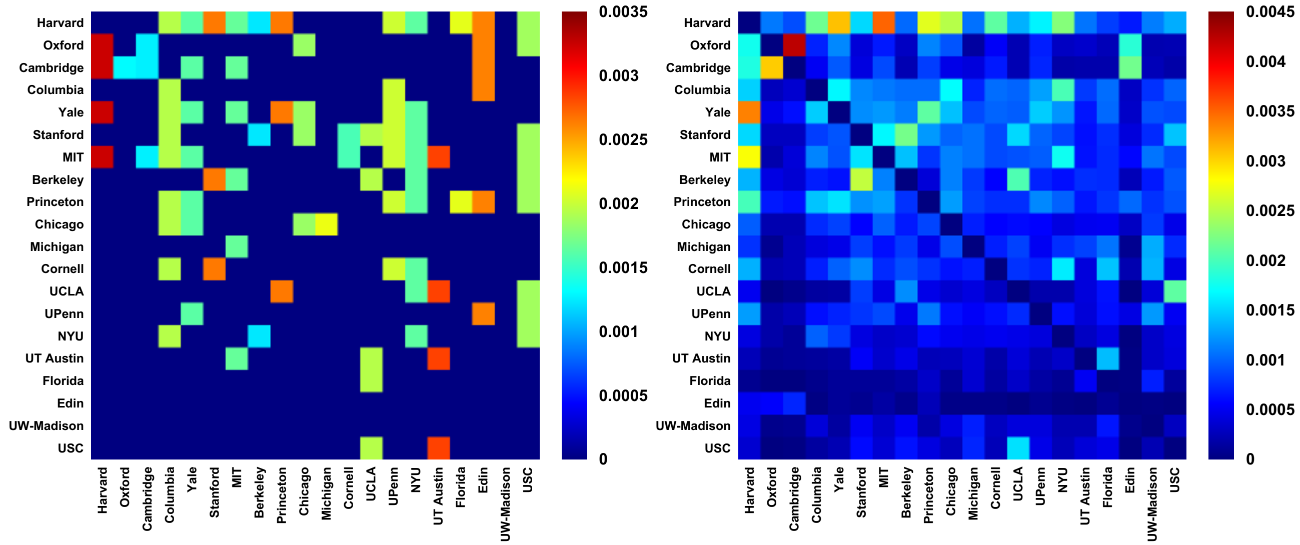

Fig. 9. Matrices $G_{\rm rr}$ (left panel) and $G_{\rm qrnd}$ (right panel) for top 20 ENWRWU (Tab. 4). The universities are ordered by their PageRank index as in Table 4. The matrix weights are $W_{\rm rr} = 0.00877$ and $W_{\rm qrnd} = 0.01381$. Color shows the strength of matrix elements. The same components for top 20 universities of FRWIKI, DEWIKI and RUWIKI are available in Supporting Information at [30,31].

Table 6. List of top 20 PageRank universities of French edition WRWU2017. The color code corresponds to the country location of universities: blue for US, violet for UK, red for FR, green for CA and yellow for BE.

Rank	University	Rank	University
1st	**Harvard University**	11th	**University Laval**
2nd	**University of Oxford**	12th	**Pantheon-Sorbonne University**
3rd	**École polytechnique**	13th	**Princeton University**
4th	**University of Cambridge**	14th	**University of California, Berkeley**
5th	**École normale supérieure**	15th	**Paris-Sorbonne University**
6th	**Massachusetts Institute of Technology**	16th	**Université libre de Bruxelles**
7th	**Yale University**	17th	**University of Montreal**
8th	**Columbia University**	18th	**Université catholique de Louvain**
9th	**Stanford University**	19th	**Paris Nanterre University**
10th	**École pratique des hautes études**	20th	**University of Chicago**

thus MIT, Harvard, Yale form one group of east coast, Oxford, Cambridge and Edinburgh are friends inside UK. However, there are also inter-regional friends formed by Princeton, UCLA, USC, UT Austin and proximity between Stanford, Berkeley and Cornell. The direct links shown in black play an important role but indirect links shown in red are also present and significant like e.g. MIT being indirect friend of Stanford, Harvard being indirect friend of Chicago and Oxford, and Edinburgh, Princeton being indirect friend of Oxford.

We also consider countries which are friends of each regional university leader as shown in Figure 11. For this we consider the matrix elements of $G_{\rm rr} + G_{\rm qrnd}$ constructed for top 20 universities of Table 4 and 85 countries listed in Table 5. For each regional leader university we select top 4 friendliest countries. The network of country-university friends is presented in Figure 11 with countries marked by colors corresponding to mostly spoke language. Thus, we see that countries friends of Oxford are mostly Arab and English speaking countries; countries friends of Harvard are dominantly from English speaking countries excepting one Hebrew speaking country; friends of Stanford are two Spanish speaking countries, one English and one Hebrew; Chicago is mostly diversified having English, Hindi, Chinese and Spanish speaking friends.

6.2 Top 20 universities in French Wikipedia edition

Here, we consider the cultural view of FRWIKI on top 20 universities and their interactions. We select from FRWIKI2017 top PageRank universities listed in Table 6. These universities belong to 5 countries (US, UK, FR, CA and BE) marked by corresponding color. Top PageRank university of each country is considered as a country leader. Similar to the case of Figure 10 we obtain from the reduced Google matrix components $G_{\rm rr} + G_{\rm qrnd}$ of these 20 universities the network of friends shown in Figure 12.

The obtained network of friends of Figure 12 shows clear cluster of universities inside their own countries. However, the inter-country links are well present and they are mainly indirect links (shown in red) pointing toward

Table 7. List of top 20 PageRank universities of German edition WRWU2017. The color code corresponds to the country location of universities: green for DE, blue for US, violet for UK and black for AT.

Rank	University	Rank	University
1st	**Ludwig Maximilian University of Munich**	11th	**University of Freiburg**
2nd	**Humboldt University of Berlin**	12th	**University of Cologne**
3rd	**University of Göttingen**	13th	**University of Münster**
4th	**Heidelberg University**	14th	**University of Oxford**
5th	**Free University of Berlin**	15th	**University of Hamburg**
6th	**University of Vienna**	16th	**Goethe University Frankfurt**
7th	**University of Tübingen**	17th	**University of Cambridge**
8th	**Harvard University**	18th	**University of Marburg**
9th	**University of Bonn**	19th	**University of Kiel**
10th	**Leipzig University**	20th	**University of Jena**

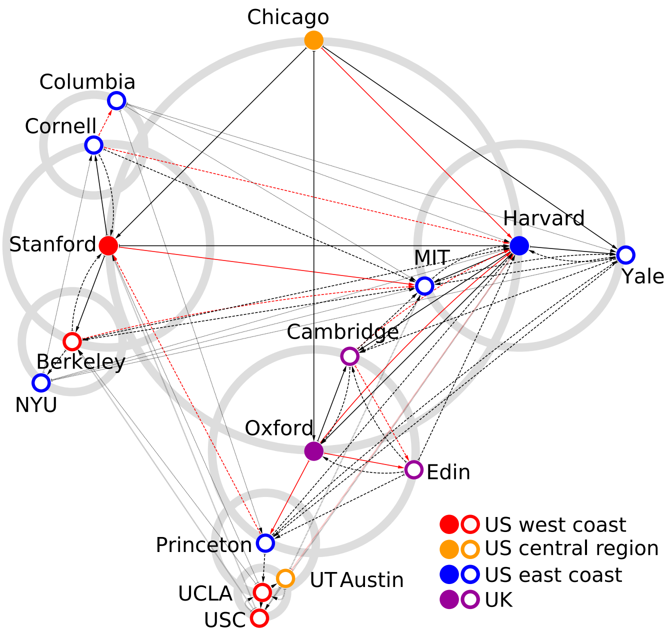

Fig. 10. Network of friends of top 20 PageRank universities of ENWRWU obtained from matrix of direct and indirect links $G_{\rm rr} + G_{\rm qrnd}$. Color filled nodes are regional leaders. Red links are purely hidden links, i.e., no corresponding adjacency matrix entry. We obtain 4 friendship levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines and from 4th level by "\" symbol lines.

English speaking universities. Thus, Princeton is indirect friend of Oxford, ULaval and ULB; Yale and Harvard are indirect friends of ULaval. Since we consider FRWIKI there are 6 French universities being the next in number after US with 7 universities among top 20. However, US universities are strongly linked with universities of Canada, Belgium and UK while French universities are weakly linked to other countries. This FRWIKI-analysis demonstrates certain world isolation of French universities. Also from FRWIKI point of view, the leading English speaking universities in Figure 12 form an invariant subset from which a random surfer cannot escape: the friends of these universities are uniquely English speaking universities.

6.3 Top 20 universities in German Wikipedia edition

Top 20 PageRank universities are given in Table 7. The network of friends, constructed for these 20 nodes from the matrix components $G_{\rm rr} + G_{\rm qrnd}$ of the reduce Google matrix, is shown in Figure 13. A specific point of these top 20 universities is that there are only 3 non-German-speaking universities (Harvard, Oxford and Cambridge). This is due to the already discussed feature of DEWIKI which gives strong preference to German universities (see right panel of Fig. 1). The universities belong only to 4 countries AT, DE, UK and US. The main cluster are formed around LMU Munich and Vienna however GU Frankfurt, Marburg and Tubingen are closely linked with Oxford and Cambridge; Hamburg is linked with Harvard. It should be pointed that there is a dominance of indirect links which are also linking different countries.

6.4 Top 20 universities in Russian Wikipedia edition

Top 20 PageRank universities are given in Table 8. The network of friends, constructed for these 20 nodes from the matrix components $G_{\rm rr} + G_{\rm qrnd}$ of the reduce Google matrix, is shown in Figure 14. Among these 20 universities there 8 from US, 5 from Russia, 2 from Ukraine, 2 from Germany (its former DDR part), 2 from UK and 1 from Austria so that there are 6 different countries. The clusters of universities are mainly linked with their own countries even if there is very close proximity between UK and US even if Berkeley and Chicago are in the circle proximity of Vienna. The main intercountry links are mainly indirect (except direct links between Kiev pointing to Moscow and St. Petersburg which were all inside former USSR). It is interesting to note that German university, belonging to the former DDR part of Germany, has strong links with Russian universities, showing that the links inside Soviet block are still significant even if Wikipedia had been created well after disappearance of DDR.

6.5 Comparison between English, French, German and Russian Wikipedia editions

In this subsection, we perform a comparison of different cultural views of DEWIKI, ENWIKI, FRWIKI and

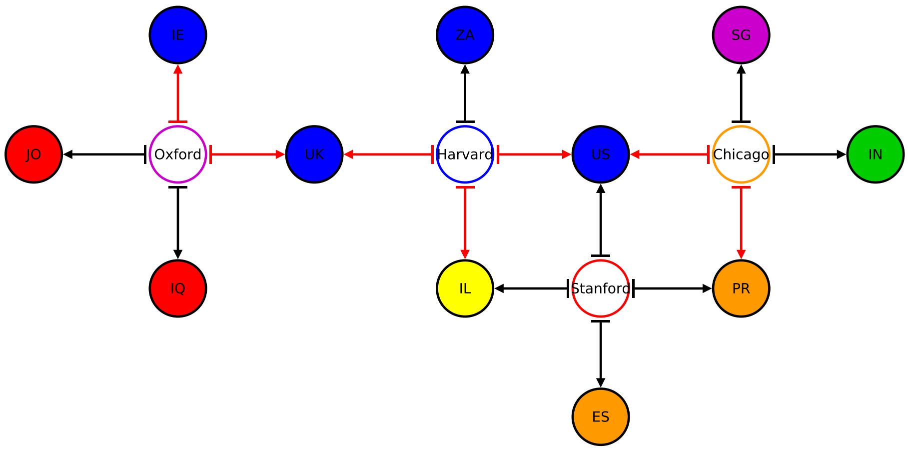

Fig. 11. Network of friends from $G_{\rm rr} + G_{\rm qrnd}$ associated to the top 20 ENWRWU and 85 countries listed in Table 5. For each regional leaders, Stanford University, University of Chicago, Harvard University and University of Oxford, the four strongest links to one of the 85 countries listed in Table 5 are presented. Universities (countries) are represented by empty (full) nodes. The color code for countries depends on the main spoken language: **blue** for English, **red** for Arabic, **orange** for Spanish, **violet** for Chinese, **green** for Hindi and **yellow** for Hebrew. Red links are purely indirect links and black ones are from direct links.

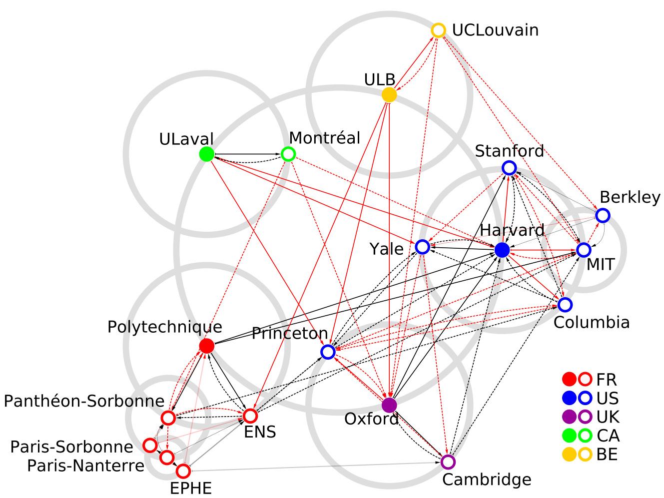

Fig. 12. Same as in Figure 10 for PageRank top 20 universities of FRWIKI2017 from Table 6. Color filled nodes are country leaders. Red links are purely hidden links, i.e., no corresponding adjacency matrix entry. We obtain 4 friendship levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines, and from 4th level by "\" symbol lines.

RUWIKI on top universities. With this aim, we take 20 PageRanked universities of each of these editions. This gives us 52 different universities presented in these editions. The list of them is given in Table 9. Each of these universities is attributed to its own foundation country shown by colors in this table. There are 9 different countries: US, FR, DE, UK, CA, RU, AT, BE and UA.

Then we perform the reduced Google matrix analysis for these 52 universities for each edition constructing $G_{\rm R}$ and its 3 components. The matrices $G_{\rm R}$ for each edition are shown in Figure 15. The $G_{\rm pr}$, $G_{\rm rr}$, $G_{\rm qrnd}$ matrix components are available at [30,31] with their weights which

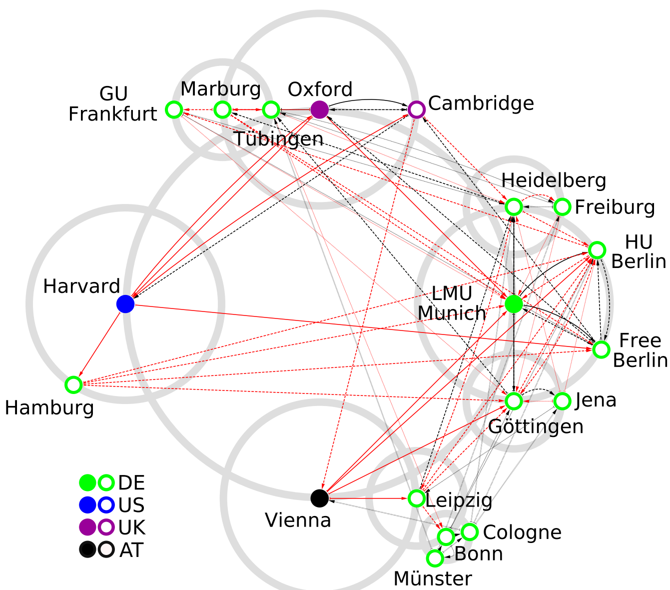

Fig. 13. Same as in Figure 10 for PageRank top 20 universities of DEWIKI2017 from Table 7. Color filled nodes are country leaders. Red links are purely hidden links, i.e., no corresponding adjacency matrix entry. We obtain 4 friendship levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines and from 4th level by "\" symbol lines.

are similar to those given in Figure 5; the weights of direct and indirect links are comparable. From Figure 15 we see that each edition have its own view on these 52 universities. Indeed, there is a clear tendency that edition rank higher universities belonging to the countries with edition language, e.g. RUWIKI places Moscow and St. Petersburg universities on top PageRank positions with a similar situation for DEWIKI. Using the matrix components of

Table 8. List of top 20 PageRank universities of German edition WRWU2017. The color code corresponds to the country location of universities: red for RU, blue for US, violet for UK, green for DE, black for AT and yellow for UA.

Rank	University	Rank	University
1st	**Moscow State University**	11th	**Kazan Federal University**
2nd	**Saint Petersburg State University**	12th	**National University of Kharkiv**
3rd	**Harvard University**	13th	**Stanford University**
4th	**University of Oxford**	14th	**Princeton University**
5th	**University of Cambridge**	15th	**University of Chicago**
6th	**Massachusetts Institute of Technology**	16th	**Higher School of Economics**
7th	**Yale University**	17th	**Bauman Moscow State Technical University**
8th	**Columbia University**	18th	**Leipzig University**
9th	**Kyiv University**	19th	**University of Vienna**
10th	**Humboldt University of Berlin**	20th	**University of California, Berkeley**

Table 9. List of the 52 different universities appearing in the top 20s of the EN, FR, DE, RU WRWU. Color code corresponds to country: **US**, **FR**, **DE**, **UK**, **CA**, **RU**, **AT**, **BE** and **UA**. Universities are ordered by countries (country groups are ordered according to PageRanking of 2017 English Wikipedia). Within country groups universities are ordered according to PageRanking of 2017 English Wikipedia.

Rank	University	Rank	University
1	**Harvard University**	28	**Heidelberg University**
2	**Columbia University**	29	**Free University of Berlin**
3	**Yale University**	30	**University of Tübingen**
4	**Stanford University**	31	**University of Bonn**
5	**Massachusetts Institute of Technology**	32	**University of Freiburg**
6	**University of California, Berkeley**	33	**University of Cologne**
7	**Princeton University**	34	**University of Münster**
8	**University of Chicago**	35	**University of Hamburg**
9	**University of Michigan**	36	**Goethe University Frankfurt**
10	**Cornell University**	37	**University of Marburg**
11	**University of California, Los Angeles**	38	**University of Kiel**
12	**University of Pennsylvania**	39	**University of Jena**
13	**New York University**	40	**University of Oxford**
14	**University of Texas at Austin**	41	**University of Cambridge**
15	**University of Florida**	42	**University of Edinburgh**
16	**University of Wisconsin Madison**	43	**University Laval**
17	**University of Southern California**	44	**University of Montreal**
18	**École polytechnique**	45	**Moscow State University**
19	**École normale supérieure**	46	**Saint Petersburg State University**
20	**École pratique des hautes études**	47	**Kazan Federal University**
21	**Panthéon-Sorbonne University**	48	**Bauman Moscow State Technical University**
22	**Paris-Sorbonne University**	49	**University of Vienna**
23	**Paris Nanterre University**	50	**Université Libre de Bruxelles**
24	**Humboldt University of Berlin**	51	**Kyiv University**
25	**Leipzig University**	52	**National University of Kharkiv**
26	**Ludwig Maximilian University of Munich**		
27	**University of Göttingen**		

$G_{\rm rr} + G_{\rm qrnd}$ we analyze the network of friend of 52 universities from the view point of ENWIKI, FRWIKI, DEWIKI and RUWIKI. The approach is the same as those used in network of friends discussed in the previous subsection.

The network of top friends for 52 universities in ENWIKI is shown in Figure 16. We see that the majority of links are indirect (red) comparing to the direct links (black). As expected two clusters of English speaking universities are well visible; in fact as in Figure 12, these English speaking universities form an invariant subspace from which a random surfer cannot escape. These universities act as an attractor subset in this friendship network. Chicago is located aside as it was already visible in previous subsection (see Fig. 10). A compact cluster of German universities is also well visible. We point that there are only 2 isolated French universities among top friends appearing in Figure 16; no university from other countries points toward these 2 universities, and these 2 universities point exclusively toward UK/US universities. In total this network of top friends has 13 US universities, 6 of DE, 3 of UK, 3 of RU, 2 of UA, 2 of CA, 2 of FR. Since the network is obtained from ENWIKI it is

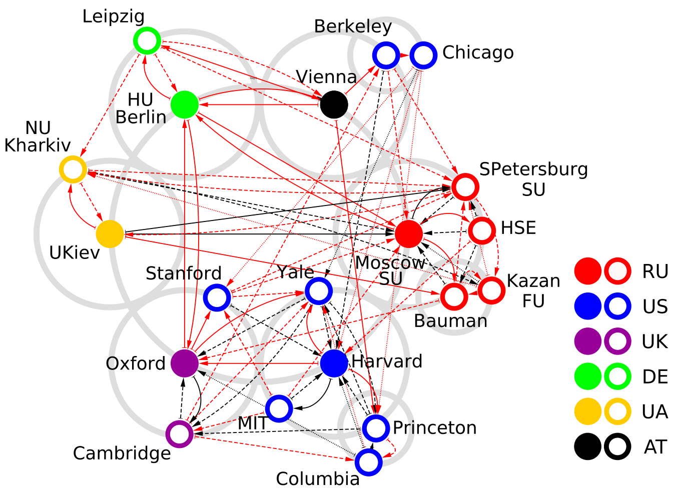

Fig. 14. Same as in Figure 10 for PageRank top 20 universities of RUWIKI2017 from Table 8. Color filled nodes are country leaders. Reduced network from top 20 RUWRWU $G_{\rm rr}+G_{\rm qrnd}$. Color filled nodes are regional leaders. Red links are purely hidden links, i.e., no corresponding adjacency matrix entry. We obtain 3 acquaintance levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines and from 3rd level by doted lines.

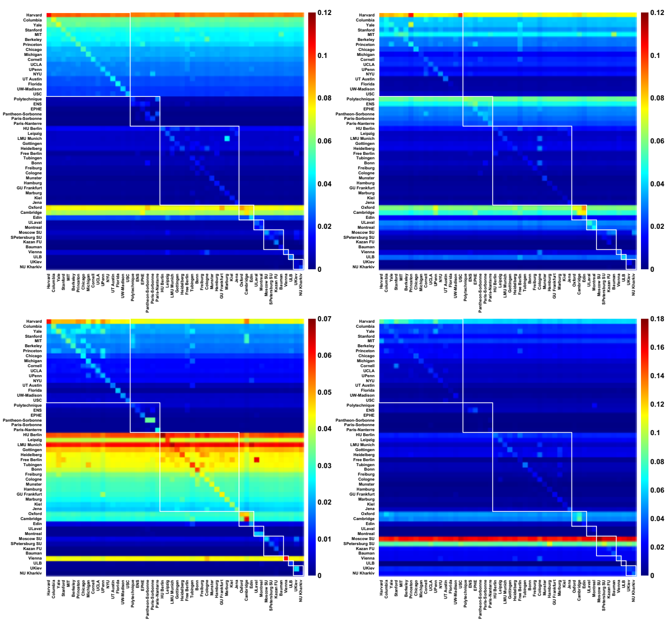

Fig. 15. Reduced Google matrix $G_{\rm R}$ for universities listed in Table 9 computed from EN (top left), FR (top right), DE (bottom left) and RU (bottom right) Wikipedia editions; for each edition nodes have the same order as in Table 9. The images of components $G_{\rm pr}$, $G_{\rm rr}$, $G_{\rm qrnd}$ are given in [30,31].

understandable that US universities (with UK ones) form the majority. However, German universities show their strength and significant influence. Comparing to them French university group is small and not significant being placed behind Russian universities. This network clearly shows the weak representation and influence of French universities that reflects a certain reality.

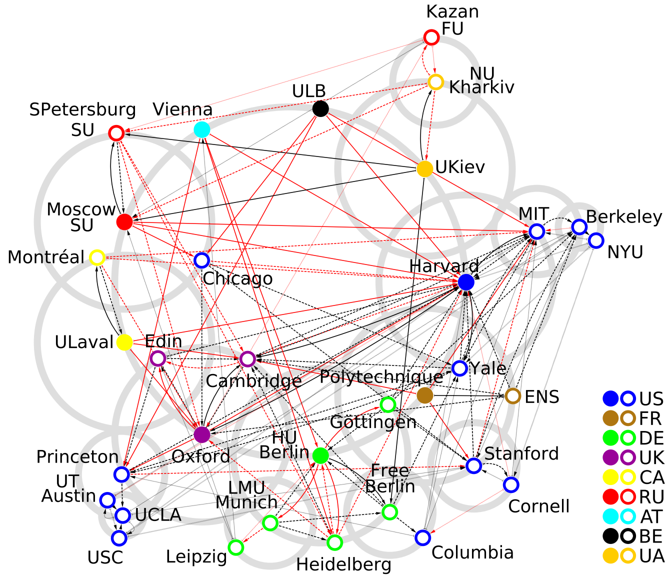

Fig. 16. Network of friends of 52 universities listed in Table 9 computed from $G_{\rm rr}+G_{\rm qrnd}$ of ENWIKI. Color filled nodes are country leaders (same colors as in the Tab. 9). Red links are purely hidden links, i.e., no corresponding adjacency matrix entry. We obtain 4 acquaintance levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines and from 4th level by "\" symbol lines.

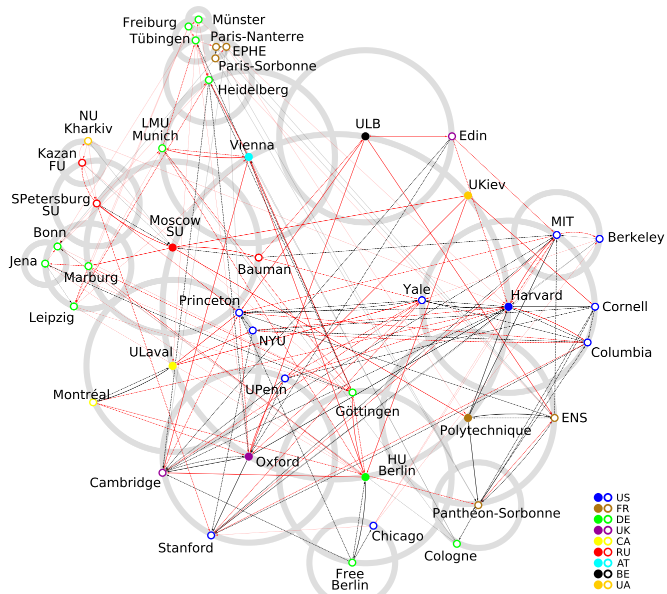

Fig. 17. Same as in Figure 16 but from FRWIKI.

The network of top friends from FRWIKI is shown in Figure 17. Here, we have the dominance of 13 German universities, followed by 11 of US, 6 of France and 4 of Russia. Still the indirect links play a dominant or comparable role with the direct links. In this network, the cluster structure is less visible, however as in Figures 12 and 16 the cluster of US–UK universities is central and acts as an attractor subset.

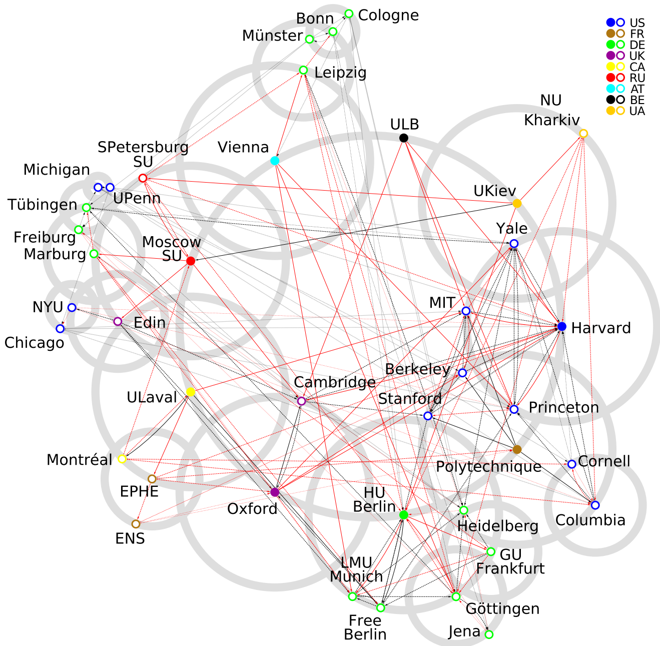

Fig. 18. Same as in Figure 16 but from DEWIKI.

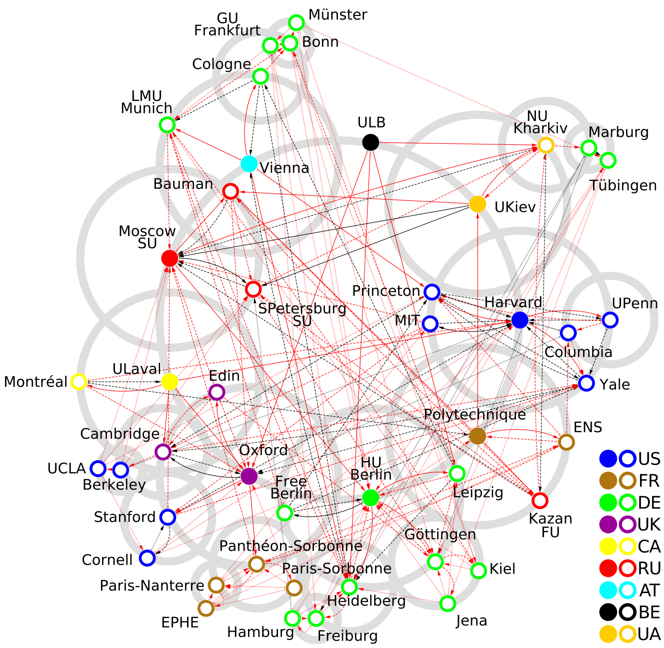

Fig. 19. Same as in Figure 16 but from RUWIKI.

Figure 18 shows the network of top friends from DEWIKI. Here, naturally, the dominance of 14 German universities remains, to be compared with 11 of US, 3 of France, 3 of UK and 2 of Russia. In global DE universities are distributed over 3 clusters, and US over 2 clusters.

The network of top friends from RUWIKI is shown in Figure 19. The dominance of German universities is well present here with 16 of them followed by 10 of US, 4 of Russia and 3 of UK. Three major hubs are clearly visible a German one centered around HU Berlin/Heidelberg/Göttingen, an US–UK one centered around Harvard and Oxford, and a Russian one centered around Moscow SU.

The analysis of this subsection allows to establish most close links between top world universities. It also shows the dominance of US and German universities.

7 Reduced Google matrix averaged over 24 Wikipedia editions

Above we have considered ranking and interactions from a view point of a given edition. Using the reduced Google matrix approach it is possible to perform an averaging over all 24 editions thus determining the averaged cultural view on selected universities. With this aim, for each of the 24 Wikipedia editions listed in Table 1, we select the subset of articles devoted to the 100 first PageRank universities of WRWU (this list is given in [30,31] and also in Tab. 10). Then we compute the corresponding reduced Google matrix $\bar{G}_{\rm R}$ averaged over 24 Wikipedia editions. The averaging is defined by the relation

$$\bar{G}_{\rm R} = \frac{1}{24}\sum_E G_{\rm R}^{(E)} \tag{6}$$

where $G_{\rm R}^{(E)}$ is the reduced Google matrix (4) of the Wikipedia edition E. Each one of the 24 reduced Google matrices is written in the same basis corresponding to the ordered PageRank list of 100 universities. For a given edition E, reduced Google matrix entries corresponding to a link pointing toward an absent university in edition E are set to 0 and reduced Google matrix columns corresponding to absent universities in edition E are filled with 1/100 entries. These contributions from absent universities are added to the $G_{\rm pr}^{(E)}$ matrix components of the full reduced Google matrices $G_{\rm R}^{(E)}$.

We note that the averaging of 24 $G_{\rm R}^{(E)}$ matrices with equal weights gives us again the reduced Google matrix which performs an averaging over different cultural views of 24 editions.

The PageRank vector computed from the averaged reduced Google matrix $\bar{G}_{\rm R}$ is presented in Table 10. We see that the rank order is changed comparing to the Θ-averaging (5) with the top 10 PageRank universities given in Table 2 (list of top 100 is given in [30,31]). We see that Harvard takes the first position instead of the third one in Table 2 and then Oxford and Cambridge are moved to second and third positions in Table 10. The top ten universities of Table 10 have overlap of 100% with PageRank WRWU of Table 2 and 90% with ARWU of Table 3. It can be discussed what ranking averaging over 24 cultural view of editions is more appropriate: with Θ-averaging or with averaging of $G_{\rm R}^{(E)}$. We think that both approaches are useful: in Θ-averaging all PageRanking vectors are completely independent while averaging of $G_{\rm R}^{(E)}$ introduces some additional links which are not present in certain network editions.

The obtained averaged Reduced Google matrix $\bar{G}_{\rm R}$ and its three components are shown in Figure 20. In global the

Table 10. Top100 2017 WRWU ordered according to averaged reduced Google matrix.

Rank	PageRank value	University	Rank	PageRank value	University
1	0.0633191	Harvard University	51	0.00659655	Univ. of Colorado Boulder
2	0.0528587	Univ. of Oxford	52	0.00657266	Univ. of Glasgow
3	0.0518905	Univ. of Cambridge	53	0.00636839	Univ. of Toronto
4	0.0339304	MIT[a]	54	0.0063255	Stockholm University
5	0.0301911	Columbia University	55	0.00624184	Univ. of Tübingen
6	0.0283041	Yale University	56	0.00609986	Univ. of Texas at Austin
7	0.0261455	Stanford University	57	0.00593539	Univ. of Virginia
8	0.024318	UC Berkeley[b]	58	0.00584412	Imperial College London
9	0.0229394	Princeton University	59	0.00582829	Carnegie Mellon University
10	0.0215136	Univ. of Chicago	60	0.00579437	Univ. of Bonn
11	0.0197203	Univ. of Copenhagen	61	0.00570673	Univ. of Minnesota
12	0.0168679	HU Berlin[c]	62	0.00567465	Keio University
13	0.0160439	Uppsala university	63	0.00557384	Univ. of Helsinki
14	0.0148231	Univ. of Tokyo	64	0.00548871	King's College London
15	0.0135633	Moscow State University	65	0.0054485	Univ. of Florida
16	0.0127305	Cornell University	66	0.00538279	Univ. of Zurich
17	0.0126064	HUJI[d]	67	0.00536546	Univ. of Manchester
18	0.0125732	Univ. of Pennsylvania	68	0.00523928	McGill University
19	0.0120329	UCLA[e]	69	0.00507791	Free University of Berlin
20	0.011732	Leiden University	70	0.00505635	Univ. of Washington
21	0.011246	Caltech[f]	71	0.00505447	Univ. of Illinois U.-C.
22	0.0112404	New York University	72	0.00497258	Brown University
23	0.0112273	Univ. of Vienna	73	0.00491403	Univ. of Wisconsin-Madison
24	0.0104997	Univ. of Edinburgh	74	0.00485964	Northwestern University
25	0.0103698	Jagiellonian University	75	0.00480294	Univ. of Coimbra
26	0.0101557	Univ. of Bologna	76	0.00479832	Univ. of Oslo
27	0.0100089	Univ. of Göttingen	77	0.00477973	Univ. of Padua
28	0.00987766	Heidelberg University	78	0.00476805	Georgetown University
29	0.00982921	Univ. of Michigan	79	0.00475634	UNAM[l]
30	0.00974263	Lund University	80	0.00468635	Boston University
31	0.00929623	LSE[g]	81	0.0045985	Ohio State University
32	0.00918967	Johns Hopkins University	82	0.00458516	Michigan State University
33	0.00909002	Univ. of Warsaw	83	0.00452351	Univ. of Geneva
34	0.00902656	Seoul National University	84	0.00451385	Univ. of Marburg
35	0.00877768	Leipzig University	85	0.00433353	Univ. of Salamanca
36	0.00832413	Univ. of Munich[h]	86	0.0042273	Univ. of Freiburg
37	0.00791791	Waseda University	87	0.00418341	Univ. of Arizona
38	0.0076835	Univ. College London	88	0.00417181	Univ. of Jena
39	0.00751886	Duke University	89	0.00415139	MLU[m]
40	0.00718132	Sapienza[i]	90	0.00401368	Univ. of St Andrews
41	0.00711981	ETH Zurich	91	0.00398415	TU Berlin[n]
42	0.0071081	USC[j]	92	0.00391916	UNC Chapel Hill[o]
43	0.00693105	École Polytechnique	93	0.00390789	Univ. of Tartu
44	0.00692597	Peking University	94	0.00388656	TU Munich[p]
45	0.00682986	Al-Azhar University	95	0.00385376	Univ. of Sydney
46	0.00682254	École Normale Supérieure	96	0.00384341	UC San Diego[q]
47	0.00680075	Kyoto University	97	0.00371085	Trinity College, Dublin
48	0.00666809	Charles University	98	0.00368454	Indiana University
49	0.00666454	SPbU[k]	99	0.00355122	University of Notre Dame
50	0.00662585	Utrecht University	100	0.00353878	University of Kiel

[a]Massachusetts Institute of Technology, [b]University of California, Berkeley, [c]Humboldt University of Berlin, [d]Hebrew University of Jerusalem, [e]University of California, Los Angeles, [f]California Institute of Technology, [g]London School of Economics, [h]Ludwig Maximilian University of Munich, [i]Sapienza University of Rome, [j]University of Southern California, [k]Saint Petersburg State University, [l]National Autonomous University of Mexico, [m]Martin Luther University of Halle-Wittenberg, [n]Technical University of Berlin, [o]University of North Carolina at Chapel Hill, [p]Technical University of Munich, [q]University of California, San Diego.

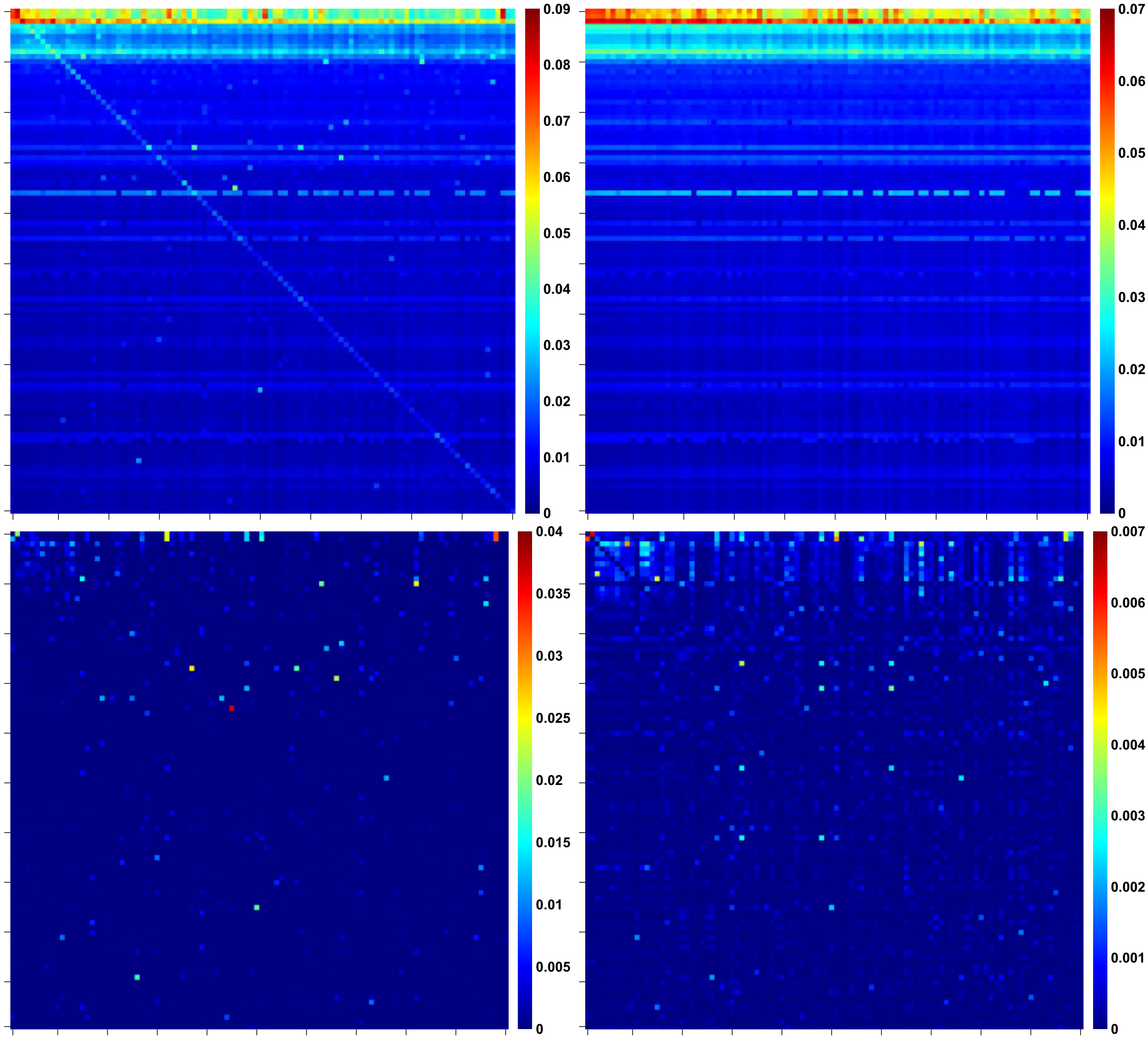

Fig. 20. Reduced Google matrix $\bar{G}_{\mathrm{R}}$ for PageRank top 100 universities in WRWU (top 10 is in Tab. 2, top 100 is in [30,31]) averaged over 24 Wikipedia editions. Matrix entries correspond to universities ordered according the top 100 WRWU PageRank order. The full reduced Google matrix $\bar{G}_{\mathrm{R}}$ is presented in top left panel, $\bar{G}_{\mathrm{pr}}$ in top right panel, $\bar{G}_{\mathrm{rr}}$ in bottom left panel and $\bar{G}_{\mathrm{qrnd}}$ in bottom right panel. The matrix weights are $W_{\mathrm{R}} = 1$, $W_{\mathrm{pr}} = 0.957$, $W_{\mathrm{rr}} = 0.019$, $W_{\mathrm{qr}} = 0.024$, $W_{\mathrm{qrnd}} = 0.015$.

matrix structure is similar to those of individual editions discussed above. The component $\bar{G}_{\mathrm{pr}}$ has the dominant weight but it is rather close to the columns of PageRank vector and hence the interesting links are determined by the components $\bar{G}_{\mathrm{rr}}$ and $\bar{G}_{\mathrm{qr}}$ which have comparable weights. It is well seen that there are indirect links which are not present between direct ones.

From the matrix of direct and indirect links $\bar{G}_{\mathrm{rr}} + \bar{G}_{\mathrm{qrnd}}$ we construct the interaction friendship network between above considered 100 universities divided by certain groups. Such a network takes into account cultural views of all 24 editions. We now show all links by the same black color since after averaging over 24 editions there is a significant mixture of direct and indirect links.

In our first division, we mark universities by foundation time (century) periods: red for foundation years from 1000 to 1300 AD, blue from 1300 to 1600 AD, green from 1600 to 1800 AD and black from 1800 to 2000 AD. Each time period has its leader taken as a university with highest rank position in this period. The resulting network of friends in shown in Figure 21. This network shows an interesting evolution of interactions between universities through 10 centuries: the cluster of universities founded in 11th–13h centuries, marked in red, is formed mainly by UK and Italian universities (one from Spain, group leader is Oxford). This cluster transfers its influence via interaction and links to next 14th–16th centuries universities, marked in blue, which are mainly from northern

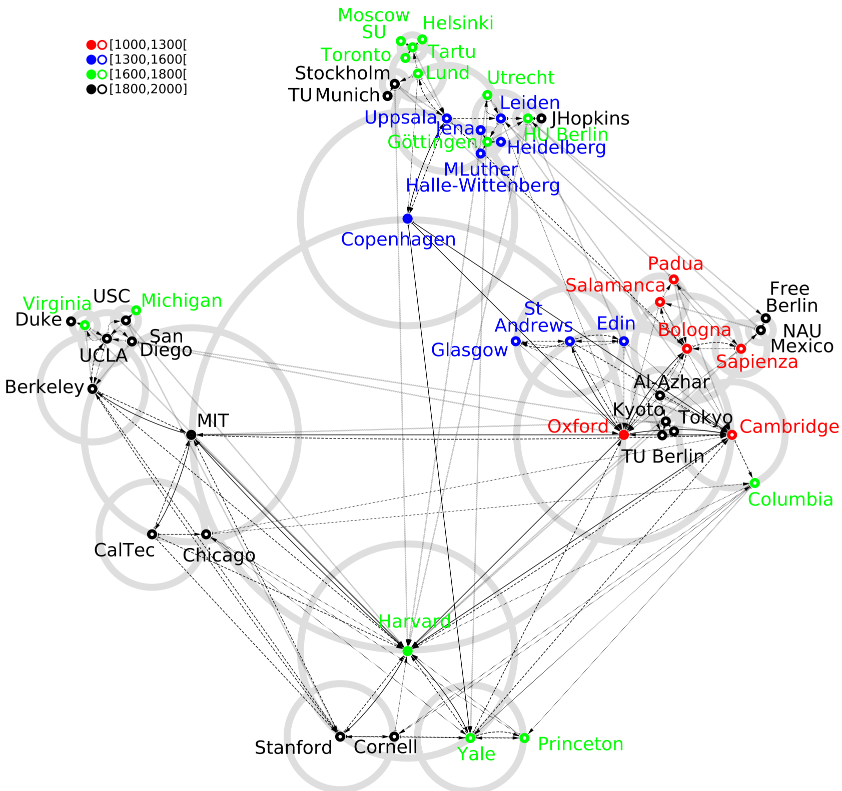

Fig. 21. Century reduced friendship network constructed for universities of PageRank top 100 list of WRWU (see [30,31], top 10 are in Tab. 2), computed from $\bar{G}_{\rm rr} + \bar{G}_{\rm qrnd}$ averaged over 24 Wikipedia editions. Color marks university founded at the same time (century) period given in years; color filled circles are time period leaders, open circles of the same color are universities from the same time period. We show 5 friendship levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines and from 4th or 5th level by "\" symbol lines.

countries including Scotland, Denmark, Germany, Sweden and Netherlands (group leader is Copenhagen). The influence of these universities is transferred to 17th–18th centuries universities, marked in green, being mainly near the blue cluster and located in the same countries with addition of Moscow in Russia, Tartu in Estonia, Helsinki in Finland; another group of green universities of this time period is linked with Oxford and Cambridge and is located mainly on US east coast (Harvard, Yale, Princeton; Columbia is directly linked to Cambridge). The university of next centuries 19th–20th, marked in black (MIT is group leader), are mainly located in US but new universities of this time period appear also in Japan (Tokyo, Kyoto), Egypt (Al-Azhar), Germany (TU Munich, TU Berlin, Free Berlin) and Sweden (Stockholm). Thus, the obtained friendship network provides a compact description of world universities development through 10 centuries taking into account the balanced view of 24 cultures presented by Wikipedia editions.

In our second division, we mark universities by continent location: blue for America, red for Europe, green for Asia (Russia is attributed to Asia), yellow for Oceania and black for Africa. Again color group leaders are marked by full circles. The friendship network obtained from $\bar{G}_{\rm rr} + \bar{G}_{\rm qrnd}$ is shown in Figure 22. The network has two large clusters of European universities (red) and US universities (blue). The group of university in Asia (green) is mainly linked between themselves having secondary links with Europe and US. Oceania (Sydney) and Africa (Al-Azhar) are represented only by one university. This network structure clearly shows the influence of European and US universities with emerging group of new group of Asian universities with strong internal links.

8 Discussion

In this work, we performed analysis of ranking and interactions of world universities from directed networks of 24 Wikipedia editions dated by May 2017. Our results show that obtained WRWU2017 with PageRank algorithm averaged over 24 editions gives a reliable ranking of universities with 60% overlap with top 100 of ARWU2017 (Shanghai ranking) [3]. At the same time WRWU2017 highlights in a stronger way the significance of historical path of a given university over centuries. There are certain changes in WRWU2017 version comparing to WRWU2013 version demonstrating appearance of new universities with time evolving and with the increase of the number of Wikipedia articles in the 24 selected editions. A comparison of WRWU and ARWU ranking positions for specific universities (e.g. Rice University) shows that the Wikipedia visibility can be significantly improved in certain cases.

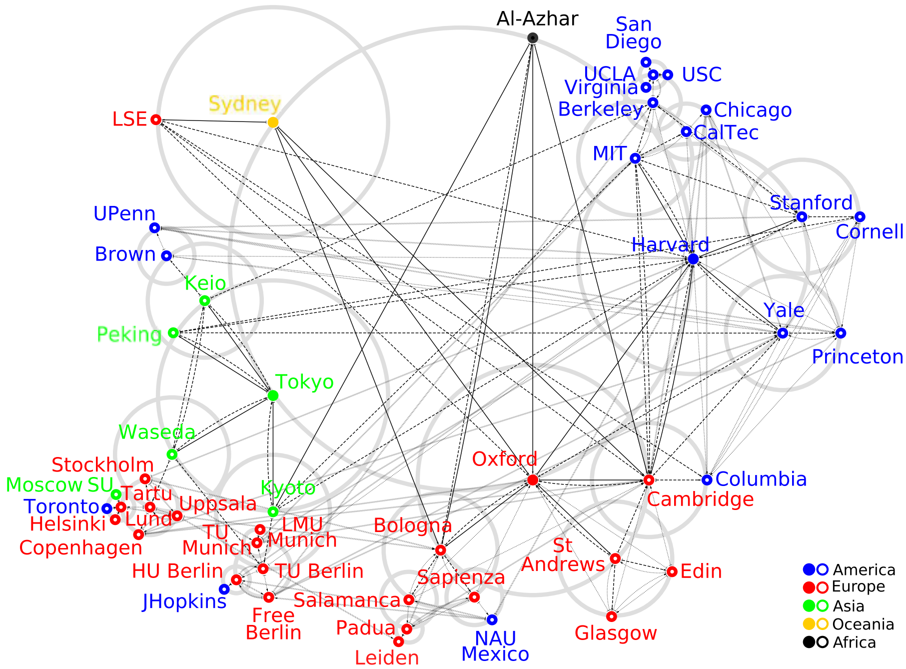

Fig. 22. Continent reduced friendship network constructed for universities of PageRank top 100 list of WRWU (see [30,31], top 10 are in Tab. 2), computed from $\bar{G}_{\rm rr} + \bar{G}_{\rm qrnd}$ averaged over 24 Wikipedia editions. Color marks university of the same continent; color filled circles are continent leaders, open circles of the same color are universities from the same continent. We obtain 5 friendship levels (gray circles). Links originating from 1st level universities are presented by solid lines, from 2nd level by dashed lines, from 3rd level by doted lines and from 4th or 5th level by "\" symbol lines.

We also performed an additional analysis based on the reduced Google matrix (REGOMAX) algorithm [26,27]. This approach allowed us to establish direct and indirect links between universities and world countries. As a result we obtain the sensitivity and influence of specific universities on world countries as it is seen by Wikipedia. The REGOMAX method allows to perform a democratic and uniform averaging over cultural views of 24 language editions and obtain a balanced cultural view on the interactions of top world universities through ten centuries of their historical development as well as their influence over continents.

In our studies, we used only network information about 24 Wikipedia editions. The obtained WRWU results are compared with Shanghai ARWU results confirming the that WRWU provides the reliable results. At the same time we find that it is important to take into account all 24 cultural views of different editions. Indeed, the results shown in Figure 1 show rather different views on importance of world universities in English, French and German editions. We think that the cultural views are really different and it is not so simple to take them correctly into account. Thus, ARWU uses only certain global criteria including the publications in Nature and Science journals which have more pronounced orientation to English speaking universities. In this sense, we estimate that averaging over different Wikipedia editions provides more balanced results from different cultures. This point is especially clear with ranking of historical figures [15] where English edition places on top positions USA presidents while the averaged ranking over 24 editions gives world known historical figures not related to USA politicians. Thus, WRWU provides a more balanced view also on world universities. It is not easy to find another open public database where the network analysis can be applied in efficient way. A possible source can be the global Wikipedia network where all editions (e.g. 24 of them) form one global network. However, the generation of such a global Wikipedia with translational links requires separate detailed investigations.

Finally, we stress that the WRWU method is independent of various personal opinions being based on purely mathematical and statistical analysis of the Wikipedia database. We think that this approach can be very complimentary to ARWU and other university rankings. We note that Wikipedia articles of universities usually appear in the top line (or lines) of Google search. As a result the world visibility of a given university can be publicly and freely broadcast all over the world increasing visibility of certain universities. We estimate that the improvement of Wikipedia articles of certain universities (e.g. we found a low visibility of French universities) can be an efficient way to increase their world visibility, attractivity and influence. Such an improvement is rather inexpensive and can be performed by a small group of researchers and students having knowledge in languages, history, computer and network sciences. We think that this approach can be complementary to various government projects which aim to increase visibility of national universities (like e.g. [4,5]).

This work was supported by the French "Investissements d'Avenir" program, project ISITE-BFC (contract ANR-15-IDEX-0003) and by the Bourgogne Franche-Comté Region 2017-2020 APEX project (conventions 2017Y-06426, 2017Y-06413, 2017Y-07534; see http://perso.utinam.cnrs.fr/~lages/apex/). The research of DLS is supported in part by the Programme Investissements d'Avenir ANR-11-IDEX-0002-02, reference ANR-10-LABX-0037-NEXT (project THETRACOM).

Author contribution statement

All authors equally contributed to all stages of this work.

References

1. UNESCO Liaison Office 24 July 2017, New York Office, Available: www.unesco.org/new/en/unesco-liaison-office-in-new-york/about-this-office/single-view/news/unesco_emphasizes_the_role_of_universities_and_higher_educat/ (Accessed July 2018)
2. E. Hazelkorn, *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence* (Palgrave Macmillan, New York, 2015)
3. Academic Ranking of World Universities. Available: http://www.shanghairanking.com/ (Accessed July 2018)
4. Enseignement supérieur et recherche, Investissements d'avenir. Available: http://www.enseignementsup-recherche.gouv.fr/pid24578/investissements-d-avenir.html. (Accessed July 2018)
5. Russian Academic Excellence Project. Available: http://5top100.ru/ (Accessed July 2018)
6. Times Higher Education World University Ranking. Available: https://www.timeshighereducation.com (Accessed July 2018)
7. U-Multirank of European Union. Available: http://www.umultirank.org/ (Accessed July 2018)
8. IREG Observatory on Academic Ranking and Excellence. Available: http://ireg-observatory.org/en/ (Accessed July 2018)
9. H. Jöns, M. Hoyler, Geoforum **46**, 45 (2013)
10. A. Rauhvargers, *Global University Rankings and Their Impact – Report II* (European University Association, Belgium, 2013)
11. D. Docampo, L. Cram, Scientometrics **98**, 1347 (2014)
12. A.O. Zhirov, O.V. Zhirov, D.L. Shepelyansky, Eur. Phys. J. B **77**, 523 (2010)
13. Y.-H. Eom, K.M. Frahm, A. Benczur, D.L. Shepelyansky, Eur. Phys. J. B **86**, 492 (2013)
14. J. Lages, A. Patt, D.L. Shepelyansky, Eur. Phys. J. B **89**, 69 (2016)
15. Y.-H. Eom, P. Aragon, D. Laniado, A. Kaltenbrunner, S. Vigna, D.L. Shepelyansky, PLoS One **10**, e0114825 (2015)
16. EC FET Open project NADINE Available: www.quantware.ups-tlse.fr/FETNADINE/ (Accessed July 2018)
17. S. Brin, L. Page, Comput. Netw. ISDN Syst. **30**, 107 (1998)
18. A.M. Langville, C.D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Princeton University Press, Princeton, 2006)
19. L. Ermann, K.M. Frahm, D.L. Shepelyansky, Rev. Mod. Phys. **87**, 1261 (2015)
20. L. Ermann, K.M. Frahm, D.L. Shepelyansky, Scholarpedia **11**, 30944 (2016)
21. R.A. Pagel (2016) Available: http://librarylearningspace.com/ruths-rankings-17-wikipedia-google-scholar-sources-university-rankings-influence-popularity-open-bibliometrics/ (Accessed July 2018)
22. Web page WRWU 2013, Available: http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/ (Accessed July 2018)
23. G. Katz, L. Rokach, World Wide Web **20**, 1153 (2017)
24. K. Ban, M. Perc, Z. Levnajic, R. Soc. Open Sci. **4**, 171217 (2017)
25. K.M. Frahm, D.L. Shepelyansky, *Wikipedia networks of 24 editions of 2017*, Available: http://www.quantware.ups-tlse.fr/QWLIB/24wiki2017 (Accessed July 2018)
26. K.M. Frahm, D.L. Shepelyansky, Reduced Google matrix, arXiv:1602.02394 [physics.soc] (2016)
27. K.M. Frahm, K. Jaffrès-Runser, D.L. Shepelyansky, Eur. Phys. J. B **89**, 269 (2016)
28. S. El Zant, K.M. Frahm, K. Jaffres-Runser, D.L. Shepelyansky, Eur. Phys. J. B **91**, 7 (2018)
29. J. Lages, D.L. Shepelyansky, A. Zinovyev, PLoS One **13**, e0190812 (2018)
30. Web page WRWU 2017, Available: http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/ (Accessed July 2018)
31. C. Coquidé, J. Lages, D.L. Shepelyansky, World influence and interactions of universities from Wikipedia networks, arXiv:1809.00332 [cs.SI] (2018)
32. A.D. Chepelianskii, Towards physical laws for software architecture, arXiv:1003.5455 [cs.SE] (2010)
33. Country codes ISO 3166-1 alpha-2. Available: https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2 (Accessed July 2018)
34. *Jenks natural breaks optimization* entry in English Wikipedia, Available: https://en.wikipedia.org/wiki/Jenks_natural_breaks_optimization (Accessed June 2018)
35. S. El Zant, K. Jaffres-Runser, K.M. Frahm, D.L. Shepelyansky, PLoS One **13**, e0201397 (2018)
36. P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Genome Res. **13**, 2498 (2003)