Two-dimensional ranking of Wikipedia articles

A.O.Zhirov
Novosibirsk State University, 630090 Novosibirsk, Russia

O.V.Zhirov
Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia

D.L.Shepelyansky
Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France
(Dated: June 22, 2010)

The Library of Babel, described by Jorge Luis Borges, stores an enormous amount of information. The Library exists *ab aeterno*. Wikipedia, a free online encyclopaedia, becomes a modern analogue of such a Library. Information retrieval and ranking of Wikipedia articles become the challenge of modern society. We analyze the properties of two-dimensional ranking of all Wikipedia English articles and show that it gives their reliable classification with rich and nontrivial features. Detailed studies are done for countries, universities, personalities, physicists, chess players, Dow-Jones companies and other categories.

PACS numbers: 89.75.Fb,89.75.Hc,89.20.Hh

The *Encyclopédie* [1] accumulates the available human knowledge making it accessible to all *citoyennes*. In this way the *Encyclopédie* becomes one of the most powerful catalysts of modern development of science and society [2]. This process of knowledge transfer becomes enormously accelerated with the appearance of Wikipedia [3], a free online encyclopaedia, which current size overcomes 6 millions English entries [3, 4]. Wikipedia comes close to Encyclopaedia Britanica [5] in terms of the accuracy of its science entries [6] overcoming the later by far in an enormous amount of available information. The classification and ranking of this information becomes the great challenge. The statistical analysis of directed network generated by links between Wikipedia articles [7–9] established their scale-free properties showing certain similarities with the World Wide Web and other scale-free networks [10–14]. Here we apply a two-dimensional ranking algorithm, proposed recently [15], and classify all Wikipedia articles in English by their degree of communication and popularity. This ranking allows to select articles in a new way, giving e.g. more preference to communicative and artistic sides of a personality compared to popularity and political aspects stressed by the PageRank algorithm [16]. With a good reliability the ranking of Wikipedia articles reproduces the well established classifications of countries [17], universities [18], personalities [19], physicists, chess players, Dow-Jones companies [20] and other categories.

At August 18, 2009 we downloaded from [4] the latest English Wikipedia snapshot and by crawling determined links between all $N_{tot} = 6855098$ entries. We keep for our study only entries which are linked with other entries and eliminate categories from consideration. After that we obtain a directed network of $N = 3282258$ articles. The multiplicity of links from one article to another is taken into account. The distributions of ingoing and outgoing links are well described by a power law $w_{in,out}(k) \propto 1/k^{\mu_{in,out}}$ with the exponents $\mu_{in} = 2.09 \pm 0.04$ and $\mu_{out} = 2.76 \pm 0.06$ (see Supplementary Material) being in a good agreement with the previous studies of Wikipedia [7–9] and the World Wide Web (WWW) [14, 21, 22].

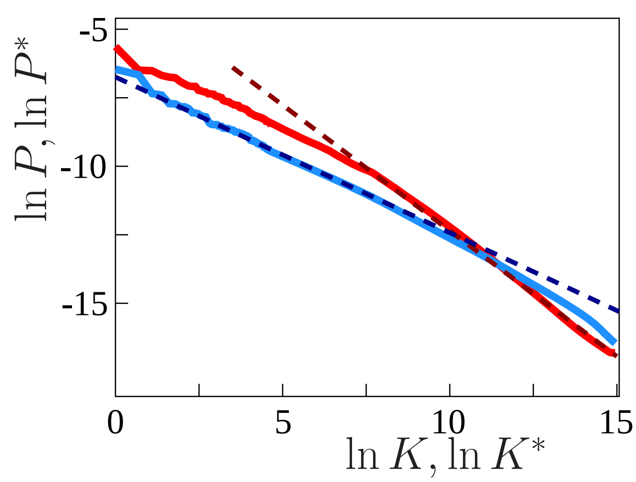

FIG. 1: Dependence of probability of PageRank P (red curve) and CheiRank P^* (blue curve) on the corresponding rank indexes K and K^*. The straight dashed lines show the power law dependence with the slope $\beta = 0.92; 0.57$ respectively, corresponding to $\beta = 1/(\mu_{in,out} - 1)$.

Due to the similarity with the WWW it is natural to construct the Google matrix of Wikipedia using the procedure described in [14, 16]:

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

where the matrix S is obtained by normalizing to unity all columns of the adjacency matrix, and replacing columns with zero elements by $1/N$, N being the network size. The damping parameter α in the WWW context describes the probability to jump to any node for a random surfer. For Wikipedia this parameter can describe the probability to modify an article that affects the overall ranking. The value $\alpha = 0.85$ gives a good classification [14] for WWW and thus we also use this value here. The matrix G belongs to the class of Perron-Frobenius operators [14], its largest eigenvalue is $\lambda = 1$ and other eigenvalues have $|\lambda| \leq \alpha$. The right eigenvector at $\lambda = 1$ gives the probability $P(i)$ to find a random surfer at site i and is called the PageRank. Once the PageRank is found, Wikipedia articles are sorted by decreasing $P(i)$, the article rank in this index $K(i)$ reflects the article relevance. The PageRank dependence on K is well described by a power law $P(K) \propto 1/K^{\beta}$ (see Fig. 1) with $\beta \approx 0.9$ that is consistent with the relation $\beta = 1/(\mu_{in} - 1)$ corresponding to the proportionality of PageRank to its in-degree w_{in}.

In addition to the PageRank, following the approach proposed recently for Linux Kernel procedure call network [15], we also consider the ranking of articles obtained from the Google matrix G^* built from the adjacency matrix with inverted direction of links using $\alpha = 0.85$. The eigenvector $P^*(i)$ of G^* at $\lambda = 1$, introduced in [15], allows to *chercher l'information* in a new way and we call it CheiRank. It gives additional ranking of articles in decreasing order of $P^*(i)$ with rank index $K^*(i)$. Our data, shown in Fig. 1, give a power law dependence $P^* \propto 1/K^{*\beta}$ with $\beta \approx 0.6$ corresponding to the relation $\beta = 1/(\mu_{out} - 1)$ similar to the one found for the PageRank. Both ranks are normalized to unity.

While PageRank characterizes a degree of knowledge and popularity of a given site i, CheiRank highlights its communication, influence and connectivity degree. These ranks have certain analogy to authorities and hubs appearing in the HITS algorithm [23] but the HITS is query dependent while the rank probabilities $P(i)$ and $P^*(i)$ classify all sites of the network. In this way the ranking of sites becomes two-dimensional (2D). The density distribution $W(K, K^*) = dN_i/dKdK^*$ of Wikipedia articles N_i in the plane $(\ln K, \ln K^*)$ is shown in Fig. 2 ($\sum_i W(K, K^*) = 1$). In contrast to the Linux network [15] the Wikipedia network has a maximum of density along the line $\ln K^* \approx 5 + (2 \ln K)/3$ that shows a strong correlation between P and P^*. Indeed, for Wikipedia we find the correlator $\kappa = N \sum_i P(i)P^*(i) - 1 = 4.08$ being much larger than for the Linux network where $\kappa \approx -0.05$. Due to these correlations the distribution $W(K, K^*)$ is absolutely different from the one given by independent product probabilities for PageRank index K and CheiRank index K^* which is homogeneous along lines $\ln(K^*) + \ln(K) = const$ (Fig 3d). The value of correlator κ for Wikipedia is comparable with the one of Cambridge University network ($\kappa = 3.79$ [15]) but the probability distributions are different. We note that for Wikipedia the distribution $W(K, K^*)$, taken along a given line in $(\ln K, \ln K^*)$, is log-normal and has a Gaussian form with a certain width as it is shown in Fig. 3.

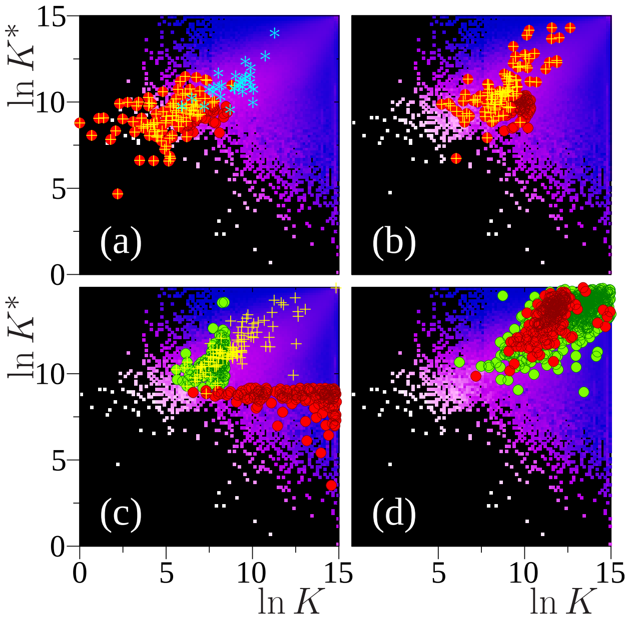

FIG. 2: Density distribution $W(K, K^*) = dN_i/dKdK^*$ of Wikipedia articles in the plane of PageRank and CheiRank indexes $(\ln K, \ln K^*)$ shown by color with blue for minimum and white for maximum (black for zero); (a) red points show top 100 countries from 2DRank, yellow pluses show top 100 countries from SJR country documentation ranking for 1996-2008 [17], cyan stars mark 30 Dow-Jones companies [20]; (b) red points show top 100 universities from 2DRank, yellow pluses show top 100 universities from Shanghai ranking of 2009 [18]; (c) green/red points show top 100 personalities from PageRank/CheiRank, yellow pluses show top 100 personalities from [19]; (d) green and red points show ranks of 758 physicists, red points mark 193 Nobel laureates.

The difference between PageRank and CheiRank is clearly seen from the names of articles with highest rank (detailed data for all ranks and all categories considered are given in Supplementary Material and [24]). For PageRank we have 1. *United States*, 2. *United Kingdom*, 3. *France* while for CheiRank we find 1. *Portal:Contents/Outline of knowledge/Geography and places*, 2. *List of state leaders by year*, 3. *Portal:Contents/Index/Geography and places*. Clearly PageRank selects first articles on a broadly known subject with a large number of ingoing links while CheiRank selects first highly communicative articles with many outgoing links. Since the articles are distributed in 2D they can be ranked in various ways corresponding to projec-

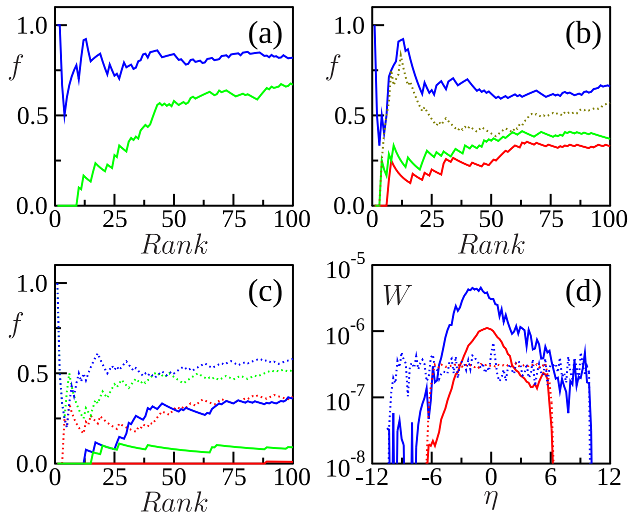

FIG. 3: Dependence of overlap fraction f of top Wikipedia articles on (a) SJR country documentation rank for 1996-2008 [17], (b) Shanghai university rank of 2009 [18], (c) Hart's personality rank [19] for Wikipedia PageRank K (blue), 2DRank K_2 (green) and CheiRank (red). Order of top 100 countries is the same for 2DRank and CheiRank. The values of f for Nobel laureate physicists are shown in panel (c) by dotted curves with the same color choice (see text for details). We also show f for the rank given by advanced Google search of a word *university* in the Wikipedia English domain at May 19, 2010 (gray dotted curve in panel (b)). Panel (d) shows W density dependence on a parametric variable η which parametrizes the straight line $\ln K = x_0 + \eta/2$, $\ln K^* = x_0 - \eta/2$ with $x_0 = 10$ (blue curve)), 12 (red curve); dotted curves show densities from independent probabilities.

tion of 2D set on a line. The horizontal and vertical lines correspond to PageRank and CheiRank. We also introduce 2DRank $K_2(i)$ constructed by increasing $K \to K+1$ and increasing 2DRank index $K_2(i)$ by one if the new entry is present in the list of first $K^* < K$ entries of CheiRank, then the one unit step is done in K^* and K_2 is increased by one if the new entry is present in the list of first $K < K^*$ entries of CheiRank. The articles with highest 2DRank are 1. *India*, 2. *Singapore*, 3. *Pakistan*. Thus, these articles are most known/popular and most communicative at the same time. The ranking of Wikipedia articles by PageRank and HITS authorities has been discussed in a literature (see e.g. [25, 26]) but 2D analysis was never done before.

To understand the properties of three ranks K, K_2, K^* in a better way we consider the three main categories of articles about countries, universities and personalities. The location of 100 top ranks taken according to [17–19] respectively are shown in the Wikipedia rank plane $(\ln K, \ln K^*)$ in Fig. 2a,b,c. In average the points are distributed along the band of maximal density of W in this plane. The same is valid for Dow-Jones companies (Fig. 2a) and physicists (Fig. 2d) taken from the Wikipedia List of physicists (with few additions). We also show 100 top countries (Fig. 2a), 100 top universities (Fig. 2b) according to 2DRank and 100 top personalities according to PageRank and CheiRank (Fig. 2c).

We discuss the main features of three ranks K, K_2, K^* starting from countries. The first three countries are 1. *United States*, 2. *United Kingdom*, 3. *France* for PageRank and 1. *India*, 2. *Singapore*, 3. *Pakistan* for 2DRank and CheiRank. To determine in a quantitative way how accurate is this ranking we introduce the overlap fraction f defined as the relative number of same entries inside the first K_s entries of selected rank of countries [17] and first K_s countries according to K, K_2, K^* ranking. The dependence $f(K_s)$ is shown in Fig. 3a. We see that the Wikipedia PageRank reproduces in average about 80% of top countries selected by [17]. CheiRank and 2DRank do not reproduce the initial values of rank [17] but reach around 70% at maximum $K_s = 100$. Indeed, CheiRank and 2DRank place at the top countries which are not so influential as United States but which are strongly connective due to historical (e.g. Egypt is at position 13) or tourist reasons (e.g. Thailand at , Malaysia at 7).

According to Wikipedia the top universities are 1.*Harvard University*, 2. *University of Oxford*, 3. *University of Cambridge* in PageRank; 1. *Columbia University*, 2. *University of Florida*, 3. *Florida State University* in 2DRank and CheiRank. The overlap fraction $f(K_s)$ for Shanghai university ranking [18] is shown in Fig 3b. The Wikipedia PageRank reproduces in average around 70% of ranking [18] that is about 10% higher than gives the advanced Google search. CheiRank and 2DRank at maximum give around 25% and 35%. These ranks highlight connectivity degree of universities that leads to appearance of significant number of arts, religious and military specialized colleges (12% and 13%) while PageRank has only 1% of them. They introduce also a larger number of relatively small universities. We argue that such colleges and universities keeps links to their alumni in a significantly better way that increases these ranks.

For personalities the Wikipedia ranking gives 1. *Napoleon I of France*, 2. *George W. Bush*, 3. *Elizabeth II of the United Kingdom* for PageRank; 1.*Michael Jackson*, 2. *Frank Lloyd Wright*, 3. *David Bowie* for 2DRank; 1. *Kasey S. Pipes*, 2. *Roger Calmel*, 3. *Yury G. Chernavsky* for CheiRank. The overlap fraction $f(K_s)$ for Hart's personality ranking [19] is shown in Fig. 3c. Even for the PageRank it is at maximum 35% being around 10% for 2DRank and almost zero for CheiRank. We attribute this to a very broad distribution of personalities in 2D plane (Fig. 2c) and a large variety of human activities which we classify by 5 main categories: politics, religion, arts, science, sport. For top 100 PagRank personalities we have for these categories: 58, 10, 17, 15, 0 respectively. Clearly PageRank overestimates the significance of politicians. For 2DRank we find respectively 24, 5, 62, 7, 2. Thus this rank highlights artistic sides of human activity. For CheiRank we have 15, 1, 52, 16, 16 so that the dominant contribution comes from arts, science and sport. The interesting property of this rank is that it selects many composers, singers, writers, actors. As an interesting feature of CheiRank we note that among scientists it selects those who are not so much known to a broad public but who discovered new objects, e.g. George Lyell who discovered many Australian butterflies or Nikolai Chernykh who discovered many asteroids. CheiRank also selects persons active in several categories of human activity.

When a human activity is fixed in a more precise way then the Wikipedia ranking gives a rather reliable ordering. For example, for 758 physicists (see Fig. 2d) we find at the top: 1. *Aristotle*, 2. *Albert Einstein*, 3. *Isaac Newton* from PageRank; 1. *Albert Einstein*, 2. *Nikola Tesla*, 3. *Benjamin Franklin* from 2DRank; 1. *Hubert Reeves*, 2. *Shen Kuo*, 3. *Stephen Hawking* from CheiRank. It is clear that PageRank gives most known, 2DRank gives most known and active in other areas, CheiRank gives those who are known and contribute to popularization of science. For physicists, who lived in the time of Nobel prize and could get it, we can determine the overlap fraction f as a relative number of Nobel laureates at a given rank value K, K_2, K^* (see dotted curves in Fig. 3c with color choice as for full curves). The data give high average values of $f \approx 0.5$ for PageRank and 2DRank, and $f \approx 0.25$ for CheiRank. We note that the number of Nobel prizes is not so large and even very notable physicists remained without it (e.g. Thomas Edison, Nikola Tesla, Alexander Graham Bell are those from the top of PageRank who remained without the prize). Hence, the prediction level of $f \approx 0.5$ can be considered as rather high. The ranking of 193 Nobel laureate physicists give 1. *Albert Einstein*, 2. *Enrico Fermi*, 3. *Richard Feynman* from PageRank; 1. *Albert Einstein*, 2. *Richard Feynman*, 3. *Werner Heisenberg* from 2DRank and CheiRank.

Thus Wikipedia ranking can be applied to various categories giving rather good results without any fitting. For example, for Dow-Jones companies we have 1. *Microsoft*, 2. *IBM*, 3. *The Walt Disney Company* from PageRank and 1. *Cisco Systems*, 2. *The Walt Disney Company*, 3. *Microsoft* from CheiRank. For chess players we find 1. *Garry Kasparov*, 2. *Bobby Fischer*, 3. *Alexander Alekhine* from K and 1. *Bobby Fischer*, 2. *Alexander Alekhine*, 3. *Wilhelm Steinitz* from K^*.

On the basis of presented results we conclude that the ranking of Wikipedia articles allows to rank human knowledge in a rather reliable way. The 2D ranking highlights the properties of articles in a new rich and fruitful manner. We think that such type of 2D ranking will find further useful applications for various complex networks including the WWW.

Wikipedia becomes the largest library of human knowledge. "The Library exists *ab aeterno*" declared Jorge Luis Borges [27]. Thus, the ranking of this enormous amount of knowledge becomes a formidable challenge and we think that the 2D ranking will play for this task a useful and important role.

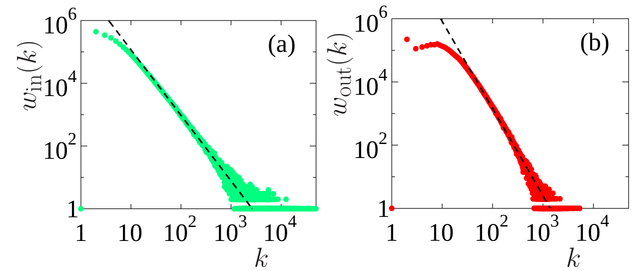

FIG. S-1: Distribution $w_{in,out}(k)$ of ingoing (a) and outgoing (b) links for $N = 3282258$ Wikipedia English articles. The straight dashed fit line shows the slope with $\mu_{in} = 2.09 \pm 0.04$ (a) and $\mu_{out} = 2.76 \pm 0.06$ (b).

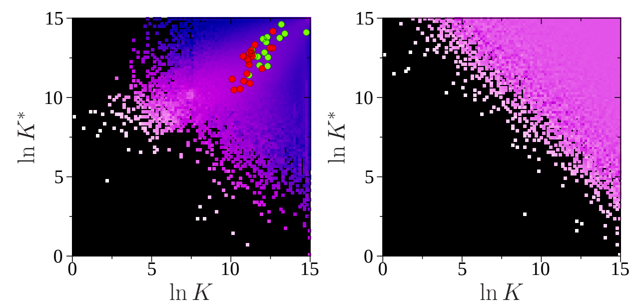

FIG. S-2: Density distribution of Wikipedia articles $W(K, K^*) = dN_i/dKdK^*$ in the plane of PageRank and CheiRank indexes $(\ln K, \ln K^*)$, 34 green and red points show distribution of chess players, world champions are marked in red (left panel). Right panel shows density distribution of $N = 3282258$ articles obtained with independent probability distributions of P and P^* given by the dependence of Fig. 1. Color choice is as in Fig. 2.

SUPPLEMENTARY MATERIAL

The distribution $w_{in,out}(k)$ of ingoing and outgoing links k for Wikipedia English articles is shown in Fig. S-1.

Density distribution of Wikipedia articles $W(K, K^*) = dN_i/dKdK^*$ is shown in Fig. S-2a for $\alpha = 0.85$. Density is computed over equidistant grid in $(\ln K, \ln K^*)$ with 100×100 cells, color shows average value of W in each cell, the normalization condition is $\sum_{K,K^*} W(K, K^*) = 1$. The densities obtained from the product of independent probabilities of P and P^* generated by the distributions of Fig. 1 give very different density $W(K, K^*)$ with W being homogeneous along lines $\ln K^* = -\ln K + const$. The distribution of 34 chess players in the plane $(\ln K, \ln K^*)$ is also shown in Fig. S-2.

Dependencies of the correlator κ on damping parameters α and α^* are shown in Fig. S-3 (here α^* is the

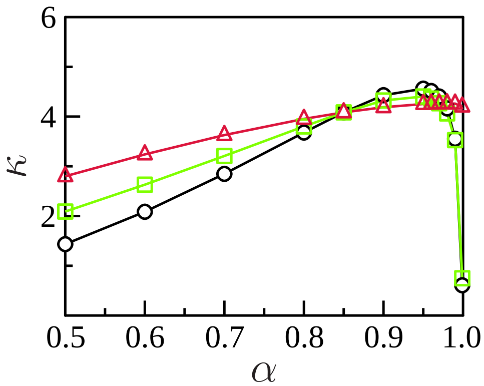

FIG. S-3: Dependence of correlator κ on α. Black circles, green squares and red triangles correspond to cases ($\alpha = \alpha^*$), (fixed $\alpha = 0.85$ and $0.5 \leq \alpha^* \leq 1$) and (fixed $\alpha^* = 0.85$ and $0.5 \leq \alpha \leq 1$), respectively. In the main part of the paper we use only the case $\alpha = \alpha^* = 0.85$.

damping parameter for the Google matrix G^* with inverted direction of links).

Below we give the listings of Wikipedia articles in their rank order according to the three ranks discussed in the paper for the categories discussed. More detailed data are available at [24].

L1) Top 20 Wikipedia articles:

PageRank: 1. United States, 2. United Kingdom, 3. France, 4. Germany, 5. England, 6. Canada, 7. World War II, 8. Australia, 9. India, 10. Japan, 11. English Language, 12.Italy, 13. Association football, 14. 2007, 15. London, 16. Poland, 17. Animal, 18. List of sovereign states, 19. 2008, 20. New York City.

2DRank: 1. India, 2. Singapore, 3. Pakistan, 4. Brazil, 5. Columbia University, 6. Thailand, 7. Amphibian, 8. Chile, 9. Fossil, 10. Malaysia, 11. Yorkshire, 12. New York Yankees, 13. Virginia, 14. Iceland, 15. California, 16. England, 17. New Jersey, 18. Michigan, 19. Hong Kong, 20. Municipalities of Switzerland.

CheiRank: 1. Portal:Contents/Outline of knowledge/Geography and places, 2. List of state leaders by year, 3. Portal:Contents/Index/Geography and places, 4. Lists of country-related topics, 5. Portal:Mathematics/Index, 6. List of United Kingdom locations, 7. Navy Office of Community Outreach, 8. Portal:Spaceflight/Status, 9. List of Tachinidae genera and species, 10. Village Development Committee, 11. Tachinidae, 12. Navy Weeks, 13. Lists of birds by region, 14. Outline of Africa, 15. List of Bulbophyllum species, 16. Bulbophyllum, 17. Index of Thailand-related articles, 18. Portal:Middle-earth/Pages, 19. Portal:Trains/Anniversaries, 20. Portal:Contents/Outline of knowledge/History and events.

L2) Top 20 articles in category countries:

PageRank: 1. United States, 2. United Kingdom, 3. France, 4. Germany, 5. Canada, 6. Australia, 7. India, 8. Japan, 9. Italy, 10. Poland, 11. Spain, 12. Russia, 13. Netherlands, 14. Brazil, 15. New Zealand, 16. Sweden, 17. Romania, 18. Switzerland, 19. Mexico, 20. Norway.

2DRank and CheiRank give the same order: 1. India, 2. Singapore, 3. Pakistan, 4. Brazil, 5. Thailand, 6. Chile, 7. Malaysia, 8. Iceland, 9. Hong Kong. 10. Canada, 11. Peru, 12. Afghanistan, 13. Egypt, 14. Armenia, 15. Argentina, 16. United Kingdom, 17. Colombia, 18. Finland, 19. Benin, 20. Bolivia.

United States are at rank 41. Top 100 countries have the same order in 2DRank and CheiRank. We attribute this to a rather sparce distance in CheiRank index between countries. SJR country documentation rank for 1996-2008 is available at [17].

L3) Top 20 articles in category universities:

PageRank: 1. Harvard University, 2. University of Oxford, 3. University of Cambridge, 4. Columbia University, 5. Yale University, 6. Massachusetts Institute of Technology, 7. Stanford University, 8. University of California, Berkeley, 9. Princeton University, 10. Cornell University, 11. University of Chicago, 12. University of Michigan, 13. University of California, Los Angeles, 14. University of Pennsylvania, 15. New York University, 16. University of Texas at Austin, 17. University of Toronto, 18. University of Southern California, 19. University of Virginia, 20. University of Florida.

2DRank: 1. Columbia University, 2. University of Florida, 3. Florida State University, 4. University of California, Berkeley, 5. Northwestern University, 6. Brown University, 7. University of Southern California, 8. Carnegie Mellon University, 9. Massachusetts Institute of Technology, 10. University of Michigan, 11. Georgetown University, 12. Juilliard School, 13. University of Pittsburgh, 14. Amherst College, 15. Cornell University, 16. Durham University, 17. Rutgers University, 18. Monash University, 19. The University of Western Ontario, 20. University of Toronto.

CheiRank: 1. Columbia University, 2. University of Florida, 3. Florida State University, 4. Brooklyn College, 5. Amherst College, 6. The University of Western Ontario, 7. University of Sheffield, 8. University of California, Berkeley, 9. Northwestern University, 10. Northeastern University, 11. Brown University, 12. Queen's University of Belfast (Northern Ireland Parliament constituency), 13. Fairfield University, 14. University of Southern California, 15. Carnegie Mellon University, 16. Grambling State University, 17. Massachusetts Institute of Technology, 18. University of Michigan, 19. Georgetown University, 20. University of Pittsburgh.

Rank given by advanced Google search in Wikipedia English domain at May 19, 2010: 1. Princeton University, 2. Columbia University, 3. University of Oxford, 4. Stanford University, 5. Harvard University, 6. University of Cambridge, 7. Cornell University, 8. University of California, Berkeley, 9. Yale University, 10. University of Virginia, 11. Northwestern University, 12. University of Chicago, 13. Brown University, 14. University of Michigan, 15. Rutgers University, 16. University of Washington, 17. Indiana University, 18. University of Minnesota, 19. Howard University, 20. Leiden University.

Shanghai university rank of 2009 is available at [18].

L4) Top 20 articles in category personalities:

PageRank: 1. Napoleon I of France, 2. George W. Bush, 3. Elizabeth II of the United Kingdom, 4. William Shakespeare, 5. Carl Linnaeus, 6. Adolf Hitler, 7. Aristotle, 8. Bill Clinton, 9. Franklin D. Roosevelt, 10. Ronald Reagan, 11. Barack Obama, 12. Richard Nixon, 13. George Washington, 14. Joseph Stalin, 15. Abraham Lincoln, 16. John F. Kennedy, 17. Muhammad, 18. Winston Churchill, 19. Henry VIII of England, 20. Alexander the Great.

2DRank: 1. Michael Jackson, 2. Frank Lloyd Wright, 3. David Bowie, 4. Hillary Rodham Clinton, 5. Charles Darwin, 6. Stephen King, 7. Richard Nixon, 8. Isaac Asimov, 9. Frank Sinatra, 10. Elvis Presley, 11. Edward Elgar, 12. Stephen Sondheim, 13. Agatha Christie, 14. Pope John Paul II, 15. Robert A. Heinlein, 16. Adolf Hitler, 17. Madonna (entertainer), 18. Ozzy Osbourne, 19. John McCain, 20. Jesus.

CheiRank: 1. Kasey S. Pipes, 2. Roger Calmel, 3. Yury G. Chernavsky, 4. Josh Billings (pitcher), 5. George Lyell, 6. Landon Donovan, 7. Marilyn C. Solvay, 8. Matt Kelley, 9. Johann Georg Hagen, 10. Chikage Oogi, 11. Bobbie Vaile, 12. Rosie Malek-Yonan, 13. Blythe McGarvie, 14. Djoko Hardono, 15. Cristina Bella, 16. Sid Deuce, 17. Joey Hamilton, 18. Kiki Dee, 19. Carlos Francis, 20. Percy Jewett Burrell.

Hart's personality rank is available at [19] and at http://www.adherents.com/adh_influ.html .

L5) Top 20 articles in category physicists:

PageRank: 1. Aristotle, 2. Albert Einstein, 3. Isaac Newton, 4. Thomas Edison, 5. Benjamin Franklin, 6. Gottfried Leibniz, 7. Avicenna, 8. Carl Friedrich Gauss, 9. Galileo Galilei, 10. Nikola Tesla, 11. Andre-Marie Ampere, 12. Michael Faraday, 13. Leonhard Euler, 14. Alexander Graham Bell, 15. James Clerk Maxwell, 16. Archimedes, 17. Blaise Pascal, 18. Stephen Hawking, 19. Enrico Fermi, 20. Johannes Kepler.

2DRank: 1. Albert Einstein, 2. Nikola Tesla, 3. Benjamin Franklin, 4. Avicenna, 5. Isaac Newton, 6. Thomas Edison, 7. Stephen Hawking, 8. Gottfried Leibniz, 9. Richard Feynman, 10. Aristotle, 11. Alhazen (Ibn al-Haytham) Iraq, 12. Werner Heisenberg, 13. Heinrich Hertz, 14. Johannes Kepler, 15. Galileo Galilei, 16. Shen Kuo, 17. Abu Rayhan Biruni– Persian, 18. Alexander Graham Bell, 19. Robert Hooke, 20. Michael Faraday.

CheiRank: 1. Hubert Reeves, 2. Shen Kuo, 3. Stephen Hawking, 4. Nikola Tesla, 5. Albert Einstein, 6. Arthur Stanley Eddington, 7. Richard Feynman, 8. John Joseph Montgomery, 9. Josiah Willard Gibbs, 10. Heinrich Hertz, 11. Benjamin Franklin, 12. Edwin Hall, 13. Avicenna, 14. Isaac Newton, 15. Thomas Edison, 16. Michio Kaku, 17. Abu Rayhan Biruni– Persian, 18. Abdul Qadeer Khan, 19. Werner Heisenberg, 20. Gottfried Leibniz.

L6) Top 20 articles in category Nobel laureate physicists (physicists who got any Nobel prize):

PageRank: 1. Albert Einstein, 2. Enrico Fermi, 3. Richard Feynman, 4. Max Planck, 5. Guglielmo Marconi, 6. Werner Heisenberg, 7. Marie Curie, 8. Niels Bohr, 9. Paul Dirac, 10. J.J.Thomson, 11. Max Born, 12. John Strutt, 3rd Baron Rayleigh, 13. Andrei Sakharov, 14. Pierre Curie, 15. Subrahmanyan Chandrasekhar, 16. Wolfgang Pauli, 17. Lev Landau, 18. Eugene Wigner, 19. Albert Abraham Michelson, 20. Abdus Salam.

2DRank: 1. Albert Einstein, 2. Richard Feynman, 3. Werner Heisenberg, 4. Enrico Fermi, 5. Max Born, 6. Marie Curie, 7. Wolfgang Pauli, 8. Max Planck, 9. Eugene Wigner, 10. Paul Dirac, 11. Guglielmo Marconi, 12. Abdus Salam, 13. Hans Bethe, 14. Andrei Sakharov, 15. Steven Chu, 16. Niels Bohr, 17. J.J.Thomson, 18. Steven Weinberg, 19. Peter Debye, 20. Subrahmanyan Chandrasekhar.

CheiRank: 1. Albert Einstein, 2. Richard Feynman, 3. Werner Heisenberg, 4. Brian David Josephson, 5. Abdus Salam, 6. C.V.Raman, 7. Peter Debye, 8. Enrico Fermi, 9. Wolfgang Pauli, 10. Steven Weinberg, 11. Max Born, 12. Eugene Wigner, 13. Marie Curie, 14. Luis Walter Alvarez, 15. Percy Williams Bridgman, 16. Roy J. Glauber, 17. Max Planck, 18. Paul Dirac, 19. Guglielmo Marconi, 20. Hans Bethe.

L7) Top 30 articles in category chess players:

PageRank: 1. Garry Kasparov, 2. Bobby Fischer, 3. Alexander Alekhine, 4. Anatoly Karpov, 5. Emanuel Lasker, 6. Mikhail Botvinnik, 7. Vladimir Kramnik, 8. Viswanathan Anand, 9. Paul Keres, 10. Boris Spassky, 11. Veselin Topalov, 12. Wilhelm Steinitz, 13. Tigran Petrosian, 14. Max Euwe, 15. David Bronstein, 16. Mikhail Tal, 17. Viktor Korchnoi, 18. Vasily Smyslov, 19. Samuel Reshevsky, 20. Bent Larsen, 21. Jose Raul Capablanca, 22. Boris Gelfand, 23. Gata Kamsky, 24. Alexei Shirov, 25. Mark Taimanov, 26. Magnus Carlsen, 27. Efim Geller, 28. Ruslan Ponomariov, 29. Rustam Kasimdzhanov, 30. Alexander Khalifman.

2DRank: 1. Bobby Fischer, 2. Alexander Alekhine, 3. Emanuel Lasker, 4. Garry Kasparov, 5. Wilhelm Steinitz, 6. Paul Keres, 7. Mikhail Botvinnik, 8. Jose Raul Capablanca, 9. Bent Larsen, 10. Boris Spassky, 11. Viswanathan Anand, 12. Magnus Carlsen, 13. Vladimir Kramnik, 14. Efim Geller, 15. Samuel Reshevsky, 16. Anatoly Karpov, 17. Mikhail Tal, 18. Viktor Korchnoi, 19. Max Euwe, 20. Veselin Topalov, 21. Gata Kamsky, 22. Tigran Petrosian, 23. David Bronstein, 24. Rustam Kasimdzhanov, 25. Ruslan Ponomariov, 26. Vasily

Smyslov, 27. Alexei Shirov, 28. Boris Gelfand, 29. Wolfgang Unzicker, 30. Mark Taimanov.

CheiRank: 1. Bobby Fischer, 2. Alexander Alekhine, 3. Wilhelm Steinitz, 4. Emanuel Lasker, 5. Garry Kasparov, 6. Paul Keres, 7. Mikhail Botvinnik, 8. Jose Raul Capablanca, 9. Magnus Carlsen, 10. Bent Larsen, 11. Boris Spassky, 12. Viswanathan Anand, 13. Vladimir Kramnik, 14. Efim Geller, 15. Samuel Reshevsky, 16. Anatoly Karpov, 17. Mikhail Tal, 18. Viktor Korchnoi, 19. Max Euwe, 20. Veselin Topalov, 21. Gata Kamsky, 22. Tigran Petrosian, 23. David Bronstein, 24. Rustam Kasimdzhanov, 25. Ruslan Ponomariov, 26. Vasily Smyslov, 27. Alexei Shirov, 28. Boris Gelfand, 29. Wolfgang Unzicker, 30. Mark Taimanov.

L8) Ranking of 30 Dow-Jones companies:

PageRank: 1. Microsoft, 2. IBM, 3. The Walt Disney Company, 4. Intel Corporation, 5. Hewlett-Packard, 6. General Electric, 7. Mcdonald's, 8. Boeing, 9. At&t, 10. Cisco Systems, 11. DuPont, 12. ExxonMobil, 13. Procter & Gamble, 14. Bank of America, 15. Verizon Communications, 16. JPMorgan Chase, 17. American Express, 18. Pfizer, 19. The Coca-Cola Company, 20. American Express, 21. Chevron Corporation, 22. 3M, 23. Merck & Co., 24. The Home Depot, 25. Alcoa, 26. Johnson & Johnson, 27. Kraft Foods, 28. Caterpillar Inc., 29. United Technologies Corporation, 30. The Travelers Companies.

2DRank and CheiRank have the same order: 1. Cisco Systems, 2. The Walt Disney Company, 3. Microsoft, 4. Kraft Foods, 5. IBM, 6. At&t, 7. Hewlett-Packard, 8. Pfizer, 9. Intel Corporation, 10. ExxonMobil, 11. Caterpillar Inc., 12. DuPont, 13. General Electric, 14. American Express, 15. Johnson & Johnson, 16. Boeing, 17. Wal-Mart, 18. Bank of America, 19. Verizon Communications, 20. JPMorgan Chase, 21. Merck & Co., 22. The Coca-Cola Company, 23. 3M, 24. Procter & Gamble, 25. The Home Depot, 26. Mcdonald's, 27. Alcoa, 28. Chevron Corporation, 29. United Technologies Corporation, 30. The Travelers Companies.

[1] *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*, Eds. D. Diderot and J. R. d'Alembert, A.F. Le Breton *et al.* Publ., Paris (1751 - 1772).

[2] P. Blom, *Enlightening the world: Encyclopédie, the book that changed the course of history*, Palgrave Macmillan, N. Y., (2005).

[3] http://www.wikipedia.org

[4] http://download.wikimedia.org/wikipedia/

[5] Encyclopaedia Britanica http://www.britannica.com/ (2010).

[6] J.Giles, *Internet encyclopaedias go head to head*, Nature **438**, 900 (2005).

[7] V. Zlatic, M. Bozicevic, H. Stefancic, and M. Domazet, *Wikipedias: collaborative web-based encyclopedias as complex network*, Phys. Rev. E **74**, 016115 (2006).

[8] A. Capocci, V.D.P. Servedio, F. Colaiori, L.S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, *Preferential attachment in the growth of social networks: the internet encyclopedia Wikipedia*, Phys. Rev. E **74**, 036116 (2006).

[9] L. Muchnik, R. Itzhack, S. Solomon, and Y. Louzon, *Self-emergence of knowledge trees: extraction of the Wikipedia hierarchies*, Phys. Rev. E **76**, 016106 (2007).

[10] D.J.Watts and S.H.Strogatz, *Collective dynamics of "small-world" networks (1998)*, Nature **393**, 440 (1998).

[11] M. E. J. Newman, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).

[12] R. Albert, A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74**, 47(2002).

[13] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks*, Oxford University Press (Oxford, 2003).

[14] A. M. Langville and C. D. Meyer, *Google's PageRank and beyond: the science of search engine rankings*, Princeton University Press (Princeton, 2006).

[15] A. D. Chepelianskii, *Towards physical laws for software architecture*, arXiv:1003.5455[cs.SE] (2010).

[16] S. Brin and L. Page, *The anatomy of a largescale hypertextual web search engine*, Computer Networks and ISDN Systems **30**, 107 (1998).

[17] http://www.scimagojr.com/countryrank.php

[18] http://www.arwu.org/ARWU2009.jsp

[19] M.H. Hart, *The 100: ranking of the most influential persons in history*, Citadel Press, N.Y. (1992).

[20] http://en.wikipedia.org/wiki/Dow_Jones_Industrial_Average

[21] D. Donato, L. Laura, S. Leonardi and S. Millozzi, *Large scale properties of the Webgraph*, Eur. Phys. J. B **38**, 239 (2004).

[22] G. Pandurangan, P. Raghavan and E. Upfal, *Using PageRank to characterize Web structure*, Internet Math. **3**, 1 (2005).

[23] J. Kleinberg, *Authoritative sources in a hyperlinked environment*, Jour. ACM **46**, 604 (1999).

[24] http://www.quantware.ups-tlse.fr/QWLIB/2drankwikipedia/

[25] F. Bellomi and R. Bonato, *Network analysis for Wikipedia*, Proceedings of Wikimania 2005, The First International Wikimedia Conference, Frankfurt, Germany, http://www.fran.it/blog/2005/08/network-analisis-for-wikipedia.htm

[26] Y. Ganjisaffar, S. Javanmardi, and C. Lopes, *Review-based Ranking of Wikipedia Articles*, in Proceedings of the International Conference on Computational Aspects of Social Networks, Fontainebleau, France, p.98 (2009).

[27] J.L. Borges, *The Library of Babel* in *Ficciones*, Grove Press, N.Y. (1962).