Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles

Young-Ho Eom[1], Dima L. Shepelyansky[1,*]

1 *Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France*

* Webpage: www.quantware.ups-tlse.fr/dima

Abstract

How different cultures evaluate a person? Is an important person in one culture is also important in the other culture? We address these questions via ranking of multilingual Wikipedia articles. With three ranking algorithms based on network structure of Wikipedia, we assign ranking to all articles in 9 multilingual Editions of Wikipedia and investigate general ranking structure of PageRank, CheiRank and 2DRank. In particular, we focus on articles related persons, identify top 30 persons for each rank among different Editions and analyze distinctions of their distributions over activity fields such as politics, art, science, religion, sport for each Edition. We find that local heroes are dominant but also global heroes exist and create an effective network representing entanglement of cultures. The Google matrix analysis of network of cultures shows signs of the Zipf law distribution. This approach allows to examine diversity and shared characteristics of knowledge organization between cultures. The developed computational, data driven approach highlights cultural interconnections in a new perspective.

Dated: June 26, 2013

1 Introduction

Wikipedia, the online collaborative encyclopedia, is an amazing example of human collaboration for knowledge description, characterization and creation. Like the Library of Babel, described by Jorge Luis Borges [1], Wikipedia goes to accumulate the whole human knowledge. Since every behavioral 'footprint' (log) is recorded and open to anyone, Wikipedia provides great opportunity to study various types of social aspects such as opinion consensus [2], language complexity [3] and collaboration structure [4]. A remarkable feature of Wikipedia is its existence in various language Editions. In a first approximation we can attribute each language to an independent culture, leaving for future refinements of cultures inside one language. Although Wikipedia has a neutral point of view policy, cultural bias or reflected cultural diversity is inevitable since knowledge and knowledge description are also affected by culture like other human behaviors [6–8]. Thus cultural bias of contents [5] becomes an important issue. However, the cross-cultural difference between Wikipedia Editions can be also a valuable opportunity for a cross-cultural empirical study with 'big data' approach. Recent steps in this direction, done for biographical networks of Wikipedia, have been reported in [9].

Here we address the question of how importance (ranking) of an article in Wikipedia depends on cultural diversity. In particular, we consider articles about persons (personalities). For instance, is an important person in English Wikipedia is also important in Korean Wikipedia? How about French? Since Wikipedia is the product of collective intelligence, the ranking of articles about persons is a collective evaluation of the persons by Wikipedia users. For the ranking of Wikipedia articles we use PageRank algorithm of Brin and Page [10], CheiRank and 2Drank algorithms used in [11–13], which allow to characterize the information flows with ingoing and outgoing links. We also analyze the distribution of top ranked persons over main human activities attributed to politics, science, art, religion, sport, etc (all others), extending the approach developed in [12, 14] to multiple cultures (languages). The comparison of different cultures shows that they have distinct dominance of these activities.

We attribute belongings of top ranked persons at each Wikipedia language to different cultures (native languages) and in this way construct the network of cultures. The Google matrix analysis of this network allows us to find interconnections and entanglement of cultures. We believe that our computational approach of 'big data' analysis of large-scale Wikipedia networks, combined with comparative distinctions of different languages, generates novel insights on cultural diversity.

2 Method

We consider Wikipedia as a network of articles. Each article corresponds to a node of the network and hyperlinks between articles correspond to links of the network. For a given network, we can define adjacency matrix A_{ij}. If there is a link (one or more quotations) from node (article) j to node (article) i then $A_{ij} = 1$, otherwise, $A_{ij} = 0$. The out-degree $k_{out}(j)$ is the number of links from node j to other nodes and the in-degree $k_{in}(j)$ is the number of links to node j from other nodes.

Google matrix

The matrix S_{ij} of Markov chain transitions is constructed from adjacency matrix A_{ij} by normalizing sum of elements of each column to unity ($S_{ij} = A_{ij}/\sum_i A_{ij}$, $\sum_i S_{ij} = 1$) and replacing columns with only zero elements (*dangling nodes*) by $1/N$, with N being the matrix size. Then the Google matrix of this directed network has the form [10, 15]:

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

In the WWW context the damping parameter α describes the probability $(1 - \alpha)$ to jump to any article (node) for a random surfer. The matrix G belongs to the class of Perron-Frobenius operators, it naturally appears in dynamical systems [16]. The right eigenvector at $\lambda = 1$, which is called the PageRank, has real non-negative 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. For large size networks the PageRank vector and several other eigenvectors can be numerically obtained using the powerful Arnoldi algorithm as described in [17]. The PageRank vector can be also obtained by a simple iteration method [15]. We use here the standard value of $\alpha = 0.85$ [15].

To rank articles of Wikipedia, we use three ranking algorithms based on network structure of Wikipedia articles. Detail description of these algorithms and their use for English Wikipedia articles are given in [12–14, 17].

PageRank algorithm

PageRank algorithm is originally introduced for Google web search engine to rank web pages of the World Wide Web (WWW) [10]. Currently PageRank is widely used to rank nodes of network systems including scientific papers [18], social network services [19] and even biological systems [20]. Here we briefly outline the iteration method of PageRank computation. The PageRank vector $P(i, t)$ of a node i at iteration t in a network of N nodes is given by

$$P(i, t) = \sum_j G_{ij} P(j, t - 1) \ , \ P(i, t) = (1 - \alpha)/N + \alpha \sum_j A_{ij} P(j, t - 1)/k_{out}(j) \ . \tag{2}$$

The stationary state $P(i)$ of $P(i, t)$ is the PageRank of node i. More detail information about PageRank algorithm is described in [15]. Ordering all nodes by their decreasing probability $P(i)$ we obtain the PageRank index $K(i)$.

The essential idea of PageRank algorithm is to use a directed link as a weighted 'recommendation'. Like academic citation, more cited nodes are considered more important. In addition, recommendations by important articles are more important. So high PageRank nodes in the network have many incoming links from other nodes or incoming links from high PageRank nodes.

CheiRank algorithm

While the PageRank algorithm uses information of ingoing links to node i, CheiRank algorithm considers information of outgoing links from node i [11–13]. Thus CheiRank is complementary to PageRank in order to rank nodes in directed networks. The CheiRank vector $P^*(i,t)$ of a node at iteration time t is given by

$$P^*(i) = (1-\alpha)/N + \alpha \sum_j A_{ji} P^*(j)/k_{in}(j) \tag{3}$$

We also point out that the CheiRank is the right eigenvector with maximal eigenvalue $\lambda = 1$ satisfying the equation $P^*(i) = \sum_j G^*_{ij} P^*(j)$, where the Google matrix G^* is built for the network with inverted directions of links via the standard definition of G given above.

Like for PageRank, we consider the stationary state $P^*(i)$ of $P^*(i,t)$ as the CheiRank probability of node i at $\alpha = 0.85$. High CheiRank nodes in the network have many outgoing links from a considered node or outgoing links to high CheiRank nodes. Ordering all nodes by their decreasing probability $P^*(i)$ we obtain the CheiRank index $K^*(i)$.

We note that PageRank and CheiRank naturally appear in the world trade network corresponding to import and export in a commercial exchange between countries [21].

2DRank algorithm

With PageRank $P(i)$ and CheiRank $P^*(i)$ probabilities, we can assign PageRank ranking $K(i)$ and CheiRank ranking $K^*(i)$ to each article, respectively. From these two ranks, we can construct 2-dimensional plane of K and K^*. The two dimensional ranking K_2 is defined by counting nodes in order of their appearance on ribs of squares in (K, K^*) plane with their size growing from $K = 1$ to $K = N$ [12]. Briefly, nodes with high PageRank and CheiRank both get high 2DRank ranking.

Data

We consider 9 Editions of Wikipedia including English (EN), French (FR), German (DE), Italian (IT), Spanish (ES), Dutch (NL), Russian (RU), Hungarian (HU) and Korean (KO). Since Wikipedia have various language Editions and language is a most fundamental part of culture, the cross-edition study of Wikipedia can give us insight on cultural diversity. The overview summary of parameters of each Wikipedia is represented in Table 1.

3 Results

We investigate ranking structure of articles and identify global properties of PageRank and CheiRank vectors. The detailed analysis is done for top 30 persons obtained from the global list of ranked articles for each of 9 languages. The distinctions and common characteristics of cultures are analyzed by attributing top 30 persons in each language to human activities listed above and to their native language.

General ranking structure

We calculate PageRank and CheiRank probabilities and indexes for all networks of considered Wikipedia Editions. The PageRank and CheiRank probabilities as functions of ranking indexes are shown in Fig. 1. The decay is compatible with an approximate algebraic decrease of a type $P \sim 1/K^{\beta}$, $P^* \sim 1/K^{*\beta}$ with $\beta \sim 1$ for PageRank and $\beta \sim 0.6$ for CheiRank. These values are similar to those found for the English Wikipedia of 2009 [12]. This difference of β values originates from asymmetric nature between in-degree and out-degree degree distributions, since PageRank is based on ingoing degree while CheiRank is based on outgoing degree. Ingoing degree distribution of a Wikipedia Editions is broader than outgoing degree distribution of the same Edition. Indeed, the CheiRank probability is proportional to frequency of outgoing links which has a more rapid decay compared to ingoing one (see discussion in [12]). The PageRank and CheiRank across Editions are similar to each other even if there are stronger fluctuations of P^* related to stronger fluctuations of outgoing links [14].

The top article of PageRank is usually *USA* or the name of country of a given language (FR, RU, KO). For NL we have at the top *beetle, species, France*. The top articles of CheiRank are various listings.

The correlation between PageRank and CheiRank vectors can be characterized by the correlator κ [11–13] defined by

$$\kappa = N \sum_i P(i) P^*(i) - 1 \tag{4}$$

The value of correlator for each Wikipedia Edition is represented in Table 1. All correlators are positive and distributed in the interval $(1, 8)$.

Since each article has its PageRank ranking K and CheiRank ranking K^*, we can assign two dimensional coordinates to all the articles. Fig. 2 shows the density of articles in the two dimensional plane (K, K^*) for each Wikipedia Edition. The density is computed for 100×100 logarithmically equidistant cells which cover the whole plane (K, K^*). The density plot represents the locations of articles in the plane. We can observe high density of articles around line $K = K^* + const$ that indicates the positive correlation between PageRank and CheiRank. However, there are only a few articles within the region of top both PageRank and CheiRank indexes. We also observe the tendency that while high PageRank articles ($K < 100$) have intermediate CheiRank ($10^2 < K^* < 10^4$), high CheiRank articles ($K^* < 100$) have broad PageRank rank values.

Ranking of articles for persons

We choose top 30 articles about persons for each Edition and each ranking. In Fig. 2, they are shown by red circles (PageRank), green squares (2DRank) and cyan triangles (CheiRank). We assign local ranking $R_{E,A}$ $(1 \ldots 30)$ to each person in the list of top 30 persons for each Edition E and ranking algorithm A. An example of $E = EN$ and $A = PageRank$ are given in Table 2.

From the lists of top persons, we identify the "fields" of activity for each top 30 rank person in which he/she is active on. We categorize six activity fields - politics, art, science, religion, sport and etc. As shown in Fig. 3, for PageRank, politics is dominant and science is secondarily dominant. The only exception is Dutch where science is the dominant activity field. In case of

2DRank, art becomes dominant and politics is secondarily dominant. In case of CheiRank, art and sport are dominant fields. Thus for example, in CheiRank top 30 list we find astronomers who discovered a lot of asteroids, e.g. Karl Wilhelm Reinmuth (4th position in RU and 7th in DE), who was a prolific discoverer of about 400 of them. As a result, his article contains a long listing of asteroids discovered by him giving him a high CheiRank.

The change of activity priority for different ranks is due to the different balance between ingoing and outgoing links there. Thus politicians are well known and are pointed by many articles but they are not very communicative since they rarely point to other articles. In contrast, persons of science, art and sport are more communicative because of listings of insects, planets, asteroids they discovered, or listings of song albums or sport competitions they gain.

Next we investigate distributions over "cultures" to which persons belong to. We determined the culture of person based on the language the person mainly used (mainly native language). We consider 10 culture categories - EN, FR, DE, IT, ES, NL, RU, HU, KO and WR. Here "WR" category represents all other cultures which do not belong to considered 9 Wikipedia Editions. Comparing with the culture of persons at various Editions, we can assign "locality" to each 30 top rank persons for a given Wikipedia Edition and ranking algorithm. For instance, as shown in Table 2, *George W. Bush* belongs to "Politics", "English" and "Local" for English Wikipedia and PageRank, *Jesus* belongs to “Religion”, “World” WR and “Non-local”.

As shown in Fig. 4, regardless of ranking algorithms, main part of top 30 ranking persons of each Edition belong to the culture of the Edition (usually about 50%). For example, high PageRank persons in English Wikipedia are mainly English (53.3%). It is notable that top ranking persons in Korean Wikipedia are not only mainly Korean (56.7%) but also the most top ranking non Korean persons in Korean Wikipedia are Chinese and Japanese (20%). Although there is a strong tendency that each Edition favors its own persons, there is also overlap between Editions. For PageRank, on average, 23.7 percent of top persons are overlapped while for CheiRank , the overlap is quite low, only 1.3 percent. For 2DRank, the overlap is 6.3 percent. The overlap of list of top persons implies the existence of cross-cultural 'heroes'.

To understand difference between local and non-local top persons for each Edition quantitatively, we consider the PageRank case because it has a large fraction of non-local top persons. From Eq. (2), a citing article j contributes $\langle P(j)/k_{out}(j)\rangle$ to PageRank of a node i. So the PageRank $P(i)$ can be high if the node i has many incoming links from citing articles j or it has incoming links from high PageRank nodes j with low out-degree $k_{out}(j)$. Thus we can identify origin of each top person's PageRank using the average PageRank contribution $\langle P(j)/k_{out}(j)\rangle$ by nodes j to person i and average number of incoming links (in-degree) $k_{in}(i)$ of person i .

As represented in Table 3, on average, except NL and HU, local top persons have more incoming links than non-local top persons but the PageRank contribution of the corresponding links are lower than links of non-local top persons. This indicates that local top persons are cited more than non-local top persons but non-local top persons are cited in a more efficient way (cited by important articles or by articles which don't have many citing links)

Global and local heroes

Based on cultural dependency on rankings of persons, we can identify global and local heroes in the considered Wikipedia Editions. However, for CheiRank the overlap is very low and our statistics is not sufficient for selection of global heroes. Hence we consider only PageRank and 2DRank cases. We determine the local heroes for each ranking and for each Edition as top persons of the given ranking who belongs to the same culture as the Edition. Top 3 local heroes for each ranking and each Edition are represented in Table 4 (PageRank), Table 5 (CheiRank) and Table 6 (2DRank), respectively.

In order to identify global heroes, we define ranking score $\Theta_{P,A}$ for each person P and each ranking algorithm A. Since every person in the top person list has relative ranking $R_{P,E,A}$ for each Wikipedia Edition E and ranking algorithm A (For instance, in Table 2, $R_{Napoleon,EN,PageRank} = 1$). The ranking score $\Theta_{P,A}$ of a person P is give by

$$\Theta_{P,A} = \sum_{E} (31 - R_{P,E,A}) \tag{5}$$

According to this definition, a person who appears more often in the lists of Editions and has top ranking in the list gets high ranking score. By sorting this ranking score for each algorithm, we have a list of global heroes for each algorithm. The result is shown in Table 7. Napoleon is the 1st global hero by PageRank and Micheal Jackson is the 1st global hero by 2DRank.

Network of cultures

To characterize the entanglement and interlinking of cultures we use the data of Fig. 4 and from them construct the network of cultures. The image of networks obtained from top 30 persons of PageRank and 2DRank listings are shown in Fig. 5 (we do not consider CheiRank case due to small overlap of persons resulting in a small data statistics). The weight of directed Markov transition from a culture A to a culture B is given by a number of persons of a given culture B (e.g FR) appearing in the list of top 30 persons of PageRank (or 2DRank) in a given culture A (e.g. EN). Thus e.g. for transition from EN to FR we find a weight 2 (2 French persons in top 30 persons of English Wikipedia). The transitions inside each culture (persons of the same language as language Edition) are omitted since we are analyzing the interlinks between cultures. Then the Google matrix of cultures is constructed by the standard rule for the directed networks of cultures (sum in each column is equal to unity, $\alpha = 0.85$). Even if this network has only 10 nodes we still can find for it PageRank and CheiRank probabilities P and P^* and corresponding indexes K and K^*. The matrix elements of G matrix, written in order of index K, are shown in Fig. 6 for the corresponding networks of cultures presented in Fig. 5.

The decay of PageRank and CheiRank probabilities with the indexes K, K^* are shown in Fig. 7 for the culture networks of Fig. 5. Even if the network size is very small we find that this decay is in a satisfactory agreement with the Zipf law [22] showing a decay $1/K$. We make a conjecture that the Zipf law should work in a better way for a larger number of multilingual Wikipedia Editions which have now about 250 languages.

The distributions of cultures on the PageRank- CheiRank plane (K, K^*) are shown in Fig. 8. For the network of cultures constructed from top 30 PageRank persons we obtain the following ranking. The node WR is located at the top PageRank $K = 1$ and stays at the last CheiRank position $K^* = 10$. This happens due to the fact that such persons as *Carl Linnaeus, Jesus, Aristotle, Plato, Alexander the Great, Muhammad* are not native for our 9 Wikipedia Editions so that we have many nodes pointing to WR node, while WR has no outgoing links. The next node in PageRank is FR node at $K = 2, K^* = 5$, then DE node at $K = 3, K^* = 4$ and only then we find EN node at $K = 4, K^* = 7$. The node EN is not at all at top PageRank positions since it has many American politician which do not count for links between cultures. After the world WR the top position is taken by French (FR) and then German (DE) cultures which have strong links inside the continental Europe.

However, the ranking is drastically changed when we consider top 30 2DRank persons. Here, the dominant role is played by art and science with singers, artists and scientists. The world WR here remains at the same position at $K = 1, K^* = 10$ but then we obtain English EN ($K = 2, K^* = 1$) and German DE ($K = 3, K^* = 5$) cultures while FR is moved to $K = K^* = 7$.

4 Discussion

We investigated cross-cultural diversity of Wikipedia via ranking of Wikipedia articles. Even if the used ranking algorithms are purely based on network structure of Wikipedia articles, we find cultural distinctions and entanglement of cultures obtained from the multilingual Editions of Wikipedia.

In particular, we analyze raking of articles about persons and identify activity field of persons and culture which persons belong to. Politics is dominant in top PageRank persons, art is dominant in top 2DRank persons and in top CheiRank persons art and sport are dominant. We find that each Wikipedia Edition favors its own persons, who have same cultural background, but there are also cross-cultural non-local heroes, and even "global heroes". We establish that local heroes are cited more often but non-local heroes on average are cited by more important articles.

Attributing top persons of the ranking list to different cultures we construct the network of cultures and characterize entanglement of cultures on the basis of Google matrix analysis of this directed network.

We considered only 9 Wikipedia Editions selecting top 30 persons in a "manual" style. It would be useful to analyze a larger number of Editions using an automatic computerized selection of persons from prefabricated listing in many languages developing lines discussed in [9]. This will allow to analyze a large number of persons improving the statistical accuracy of links between different cultures.

The importance of understanding of cultural diversity in globalized world is growing. Our computational, data driven approach can provide a quantitative and efficient way to understand diversity of cultures by using big data of created by millions of Wikipedia users. We believe that our results shed a new light on how knowledge is organized and spread by different cultures.

Acknowledgments

We thank Sebastiano Vigna [23] who kindly provided to us the network data of 9 Wikipedia Editions, collected in the frame of FET NADINE project. This research is supported in part by the EC FET Open project "New tools and algorithms for directed network analysis" (NADINE No 288956).

Supporting Information

Supplementary methods, tables, ranking lists and figures are available at http://www.quantware.ups-tlse.fr/QW
data sets are available at [23].

References

1. Borges JL (1962) *The Library of Babel* in *Ficciones*, Grove Press, New York

2. Kaltenbrunner A, Laniado D (2012) *There is no deadline - time evolution of Wikipedia discussions*, Proc. of the 8th Intl. Symposium on Wikis and Open Collaboration, WikiSym12, Linz

3. Yasseri T, Kornai A, Kertész J (2012) *A practical approach to language complexity: a Wikipedia case study* PLoS ONE, **7**: e48386

4. Brandes U, Kenis P, Lerner U, van Raaij D (2009) *Network analysis of collaboration structure in Wikipedia* Proc. 18th Intl. Conf. WWW, :731

5. Callahan ES, Herriing SC (2011) *Cultural bias in Wikipedia content on famous persons*, Journal of the American society for information science and technology **62**: 1899

6. Norenzayan A (2011) *Explaining human behavioral diversity*, Science, **332**: 1041

7. Gelfand MJ, Raver JL, Nishii L, Leslie LM, Lun J *et al.* (2011) *Differences between tight and loose cultures: a 33-nation study*, Science, **332**: 1100

8. UNESCO World Report (2009) *Investing in cultural diversity and intercultural dialogue*, Available: http://www.unesco.org/new/en/culture/resources/report/the-unesco-world-report-on-cultural-diversity

9. Aragon P, Laniado D, Kaltenbrunner A, Volkovich Y (2012) *Biographical social networks on Wikipedia*, arXiv:1204.3799v2[cs.SI]

10. Brin S, Page L (1998) *The anatomy of a large-scale hypertextual Web search engine* Computer Networks and ISDN Systems **30**: 107

11. Chepelianskii AD (2010) *Towards physical laws for software architecture* arXiv:1003.5455 [cs.SE]

12. Zhirov AO, Zhirov OV, Shepelyansky DL (2010) *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77**: 523

13. Ermann L, Chepelianskii AD, Shepelyansky DL (2012) *Toward two-dimensional search engines*, J. Phys. A: Math. Theor. **45**: 275101

14. Eom YH, Frahm KM, Benczúr A, Shepelyansky DL (2013) *Time evolution of Wikipedia network ranking* arXiv:1304.6601 [physics.soc-ph]

15. Langville AM, Meyer CD (2006) *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton

16. Brin M, Stuck G (2002) *Introduction to dynamical systems*, Cambridge Univ. Press, Cambridge, UK

17. Ermann L, Frahm KM, Shepelyansky DL (2013) *Spectral properties of Google matrix of Wikipedia and other networks*, Eur. Phys. J. D **86**: 193

18. Chen P, Xie H, Maslov S, Redner S (2007) *Finding scientific gems with Googleś PageRank algorithm* Jour. Informetrics, **1**: 8

19. Kwak H, Lee C, Park H, Moon S (2010) *What is Twitter, a social network or a news media?*, Proc. 19th Int. Conf. WWW2010, :591, ACM, New York

20. Kandiah V, Shepelyansky DL (2013) *Google matrix analysis of DNA sequences*, PLoS ONE **8(5)**: e61519

21. Ermann L, Shepelyansky DL (2011) *Google matrix of the world trade network*, Acta Physica Polonica A **120(6A)**, A158

22. Zipf GK (1949) *Human behavior and the principle of least effort*, Addison-Wesley, Boston

23. Vigna S, personal web page, Available: http://vigna.dsi.unimi.it/

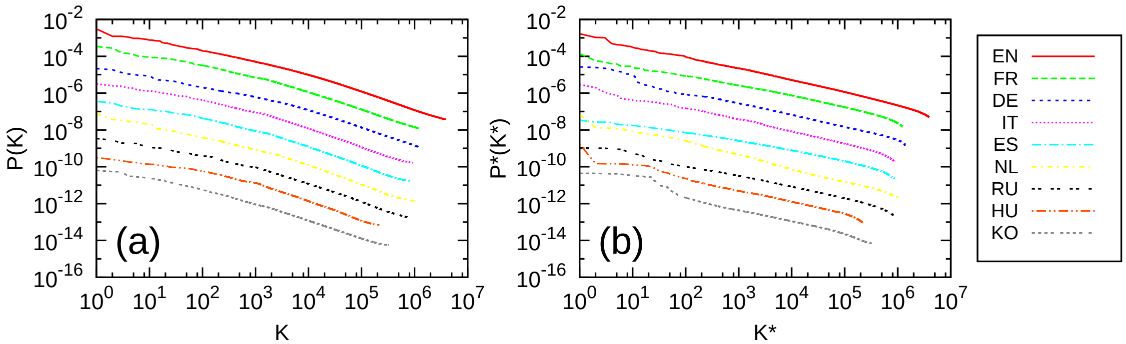

Figure 1. (a) PageRank probability $P(K)$ as function of PageRank index K; (b) CheiRank probability $P^*(K^*)$ as function of CheiRank index K^*. For a better visualization each PageRank P and CheiRank P^* curve is shifted down by a factor 10^0 (EN), 10^1 (FR), 10^2 (DE), 10^3 (IT), 10^4 (ES), 10^5 (NL), 10^6 (RU), 10^7 (HU), 10^8 (KO).

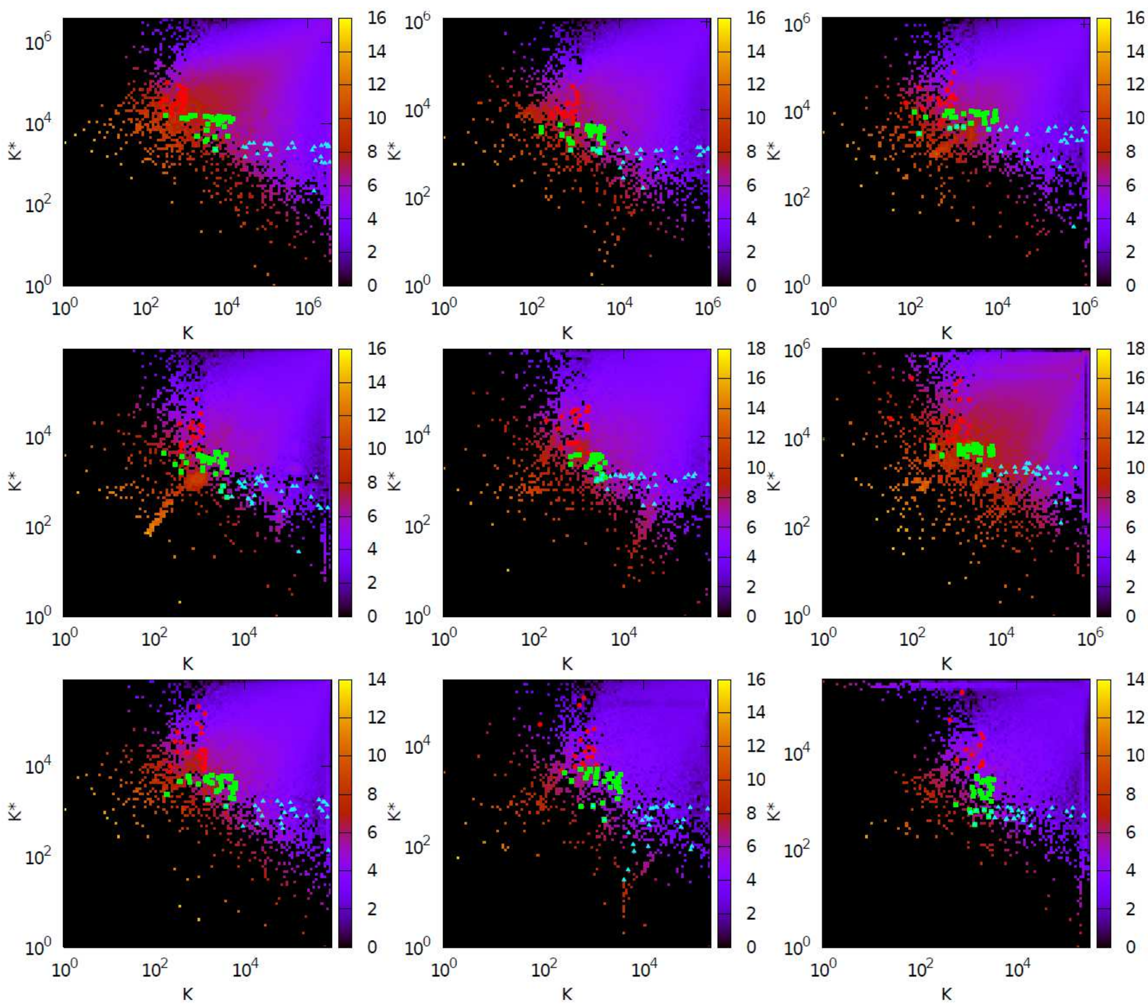

Figure 2. Density of Wikipedia articles in the PageRank ranking K versus CheiRank ranking K^* plane for each Wikipedia Edition. The red points are top PageRank articles of persons, the green points are top 2DRank articles of persons and the cyan points are top CheiRank articles of persons. Panels show: English (top-left), French (top-center), German (top-right), Italian (middle-left), Spanish (middle-center), Dutch (middle-left), Russian (bottom-left), Hungarian (bottom-center), Korean (bottom-right). Color bars shown natural logarithm of density, changing from minimal nonzero density (dark) to maximal one (white), zero density is shown by black.

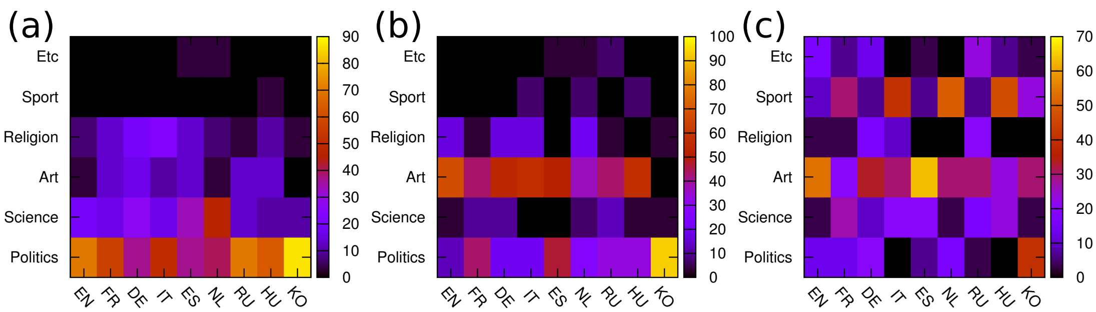

Figure 3. Distribution of top 30 persons in each rank over activity fields for each Wikipedia Edition. Panels correspond to (a) PageRank, (b) 2DRank, (3) CheiRank. The color bar shows the values in percents.

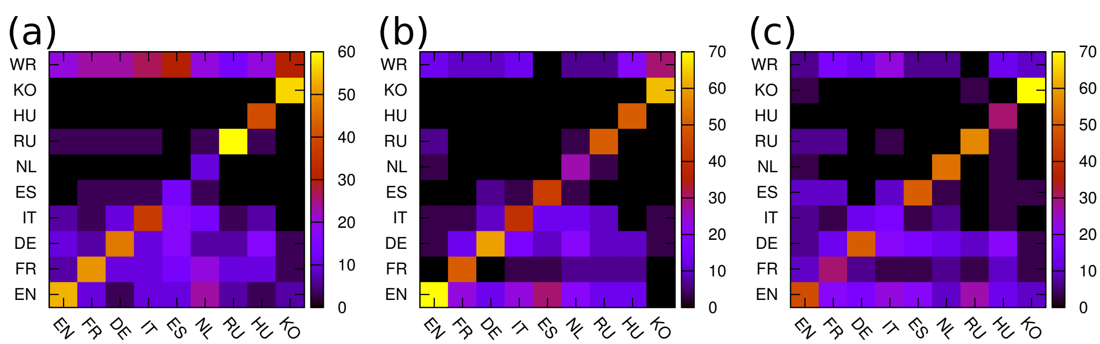

Figure 4. Distributions of top 30 persons over different cultures corresponding to Wikipedia Editions, "WR" category represents all other cultures which do not belong to considered 9 Wikipedia Editions. Panels show ranking by (a) PageRank, (b) 2DRank, (3) CheiRank. The color bar shows the values in percents.

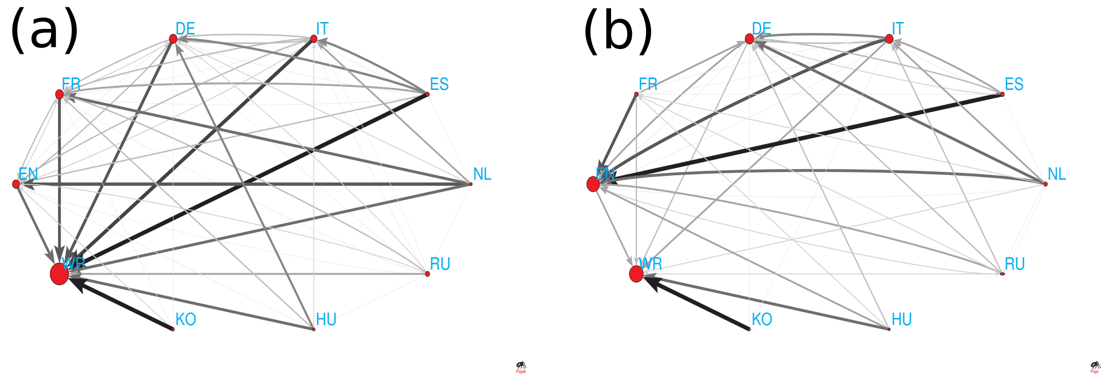

Figure 5. Network of cultures obtained from 9 Wikipedia languages and the remaining world (WR) selecting 30 top persons of PageRank (a) and 2DRank (b) in each culture. The link width and darkness are proportional to a number of foreign persons quoted in top 30 of a given culture, the link direction goes from a given culture to cultures of quoted foreign persons, quotations inside cultures are not considered. The size of nodes is proportional to their PageRank.

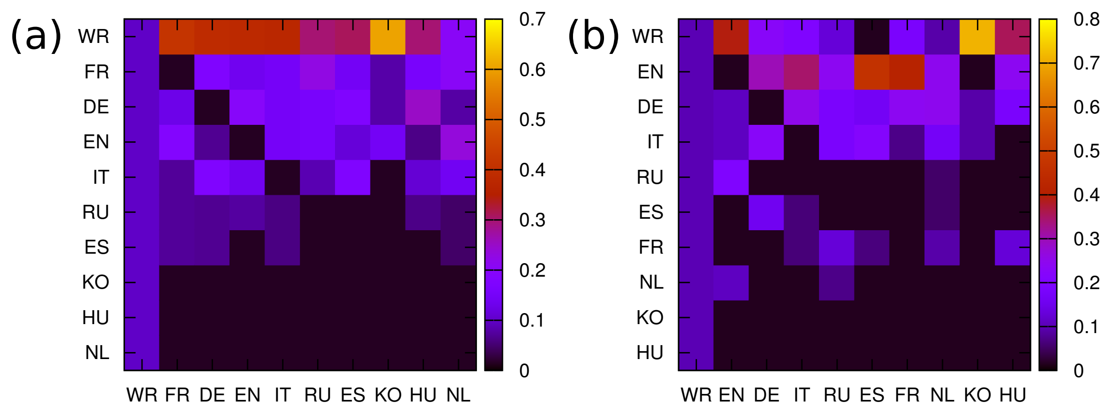

Figure 6. Google matrix of network of cultures from Fig. 5, shown respectively for panels $(a), (b)$. The matrix elements G_{ij} are shown by color at the damping factor $\alpha = 0.85$, index j is chosen as the PageRank index K of PageRank vector so that the top cultures with $K = K' = 1$ are located at the top left corner of the matrix.

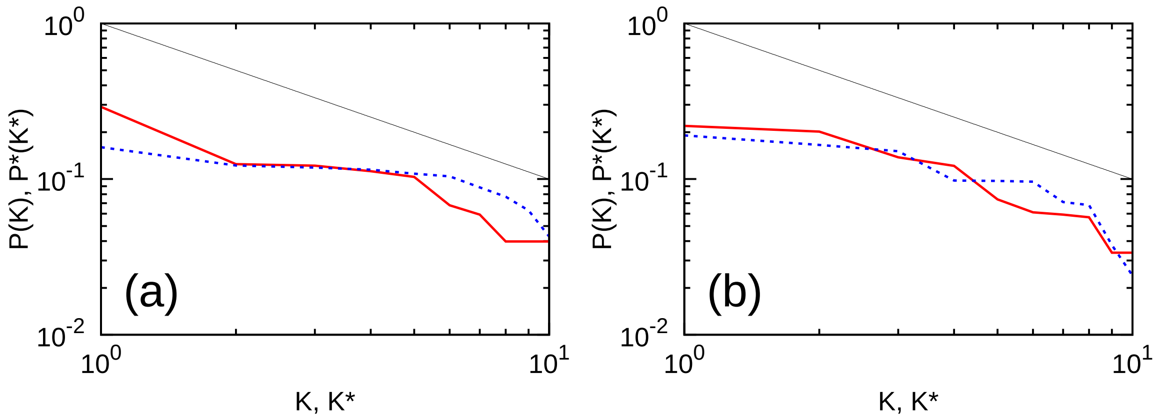

Figure 7. Dependence of probabilities of PageRank P (red) and CheiRank P^* (blue) on corresponding indexes K and K^*. The probabilities are obtained from the network and Google matrix of cultures shown in Fig. 5 and Fig. 6 for corresponding panels $(a),(b)$. The straight lines indicate the Zipf law $P \sim 1/K; P^* \sim 1/K^*$.

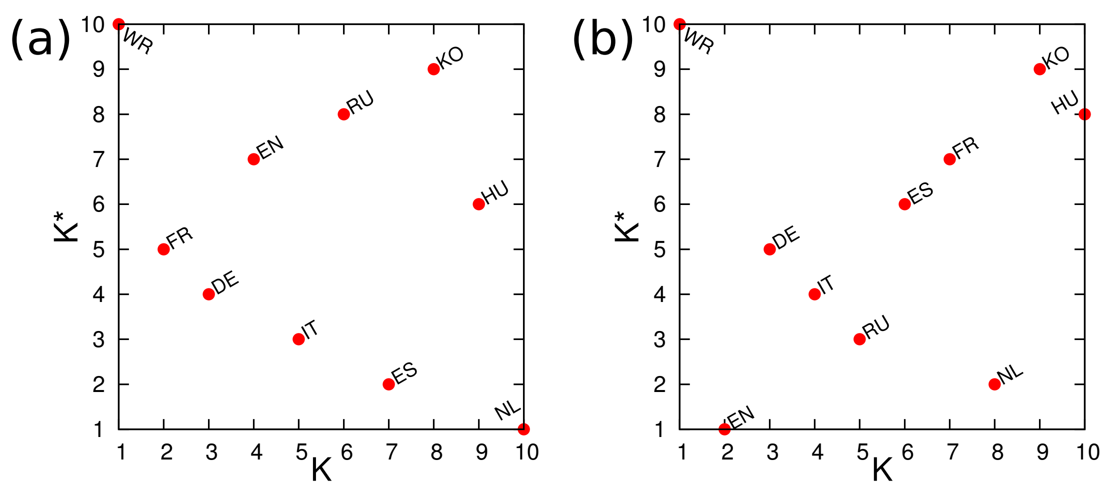

Figure 8. PageRank versus CheiRank plane of cultures with corresponding indexes K and K^* obtained from the network of cultures for corresponding panels $(a),(b)$.

Table 1. Considered Wikipedia networks from language Editions: English (EN), French (FR), German (DE), Italian (IT), Spanish (ES), Dutch (NL), Russian (RU), Hungarian (HU), Korean (KO). Here N_A is number of articles, N_L is number of hyperlinks between articles, κ is the correlator between PageRank and CheiRank. Date represents the time in which data are collected.

Edition	N_A	N_L	κ	Date
EN	3920628	92878869	3.905562	Mar. 2012
FR	1224791	30717338	3.411864	Feb. 2012
DE	1396293	32932343	3.342059	Mar. 2012
IT	917626	22715046	7.953106	Mar. 2012
ES	873149	20410260	3.443931	Feb. 2012
NL	1034912	14642629	7.801457	Feb. 2012
RU	830898	17737815	2.881896	Feb. 2012
HU	217520	5067189	2.638393	Feb. 2012
KO	323461	4209691	1.084982	Feb. 2012

Table 2. Example of list of top 10 person by PageRank for English Wikipedia with their field of activity and native language.

$R_{EN,PageRank}$	Person	Field	Culture	Locality
1	Napoleon	Politics	FR	Non-local
2	Carl Linnaeus	Science	WR	Non-local
3	George W. Bush	Politics	EN	Local
4	Barack Obama	Politics	EN	Local
5	Elizabeth II	Politics	EN	Local
6	Jesus	Religion	WR	Non-local
7	William Shakespeare	Art	EN	Local
8	Aristotle	Science	WR	Non-local
9	Adolf Hitler	Politics	DE	Non-local
10	Bill Clinton	Politics	EN	Local

Table 3. Average PageRank and average incoming degree of local and non-local heroes i for each Edition. $\langle P(j)/k(j)_{out}\rangle_L$ and $\langle P(j)/k(j)_{out}\rangle_{NL}$ are average PageRank contribution of hero i by articles j which cite local heroes and non-local heroes i respectively. $\langle k(i)_{in}\rangle_L$ and $\langle k(i^*)_{in}\rangle_{NL}$ are average number of incoming degree $k(i)_{in}$ and $k(i^*)_{in}$ of local hero i and non-local i^*, respectively.

Edition	N_{Local}	$\langle P(j)/k(j)_{out}\rangle_L$	$\langle P(j)/k(j)_{out}\rangle_{NL}$	$\langle k(i)_{in}\rangle_L$	$\langle k(i^*)_{in}\rangle_{NL}$
EN	16	1.53×10^{-8}	2.19×10^{-8}	5.9×10^{3}	4.3×10^{3}
FR	15	4.03×10^{-8}	5.56×10^{-8}	2.9×10^{3}	2.4×10^{3}
DE	14	4.28×10^{-8}	4.43×10^{-8}	2.7×10^{3}	2.9×10^{3}
IT	11	7.86×10^{-8}	8.44×10^{-8}	1.7×10^{3}	2.0×10^{3}
ES	4	5.66×10^{-8}	9.93×10^{-8}	2.0×10^{3}	1.5×10^{3}
NL	3	9.33×10^{-7}	4.24×10^{-7}	6.8×10^{2}	1.0×10^{3}
RU	18	7.27×10^{-8}	9.60×10^{-8}	1.9×10^{3}	1.8×10^{3}
HU	12	3.51×10^{-7}	3.19×10^{-7}	7.8×10^{2}	8.1×10^{2}
KO	17	1.71×10^{-7}	6.06×10^{-7}	5.8×10^{2}	3.3×10^{2}

Table 4. List of local heroes by PageRank for each Wikipedia Edition. All names are represented by article titles in English Wikipedia.

Edition	1st	2nd	3rd
EN	George W. Bush	Barack Obama	Elizabeth II
FR	Napoleon	Louis XIV of France	Charles de Gaulle
DE	Adolf Hitler	Martin Luther	Immanuel Kant
IT	Augustus	Dante Alighieri	Julius Caesar
ES	Charles V, Holy Roman Emperor	Philip II of Spain	Francisco Franco
NL	Ernst Mayr	William I of the Netherlands	Beatrix of the Netherlands
RU	Peter the Great	Joseph Stalin	Alexander Pushkin
HU	Matthias Corvinus	Szentágothai János	Stephen I of Hungary
KO	Gojong of the Korean Empire	Sejong the Great	Park Chung-hee

Table 5. List of local heroes by CheiRank for each Wikipedia Edition. All names are represented by article titles in English Wikipedia.

Edition	1st	2nd	3rd
EN	C. H. Vijayashankar	Matt Kelley	William Shakespeare (inventor)
FR	Jacques Davy Duperron	Jean Baptiste Eblé	Marie-Magdeleine Aymé de La Chevrelière
DE	Harry Pepl	Marc Zwiebler	Eugen Richter
IT	Nduccio	Vincenzo Olivieri	Mina (singer)
ES	Che Guevara	Arturo Mercado	Francisco Goya
NL	Hans Renders	Julian Jenner	Marten Toonder
RU	Aleksander Vladimirovich Sotnik	Aleksei Aleksandrovich Bobrinsky	Boris Grebenshchikov
HU	Csernus Imre	Kati Kovács	Pléh Csaba
KO	Lee Jong-wook (baseball)	Kim Dae-jung	Kim Kyu-sik

Table 6. List of local heroes by 2DRank for each Wikipedia Edition. All names are represented by article titles in English Wikipedia

Edition	1st	2nd	3rd
EN	Frank Sinatra	Paul McCartney	Michael Jackson
FR	François Mitterrand	Jacques Chirac	Honoré de Balzac
DE	Adolf Hitler	Otto von Bismarck	Ludwig van Beethoven
IT	Giusppe Garibaldi	Raphael	Benito Mussolini
ES	Simón Bolívar	Francisco Goya	Fidel Castro
NL	Albert II of Belgium	Johan Cruyff	Rembrandt
RU	Dmitri Mendeleev	Peter the Great	Yaroslav the Wise
HU	Stephen I of Hungary	Sándor Petőfi	Franz Liszt
KO	Gojong of the Korean Empire	Sejong the Great	Park Chung-hee

Table 7. List of global heroes by PageRank and 2DRank for all 9 Wikipedia Editions. All names are represented by article titles in English Wikipedia. Here, Θ_A is the ranking score of the algorithm A (5); N_A is the number of appearances of a given person in the top 30 rank for all Editions.

Rank	PageRank global heroes	Θ_{PR}	N_A	2DRank global heroes	Θ_{2D}	N_A
1st	Napoleon	259	9	Micheal Jackson	119	5
2nd	Jesus	239	9	Adolf Hitler	93	6
3rd	Carl Linnaeus	235	8	Julius Caesar	85	5
4th	Aristotle	228	9	Pope Benedict XVI	80	4
5th	Adolf Hitler	200	9	Wolfgang Amadeus Mozart	75	5
6th	Julius Caesar	161	8	Pope John Paul II	71	4
7th	Plato	119	6	Ludwig van Beethoven	69	4
8th	Charlemagne	111	8	Bob Dylan	66	4
9th	William Shakespeare	110	7	William Shakespeare	57	3
10th	Pope John Paul II	108	6	Alexander the Great	56	3