# Top 100 historical figures of Wikipedia

Dima Shepelyansky

# Top 100 historical figures of Wikipedia

D.L.Shepelyansky[1]

[1]Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France
(Dated: August 13, 2015)

The top 100 historical figures of Wikipedia are determined on the basis of statistical methods and mathematical algorithms like PageRank, CheiRank and 2DRank applied to networks of Wikipedia in up to 24 languages. In the statistical respects this top 100 list is of differs from the historical, cultural and other type arguments used by such historians like Michael H. Hart. The various mathematical methods and results obtained by different groups are described below. In spite or the mathematical and statistical grounds of those approaches they have overlap of about 43 percent with the top 100 list of Hart. The distributions of top PageRank historical figures over world countries (Fig. 1) and 35 centuries of human history are analyzed. This short popular note presents overview of results and methods of different groups of the world.

## APPROACHES OF DIFFERENT GROUPS

Top 100 historical figures have been selected by Michael H. Hart [1] on the basis of various historical reasons. The ranking of top historical figures on the basis of statistical and mathematical algorithms is described below.

The early ranking of top people of Wikipedia was done on the basis of PageRank algorithm and HITS algorithm for English Wikipedia edition (2005) by F.Belloni and R.Bonato [2]. For top people of PageRank they found Jesus, Paul the Apostole, Saint Peter and for HITS George W. Bush, Adolf Hitler, Bill Clinton.

Later studies of Quantware group analyzed English Wikipedia edition Aug 2009 using PageRank, CheiRank and 2DRank algorithms [3]. The top persons found are: Napoleon, George W. Bush, Elizabeth II for PageRank; Michael Jackson, Frank Lloyd Wright, David Bowie for 2DRank; Kasey S. Pipes, Roger Calmel, Yury G. Chernavsky for CheiRank. For this study the distributions of top 100 historical figures of PageRank, CheiRank and Hart's list are shown in Fig. 2.
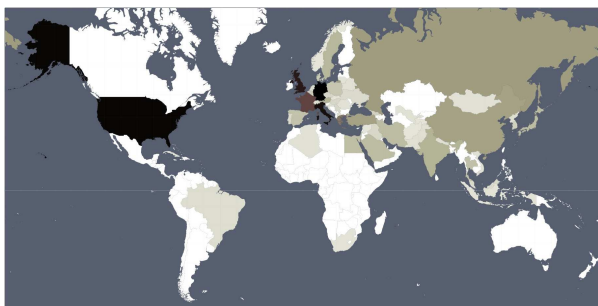


FIG. 1: World map of top 100 historical figures of 24 Wikipedia editions from PageRank for the period from BC 15th century to AD 20th century (darkness is proportional to a number of appearances of top persons born in a given country of AD 21st century geographical borders). After [9].
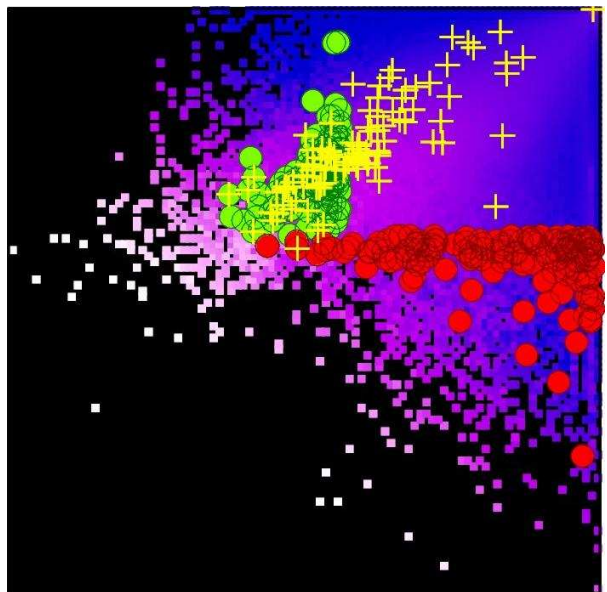


FIG. 2: Density distribution of Wikipedia English articles (Aug 2009) in the plane of PageRank and CheiRank indexes shown by color with blue for minimum and white for maximum (black for zero); green/red points show top 100 personalities from PageRank/CheiRank, yellow pluses show top 100 personalities from Hart's book, number of articles $N = 3282257$. After [3].

Time evolution of Wikipedia ranking of historical figures was investigated for English Wikipedia editions for 2003 - 2011 using the approach developed for Wikipedia Aug 2009 [4]. The distribution over fields of human activity was established there for various years.

Independently, a study with 15 largest Wikipedia language editions was done by Barcelona Media group [5]. This group considered network of links between biographical articles of Wikipedia. However, a number of such biographical articles is relatively small compared to the total number of articles of a given edition that led to fluctuating ranking results.

The investigations of 9 Wikipedia editions have been reported in [6] producing a reliable ranking of top 30 persons for each edition. However, a selection of historical figures from the whole list of ranked edition articles was done manually that was restricting efficiency of the approach.

In parallel, the Stony-Brook group performed ranking of English Wikipedia edition combining PageRank method with other methods [7]. This group found the top figures: Jesus, Napoleon. Muhammad. However, even if this group used the public Wikipedia database the whole list of their top 100 people is not publicly available.

The Pantheon MIT project produced the ranking list of top 100 persons using all language editions of Wikipedia counting number of editions and clicks on an article about a given person [8]. This group found at the top: Aristotle, Plato, Jesus.

A list of the top 100 historical figures was created from Wikipedia pages in 24 different languages, using computer algorithms to analyze the importance of people based on the links to those people's pages [9]. All data of this study are available at [10].

## RANKING METHODOLOGY

The researchers used several different page-ranking algorithms, including Google PageRank, 2DRank, and CheiRank (see [3]). They retrieved data from the text of Wikipedia pages in the 24 languages, and applied the algorithms to the data to create culturally-specific list of influential people, as well as a list across all the cultures examined in the project.

Among the data elements specifically targeted as indicators of importance were each person's birth country, date of birth, century of birth, and quantity of hyperlinks. In the case of hyperlinks for people's Wikipedia pages, both links to a person's page and links from that person's page were included in the analysis. Other methods are described at [7] and [8], they are not directly related to link analysis, network theory and Markov chains.

## RESULTS

For the global list of 24 editions of Wikipedia [9], the top 10 historical figures, identified by averaging over PageRank lists, were as follows:

- 1. Carl Linnaeus
- 2. Jesus
- 3. Aristotle
- 4. Napoleon
- 5. Hitler

- 6. Julius Caesar
- 7. Plato
- 8. William Shakespeare
- 9. Albert Einstein
- 10. Elizabeth II

The top global persons of 2DRank are Adolf Hitler, Michael Jackson, Madonna (entertainer). The top women of human history are Elizabeth II, Mary (mother of Jesus), Queen Victoria for PageRank list and Madonna (entertainer), Elizabeth II, Mary (mother of Jesus) for 2DRank list. Top 100 historical figures for 24 Wikipedia editions are available at [10]. The distribution of these historical figures over the world map of countries is shown in Fig. 1 according to the birth place on the actual world map.

The overlap of top 100 people of Quantware, Stony-Brook, MIT Pantheon groups with the Hart list is found to be on a level of 42-44 percents. This shows that the mathematical methods of determination of top 100 historical figures of humanity via Wikipedia database give the reliable results.

## DISCUSSION

Discussion of Wikipedia ranking of historical figures has been highlighted in public press including The New Scientist, The Guardian, The Washington Post, Uppsala Universitet, EC Cordis, CNRS INP news, Le Figaro (links to these highlights are available at [11]).

In the frame of this discussion an article "Top 100 historical figures" had been created at English Wikipedia by an unknown user (HarryBoston) in June 2014. After discussion at Wikipedia it was transfered to "Top 100 historical figures of Wikipedia" with certain updates presented here. In 2014 the list of Wikipedia contributers was: Animalparty, Frmorrison, HarryBoston, OccultZone, Roscelese, Shepelyansky, Spinningspark, Topbanana. In this form with some updates this Wikipedia article was attracting about 2000 views per month. In June 2015 Wikipedia administration decided to eliminated this article and without long discussions it was canceled at Wikipedia.

---

[1] M.H. Hart, *The 100: A Ranking of the Most Influential Persons in History*, ISBN 978-0-8065-1068-2, Carol Publishing Group (1978)
[2] F.Belloni and R.Bonato, *Network analysis for Wikipedia*, Proceedings of Wikimania 2005, The First International Wikimedia Conference. Frankfurt, Germany. Retrieved September 16, 2006"

(http://www.fran.it/blog/2005/08/network-analisis-for-wikipedia.html)

[3] A.O.Zhirov, O.V.Zhirov and D.L.Shepelyansky, *Two-dimensional ranking of Wikipedia articles*, Eur. Phys. J. B **77**, 523 (2010); arxiv:1006.4270[cs.IR]

[4] Y.-H.Eom, K.M.Frahm, A.Benczur and D.L.Shepelyansky, *Time evolution of Wikipedia network ranking*, Eur. Phys. J. B v.86, p.492 (2013); arXiv:1304.6601 [physics.soc-ph]

[5] P.Aragon, A.Kaltenbrunner, D.Laniado and Y.Volkovich, *Biographical Social Networks on Wikipedia - A cross-cultural study of links that made history*, Proceedings of WikiSym, 2012; arXiv:1204.3799(cs.SI) (2012)

[6] Y.-H.Eom and D.L.Shepelyansky, *Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles*, PLoS ONE **8(10)**, e74554 (2013); arXiv:1306.6259

[cs.SI]

[7] S.Skiena and C.B.Ward, *Who's Bigger?: Where Historical Figures Really Rank*, Cambridge University Press. ISBN 978-1-107041370 (2014)

[8] The Pantheon MIT project "The Pantheon MIT project" (http://pantheon.media.mit.edu)

[9] Y.-H.Eom, P.Aragon, D.Laniado, A.Kaltenbrunner, S.Vigna and D.L.Shepelyansky, *Interactions of cultures and top people of Wikipedia from ranking of 24 language editions*, PLoS ONE **10(3)**, e0114825 (2015); arXiv:1405.7183[cs.SI] (2014)

[10] http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/

[11] http://www.quantware.ups-tlse.fr/FETNADINE/press.htm