

# Interactions of cultures and top people of Wikipedia from ranking of 24 language editions

Young-Ho Eom<sup>1</sup>, Pablo Aragón<sup>2</sup>, David Laniado<sup>2</sup>, Andreas Kaltenbrunner<sup>2</sup>, Sebastiano Vigna<sup>3</sup>, Dima L. Shepelyansky<sup>1,\*</sup>

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

<sup>2</sup> *Barcelona Media Foundation, Barcelona, Spain*

<sup>3</sup> *Dipartimento di Informatica, Università degli Studi di Milano, Italy*

\* Webpage: [www.quantware.ups-tlse.fr/dima](http://www.quantware.ups-tlse.fr/dima)

## Abstract

Wikipedia is a huge global repository of human knowledge, that can be leveraged to investigate interwinements between cultures. With this aim we apply two methods, Markov chains and Google matrix, for the analysis of the hyperlink networks of 24 Wikipedia language editions, and rank all their articles by PageRank, 2DRank and CheiRank algorithms. Using automatic extraction of people names we obtain the top 100 historical figures for each edition and for each algorithm. We investigate their spatial, temporal, and gender distributions in dependence of their cultural origins. Our study demonstrates not only the existence of skewness with local figures, mainly recognized only in their own culture, but also the existence of global historical figures appearing in a large number of editions. By determining the birth time and place of these persons, we perform an analysis of the evolution of such figures through 35 centuries of human history for each language, thus recovering interactions and entanglement of cultures over time. We also obtain the distributions of historical figures over world countries, highlighting geographical aspects of cross-cultural links. Considering historical figures who appear in multiple editions as interactions between cultures, we construct a network of cultures and identify the most influential cultures according to such network.

Dated: May 28, 2014

## Introduction

The influence of digital media on collective opinions, social relationships, and information dynamics is growing significantly with the advances of information technology. On the other hands, understanding how collective opinions are reflected in digital media has crucial importance. Among such a media, Wikipedia, the open, free, and online encyclopedia, has crucial importance since it is not only the largest global knowledge repository but also the biggest collaborative knowledge platform on the Web. Thanks to its huge size, broad coverage and ease of use, currently Wikipedia is one of the most widely used knowledge references. However, since the beginning, there have been constant concerns about the reliability of Wikipedia because of its openness. Although professional scholars may not be affected by skewness or bias in Wikipedia, students and the public can be affected significantly [1,2]. Extensive studies have examined the reliability of contents [1–3], topic coverage [4], vandalism [5], and conflict [6–8] in Wikipedia.

Wikipedia is available in different language editions; about 287 language editions are currently active. This indicates that the same topic can be described in hundreds of articles written by different language user groups. Since language is one of the primary elements of culture [9], collective cultural bias may be reflected on the contents and organization of each Wikipedia edition. Although Wikipedia adopts a “neutral point of view” (NPOV) policy for the description of contents, aiming to provide unbiased information to the public [10], it is straightforward that each language edition presents reality from a different angle. To investigate differences and relationships among different language editions, we develop mathematical and statistical methods which treat the huge amount of information in Wikipedia excluding cultural preferences of the investigators.

Cultural bias or differences across Wikipedia editions have been investigated in previous research [11–16]. A special emphasis was devoted to persons described in Wikipedia articles [11] and their ranking [17,18]. Indeed, human knowledge, as well as Wikipedia itself, was created by people who are the main actors of its development. Thus it is rather natural to analyze a ranking of people according to the Wikipedia hyper-link network of citations between articles. A cross-cultural study of biographical articles was presented in [19], by building a network of interlinked biographies. Another approach was proposed recently in [20]: the difference in importance of historical figures across Wikipedia language editions is assessed on the basis of the global ranking of Wikipedia articles about persons. This study, motivated by the question “Is an important person in a given culture also important in other cultures?”, showed that there are strong entanglements and local biases of historical figures in Wikipedia. Indeed, the results of the study show that each Wikipedia edition favors persons belonging to same culture (language), but also that there are cross-Wikipedia top ranked persons, who can be signs of entanglement between cultures. These cross-language historical figures highlight entanglement of cultures and generate inter-culture networks demonstrating interactions between cultures [20]. Such an approach provides us novel insights on cross-cultural differences across Wikipedia editions. However, in [20] only 9 Wikipedia editions, mainly of European languages, have been considered. Thus a broader set of language editions is needed to offer a more complete view on a global scale.

We note that the analysis of persons' importance via Wikipedia becomes more and more popular. This is well visible from the appearance of new recent studies for the English Wikipedia [21] and for multiple languages [22].

Here we investigate interactions and skewness of cultures with a broader perspective, using global ranking of articles about persons in 24 Wikipedia language editions. According to Wikipedia [23] these 24 languages cover 59 percent of world population. Moreover, according to Wikipedia [24], our selection of 24 language editions covers 68 percent of the total number of 30.9 millions of Wikipedia articles in all 287 languages. These 24 editions also cover languages which played an important role in human history including Western, Asian and Arabic cultures.

On the basis of this data set we analyze spatial, temporal, and gender skewness in Wikipedia by analyzing birth place, birth date, and gender of the top ranked historical figures in Wikipedia. We identified overall Western, modern, and male skewness of important historical figures across Wikipedia editions, a tendency towards local preference (i.e. each Wikipedia edition favors historical figures born in countries speaking that edition's language), and the existence of global historical figures who are highly ranked in most of Wikipedia editions. We also constructed networks of cultures based on cross-cultural historical figures to represent interactions between cultures.

To obtain a unified ranking of historical figures for all 24 Wikipedia editions, we introduce an average ranking which gives us the top 100 persons of human history. To understand how reliable is the obtained ranking we compare it with the Hart's list of the top 100 people who, according to him, most influenced human history [25]. We note that Hart "ranked these 100 persons in order of importance: that is, according to the total amount of influence that each of them had on human history and on the everyday lives of other human beings".

## Methods

In this research, we consider each Wikipedia edition 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 an 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.

To rank the articles of a Wikipedia edition, we use two ranking algorithms based on the article network structure. Detailed descriptions of these algorithms and their use for Wikipedia editions are given in [17, 18, 26, 27]. The methods used here are described in [20]; we keep the same notations.

### Google matrix

First we construct the matrix  $S_{ij}$  of Markov transitions by normalizing the sum of the elements in each column to unity ( $S_{ij} = A_{ij} / \sum_i A_{ij}$ ,  $\sum_i S_{ij} = 1$ ) and replacing columns with zero elements by elements  $1/N$  with  $N$  being the matrix size. Then the Google matrix is given by the relation  $G_{ij} = \alpha S_{ij} + (1 - \alpha)/N$ , where  $\alpha$  is the damping factor [28]. As in [20] we use the usual value  $\alpha = 0.85$ .

## PageRank algorithm

PageRank is a widely used algorithm to rank nodes in a directed network. It was originally introduced for Google web search engine to rank web pages of the World Wide Web based on the idea of academic citations [29]. Currently PageRank is used to rank nodes of network systems from scientific papers [30] to social network services [31], world trade [32] and biological systems [33]. 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 with  $N$  nodes is given by

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

The stationary state  $P(i)$  of  $P(i, t)$  is the PageRank of node  $i$ . More detail information about the PageRank algorithm is described in [28]. Ordering all nodes by their decreasing probability  $P(i)$ , we obtain the PageRank ranking index  $K(i)$ . In qualitative terms, the PageRank probability of a node is proportional to the number of incoming links weighted according to their own probability. A random network surfer spends on a given node a time given on average by the PageRank probability.

## CheiRank algorithm

In a directed network, outgoing links can be as important as ingoing links. In this sense, as a complementary to PageRank, the CheiRank algorithm is defined and used in [17, 26, 34]. The CheiRank vector  $P^*(i, t)$  of a node at iteration time  $t$  is given by

$$P^*(i) = (1 - \alpha)/N + \alpha \sum_j A_{ji} P^*(j)/k_{in}(j) \quad (2)$$

Same as the case of PageRank, we consider the stationary state  $P^*(i)$  of  $P^*(i, t)$  as the CheiRank probability of node  $i$  with  $\alpha = 0.85$ . High CheiRank nodes in the network have large out-degree. Ordering all nodes by their decreasing probability  $P^*(i)$ , we obtain the CheiRank ranking index  $K^*(i)$ . The PageRank probability of an article is proportional to the number of incoming links, while the CheiRank probability of an article is proportional to the number of outgoing links. Thus a top PageRank article is important since other articles refer to it, while a top CheiRank article is highly connected because it refers to other articles.

## 2DRank algorithm

PageRank and CheiRank algorithms focus only on in-degree and out-degree of nodes, respectively. The 2DRank algorithm considers both types of information simultaneously to rank nodes with a balanced point of view in a directed network. Briefly speaking, nodes with both high PageRank and CheiRank get high 2DRank ranking. Consider a node  $i$  which is  $K_i$ -th ranked by PageRank and  $K^*_i$  ranked by CheiRank. Then we can assign a secondary ranking  $K'_i = \max\{K_i, K^*_i\}$  to the node. If  $K'_i < K'_j$ , then node  $j$  has lower 2DRank and vice versa. A detailed illustration and description of this algorithm is given in [17]

## Data preparation

We consider 24 editions of Wikipedia including English (EN), Dutch (NL), German (DE), French (FR), Spanish (ES), Italian (IT), Portuguese (PT), Greek (EL), Danish (DA), Swedish (SV), Polish (PL), Hungarian (HU), Russian (RU), Hebrew (HE), Turkish (TR), Arabic (AR), Persian (FA), Hindi (HI), Malaysian (MS), Thai (TH), Vietnamese (VI), Chinese (ZH), Korean (KO), and Japanese (JA). The Wikipedia data were collected in middle February 2013. The overview summary of each Wikipedia is represented in Table 1.

We understand that our selection of Wikipedia editions does not represent a complete view of all 287 languages of Wikipedia editions. However, this selection covers most of the largest language editions and allows us to perform quantitative and statistical analysis of important historical figures.

First we ranked all the articles in a given Wikipedia edition by PageRank and 2DRank algorithms, and selected biographic articles about historical figures. To identify biographical articles, we considered all articles belonging to 'Category:living people', or to 'Category:Deaths by year' or 'Category:Birth by year' or their subcategories in the English Wikipedia. In this way we obtained a list of about 1.1 million biographical articles. We identified birth place, birth date, and gender of each selected historical figure based on DBpedia [35] or a manual inspection of the corresponding Wikipedia biographical article, when for the considered historical figure no DBpedia data were available. We then started from the list of persons with their biographical article's title on the English Wikipedia, and found the corresponding titles in other language editions using the inter-language links provided by Wikipedia. Using the corresponding articles, identified by the inter-languages links in different language editions, we extracted the top 100 persons from the rankings of all Wikipedia articles of each edition. At the end, for each Wikipedia edition and for each ranking algorithm, we have information of the top 100 historical figures with their corresponding name in the English Wikipedia, their birth place and date, and their gender. All 48 lists of the top 100 historical figures in PageRank and 2DRank for the 24 Wikipedia editions and for the two ranking algorithms are represented in [36] and Supporting Information (SI). The original network data for each edition are available at [36]. The automatic extraction of persons from PageRank and 2DRank listings of articles of each edition is performed by using the above whole list of person names in all 24 editions. This method implies a significantly higher recall compared to the manual selection of persons from the ranking list of articles for each edition used in [20].

We attribute each of the 100 historical figures to a birth place at the country level, to a birth date in year, to a gender, and to a cultural group. The birth place is assigned as the current country. The cultural group of historical figures is assigned by the the most spoken language of their birth place at the current country level. For examples, if someone was born in "Constantinople" in the ancient Roman era, since the place is now Istanbul, Turkey, we assign her/his birth place as "Turkey" and since Turkish is the most spoken language in Turkey, we assign this person to the Turkish cultural group. If the birth country does not belong to any of the 24 cultures (languages) which we consider, we assign WR (world) as the culture of this person. We would like to point out that although a culture can not be defined by only language, we think that language is a suitable first-approximation of culture. All lists of top 100 historical figures with their birth place, birth date, gender, and cultural group for each Wikipedia edition

and for each ranking algorithm are represented in [36]. A part of this information is also reported in SI.

To apply PageRank and 2DRank methods, we consider each edition as the network of articles of the given edition connected by hyper-links among the articles (see the details of ranking algorithms in Section Methods). The full list of considered Wikipedia language editions is given in Table 1. Table 2 represents the top 10 historical figures by PageRank and 2DRank in the English Wikipedia. Roughly speaking, top PageRank articles imply highly cited articles in Wikipedia and top 2DRank articles imply highly cited and simultaneously highly citing articles in Wikipedia. In total, we identified 2400 top historical figures for each ranking algorithm. However, since some historical figures such as *Jesus*, *Aristotle*, or *Napoleon* appear in multiple Wikipedia editions, we have 1045 unique top PageRank historical figures and 1616 unique top 2DRank historical figures.

We should note that the extraction of persons and their information from a Wikipedia edition is not an easy task even for the English edition, and much more complicated for certain other language editions. Therefore, the above automatic method based on 1.1 million English names and their corresponding names seems to us to be the most adequate approach. Of course, it will miss people who do not have a biographical article in English Wikipedia. Cross-checking investigation is done for Korean and Russian Wikipedia, which are native for two authors, by selecting top 100 persons manually from top lists of all articles ordered by PageRank and 2DRank in both Wikipedia editions. We find that our automatic search misses on average only 2 persons from 100 top persons for these two editions (the missed names are given in SI). The errors appear due to transcription changes of names or missing cases in our name-database based on English Wikipedia. For Western languages the number of errors is presumably reduced since transcription remains close to English. Based on the manual inspection for the Korean and the Russian Wikipedia, we expect that the errors of our automatic recovery of the top people from the whole articles ordered by PageRank and 2DRank are on a level of two percent.

## Results

Above we described the methods used for the extraction of the top 100 persons in the ranking list of each edition. Below we present the obtained results describing the spatial, temporal and gender distributions of top historical figures. We also determine the global and local persons and obtain the network of cultures based on ranking of persons of a given language by other language editions of Wikipedia.

### Spatial distribution

The birth places of historical figures are attributed to the country containing their geographical location of birth according to the present geographical territories of all world countries. The list of countries appeared for the top 100 persons in all editions is given in Table 3. We also attribute each country to one of the 24 languages of the considered editions. This attribution is done according to the language spoken by the largest part of population in the given country. Thus e.g. Belgium is attributed to Dutch (NL) since the majority of the population speaks Dutch. If the main language of a country is not among our 24 languages, then this country is attributed

to an additional section WR corresponding to the remaining world (e.g. Ukraine, Norway are attributed to WR). If the birth place of a person is not known, then it is also attributed to WR. The choice of attribution of a person to a given country in its current geographic territory, and as a result to a certain language, may have some fluctuations due to historical variations of country borders (e.g. Immanuel Kant was born in the current territory of Russia and hence is attributed to Russian language). However, the number of such cases is small, being on a level of 3.5 percent (see Section “Network of cultures” below). We think that the way in which a link between person, language and country is fixed by the birth place avoids much larger ambiguity of attribution of a person according to the native language which is not so easy to fix in an automatic manner.

The obtained spatial distribution of historical figures of Wikipedia over countries is shown in Fig. 1. This averaged distribution gives the average number of top 100 persons born in a specific country as birth place, with averaging done over our 24 Wikipedia editions. Thus an average over the 24 editions gives for Germany (DE) approximately 9.7 persons in the top 100 of PageRank, being at the first position, followed by USA with approximately 9.5 persons. For 2DRank we have USA at the first position with an average of 9.8 persons and Germany at the second with an average of 8.0 persons.

Western (Europe and USA) skewed patterns are observed in both top PageRank historical figures (Fig 1. (A)) and top 2DRank historical figures (Fig. 1. (B)). This Western skewed pattern is remarkable since 11 Wikipedia editions of the 24 considered editions are not European language editions. Germany, USA, Italy, UK and France are the top five birth places of top PageRank historical figures among 71 countries. On the other hand, USA, Germany, UK, Italy and Japan are top five birth places of the top 2DRank historical figures among 91 countries.

In Fig. 2 we show the world map of countries, where color indicates the number of persons from a given country among the  $24 \times 100$  top persons for PageRank and 2DRank. Additional figures showing these distributions for different centuries are available at [36].

We also observed local skewness in the spatial distribution of the top historical figures for PageRank (2DRank) ranking algorithm as shown in Fig. 3A (in Fig. 3B). For example, 47 percent of top PageRank historical figures in the English Wikipedia were born in USA (25 percent) and UK (22 percent) and 56 percent of top historical figures in Hindi Wikipedia were born in India. A similar strong locality pattern of the top historical figures was observed in our previous research [20]. However it should be noted that in the previous study we considered the native language of the top historical figure as a criterion of locality, while in the current study we considered ‘birth place’ as criterion of locality.

Regional skewness, the preferences of Wikipedia editions for historical figures who were born in geographically or culturally related countries, is also observed. For example, 18 (5) of the top 100 PageRank historical figures in the Korean (Japanese) Wikipedia were born in China. Also 9 of the top 100 PageRank historical figures in the Persian Wikipedia were born in Saudi Arabia. The distribution of top persons from each Wikipedia edition over world countries is shown in Fig. 3A and Fig. 3B. The countries on a horizontal axis are grouped by clusters of corresponding language so that the links inside a given culture (or language) become well visible.

To observe patterns in a better way at low numbers of historical figures, we normalized each column of Fig. 3A and Fig. 3B corresponding to a given country. In this way we obtain a rescaled distribution with better visibility for each birth country level as shown in Fig. 3C and Fig. 3D,

respectively. We can observe a clear birth pattern of top PageRank historical figures born in Lebanon, Libya, Oman, and Tunisia in the case of the Arabic Wikipedia, and historical figures born in N. Korea appearing not only in the Korean but also in the Japanese Wikipedia.

In the case of the top 2DRank historical figures shown in Fig. 3B and Fig. 3D, we observe overall patterns of locality and regions being similar to the case of PageRank, but the locality is stronger.

In short, we observed that most of top historical figures in Wikipedia were born in Western countries, but also that each edition shows its own preference to the historical figures born in countries which are closely related to the corresponding language edition.

## Temporal distribution

The analysis of the temporal distribution of top historical figures is done based on their birth dates. As shown in Fig. 4A for PageRank, most of historical figures were born after the 17th century on average, which shows similar pattern with world population growth [37]. However, there are some distinctive peaks around BC 5th century and BC 1st century for the case of PageRank because of Greek scholars (*Socrates*, *Plato*, and *Herodotus*), Roman politicians (*Julius Caesar*, *Augustus*) and Christianity leaders (*Jesus*, *Paul the Apostle*, and *Mary (mother of Jesus)*). We also observe that the Arabic and the Persian Wikipedia have more historical figures than Western language Wikipedia editions from AD 6th century to AD 12th century. For the case of 2DRank in Fig. 4B, there is only one small peak around BC 1C, which is also smaller than the peak in the case of PageRank, and all the distribution is dominated by a strong growth on the 20th century.

The distributions of the top PageRank historical figures over the 24 Wikipedia editions for each century are shown in Fig. 4C. The same distribution, but normalized to unity over all editions for each century, is shown in Fig. 4E. The Persian (FA) and the Arabic (AR) Wikipedia have more historical figures than other language editions (in particular European language editions) from the 6th to the 12th century due to Islamic leaders and scholars. On the other hand, the Greek Wikipedia has more historical figures in BC 5th century because of Greek philosophers. Also most of western-southern European language editions, including English, Dutch, German, French, Spanish, Italian, Portuguese, and Greek, have more top historical figures because they have *Augustine the Hippo* and *Justinian I* in common. Similar distributions obtained from 2DRank are shown in Fig. 4D and Fig. 4F respectively.

The data of Figs. 4E,F clearly show well pronounced patterns, corresponding to strong interactions between cultures: from BC 5th century to AD 15th century for JA, KO, ZH, VI; from AD 6th century to AD 12th century for FA, AR; and a common birth pattern in EN,EL,PT,IT,ES,DE,NL (Western European languages) from BC 5th century to AD 6th century.

At a first glance from Figs. 4E,F we observe for persons born in AD 20th century a significantly more homogeneous distribution over cultures compared to early centuries. However, as noted in [20], each Wikipedia edition favors historical figures speaking the corresponding language. We investigate how this preference to same-language historical figures changes in time. For this analysis, we define two variables  $M_{L,C}$  and  $N_{L,C}$  for a given language edition  $L$  and a given century  $C$ . Here  $M_{L,C}$  is the number of historical figures born in all countries being

attributed to a given language  $L$ , and  $N_{L,C}$  is the total number of historical figures for a given century  $C$  and a given language edition  $L$ . For example, among the 21 top PageRank historical figures from the English Wikipedia, who were born in AD 20th century, two historical figures (Pope John Paul II and Pope Benedict XVI) were not born in English speaking countries. Thus in this case  $N_{EN,20} = 21$  and  $M_{EN,20} = 19$ . Fig 5 represents the ratio  $r_{L,C} = M_{L,C}/N_{L,C}$  for each edition and each century. In ancient times (i.e. before AD 5th century), most historical figures for each Wikipedia edition are not born in the same language region except for the Greek, Italian, Hebrew, and Chinese Wikipedia. However, after AD 5th century, the ratio of same language historical figures is rising. Thus, in AD 20th century, most Wikipedia editions have significant numbers of historical figures born in countries speaking the corresponding language. For PageRank persons and AD 20th century, we find that the English edition has the largest fraction of its own language, followed by Arabic and Persian editions while other editions have significantly large connections with other cultures. For the English edition this is related to a significant number of USA presidents appearing in the top 100 list (see [17, 18]). For 2DRank persons the largest fractions were found for Greek, Arabic, Chinese and Japanese cultures. These data show that even in age of globalization there is a significant dominance of local historical figures for certain cultures.

## Gender distribution

From the gender distributions of historical figures, we observe a strong male-skewed pattern across Wikipedia editions regardless of the ranking algorithm. On average, 5.2(10.1) female historical figures are observed among the 100 top PageRank (2DRank) persons for each Wikipedia edition. Fig. 6 shows the number of top female historical figures for each Wikipedia edition. Thai, Hindi, Swedish, and Hebrew have more female historical figures than the average over our 24 editions in the case of PageRank. On the other hand, the Greek and the Korean versions have a lower number of females than the average. In the case of 2DRank, English, Hindi, Thai, and Hungarian Wikipedia have more females than the average while German, Chinese, Korean, and Persian Wikipedia have less females than the average. In short, the top historical figures in Wikipedia are quite male-skewed. This is not surprising since females had little chances to be historical figures for most of human history. We compare the gender skewness to other cases such as the number of female editors in Wikipedia (9 percent) in 2011 [38] and the share of women in parliaments, which was 18.7 percent in 2012 by UN Statistics and indicators on women and men [39], the male skewness is stronger in the contents of Wikipedia [40]. However, the ratio of females among the top historical figures is growing by time as shown in Fig. 6 C. It is notable that the peak in Fig. 6C at BC 1st is due to "Mary (mother of Jesus)". In 20th century 2DRank gives a larger percent of female persons comparing to PageRank. This is due to the fact that 2DRank has a larger fraction of singers and artists comparing to PageRank (see [17, 18]) and that the fraction of female persons in these fields of activity is larger.

## Global historical figures

Above we analyzed how top historical figures in Wikipedia are distributed in terms of space, time, and gender. Now we identify how these top historical figures are distributed in each

Wikipedia edition and which are global historical figures. According to previous research [20], there are some global historical figures who are recognized as important historical figures across Wikipedia editions. We identify global historical figures based on the ranking score for a given person determined by her number of appearances and ranking index over our 24 Wikipedia editions.

Following [20], the ranking score  $\Theta_{P,A}$  of a historical figure  $P$  is given by

$$\Theta_{P,A} = \sum_E (101 - R_{P,E,A}) \quad (3)$$

Here  $R_{P,E,A}$  is the ranking of a historical figure  $P$  in Wikipedia edition  $E$  by ranking algorithm  $A$ . According to this definition, a historical figure who appears more often in the lists of top historical figures for the given 24 Wikipedia editions or has higher ranking in the lists gets a higher ranking score. Table 4 represents the top 10 global historical figures for PageRank and 2DRank. *Carl Linnaeus* is the 1st global historical figure by PageRank and *Adolf Hitler* is the 1st global historical figure by 2DRank. On the other hand, the lists of the top 10 local historical figures ordered by our ranking score for each language are represented in SI Tables S1-S25 and [36].

Our analysis suggests that there might be three groups of historical figures. Fig. 7 shows these three groups of top PageRank historical figures in Wikipedia: (i) global historical figures who appear in most of Wikipedia editions ( $N_A \geq 18$ ) and are highly ranked ( $\langle K \rangle \leq 50$ ) for each Wikipedia such as *Carl Linnaeus*, *Plato*, *Jesus*, and *Napoleon* (Right-Top of the Fig. 7A); (ii) local-highly ranked historical figures who appear in a few Wikipedia editions ( $N_A < 18$ ) but are highly ranked ( $\langle K \rangle \leq 50$ ) in the appeared Wikipedia editions such as *Tycho Brahe*, *Sejong the Great*, and *Sun Yat-sen* (Left-Top of the Fig. 7A); (iii) locally-low ranked historical figures who appear in a few Wikipedia editions ( $N_A < 18$ ) and who are not highly ranked ( $\langle K \rangle > 50$ ). Here  $N_A$  is the number of appearances in different Wikipedia editions for a given person and  $\langle K \rangle$  is the average ranking of the given persons across Wikipedia editions for each ranking algorithm. In the case of 2DRank historical figures, due to the absence of global historical figures, most of them belong to two types of local historical figures (i.e. local-highly ranked or local-lowly ranked).

Following ranking of persons via  $\Theta_{P,A}$  we determine also the top global female historical figures, presented in Table 5 for PageRank and 2DRank persons. The full lists of global female persons are available at [36] (63 and 165 names for PageRank and 2DRank).

The comparison of our 100 global historical figures with the top 100 from Hart's list [25] gives an overlap of 43 persons for PageRank and 26 persons for 2DRank. We note that for the top 100 of English Wikipedia we obtain a lower overlap of 37 (PageRank) and 4 (2DRank) persons. For the Italian Wikipedia edition we obtain an overlap of 33 (PageRank) and 12 (2DRank) persons. This shows that the consideration of 24 editions provides us with a more balanced selection of top historical figures. Our overlap of the top 100 global historical figures by PageRank with the top 100 people from Pantheon MIT ranking list [22] is 44 percent, while the overlap of this Pantheon list with Hart's list is 42 percent, a bit smaller than our overlap with Hart's list. The top 100 persons from [21] are not publicly available and we do not present comparison with this work.

The overlap of the top 100 global persons found here with the previous study [20] gives 54 and 47 percent for PageRank and 2DRank lists, respectively. However, we note that the global list in [20] was obtained from the top 30 persons in each edition while here we use the top 100 persons.

It is interesting to note that for the top 100 PageRank universities from the English Wikipedia edition the overlap with Shanghai top 100 list of universities is on a even higher level of 75 percent [17].

Finally, we note that the ranking of historical figures using the whole PageRank (or 2DRank) list of all Wikipedia articles of a given edition provides a more stable approach compared to the network of biographical articles used in [19]. Indeed, the number of nodes and links in such a biographical network is significantly smaller compared to the whole network of Wikipedia articles and thus the fluctuations become rather large. For example, from the biographical network of the Russian edition one finds as the top person *Napoleon III* (and even not *Napoleon I*) [19], who has a rather low importance for Russia. In contrast to that the present study gives us the top PageRank historical figure of the Russian edition to be *Peter the Great*, that has much more historical grounds.

## Network of cultures

We consider the selected top persons from each Wikipedia edition as important historical figures recognized by people who speak the language of a given Wikipedia edition. Therefore, if a top person from a language edition *A* appears in another edition *B*, then we can consider this as a 'cultural' influence from culture *A* to *B*. Here we consider each language as a proxy for a cultural group and assign each historical figure to one of these cultural groups based on the most spoken language of her/his birth place at the country level. For example, *Adolf Hitler* was born in modern Austria and since German language is the most spoken language in Austria, he is considered as a German historical figure in our analysis. This method may lead to some misguided results due to discrepancy between territories of country and cultures, e.g. *Jesus* was born in the modern State of Palestine (Bethlehem), which is an Arabic speaking country. Thus *Jesus* is from the Arabic culture in our analysis. Other similar examples we find are: *Charlemagne* (Belgium - Dutch), *Immanuel Kant* (Russia - Russian), *Moses* (Egypt - Arabic), *Catherine the Great* (Poland - Polish).

In total there are such 36 cases from the global PageRank list of 1045 names (these 36 names are given in SI). However, in our knowledge, the birth place is the best way to assign a given historical figure to a certain cultural background computationally and systematically and with the data we have available. In total we have only about 3.4 percent of cases which can be discussed and where a native speaking language can be a better indicator of belonging to a given culture. For the global 2DRank list of 1616 names we identified 53 similar cases where an attribution to a culture via a native language or a birth place could be discussed (about 3.3 percent). These 53 names are given in SI. About half of such cases are linked with birth places in ancient Russian Empire where people from Belarus, Lithuania and Ukraine moved to RU, IL, PL, WR. However, the percentage of such cases is small and the corresponding errors also remain small.

Based on the above assumption and following the approach developed in [20], we construct

two weighted networks of cultures (or language groups) based on the top PageRank historical figures and top 2DRank historical figures respectively. Each culture (i.e. language) is represented as a node of the network, and the weight of a directed link from culture  $A$  to culture  $B$  is given by the number of historical figures belonging to culture  $B$  (e.g. French) appearing in the list of top 100 historical figures for a given culture  $A$  (e.g. English). The persons in a given edition, belonging to the language of the edition, are not taken into account since they do not create links between cultures. In Table 6 we give the number of such persons for each language. This table also gives the number of persons of a given language among the top 100 persons of the global PageRank and 2DRank listings.

For example, there are 5 French historical figures among the top 100 PageRank historical figures of the English Wikipedia, so we can assign weight 5 to the link from English to French. Fig. 8A and Fig. 8B represent the constructed networks of cultures defined by appearances of the top PageRank historical figures and top 2DRank historical figures, respectively. In total we have two networks with 25 nodes which include our 24 editions and an additional node WR for all other world cultures.

The Google matrix  $G_{ij}$  for each network is constructed following the standard rules described in [20] and Methods Section. In a standard way we determine the PageRank index  $K$  and the CheiRank index  $K^*$  that order all cultures according to decreasing PageRank and CheiRank probabilities (see Methods and FigS1. in SI). The structure of matrix elements  $G_{KK'}$  is shown in Fig. 9.

To identify which culture (or language groups) are more influential than others, we calculated PageRank and CheiRank of the constructed networks of cultures by considering link weights. Briefly speaking, a culture has high PageRank (CheiRank) if it has many ingoing (outgoing) links from (to) other cultures (see Methods). The distribution of cultures on PageRank-CheiRank plane is shown in Fig. 10. In both case of PageRank and 2DRank historical figures, historical figures of English culture (i.e. born in English language spoken countries) are most influential (highest PageRank) and German culture is the second one (Fig. 10A,B). Here we consider the historical figures for the whole range of centuries. Fig. 10 represents the detailed features of how each culture is located on the plane of PageRank ranking  $K$  and CheiRank ranking  $K^*$  based on the top PageRank historical figures (Fig. 10A) and top 2DRank historical figures (Fig. 10B). Here  $K$  indicates the ranking of a given culture ordered by how many of its own top historical figures appear in other Wikipedia editions, and  $K^*$  indicates the ranking of a given culture according to how many of the top historical figures in the considered culture are from other cultures. As described above, English is on ( $K = 1$ ,  $K^* = 19$ ) and German is on ( $K = 2$ ,  $K^* = 21$ ) in the case of PageRank historical figures (Fig. 10A). In the case of 2DRank historical figures, English is on ( $K = 1$ ,  $K^* = 14$ ) and German is on ( $K = 2$ ,  $K^* = 9$ ).

It is important to note that there is a significant difference compared to the previous study [20]: there, only 9 editions had been considered and the top positions were attributed to the world node WR which captured a significant fraction of the top persons. This indicated that 9 editions are not sufficient to cover the whole world. Now for 24 editions we see that the importance of the world node WR is much lower (it moves from  $K = 1$  for 9 editions [20] to  $K = 4$  and 3 in Fig. 10A and Fig. 10B). Thus our 24 editions cover the majority the world. Still it would be desirable to add a few additional editions (e.g. Ukraine, Baltic Republics, Serbia etc.) to fill certain gaps.

It is interesting to note that the ranking plane of cultures ( $K, K^*$ ) changes significantly in time. Indeed, if we take into account only persons with a birth date before the 19th century then the ranking is modified with EN going to 4th (Fig. 10C for PageRank figures) and 6th position (Fig. 10C for 2DRank figures) while the top positions are taken by IT, DE, FR and DE, IT, AR, respectively.

At the same time we may also argue that for cultures it is important not only to be cited but also to be communicative with other cultures. To characterize communicative properties of nodes on the network of cultures shown in Fig. 8 we use again the concepts of PageRank, CheiRank and 2DRank for these networks as described in Methods and [20]. Thus, for the network of cultures of Fig. 8, the 2DRank index of cultures highlights their influence in a more balanced way taking into account their importance (incoming links) and communicative (outgoing links) properties in a balanced manner.

Thus we find for all centuries at the top positions Greek, Turkish and Arabic (for PageRank persons) and French, Russian and Arabic (for 2DRank persons). For historical figures before the 19th century, we find respectively Arabic, Turkish and Greek (for PageRank) and Arabic, Greek and Hebrew (for 2DRank). The high position of Turkish is due to its close links both with Greek culture in ancient times and with Arabic culture in more recent times. We see also that with time the positions of Greek in 2DRank improves due to a global improved ranking of Western cultures closely connected with Greece.

## Discussion

By investigating birth place, birth date, and gender of important historical figures determined by the network structure of Wikipedia, we identified spatial, temporal, and gender skewness in Wikipedia. Our analysis shows that most important historical figures across Wikipedia language editions are born in Western countries after the 17th century, and are male. Also, each Wikipedia edition highlights local figures so that most of its own historical figures are born in the countries which use the language of the edition. The emergence of such pronounced accent to local figures seems to be natural since there are more links and interactions within one culture. This is also visible from the fact that in many editions the main country for the given language is at the first PageRank position among all articles (e.g. Russia in RU edition) [20]. Despite such a locality feature, there are also global historical figures who appear in most of the considered Wikipedia editions with very high rankings. Based on the cross-cultural historical figures, who appear in multiple editions, we can construct a network of cultures which describes interactions and entanglement between cultures.

It is very difficult to describe history in an objective way and due to that it was argued that history is "an unending dialogue between the past and present" [41]. In a similar way we can say that history is an unending dialogue between different cultural groups. The computational and data-mining approach, based on rank vectors of Google matrix of Wikipedia, provides us a purely statistical analysis of interactions and entanglement of cultures. We find that this approach gives a reliable extraction of the most influential historical figures allowing to determine in a statistical and mathematical way the entanglement and interactions of cultures. Our approach recovers 43 percent of historical figures selected by historical studies [25], which demonstrates its reliability.

Thus, such a mathematical analysis of local and global historical figures can be a useful step to the understanding of local and global history and interactions of world cultures.

Influence of digital media on information dissemination and social collective opinions among the public is growing fast. Our research across Wikipedia language editions suggests a rigorous mathematical way, based on Markov chains and Google matrix, for recognition of important historical figures and analysis of interactions of cultures at different historical periods and in different world regions. We think that a further extension of this approach to a larger number of Wikipedia editions will provide a more detailed and balanced analysis of interactions of world cultures.

## Acknowledgments

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

Supporting Information file presents Figure S1 with PageRank and CheiRank probabilities for networks of cultures of Fig. 8, lists of top 100 global PageRank and 2DRank names; Tables S1-S25 of top 10 names of given language and remained world from the global PageRank and 2DRank ranking lists of persons ordered by the score  $\Theta_{P,A}$  of Eq.(3). Additional lists of all 100 ranked names for all 24 Wikipedia editions and corresponding network link data for each edition are given at [36].

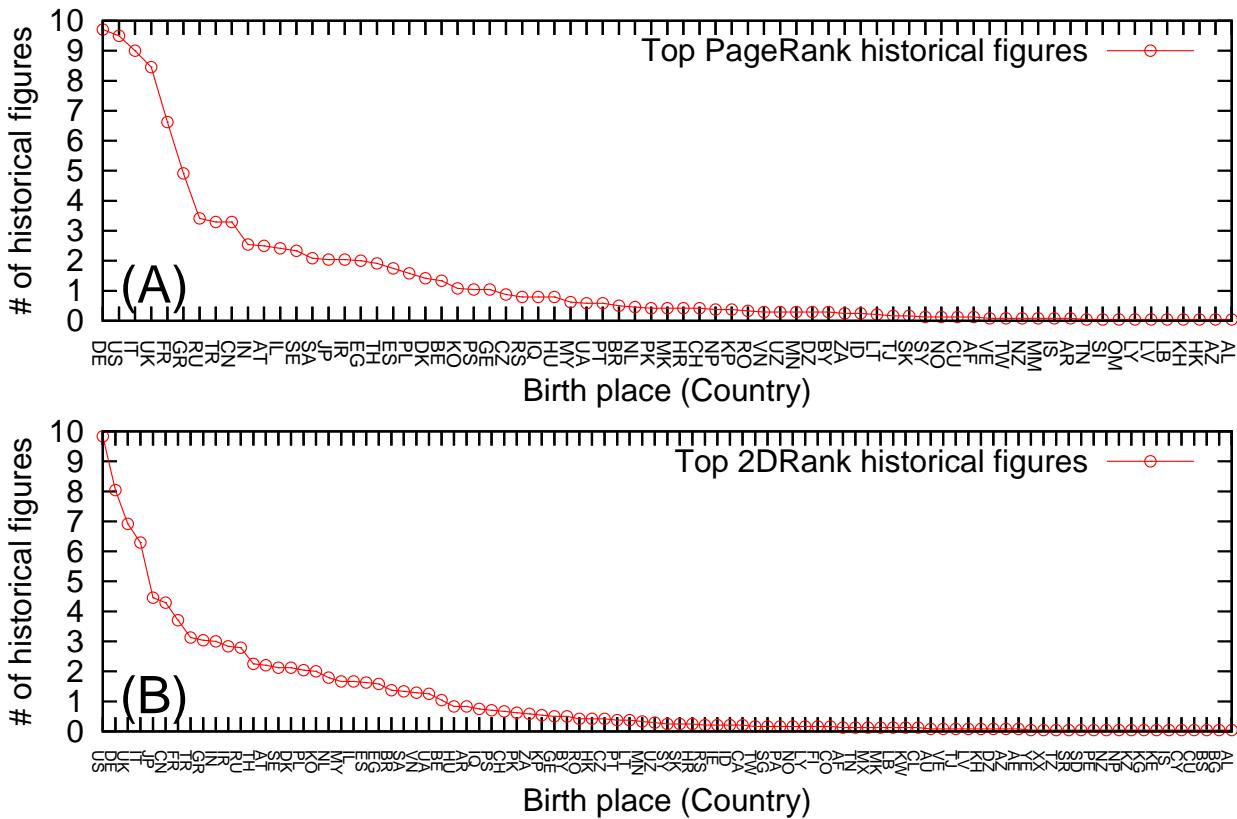
## References

1. Rosenzweig R (2006) *Can history be open source? Wikipedia and the future and the past*, Journal of American History **93**(1), 117
2. Lavsa SM, Corman SL, Culley CM, Pummer TL (2011) *Reliability of Wikipedia as a medication information source for pharmacy students*, Currents in Pharmacy Teaching and Learning **3**(2): 154-158
3. Giles J (2005) *Internet encyclopedia go head to head*, Nature, **438**, 900
4. Kittur A, Chi EH, Suh B (2009) *What's in Wikipedia?: mapping topics and conflict using socially annotated category structure*, In Proc. of SIGCHI Conference on Human Factors in Computing Systems, CHI'09, ACM, New York
5. Priedhorsky R, Chen J, Lam STK, Panciera K, Terveen L and Riedl J (2007). *Creating, Destroying, and Restoring Value in Wikipedia*, In Proceedings of the Intl. Conf. on Supporting Group Work, 295, ACM, New York

6. Yasseri T, Sumi R, Rung A, Kornai A, Kertész J (2012) *Dynamics of Conflicts in Wikipedia*, PLoS ONE **7**(6): e38869
7. Yasseri T, Spoerri A, Graham M, Kertész J (2013) *The most controversial topics in Wikipedia: a multilingual and geographical analysis* arXiv:1305.5566 [physics.soc-ph]
8. Laniado D, Tasso R, Volkovich Y, Kaltenbrunner A (2011) *When the wikipedians talk: Network and tree structure of Wikipedia discussion pages*, Proc. of ICWSM
9. 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>
10. Wikipedia: Neutral point of view. Retrieved May 12, 2014 from [http://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)
11. Callahan ES, Herrington SC (2011) *Cultural bias in Wikipedia content on famous persons*, Journal of the American society for information science and technology **62**: 1899
12. Hecht B, Gergle D (2009) *Measuring self-focus bias in community-maintained knowledge repositories* Proc. of the Fourth Intl Conf. Communities and technologies, ACM, New York :11
13. Hecht B, Gergle D (2010) *The Tower of Babel Meets Web 2.0: User-Generated Content and Its Applications in a Multilingual Context*, Proc. of SIGCHI Conference on Human Factors in Computing Systems, CHI'10, Atlanta, ACM, New York
14. Nemoto K, Gloor PA (2011) *Analyzing cultural differences in collaborative innovation networks by analyzing editing behavior in different-language Wikipedias* Procedia - Social and Behavioral Sciences **26**: 180
15. Warncke-Wang M, Uduwage A, Dong Z, Riedl J (2012) *In search of the ur-Wikipedia: universality, similarity, and translation in the Wikipedia inter-language link network*, Proceedings of the 8th Intl. Symposium on Wikis and Open Collaboration (WikiSym 2012), ACM, New York
16. Massa P, Scrinzi F (2012) *Manypedia: Comparing language points of view of Wikipedia communities*, Proceedings of the 8th Intl. Symposium on Wikis and Open Collaboration (WikiSym 2012), ACM, New York Warncke-Wang M, Uduwage A, Dong Z, Riedl J (2012) *In search of the ur-Wikipedia: universality, similarity, and translation in the Wikipedia inter-language link network*, Proceedings of the 8th Intl. Symposium on Wikis and Open Collaboration (WikiSym 2012), ACM, New York No 20
17. Zhirov AO, Zhirov OV, Shepelyansky DL (2010) *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77**: 523

18. Eom YH, Frahm KM, Benczur A, Shepelyansky DL (2013) *Time evolution of Wikipedia network ranking* Eur. Phys. J. B, **86**:482
19. Aragón P, Laniado D, Kaltenbrunner A, Volkovich Y (2012) *Biographical social networks on Wikipedia: a cross-cultural study of links that made history*, Proc. of the 8th Intl. Symposium on Wikis and Open Collaboration (WikiSym 2012), ACM, New York No 19;
20. Eom YH, Shepelyansky DL (2013) *Highlighting Entanglement of Cultures via Ranking of Multilingual Wikipedia Articles*, PLoS ONE, **8**(10): e74554
21. Skiena S, Ward CB (2013) *Who is Bigger?: Where Historical Figures Really Rank*, Cambridge University Press, Cambridge UK
22. MIT Pantheon project. Available: <http://pantheon.media.mit.edu>. Accessed 2014 May 12.
23. Wikipedia: List of languages by number of native speakers. Retrieved May 12, 2014 from [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers)
24. Wikipedia: Wikipedia . Retrieved May 12, 2014 from <http://en.wikipedia.org/wiki/Wikipedia>
25. Hart MH (1992) *The 100: ranking of the most influential persons in history*, Citadel Press, N.Y.
26. Ermann L, Chepelianskii AD, Shepelyansky DL (2012) *Toward two-dimensional search engines*, J. Phys. A: Math. Theor. **45**: 275101
27. Ermann L, Frahm KM, Shepelyansky DL (2013) *Spectral properties of Google matrix of Wikipedia and other networks*, Eur. Phys. J. D **86**: 193
28. Langville AM, Meyer CD (2006) *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton
29. Brin S, Page L (1998) *The anatomy of a large-scale hypertextual Web search engine* Computer Networks and ISDN Systems **30**: 107
30. Chen P, Xie H, Maslov S, Redner S (2007) *Finding scientific gems with Google's PageRank algorithm* Jour. Informetrics, **1**: 8
31. Kwak H, Lee C, Park H, Moon S (2010) *What is Twitter, a social network or a news media?*, Proc. 19th Int. Conf. WWW2010, ACM, New York :591
32. Ermann L, Shepelyansky DL (2011) *Google matrix of the world trade network*, Acta Physica Polonica A **120(6A)**, A158
33. Kandiah V, Shepelyansky DL (2013) *Google matrix analysis of DNA sequences*, PLoS ONE **8(5)**: e61519

34. Chepelianskii AD (2010) *Towards physical laws for software architecture* arXiv:1003.5455 [cs.SE]
35. <http://dbpedia.org>. Accessed 2014 May 12
36. Top wikipeople. Available: <http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/>. Accessed 2014 May 12.
37. United States Census Bureau. Retrieved May 12, 2014 from [http://www.census.gov/population/international/data/worldpop/table\\_history.php](http://www.census.gov/population/international/data/worldpop/table_history.php)
38. Wikipedia: Wikipedians. Retrieved May 12, 2014 from <http://en.wikipedia.org/wiki/Wikipedia:Wikipedians>
39. Statistics and indicators on women and men by United Nation. <http://unstats.un.org/unsd/Demographic/products/indwm/> (accesible May 12, 2014)
40. Lam STK, Uduwage A, Dong Z, Sen S (2011) *WP:clubhouse?: an exploration of Wikipedia's gender imbalance* Proc. of the 7th Intl. Symposium on Wikis and Open Collaboration, WikiSym'11, Moutain View
41. Carr EH (1961) *What is History?*, Vintage Books, New York



**Figure 1.** Birth place distribution of top historical figures averaged over 24 Wikipedia edition for (A) PageRank historical figures (71 countries) and (B) 2DRank historical figures (91 countries). Two letter country codes are represented in Table 3.



**Figure 2.** Number of appearances of historical figures of a given country, obtained from 24 lists of top 100 persons of PageRank (top panel) and 2DRank (bottom panel), shown on the world map. Color changes from zero (white) to maximum (black), it corresponds to average number of persons per country as shown in Fig. 1.

**Table 1.** Wikipedia hyperlink networks from the 24 considered language editions. Here  $N_a$  is the number of articles and  $N_L$  is the number of hyperlinks between articles. Wikipedia data were collected in middle February 2013.

Edition	Language	$N_a$	Edition	Language	$N_a$
EN	English	4212493	RU	Russian	966284
NL	Dutch	1144615	HE	Hebrew	144959
DE	German	1532978	TR	Turkish	206311
FR	French	1352825	AR	Arabic	203328
ES	Spanish	974025	FA	Persian	295696
IT	Italian	1017953	HI	Hindi	96869
PT	Portuguese	758227	MS	Malaysian	180886
EL	Greek	82563	TH	Thai	78953
DA	Danish	175228	VI	Vietnamese	594089
SV	Swedish	780872	ZH	Chinese	663485
PL	Polish	949153	KO	Korean	231959
HU	Hungarian	235212	JA	Japanese	852087

**Table 2.** List of top persons by PageRank and 2DRank for the English Wikipedia. All names are represented by article titles in the English Wikipedia.

Rank	PageRank persons	2DRank persons
1st	Napoleon	Frank Sinatra
2nd	Barack Obama	Michael Jackson
3rd	Carl Linnaeus	Pope Pius XII
4th	Elizabeth II	Elton John
5th	George W. Bush	Elizabeth II
6th	Jesus	Pope John Paul II
7th	Aristotle	Beyoncé Knowles
8th	William Shakespeare	Jorge Luis Borges
9th	Adolf Hitler	Mariah Carey
10th	Franklin D. Roosevelt	Vladimir Putin

**Table 3.** List of country code (CC), countries as birth places of historical figures, and language code (LC) for each country. LC is determined by the most spoken language in the given country. Country codes are based on country codes of Internet top-level domains and language codes are based on language edition codes of Wikipedia; WR represents all languages other than the considered 24 languages.

CC	Country	LC	CC	Country	LC	CC	Country	LC
AE	United Arab Emirates	AR	AF	Afghanistan	FA	AL	Albania	WR
AR	Argentina	ES	AT	Austria	DE	AU	Australia	EN
AZ	Azerbaijan	TR	BE	Belgium	NL	BG	Bulgaria	WR
BR	Brazil	PT	BS	Bahamas	EN	BY	Belarus	RU
CA	Canada	EN	CH	Switzerland	DE	CL	Chile	ES
CN	China	ZH	CO	Colombia	ES	CU	Cuba	ES
CY	Cyprus	EL	CZ	Czech Rep.	WR	DE	Germany	DE
DK	Denmark	DA	DZ	Algeria	AR	EG	Egypt	AR
ES	Spain	ES	FI	Finland	WR	FR	France	FR
GE	Georgia	WR	GR	Greece	EL	HK	Hong Kong	ZH
HR	Croatia	WR	HU	Hungary	HU	ID	Indonesia	WR
IE	Ireland	EN	IL	Israel	HE	IN	India	HI
IQ	Iraq	AR	IR	Iran	FA	IS	Iceland	WR
IT	Italy	IT	JP	Japan	JA	KE	Kenya	EN
KG	Kyrgyzstan	WR	KH	Cambodia	WR	KO	S. Korea	KO
KP	N. Korea	KO	KW	Kuwait	AR	KZ	Kazakhstan	WR
LB	Lebanon	AR	LT	Lithuania	WR	LV	Latvia	WR
LY	Libya	AR	MK	Macedonia	WR	MM	Myanmar	WR
MN	Mongolia	WR	MX	Mexico	ES	MY	Malaysia	MS
NL	Netherlands	NL	NO	Norway	WR	NP	Nepal	WR
NZ	New Zealand	EN	OM	Oman	AR	PA	Panama	ES
PE	Peru	ES	PK	Pakistan	HI	PL	Poland	PL
PS	State of Palestine	AR	PT	Portugal	PT	RO	Romania	WR
RS	Serbia	WR	RU	Russia	RU	SA	Saudi Arabia	AR
SD	Sudan	AR	SE	Sweden	SV	SG	Singapore	ZH
SI	Slovenia	WR	SK	Slovakia	WR	SR	Suriname	NL
SY	Syria	AR	TH	Thailand	TH	TJ	Tajikistan	WR
TN	Tunisia	AR	TR	Turkey	TR	TW	Taiwan	ZH
TZ	Tanzania	WR	UA	Ukraine	WR	UK	United Kingdom	EN
US	United States	EN	UZ	Uzbekistan	WR	VE	Venezuela	ES
VN	Vietnam	VI	XX	Unknown	WR	YE	Yemen	AR
ZA	South Africa	WR						

**Table 4.** List of global historical figures by PageRank and 2DRank for all 24 Wikipedia editions. All names are represented by the corresponding article titles in the English Wikipedia. Here,  $\Theta_A$  is the ranking score of algorithm  $A$  (3);  $N_A$  is the number of appearances of a given person in the top 100 rank for all editions.

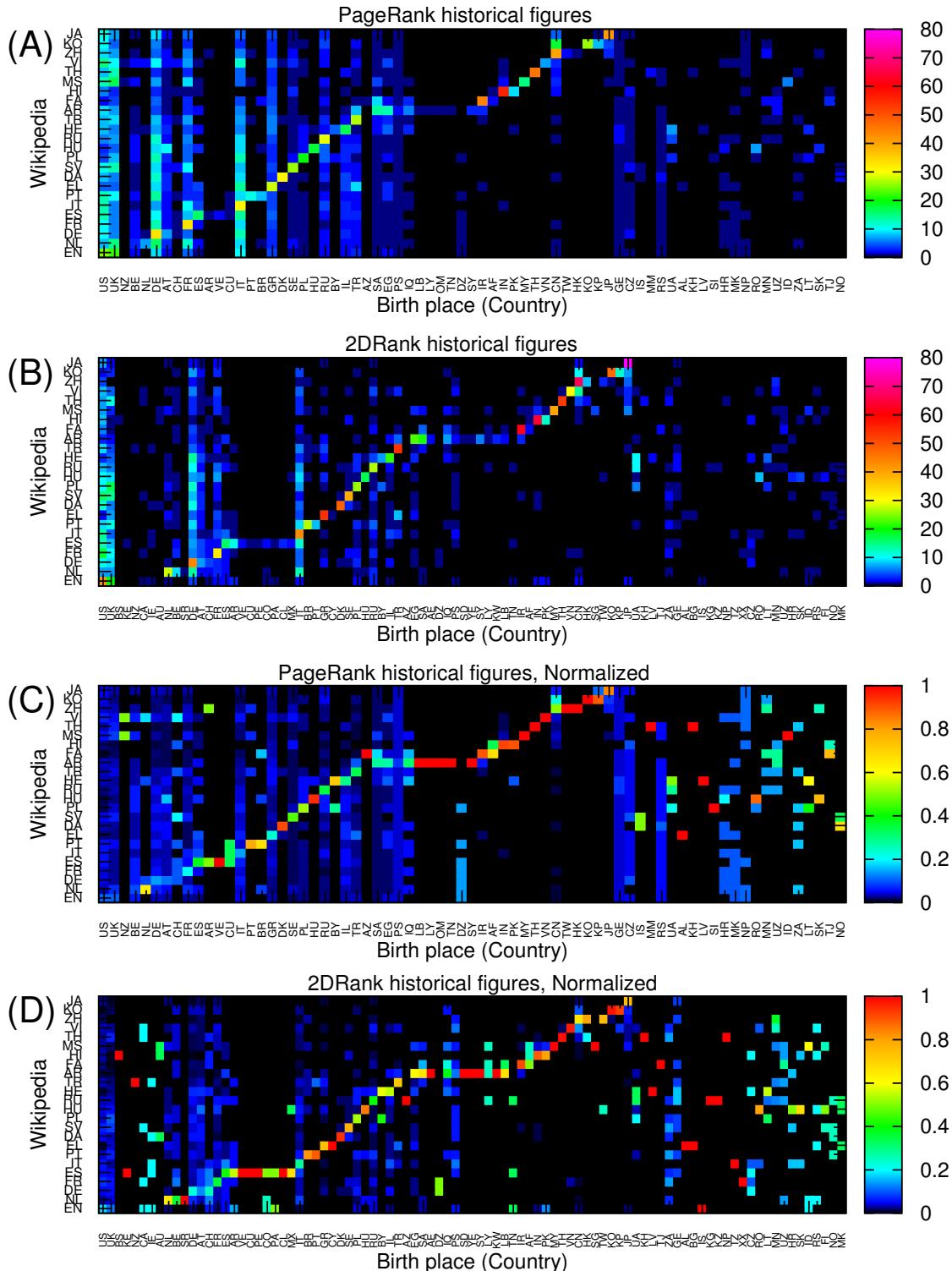
Rank	PageRank global figures	$\Theta_{PR}$	$N_A$	2DRank global figures	$\Theta_{2D}$	$N_A$
1st	Carl Linnaeus	2284	24	Adolf Hitler	1557	20
2nd	Jesus	2282	24	Michael Jackson	1315	17
3rd	Aristotle	2237	24	Madonna (entertainer)	991	14
4th	Napoleon	2208	24	Jesus	943	14
5th	Adolf Hitler	2112	24	Ludwig van Beethoven	872	14
6th	Julius Caesar	1952	23	Wolfgang Amadeus Mozart	853	11
7th	Plato	1949	24	Pope Benedict XVI	840	12
8th	William Shakespeare	1861	24	Alexander the Great	789	11
9th	Albert Einstein	1847	24	Charles Darwin	773	12
10th	Elizabeth II	1789	24	Barack Obama	754	16

**Table 5.** List of the top 10 global female historical figures by PageRank and 2DRank for all the 24 Wikipedia editions. All names are represented by article titles in the English Wikipedia. Here,  $\Theta_A$  is the ranking score of the algorithm  $A$  (Eq.3);  $N_A$  is the number of appearances of a given person in the top 100 rank for all editions. Here  $CC$  is the birth country code and  $LC$  is the language code of the given historical figure.

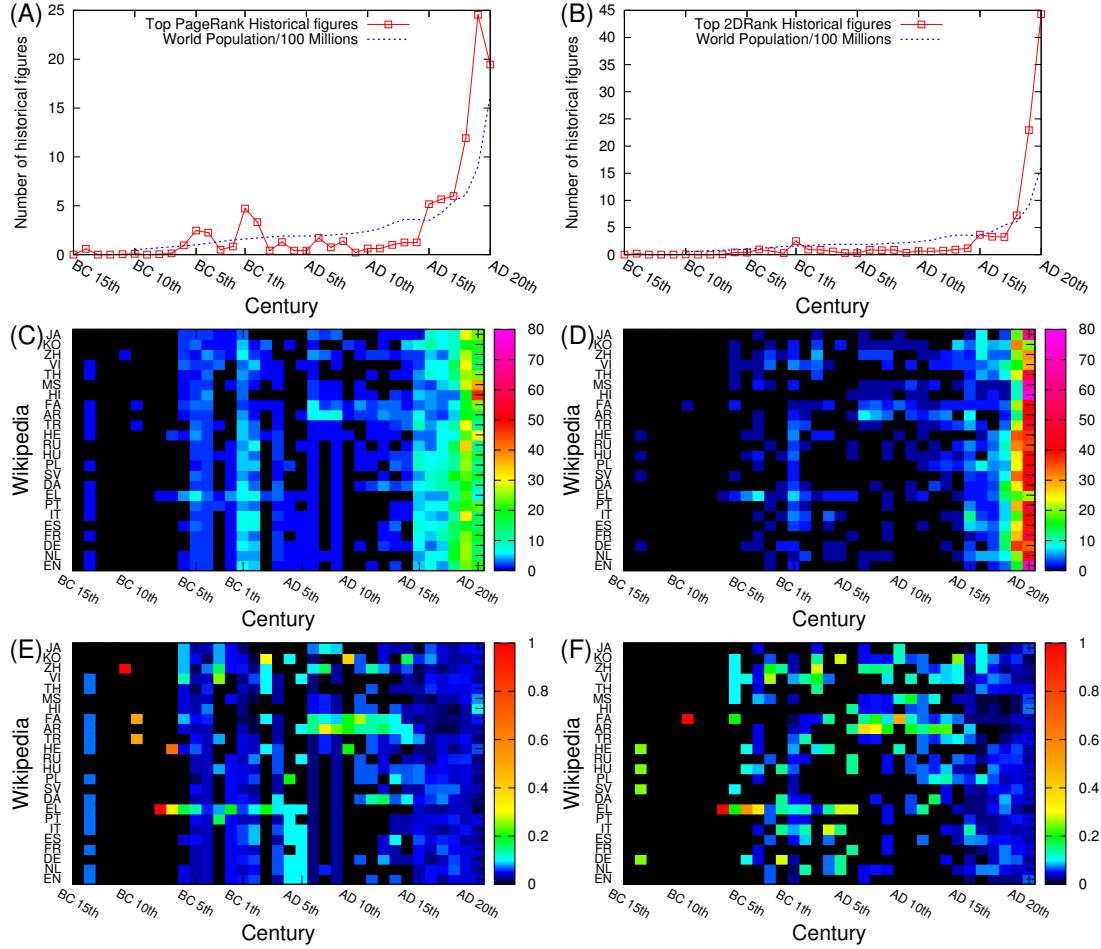
Rank	$\Theta_{PR}$	$N_A$	PageRank female figures	$CC$	Century	$LC$
1	1789	24	Elizabeth II	UK	20	EN
2	1094	17	Mary (mother of Jesus)	IL	-1	HE
3	404	12	Queen Victoria	UK	19	EN
4	234	6	Elizabeth I of England	UK	16	EN
5	128	2	Maria Theresa	AT	18	DE
6	100	1	Benazir Bhutto	PK	20	HI
7	94	1	Catherine the Great	PL	18	PL
8	91	1	Anne Frank	DE	20	DE
9	87	1	Indira Gandhi	IN	20	HI
10	86	1	Margrethe II of Denmark	DK	20	DA
Rank	$\Theta_{2D}$	$N_A$	2DRank female figures	$CC$	Century	$LC$
1	991	14	Madonna (entertainer)	US	20	EN
2	664	9	Elizabeth II	UK	20	EN
3	580	8	Mary (mother of Jesus)	IL	-1	HE
4	550	9	Queen Victoria	UK	19	EN
5	225	5	Agatha Christie	UK	19	EN
6	211	4	Mariah Carey	US	20	EN
7	206	7	Britney Spears	US	20	EN
8	200	3	Margaret Thatcher	UK	20	EN
9	191	2	Martina Navratilova	CZ	20	WR
10	175	2	Elizabeth I of England	UK	16	EN

**Table 6.** Numbers of certain historical figures for top 100 list of each language:  $N_1$  is the number of historical figures of a given language among the top 100 PageRank global historical figures;  $N_2$  is the number of historical figures of a given language among the top 100 PageRank historical figures for the given language edition;  $N_3$  is the number of historical figures of a given language among the top 100 2DRank global historical figures;  $N_4$  is the number of historical figures of a given language among the top 100 2DRank historical figures for the given language edition.

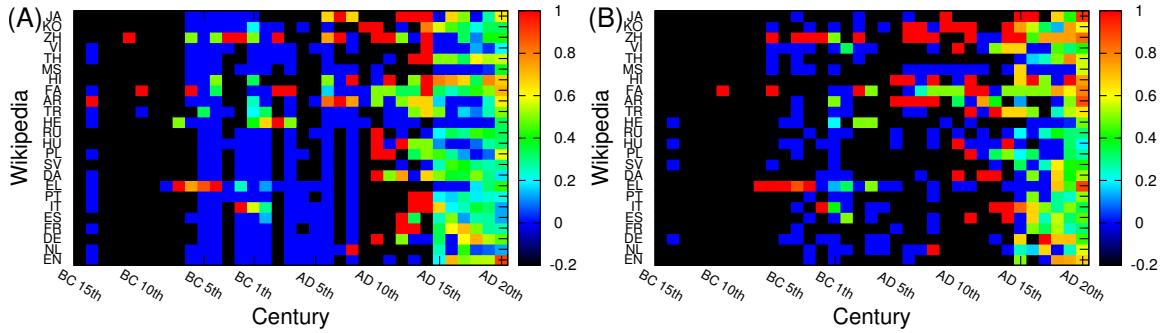
Language	$N_1$	$N_2$	$N_3$	$N_4$	Language	$N_1$	$N_2$	$N_3$	$N_4$
EN	22	47	27	64	RU	2	29	3	27
NL	2	10	4	38	HE	2	17	2	22
DE	20	41	16	55	TR	2	27	2	54
FR	8	33	3	32	AR	8	42	5	69
ES	2	20	5	39	FA	0	46	1	64
IT	11	31	9	43	HI	1	65	0	76
PT	0	19	0	35	MS	0	15	0	40
EL	5	28	2	55	TH	0	46	0	53
DA	0	31	1	48	VI	0	7	0	30
SV	1	26	1	39	ZH	5	43	6	79
PL	1	20	2	26	KO	0	34	0	59
HU	0	18	0	18	JA	0	41	4	80
WR	8	-	7	-					



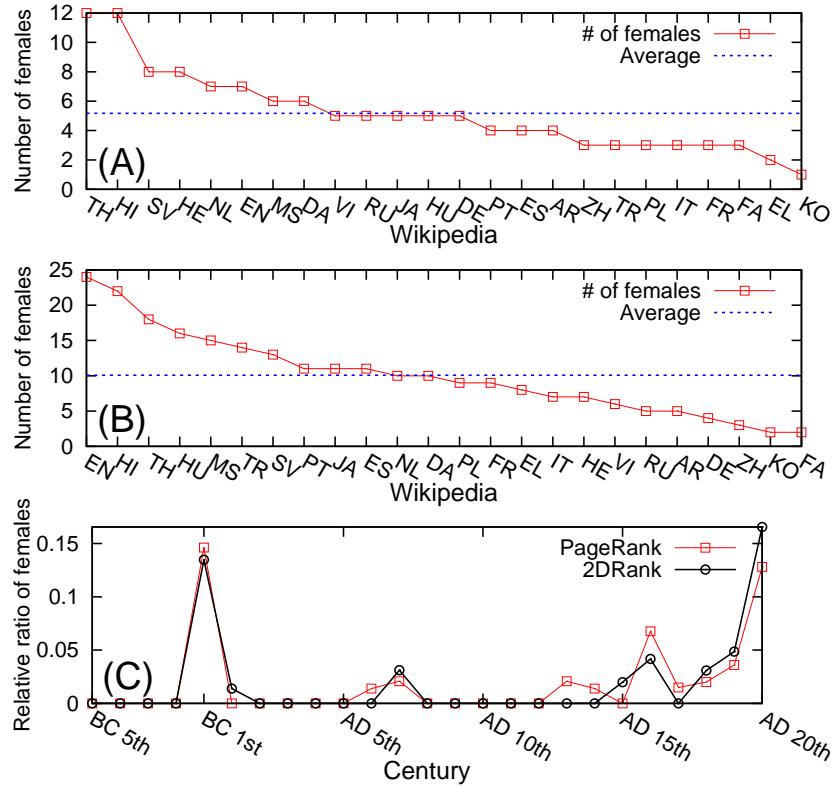
**Figure 3.** Birth place distributions over countries of top historical figures from each Wikipedia edition; two letter country codes are represented in Table 3. Panels: (A) distributions of PageRank historical figures over 71 countries for each Wikipedia edition; (B) distributions of 2DRank historical figures over 91 countries for each Wikipedia edition; (C) column normalized birth place distributions of PageRank historical figures of panel (A); (D) column normalized birth place distributions of 2DRank historical figures of panel (B).



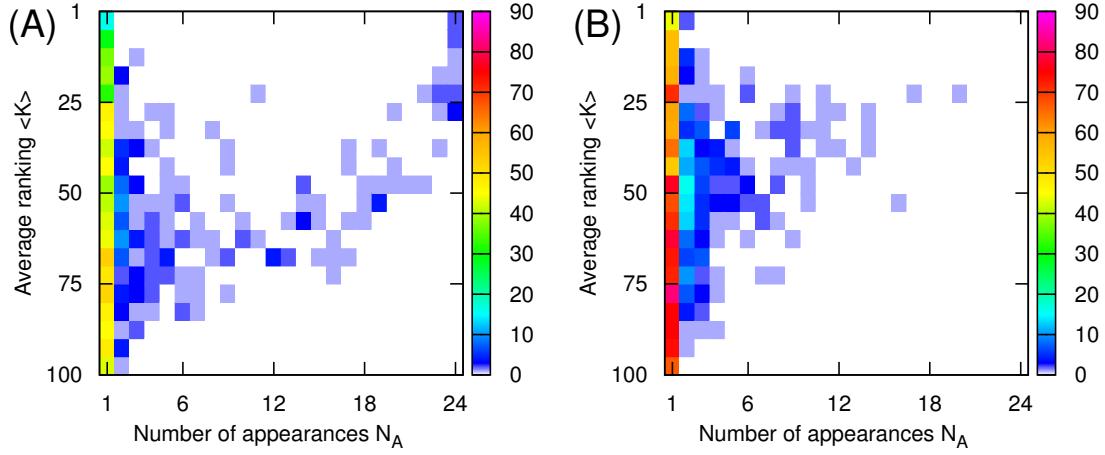
**Figure 4.** Birth date distributions of top historical figures. (A) Birth date distribution of PageRank historical figures averaged over 24 Wikipedia editions (B) Birth date distribution of 2DRank historical figures averaged over 24 Wikipedia editions (C) Birth date distributions of PageRank historical figures for each Wikipedia edition. (D) Birth date distributions of 2DRank historical figures for each Wikipedia edition. (E) Column normalized birth date distributions of PageRank historical figures for each Wikipedia edition. (F) Column normalized birth date distributions of 2DRank historical figures for each Wikipedia edition.



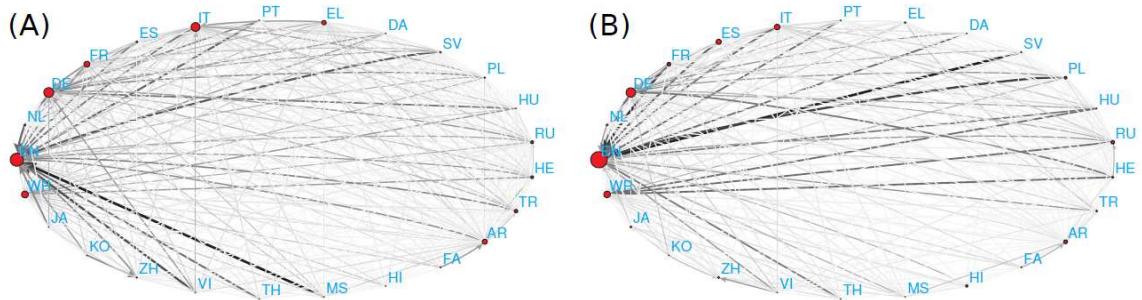
**Figure 5.** The locality property of cultures represented by the ratio  $r_{L,C} = M_{L,C}/N_{L,C}$  for each edition  $L$  and each century  $C$ . Here  $M_{L,C}$  is the number of historical figures born in countries attributed to a given language edition  $L$  at century  $C$  and  $N_{L,C}$  is the total number of historical figures in a given edition at a given century, regardless of language of their birth countries. Black color (-0.2 in the color bars) shows that there is no historical figure at all for a given edition and century; blue (0 in the color bars) shows there are some historical figures but no same language historical figures. Here (A) panel shows PageRank historical figures, and (B) panel shows 2DRank historical figures.



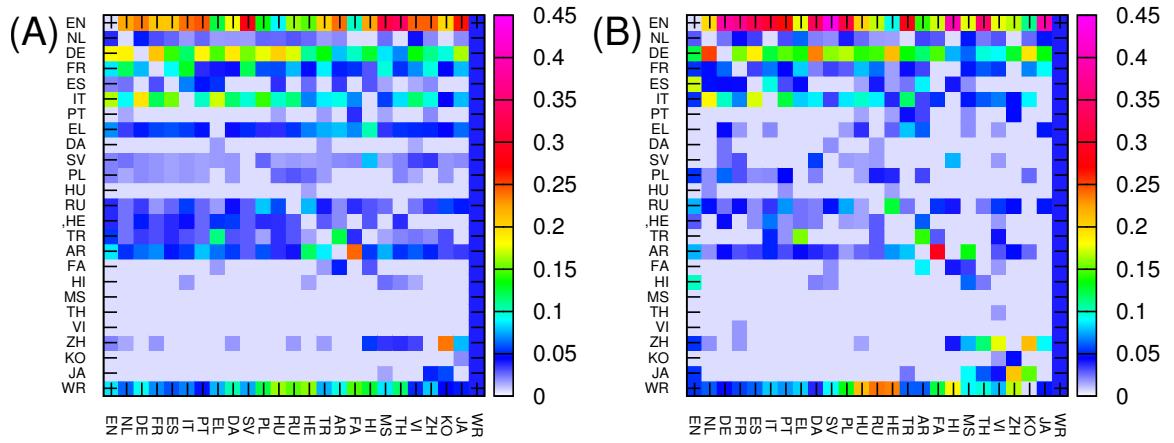
**Figure 6.** Number of females of top historical figures from each Wikipedia edition (A) Top PageRank historical figures (B) Top 2DRank historical figures. (C) The average female ratio of historical figures in given centuries across 24 Wikipedia editions.



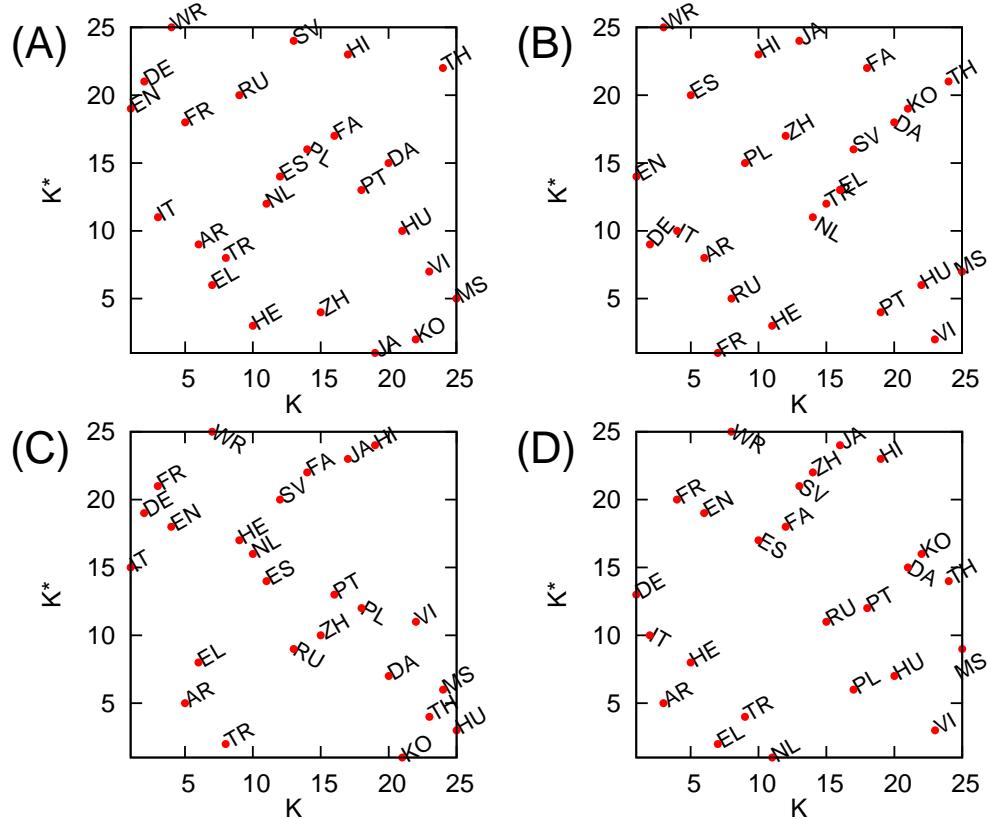
**Figure 7.** The distribution of 1045 top PageRank persons (A) and 1616 top 2DRank persons (B) as a function of number of appearances  $N_A$  of a given person and the rank  $\langle K \rangle$  of this person averaged over Wikipedia editions where this person appeared.



**Figure 8.** Network of cultures obtained from 24 Wikipedia languages and the remaining world (WR) consider (A) top PageRank historical figures and (B) 2DRank historical figures. The link width and darkness are proportional to a number of foreign historical figures quoted in top 100 of a given culture, the link direction goes from a given culture to cultures of quoted foreign historical figures, quotations inside cultures are not considered. The size of nodes is proportional to their PageRank.



**Figure 9.** Google matrix of network of cultures shown in Fig 8 respectively. The matrix elements  $G_{ij}$  are shown by color with damping factor  $\alpha = 0.85$ .



**Figure 10.** PageRank ranking versus CheiRank ranking plane of cultures with corresponding indexes  $K$  and  $K^*$  obtained from the network of cultures based on (A) all PageRank historical figures, (B) all 2DRank historical figures, (C) PageRank historical figure born before AD 19th century, and (D) 2DRank historical figure born before AD 19th century, respectively.

# SUPPORTING INFORMATION FOR:

## Interactions of cultures and top people of Wikipedia from ranking of 24 language editions

Young-Ho Eom<sup>1</sup>, Pablo Aragón<sup>2</sup>, Andreas Kaltenbrunner<sup>2</sup>, Sebastiano Vigna<sup>3</sup>, Dima L. Shepelyansky<sup>1,\*</sup>

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

<sup>2</sup> *Barcelona Media Foundation, Barcelona, Spain*

<sup>3</sup> *Dipartimento di Informatica, Università degli Studi di Milano, Italy*

\* Webpage: [www.quantware.ups-tlse.fr/dima](http://www.quantware.ups-tlse.fr/dima)

## 1 Additional data

Supplementary methods, tables, ranking lists and figures are available at  
<http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/>.

The data sets of 24 hyperlink networks are available at the same web page.

Here we present additional figures and tables for the main part of the paper.

Figure S1 shows PageRank and CheiRank probabilities for the networks of cultures shown in Fig.8.

Figure S2 shows overlap between the global list of top 100 global PageRank persons and list of Hart [23], PageRank list of English Wikipedia from [15], list of Stony-Brook [19], list of Pantheon MIT project [20].

The names of persons from top 100 missed by automatic recovery of persons are: Homer, Charles Darwin (RU PageRank); Philipp Kirkorov (RU 2DRank); Alexander the Great, Emperor Gaozu of Han, Homer (KO PageRank); Jinpyeong of Silla, Hyeonjong of Goryeo (KO 2DRank). Unfortunately, the name of Homer has been missed in the 1.1 million list of English names, other names are missed due to incompleteness and modifications of inter-language translations.

Below we give the list of global top 100 PageRank names from 24 Wikipedia editions. The names are ordered by the ranking score  $\Theta_{P,A}$  of Eq.(1). In brackets we give country of birth, century of birth, gender, and language of birth. In the same manner we also give the list of top 100 2DRank names from 24 Wikipedia editions.

We also give 24 names from global 1045 PageRank names and 40 names from 1616 global 2DRank names where a birth place language attribution differs from native language.

We also give the tables of top 10 persons in each language and also world names (tables S1 - S25) extracted from the global PageRank and 2DRank ranking lists of persons ordered by the score  $\Theta_{P,A}$  of Eq.(1).

Top 100 of global PageRank names: 1. Carl Linnaeus (SE, 18, M, SV) 2. Jesus (PS, -1, M, AR) 3. Aristotle (GR, -4, M, EL) 4. Napoleon (FR, 18, M, FR) 5. Adolf Hitler (AT, 19, M, DE) 6. Julius Caesar (IT, -1, M, IT) 7. Plato (GR, -5, M, EL) 8. William Shakespeare (UK, 16, M, EN) 9. Albert Einstein (DE, 19, M, DE) 10. Elizabeth II (UK, 20, F, EN) 11. Alexander the Great (GR, -4, M, EL) 12. Isaac Newton (UK, 17, M, EN) 13. Muhammad (SA, 6, M, AR)

14. Karl Marx (DE, 19, M, DE) 15. Joseph Stalin (GE, 19, M, WR) 16. Augustus (IT, -1, M, IT) 17. Christopher Columbus (IT, 15, M, IT) 18. Charlemagne (BE, 8, M, NL) 19. Louis XIV of France (FR, 17, M, FR) 20. George W. Bush (US, 20, M, EN) 21. Immanuel Kant (RU, 18, M, RU) 22. Barack Obama (US, 20, M, EN) 23. Mary (mother of Jesus) (IL, -1, F, HE) 24. Vladimir Lenin (RU, 19, M, RU) 25. Wolfgang Amadeus Mozart (AT, 18, M, DE) 26. Paul the Apostle (TR, 1, M, TR) 27. Charles Darwin (UK, 19, M, EN) 28. Martin Luther (DE, 15, M, DE) 29. Herodotus (TR, -5, M, TR) 30. Franklin D. Roosevelt (US, 19, M, EN) 31. Galileo Galilei (IT, 16, M, IT) 32. Pope John Paul II (PL, 20, M, PL) 33. Constantine the Great (RS, 3, M, WR) 34. Benito Mussolini (IT, 19, M, IT) 35. Cicero (IT, -2, M, IT) 36. Ren Descartes (FR, 16, M, FR) 37. Saint Peter (IL, 1, M, HE) 38. Ludwig van Beethoven (DE, 18, M, DE) 39. George Washington (US, 18, M, EN) 40. Moses (EG, -14, M, AR) 41. Johann Sebastian Bach (DE, 17, M, DE) 42. Bill Clinton (US, 20, M, EN) 43. Leonardo da Vinci (IT, 15, M, IT) 44. Johann Wolfgang von Goethe (DE, 18, M, DE) 45. Gautama Buddha (NP, -6, M, WR) 46. Winston Churchill (UK, 19, M, EN) 47. John F. Kennedy (US, 20, M, EN) 48. Charles V, Holy Roman Emperor (BE, 15, M, NL) 49. Pope Benedict XVI (DE, 20, M, DE) 50. Richard Nixon (US, 20, M, EN) 51. Sigmund Freud (CZ, 19, M, WR) 52. Ronald Reagan (US, 20, M, EN) 53. Abraham Lincoln (US, 19, M, EN) 54. Saddam Hussein (IQ, 20, M, AR) 55. Ptolemy (EG, 1, M, AR) 56. Richard Wagner (DE, 19, M, DE) 57. Diocletian (HR, 3, M, WR) 58. Queen Victoria (UK, 19, F, EN) 59. Napoleon III (FR, 19, M, FR) 60. Charles de Gaulle (FR, 19, M, FR) 61. Mao Zedong (CN, 19, M, ZH) 62. William Herschel (DE, 18, M, DE) 63. Michael Jackson (US, 20, M, EN) 64. Justinian I (MK, 5, M, WR) 65. Augustine of Hippo (DZ, 4, M, AR) 66. Ali (SA, 7, M, AR) 67. Jean-Jacques Rousseau (CH, 18, M, DE) 68. Ernst Haeckel (DE, 19, M, DE) 69. Pliny the Elder (IT, 1, M, IT) 70. Pope Gregory XIII (IT, 16, M, IT) 71. Confucius (CN, -6, M, ZH) 72. Henry VIII of England (UK, 15, M, EN) 73. Thomas Jefferson (US, 18, M, EN) 74. Francisco Franco (ES, 19, M, ES) 75. Georg Wilhelm Friedrich Hegel (DE, 18, M, DE) 76. Pierre Andr Latreille (FR, 18, M, FR) 77. Pope Paul VI (IT, 19, M, IT) 78. Gottfried Wilhelm Leibniz (DE, 17, M, DE) 79. Chiang Kai-shek (CN, 19, M, ZH) 80. John Herschel (UK, 18, M, EN) 81. Elizabeth I of England (UK, 16, F, EN) 82. J. R. R. Tolkien (ZA, 19, M, WR) 83. Socrates (GR, -5, M, EL) 84. Genghis Khan (MN, 12, M, WR) 85. Qin Shi Huang (CN, -3, M, ZH) 86. Umar (SA, 6, M, AR) 87. Philip II of Spain (ES, 16, M, ES) 88. Frederick the Great (DE, 18, M, DE) 89. Johannes Kepler (DE, 16, M, DE) 90. Emperor Wu of Han (CN, -2, M, ZH) 91. Friedrich Nietzsche (DE, 19, M, DE) 92. Plutarch (GR, 1, M, EL) 93. Thomas Edison (US, 19, M, EN) 94. Max Weber (DE, 19, M, DE) 95. Dante Alighieri (IT, 13, M, IT) 96. Ashoka (IN, -4, M, HI) 97. Tacitus (FR, 1, M, FR) 98. Ernst Mayr (DE, 20, M, DE) 99. Jean-Baptiste Lamarck (FR, 18, M, FR) 100. Elvis Presley (US, 20, M, EN).

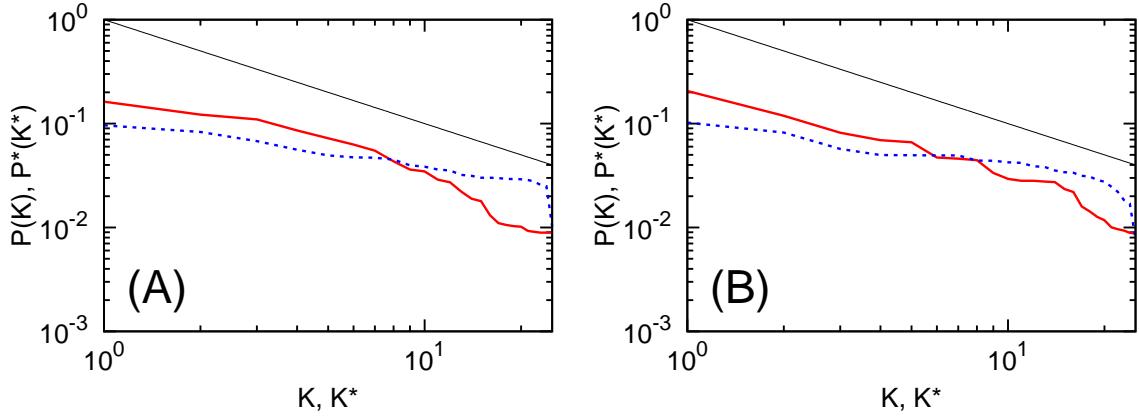
Top 100 of global 2DRank names: 1. Adolf Hitler (AT, 19, M, DE) 2. Michael Jackson (US, 20, M, EN) 3. Madonna (entertainer) (US, 20, F, EN) 4. Jesus (PS, -1, M, AR) 5. Ludwig van Beethoven (DE, 18, M, DE) 6. Wolfgang Amadeus Mozart (AT, 18, M, DE) 7. Pope Benedict XVI (DE, 20, M, DE) 8. Alexander the Great (GR, -4, M, EL) 9. Charles Darwin (UK, 19, M, EN) 10. Barack Obama (US, 20, M, EN) 11. Johann Sebastian Bach (DE, 17, M, DE) 12. Napoleon (FR, 18, M, FR) 13. Pope John Paul II (PL, 20, M, PL) 14. Julius Caesar (IT, -1, M, IT) 15. Elizabeth II (UK, 20, F, EN) 16. Albert Einstein (DE, 19, M, DE) 17. Augustus (IT, -1, M, IT) 18. Bob Dylan (US, 20, M, EN) 19. Leonardo da Vinci (IT, 15, M, IT) 20. Mary (mother of Jesus) (IL, -1, F, HE) 21. Charlemagne (BE, 8, M, NL) 22. William Shakespeare

(UK, 16, M, EN) 23. Elvis Presley (US, 20, M, EN) 24. Queen Victoria (UK, 19, F, EN) 25. John Lennon (UK, 20, M, EN) 26. George Frideric Handel (DE, 17, M, DE) 27. J. R. R. Tolkien (ZA, 19, M, WR) 28. Muhammad (SA, 6, M, AR) 29. Joseph Stalin (GE, 19, M, WR) 30. Karl Marx (DE, 19, M, DE) 31. Benito Mussolini (IT, 19, M, IT) 32. Franklin D. Roosevelt (US, 19, M, EN) 33. Michael Schumacher (DE, 20, M, DE) 34. Paul McCartney (UK, 20, M, EN) 35. Stephen King (US, 20, M, EN) 36. Henry VIII of England (UK, 15, M, EN) 37. Tokugawa Ieyasu (JP, 16, M, JA) 38. Edgar Allan Poe (US, 19, M, EN) 39. Martin Luther (DE, 15, M, DE) 40. David Bowie (UK, 20, M, EN) 41. Pope Pius XII (IT, 19, M, IT) 42. Alfred Hitchcock (UK, 19, M, EN) 43. Friedrich Nietzsche (DE, 19, M, DE) 44. Vladimir Putin (RU, 20, M, RU) 45. Christopher Columbus (IT, 15, M, IT) 46. Elton John (UK, 20, M, EN) 47. Carl Linnaeus (SE, 18, M, SV) 48. Michelangelo (IT, 15, M, IT) 49. Raphael (IT, 15, M, IT) 50. Roger Federer (CH, 20, M, DE) 51. Cao Cao (CN, 2, M, ZH) 52. Vincent van Gogh (NL, 19, M, NL) 53. Frdric Chopin (PL, 19, M, PL) 54. Steven Spielberg (US, 20, M, EN) 55. Rembrandt (NL, 17, M, NL) 56. Ali (SA, 7, M, AR) 57. Richard Wagner (DE, 19, M, DE) 58. Che Guevara (AR, 20, M, ES) 59. Nelson Mandela (ZA, 20, M, WR) 60. Isaac Asimov (RU, 20, M, RU) 61. Jules Verne (FR, 19, M, FR) 62. Toyotomi Hideyoshi (JP, 16, M, JA) 63. Winston Churchill (UK, 19, M, EN) 64. Paul the Apostle (TR, 1, M, TR) 65. Hirohito (JP, 20, M, JA) 66. 14th Dalai Lama (CN, 20, M, ZH) 67. Franz Liszt (AT, 19, M, DE) 68. Genghis Khan (MN, 12, M, WR) 69. Otto von Bismarck (DE, 19, M, DE) 70. Saint Peter (IL, 1, M, HE) 71. Charlie Chaplin (UK, 19, M, EN) 72. Liu Bei (CN, 2, M, ZH) 73. Oda Nobunaga (JP, 16, M, JA) 74. Suleiman the Magnificent (TR, 15, M, TR) 75. Cyrus the Great (IR, -6, M, FA) 76. George W. Bush (US, 20, M, EN) 77. Agatha Christie (UK, 19, F, EN) 78. Carl Friedrich Gauss (DE, 18, M, DE) 79. Louis XIV of France (FR, 17, M, FR) 80. Saddam Hussein (IQ, 20, M, AR) 81. Pablo Picasso (ES, 19, M, ES) 82. Mariah Carey (US, 20, F, EN) 83. Hans Christian Andersen (DK, 19, M, DA) 84. Plato (GR, -5, M, EL) 85. Britney Spears (US, 20, F, EN) 86. Rafael Nadal (ES, 20, M, ES) 87. George Harrison (UK, 20, M, EN) 88. Margaret Thatcher (UK, 20, F, EN) 89. Jorge Luis Borges (AR, 19, M, ES) 90. Salvador Dal (ES, 20, M, ES) 91. Peter the Great (RU, 17, M, RU) 92. Giuseppe Verdi (IT, 19, M, IT) 93. Sigmund Freud (CZ, 19, M, WR) 94. Qin Shi Huang (CN, -3, M, ZH) 95. Kangxi Emperor (CN, 17, M, ZH) 96. Martina Navratilova (CZ, 20, F, WR) 97. Charles V, Holy Roman Emperor (BE, 15, M, NL) 98. Zhuge Liang (CN, 2, M, ZH) 99. Constantine the Great (RS, 3, M, WR) 100. Muammar Gaddafi (LY, 20, M, AR)

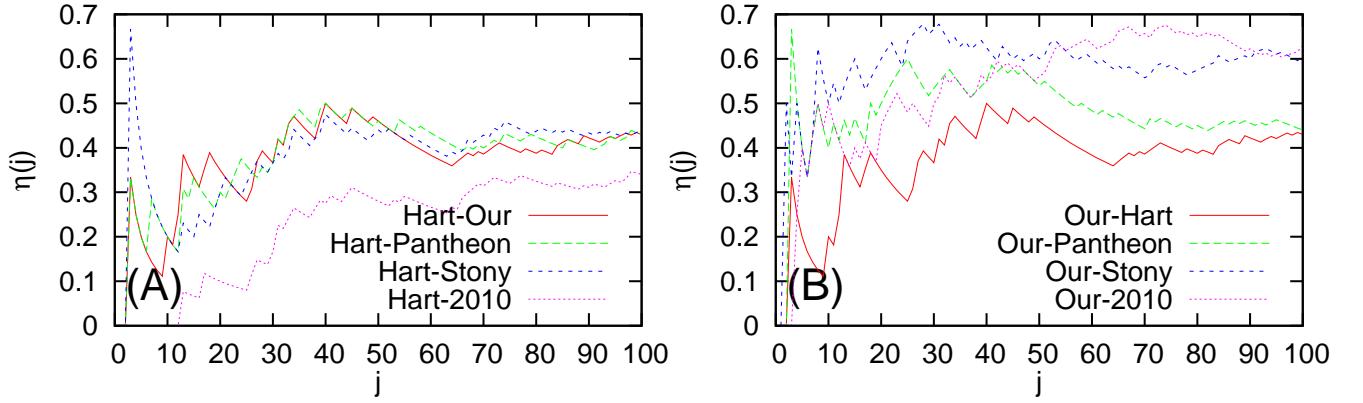
List of 36 names from the global PageRank list of 1045 names where the birth place in modern geography of countries differs from native language: Jesus (PS AR), Charlemagne (Belgium NL), Immanuel Kant (Russia RU), Moses (Egypt AR), Catherine the Great (Poland PL), Mustafa Kemal Atatürk (Greece EL), Bhumibol Adulyadej (USA EN), Christian V of Denmark (Germany DE), Józef Pilsudski (Litvania WR), Christian IX of Denmark (Germany DE), Philip V of Spain (France FR), Giuseppe Garibaldi (France FR), Muhammad al-Idrisi (Spain ES), Charles XIV John of Sweden (France FR), Leonid Brezhnev (Ukraine WR), George I of Greece (Denmark DA), Juan Carlos I of Spain (Italy IT), Leon Trotsky (Ukraine WR), Golda Meir (Ukraine WR), Valéry Giscard d'Estaing (Germany DE), Magnus IV of Sweden (Norway WR), Christian I of Denmark (Germany DE), Yitzhak Ben-Zvi (Ukraine WR), Mikhail Bulgakov (Ukraine WR); Kim Jong-il (Russia RU). Lee Myung-bak (Japan JA), Jangsu of Goguryeo (China ZH); Galyani Vadhana (UK EN), Abhisit Vejjajiva (UK EN); Matthias Corvinus (Romania WR), Ferenc Kaz-

inczy (Romania WR), György Kulin (Romania WR), Gabriel Bethlen (Romania WR), Endre Ady (Romania WR), János Arany (Romania WR), Béla Bartók (Romania WR).

List of 53 names from the global 2DRank list of 1616 names where the birth place in modern geography of countries differs from native language: Jesus (PS AR), Charlemagne (BE NL), Isaac Asimov (RU RU), Paul the Apostle (TR TR), Peter Paul Rubens (DE DE), Catherine the Great (PL PL), Julian (emperor) (TR TR), Józef Piłsudski (LT WR), Muhammad Ali of Egypt (GR EL), Juan Carlos I of Spain (IT IT), Shmuel Yosef Agnon (UA WR), Saint Joseph (PS AR), Golda Meir (UA WR), Baibars (UA WR), Levi Eshkol (UA WR), Augustine of Hippo (DZ AR), Yitzhak Ben-Zvi (UA WR), Natan Yonatan (UA WR), Edward Rydz-mig (UA WR), Immanuel Kant (RU RU), Pyotr Stolypin (DE DE), Czesław Niemen (BY RU), Moses (EG AR), Albert Camus (DZ AR), Leonid Brezhnev (UA WR), Aharon Barak (LT WR), George Orwell (IN HI), Sergei Korolev (UA WR), Garry Kasparov (AZ TR), Ibn 'Abd al-Barr (ES ES), Georges Simenon (BE NL), Ryszard Kapuściński (BY RU), Mihly Munkácsy (UA WR), Juliusz Słowacki (UA WR), Tadeusz Kościuszko (BY RU), John McCain (PA ES), Maurice, Prince of Orange (DE DE), Zbigniew Herbert (UA WR), Leon Trotsky (UA WR), Charles XIV John of Sweden (FR FR), Lee Myung-bak (JA JA), Jangsu of Goguryeo (CN ZH), Gwanggaeto the Great (CN ZH); Galyani Vadhana (UK EN), Abhisit Vejjajiva (UK EN); Matthias Corvinus (RO WR), Károly Kós (RO WR), László Németh (RO WR), Sándor Körösi Csoma (RO WR), János Bolyai (RO WR), György Kulin (RO WR), Ferenc Kazinczy (RO WR), Béla Bartók (RO WR).



**Figure S1.** Dependence of probabilities of PageRank  $P$  (red) and CheiRank  $P^*$  (blue) on corresponding indexes  $K$  and  $K^*$ . The probabilities are obtained from the network shown in Fig.7 for corresponding panels (A), (B). The straight lines indicate the Zipf's law  $P \sim 1/K$ ;  $P^* \sim 1/K^*$ .



**Figure S2.** Dependence of fraction  $\eta$  of overlaped persons on rank index of person  $j$ . (A) Comparison is done of present study (“our”), PageRank list of English Wikipedia of [15] (“2010”), Stony-Brook list [19], Pantheon MIT project [20] in respect to Hart top 100 list. (B) Same as in (A) but comparison is done in respect to present study.

**Table S1.** List of local historical figures for EN category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1861	William Shakespeare	1315	Michael Jackson
2	1789	Elizabeth II	991	Madonna (entertainer)
3	1756	Isaac Newton	773	Charles Darwin
4	1173	George W. Bush	754	Barack Obama
5	1101	Barack Obama	664	Elizabeth II
6	932	Charles Darwin	624	Bob Dylan
7	910	Franklin D. Roosevelt	556	William Shakespeare
8	656	George Washington	555	Elvis Presley
9	596	Bill Clinton	550	Queen Victoria
10	564	Winston Churchill	541	John Lennon

**Table S2.** List of local historical figures for NL category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1476	Charlemagne	569	Charlemagne
2	556	Charles V, Holy Roman Emperor	297	Vincent van Gogh
3	83	Maurice Maeterlinck	294	Rembrandt
4	81	William I of the Netherlands	190	Charles V, Holy Roman Emperor
5	78	Beatrix of the Netherlands	138	Beatrix of the Netherlands
6	61	Baruch Spinoza	98	Baruch Spinoza
7	61	Rembrandt	94	Hugo Claus
8	51	Wilhelmina of the Netherlands	91	Johan Cruyff
9	47	Juliana of the Netherlands	76	Louis Couperus
10	39	Christiaan Huygens	75	Pierre Cuypers

**Table S3.** List of local historical figures for DE category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1)

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	2112	Adolf Hitler	1557	Adolf Hitler
2	1847	Albert Einstein	872	Ludwig van Beethoven
3	1730	Karl Marx	853	Wolfgang Amadeus Mozart
4	996	Wolfgang Amadeus Mozart	840	Pope Benedict XVI
5	925	Martin Luther	733	Johann Sebastian Bach
6	700	Ludwig van Beethoven	651	Albert Einstein
7	610	Johann Sebastian Bach	540	George Frideric Handel
8	570	Johann Wolfgang von Goethe	465	Karl Marx
9	528	Pope Benedict XVI	446	Michael Schumacher
10	417	Richard Wagner	344	Martin Luther

**Table S4.** List of local historical figures for FR category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	2208	Napoleon	720	Napoleon
2	1207	Louis XIV of France	268	Jules Verne
3	724	René Descartes	221	Louis XIV of France
4	397	Napoleon III	168	Giuseppe Garibaldi
5	385	Charles de Gaulle	146	Denis Diderot
6	260	Pierre André Latreille	144	François Mitterrand
7	167	Tacitus	127	Napoleon III
8	165	Jean-Baptiste Lamarck	121	Nicolas Sarkozy
9	157	Molière	113	Claudius
10	112	Francis I of France	112	Henry IV of France

**Table S5.** List of local historical figures for ES category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	276	Francisco Franco	285	Che Guevara
2	195	Philip II of Spain	216	Pablo Picasso
3	119	Pablo Picasso	206	Rafael Nadal
4	82	Lionel Messi	199	Jorge Luis Borges
5	74	Charles III of Spain	198	Salvador Dalí
6	72	Teresa of Ávila	178	Hadrian
7	71	Miguel de Cervantes	105	Shakira
8	70	Ferdinand VII of Spain	100	Francisco Goya
9	66	Alfonso X of Castile	95	Juan Perón
10	65	Ferdinand I, Holy Roman Emperor	94	Augusto Pinochet

**Table S6.** List of local historical figures for IT category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1952	Julius Caesar	689	Julius Caesar
2	1662	Augustus	647	Augustus
3	1476	Christopher Columbus	616	Leonardo da Vinci
4	893	Galileo Galilei	464	Benito Mussolini
5	758	Benito Mussolini	339	Pope Pius XII
6	753	Cicero	330	Christopher Columbus
7	594	Leonardo da Vinci	326	Michelangelo
8	292	Pliny the Elder	322	Raphael
9	288	Pope Gregory XIII	197	Giuseppe Verdi
10	250	Pope Paul VI	172	Galileo Galilei

**Table S7.** List of local historical figures for PT category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	91	Getúlio Vargas	109	Ronaldo
2	83	Cristiano Ronaldo	100	Getúlio Vargas
3	74	John VI of Portugal	92	Juscelino Kubitschek
4	71	Luiz Inácio Lula da Silva	91	Rubens Barrichello
5	70	Pedro I of Brazil	90	Joaquim Maria Machado de Assis
6	67	Ferdinand Magellan	89	Fernando Henrique Cardoso
7	66	Maria I of Portugal	82	Luís de Camões
8	64	John I of Portugal	80	José Saramago
9	63	Pedro II of Brazil	79	John VI of Portugal
10	62	Juscelino Kubitschek	77	Oscar Niemeyer

**Table S8.** List of local historical figures for EL category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	2237	Aristotle	789	Alexander the Great
2	1949	Plato	207	Plato
3	1771	Alexander the Great	167	Aristotle
4	213	Socrates	108	Pericles
5	178	Plutarch	100	Mustafa Kemal Atatürk
6	153	Mustafa Kemal Atatürk	98	Eleftherios Venizelos
7	123	Sophocles	95	Andreas Papandreou
8	93	Aeschylus	94	Muhammad Ali of Egypt
9	86	Euripides	94	Ioannis Kapodistrias
10	84	Ioannis Kapodistrias	93	Plutarch

**Table S9.** List of local historical figures for DA category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	99	Tycho Brahe	210	Hans Christian Andersen
2	94	Ole Rømer	98	Margrethe II of Denmark
3	93	Christian IV of Denmark	95	N. F. S. Grundtvig
4	86	Margrethe II of Denmark	92	Sren Kierkegaard
5	85	Hans Christian Andersen	89	Christian IV of Denmark
6	84	Frederick IV of Denmark	88	Hans Christian Ørsted
7	80	Frederick II of Denmark	86	Anders Fogh Rasmussen
8	78	John Louis Emil Dreyer	84	Carl Nielsen
9	77	Christian VII of Denmark	83	Christian X of Denmark
10	76	Frederick III of Denmark	82	Niels Bohr

**Table S10.** List of local historical figures for SV category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	2284	Carl Linnaeus	326	Carl Linnaeus
2	125	August Strindberg	151	Ingmar Bergman
3	98	Alfred Nobel	146	Charles XII of Sweden
4	94	Gustav I of Sweden	116	Astrid Lindgren
5	93	Gustav III of Sweden	100	August Strindberg
6	86	Charles XII of Sweden	98	Carl XVI Gustaf of Sweden
7	82	Gustavus Adolphus of Sweden	92	Evert Taube
8	72	Carl XVI Gustaf of Sweden	89	Jan Myrdal
9	71	Charles XI of Sweden	88	Carl Jonas Love Almqvist
10	67	Charles IX of Sweden	83	Gustav I of Sweden

**Table S11.** List of local historical figures for PL category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	864	Pope John Paul II	693	Pope John Paul II
2	94	Catherine the Great	296	Frédéric Chopin
3	88	David Ben-Gurion	135	Catherine the Great
4	80	Casimir III the Great	98	David Ben-Gurion
5	72	Nathan Alterman	95	Bolesaw III Wrymouth
6	69	Lech Walesa	94	Andrzej Wajda
7	66	Lech Kaczyński	93	Nathan Alterman
8	63	Frédéric Chopin	91	Gerhart Hauptmann
9	60	Henryk Sienkiewicz	88	Anton Denikin
10	58	Sigismund I the Old	83	Lech Kaczyński

**Table S12.** List of local historical figures for HU category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	93	János Szentágothai	100	Stephen I of Hungary
2	91	Stephen I of Hungary	99	Sándor Petöfi
3	87	Lajos Kossuth	94	Kati Kovács
4	86	Miklós Réthelyi	93	Miklós Horthy
5	80	Béla IV of Hungary	92	Attila József
6	79	Louis I of Hungary	89	Sándor Weöres
7	75	Sándor Petöfi	86	Theodor Herzl
8	67	Miklós Horthy	83	Lajos Kossuth
9	56	Theodor Herzl	81	Miklós Radnóti
10	53	Andrew II of Hungary	77	János Kodolányi

**Table S13.** List of local historical figures for RU category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1123	Immanuel Kant	334	Vladimir Putin
2	1022	Vladimir Lenin	274	Isaac Asimov
3	156	Peter the Great	198	Peter the Great
4	130	Mikhail Gorbachev	171	Vladimir Lenin
5	101	Pyotr Ilyich Tchaikovsky	127	Yuri Gagarin
6	97	Yuri Gagarin	109	Igor Stravinsky
7	97	Alexander Pushkin	100	Menachem Begin
8	91	Vladimir Putin	99	Dmitri Mendeleev
9	89	Nikita Khrushchev	96	Aleksander Griboyedov
10	88	Alexander II of Russia	95	Shimon Peres

**Table S14.** List of local historical figures for HE category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1094	Mary (mother of Jesus)	580	Mary (mother of Jesus)
2	724	Saint Peter	240	Saint Peter
3	138	John the Baptist	171	John the Baptist
4	99	Yitzhak Rabin	99	Saint George
5	95	Yigal Amir	99	Yitzhak Rabin
6	84	Josephus	96	Ariel Sharon
7	81	Tom Segev	92	Benjamin Netanyahu
8	75	Ariel Sharon	85	Ehud Barak
9	65	Benjamin Netanyahu	82	Roni Dalumi
10	54	Herod the Great	79	Moshe Dayan

**Table S15.** List of local historical figures for TR category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	973	Paul the Apostle	252	Paul the Apostle
2	925	Herodotus	231	Suleiman the Magnificent
3	133	Strabo	172	Mehmed the Conqueror
4	117	Mehmed the Conqueror	169	Selim I
5	106	Suleiman the Magnificent	142	Abdul Hamid II
6	96	Abdul Hamid II	111	Julian (emperor)
7	93	Pausanias (geographer)	90	Recep Tayyip Erdoğan
8	83	İsmet İnönü	87	Adnan Menderes
9	79	Selim I	85	Lucian
10	79	Hesiod	84	Blent Ecevit

**Table S16.** List of local historical figures for AR category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	2282	Jesus	943	Jesus
2	1735	Muhammad	499	Muhammad
3	629	Moses	291	Ali
4	426	Saddam Hussein	219	Saddam Hussein
5	424	Ptolemy	181	Muammar Gaddafi
6	329	Augustine of Hippo	143	Hannibal
7	328	Ali	128	Saladin
8	196	Umar	128	Anwar Sadat
9	147	Anwar Sadat	117	Hosni Mubarak
10	134	Euclid	108	Yasser Arafat

**Table S17.** List of local historical figures for FA category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	110	Zoroaster	229	Cyrus the Great
2	101	Darius I	99	Zoroaster
3	100	Mahmoud Ahmadinejad	98	Mohammad Reza Pahlavi
4	97	Mohammad Reza Pahlavi	97	Mohammad Khatami
5	96	Rez Shh	96	Mir-Hossein Mousavi
6	94	Cyrus the Great	95	Ruhollah Khomeini
7	92	Ferdowsi	94	Naser al-Din Shah Qajar
8	90	Ruhollah Khomeini	93	Ali Khamenei
9	89	Naser al-Din Shah Qajar	92	Mohammad Mosaddegh
10	86	Mohammad Khatami	91	Ardashir I

**Table S18.** List of local historical figures for HI category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	168	Ashoka	126	Ashoka
2	106	Mahatma Gandhi	108	Akbar
3	100	Benazir Bhutto	99	Indira Gandhi
4	91	Vikramditya	98	Mahadevi Varma
5	90	Shivaji	96	Sanjeev Kumar
6	89	Jawaharlal Nehru	93	Amitabh Bachchan
7	88	Akbar	91	Premchand
8	87	Indira Gandhi	90	Dayananda Saraswati
9	86	Adi Shankara	89	Jaishankar Prasad
10	85	Vishnu Prabhakar	86	Adi Shankara

**Table S19.** List of local historical figures for MS category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	96	Mahathir Mohamad	100	Mahathir Mohamad
2	85	Najib Razak	99	Najib Razak
3	84	P. Ramlee	98	Anwar Ibrahim
4	81	Tunku Abdul Rahman	93	Mizan Zainal Abidin of Terengganu
5	79	Abdullah Ahmad Badawi	92	Sudirman Arshad
6	77	Muhyiddin Yassin	91	Tunku Abdul Rahman
7	74	Abdul Razak Hussein	90	Siti Nurhaliza
8	62	Anwar Ibrahim	89	Abdullah Ahmad Badawi
9	58	Hussein Onn	88	Abdul Taib Mahmud
10	37	Mizan Zainal Abidin of Terengganu	84	P. Ramlee

**Table S20.** List of local historical figures for TH category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	100	Chulalongkorn	100	Sirindhorn
2	97	Vajiravudh	98	Sirikit
3	96	Mongkut	97	Thaksin Shinawatra
4	94	Buddha Yodfa Chulaloke	94	Taksin
5	92	Nangklao	91	Pridi Banomyong
6	91	Thaksin Shinawatra	90	Yingluck Shinawatra
7	90	Damrong Rajanubhab	88	Srinagarindra
8	89	Taksin	86	Samak Sundaravej
9	88	Plaek Phibunsongkhram	82	Vajiralongkorn
10	87	Prajadhipok	80	Chao Keo Naovarat

**Table S21.** List of local historical figures for VI category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	91	Ho Chi Minh	98	Ho Chi Minh
2	71	Ngo Dinh Diem	97	Gia Long
3	62	Minh Mng	96	Minh Mng
4	46	Gia Long	94	Nguyen Hue
5	44	Bo i	86	Le Loi
6	22	Le Loi	84	Tran Hung Dao
7	15	Nhat Linh	83	Vo Nguyen Giap
8	N/A	N/A	82	Tu Duc
9	N/A	N/A	81	Le Thánh Tông
10	N/A	N/A	80	Trung Sisters

**Table S22.** List of local historical figures for ZH category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	375	Mao Zedong	306	Cao Cao
2	285	Confucius	243	14th Dalai Lama
3	244	Chiang Kai-shek	234	Liu Bei
4	197	Qin Shi Huang	192	Qin Shi Huang
5	186	Emperor Wu of Han	191	Kangxi Emperor
6	135	Cao Cao	188	Zhuge Liang
7	129	Hongwu Emperor	179	Qianlong Emperor
8	119	Qianlong Emperor	154	Mao Zedong
9	119	Kangxi Emperor	147	Hongwu Emperor
10	94	Sun Yat-sen	146	Sun Yat-sen

**Table S23.** List of local historical figures for KO category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	100	Gojong of the Korean Empire	114	Gojong of the Korean Empire
2	98	Kim Il-sung	106	Kim Il-sung
3	95	Sejong the Great	100	Park Chung-hee
4	94	Park Chung-hee	99	Kim Dae-jung
5	93	Taejong of Joseon	97	Roh Moo-hyun
6	92	Syngman Rhee	95	Sejong the Great
7	91	Yeongjo of Joseon	94	Taejo of Goryeo
8	90	Kim Dae-jung	93	Kim Young-sam
9	89	Seonjo of Joseon	92	Jeongjo of Joseon
10	86	Taejo of Joseon	90	Syngman Rhee

**Table S24.** List of local historical figures for JA category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	154	Toyotomi Hideyoshi	346	Tokugawa Ieyasu
2	153	Tokugawa Ieyasu	266	Toyotomi Hideyoshi
3	108	Hirohito	252	Hirohito
4	97	Oda Nobunaga	233	Oda Nobunaga
5	86	Emperor Meiji	140	Junichiro Koizumi
6	81	Minamoto no Yoritomo	131	Shinzō Abe
7	76	Junichiro Koizumi	112	Tsunku
8	73	Emperor Tenmu	106	Emperor Meiji
9	70	Natsume Sōseki	100	Koxinga
10	69	Akihito	97	Osamu Tezuka

**Table S25.** List of local historical figures for WR category. Here  $\Theta_A$  is the ranking score of the algorithm  $A$  defined in Eq.(1).

	$\Theta_A$	PageRank local figures	$\Theta_A$	2DRank local figures
1	1686	Joseph Stalin	529	J. R. R. Tolkien
2	842	Constantine the Great	477	Joseph Stalin
3	564	Gautama Buddha	276	Nelson Mandela
4	506	Sigmund Freud	241	Genghis Khan
5	405	Diocletian	195	Sigmund Freud
6	351	Justinian I	191	Martina Navratilova
7	219	J. R. R. Tolkien	186	Constantine the Great
8	203	Genghis Khan	173	Justinian I
9	138	Avicenna	127	Nikola Tesla
10	129	Rumi	123	Kublai Khan