

Who's Bigger? Where Historical Figures Really Rank. By Steven Skiena and Charles Ward, Cambridge University Press, Cambridge, UK, 2013, 408 pages, \$27.99.

Who was more important historically—Mary, Queen of Scots, Mary Tudor, Queen of England, or Marie Antoinette? Copernicus or Freud? Charlie Chaplin or Steven Spielberg? Wonder no longer. Thanks to the power combo of Big Data and the Wisdom of the Crowd, these and all such questions have been scientifically answered. Specifically, Steven Skiena and Charles Ward have produced a ranking of the historical importance of everyone with a Wikipedia entry—which, needless to say, is everyone who was ever anyone—from Jesus [#1] to Sagusa Ryusei [#843,790]. Consulting the index to the book, or the accompanying website (whoisbigger.com), we find that the battle of the Mary's was a photo finish: Marie Antoinette was the 125th-most-important person in history, Mary of England was 126th, and Mary, Queen of Scots, was 127th. Freud at 44 handily beat Copernicus at 74; and Chaplin at 295 clobbered Spielberg at 1079.

The list, as I mentioned, includes anyone with a Wikipedia article; for instance, my boss John Sexton, president of NYU, is the 69,747th-most-important person in the history of the world; my instructor in undergraduate topology, James Munkres, is the 195,642nd. One can easily imagine that after the next project of this kind—which would incorporate everyone with a web presence and constantly update the calculations—it would be de rigueur to list one's current ranking on one's CV, together with citation count, h-index, i10-index, and all the other numbers that reliably quantify one's life and labors.

How Is the Ranking of Person X Computed?

Skiena and Ward start with six basic statistics:

1 and 2. The PageRank of X's Wikipedia page. This measure, famous as the basis of the Google search engine, is computed from the number of Wikipedia pages that contain a link to X, weighted by the importance of the pages linking to X. Two versions of PageRank are computed: One considers all Wikipedia pages, the other biographical pages only. For instance, Linnaeus (overall rank 31) scores very high on the first measure, because every species that he named links back to him; he scores less high if one considers only biographical pages.

3. The number of times the Wikipedia page has been viewed.

4. The number of times the Wikipedia page has been modified.

5. The length of the Wikipedia article.

6. The frequency with which X is mentioned in the news.

Applying a factor analysis to these numbers reveals that there were two primary

factors. One, which Skiena and Ward call *celebrity*, is the person's current notoriety; hot rock stars, politicians in the news, and so on score high here. The other, called *gravitas*, is the measure of solid accomplishment; philosophers, scientists, classic historical figures score high here. A linear combination of celebrity and gravitas gives *fame*. Fame, however, is fleeting and declines over time; correcting for this effect, Skiena and Ward arrive at the final value for *historical significance*. They also discuss

and analyze the evolution of fame over time, using the Google Ngrams tool that reports the number of times a given name was mentioned in publications within a given range of dates. In many ways, this diachronic analysis is more interesting and more informative, though less complete, than the ranking studies.

What Is Being Measured? Given that "historical significance" is obviously entirely vague and nonquantifiable, what do these numbers actually signify? Skiena and Ward make a number of claims. The most cautious claim is that the rankings measure "the strength of historical memes" and that their study of change over time analyzes the processes that cause figures to become more and less famous. Among their normative claims are that highly ranked figures are those who are "most worth knowing" and "really belong in history textbooks." They claim further that these numbers correlate strongly with the figures' "true" importance as measured by historians. Finally, there is the tongue-in-cheek claim of the subtitle: "Where Historical Figures Really Rank."

How Accurate Are the Rankings? That's harder to say. Skiena and Ward, naturally, are very enthusiastic about their ranking. They have validated it against quite a collection of existing measures: lists put together by others, prices of autographs, answers from people asked to compare pairs of historical figures, and so on. The authors report correlations of about 0.5 with these other measures, which they argue is as good as could be expected in that the different measures don't agree with one another better than that.

Looking over the list, I had mixed feelings. On the one hand, most of the rankings, especially the comparative rankings of people in the same field, are plausible. Jesus [1], Napoleon [2], Muhammad [3], Shakespeare [4], and Lincoln [5] are important people—check; Leonardo [29], Michelangelo [86], Raphael [140], Rembrandt [189], and Titian [319] were great painters—check; and so on. The work is also impressive in some technical respects; in particular, the distinction between celebrity and gravitas and the correction for time both seem, on the whole, to work very well. (Among intellectuals, in fact, it seems to me that they *over-compensate* for time and rank pre-modern figures higher than they deserve.)

On the other hand, there are a number of significant biases and numerous rankings that, I would argue, are just indisputably wrong. To the extent that comparisons of this type are meaningful at all, it is simply

wrong to say that two of the top 20 and four of the top 41 most important people in history were Tudor or Stuart British monarchs; or that Queen Victoria, who had pretty much no political power, was the 16th-most-important person in history; or that Charles Babbage [273] and Ada Lovelace [994] were more important mathematicians than Noether [2523], Chebyshev [3571], or Grothendieck [7311]; or that all but one (Schiller [564]) of the most important poets have been anglophone; or that Francis Scott Key [1050] was the 19th-most-important poet in history.

The problems are compounded by the fact that some figures who are reasonably ranked in the database are mistakenly omitted or miscategorized in the book. Wagner [62] is missing from the list of composers, Hilbert [641] from the list of mathematicians; Tolkien [192] from the list of 20th-century novelists. Oscar Wilde [77], absurdly, is listed as the third-greatest novelist rather than, less implausibly, as the second-greatest playwright. As I was writing this review, the website was full of bugs. About a fifth of the pages did not display the statistics correctly. The web page for Queen Victoria strangely compared her ranking to New York, Toronto, San Francisco, and so on. The website included pages for "Knitting" (the activity) and for "December 6" (the date). Presumably, these are the results of misclassified pages in Wikipedia, but those who live by Wikipedia perish by Wikipedia.

Biases: As expected for a collection based in the English-language Wikipedia, there are biases in favor of English-speakers, against women, and, in descending order, in favor of the US, the UK, Western Europe, classical Greece and Rome, Eastern Europe, the Middle East, the Far East. There are also striking biases in the categories: The top 200 figures include ten classical composers and five artists, but only one historian (Herodotus [123]) (i.e., person known primarily as a historian) POSSIBLE TO JUST SAY "ONLY ONE PERSON KNOWN PRIMARILY AS A HISTORIAN (HERODOTUS [123])?". In the top 1000, we find only 11 more historians, only two of whom are of the modern era (Gibbon [573] and Tocqueville [716]), and only one computer scientist (Bill Gates [904]). Jimmy Wales, who founded Wikipedia, is #3198; Tim Berners-Lee, who created the World Wide Web, is #3931.

What Is the Use of It? There is an inherent difficulty in finding an actual use for a project of this kind. To the extent that the rankings correspond to the conventional wisdom (Jesus, Napoleon, Muhammad), we don't need the study. To the extent that they contradict the conventional wisdom (Ada Lovelace, Queen Victoria), the study seems wrong. Of course, Skiena and Ward could argue the exact reverse: To the extent that they correspond, the rankings are validated; to the extent that they differ, they offer us new insights. The problem, though, is that the new insights—i.e., about the people who are more highly ranked than expected—do not seem particularly [[interesting or insightful DEEP?]]; they just concern

people (Queen Victoria, Jules Verne, Ada Lovelace) who, for one reason and another, are much better known than their actual accomplishments would warrant.

Skiena and Ward suggest a number of uses for the rankings. One is for vetting history textbooks; Skiena discusses at length his daughter's fifth-grade history textbook, which includes some very obscure people. He proposes the substitution of other people, judged more important in his ranking. His suggestions mostly seem sensible; precisely because they are [self-evidently?] sensible, however, it is not clear why you would need the rankings to arrive at them (except to intimidate reluctant educationalists with numbers). The authors also suggest that the lower rankings of women collectively can be used to measure the neglect of women in the historical literature, though this obviously conflicts with the claim that this is an accurate measure of the true ranking. A LOT OF THIS'S IN THAT SENTENCE. IT'S KIND OF HARD TO FOLLOW

What Is the Harm in It? Against these uses, one has to weigh the harm done by a book of this kind in reinforcing the widespread and growing illusions that all questions can be answered by web mining; that fame is equivalent to a worthwhile life; and that the significance of a human life can be reduced to a number and a 25-word summary. We are awash in lists; the last thing we need is an exhaustive list of everyone judged on a single criterion, supported by pretenses to objectivity.

Bottom Line: All in all, the book seems to me bloated, both in its claims and in its length. The claim that it constitutes any kind of contribution to our understanding of what figures are historically significant seems to me entirely baseless. And the book is about ten times too long. It contains a variety of silly lists: Who is the most important person to die at age 57? Who is the most important person to be born on March 28? The discussion of the fifth-grade textbook mentioned earlier is sensible, but the point could have been made in one page rather than 30. A long history of the inductees into the "Hall of Fame for Great Americans" in the Bronx, with a year-by-year account of the honorees and the rejected candidates, is entirely uninteresting. If this book had been a 30-page research paper, with conclusions along the lines of, "We have shown that we can automatically compute historical importance using these kinds of techniques, and that the results are pretty good, with such and such kinds of bugs and biases," I would have said it was a fascinating, though useless, project, very well executed.

Ernest Davis is a professor of computer science at the Courant Institute of Mathematical Sciences, NYU.

BOOK REVIEW

By Ernest Davis