

Review of *Who's Bigger? Where Historical Figures Really Rank* by Steven Skiena and Charles Ward (Cambridge University Press, 2013, xiii+379 pps.)

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 article — which, needless to say, is everyone who was ever anyone — from #1 Jesus to #843,790 Sagusa Ryusei. Looking these up in the index of the book or on the accompanying web site <http://www.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 have said, includes everyone 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, and so on. One can easily imagine that, after the next project of this kind, which will incorporate everyone with a web presence and constantly update the calculations, it will become *de rigueur* to list one's current ranking on one's CV, together with one's 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 that considers all Wikipedia pages, the other that considers only biographical pages. 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 the person is mentioned in the news.

Applying a factor analysis to these numbers revealed that there were two primary factors here. One, which Skiena and Ward call *celebrity*, is the 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; Skiena and Ward added a correction for this effect, yielding the final value for *historical significance*.

Skiena and Ward also include a discussion and analysis of the evolution of fame over time, using the Google Ngrams tool that reports how many times a given name was mentioned in publications during 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? Since “historical significance” is obviously entirely vague and non-quantifiable, what do these numbers actually signify? Skiena and Ward make a number of different 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. Then there are normative claims: highly ranked figures are those that are “most worth knowing” (p. 13) and “really belong in history textbooks” (p. 39). Further they claim that these numbers correlate strongly with the “true” importance as measured by historians (p. 13). 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 it. They have validated it against quite a collection of other measures: Other lists that people have put together, prices of autographs, asking people to compare pairs of historical figures, and so on. They report correlations of about 0.5 with these measures, which they argue is as well as one could expect, since the different measures don’t agree with one another better than that.

Looking over it myself, I had mixed feelings. On the one hand, most of the rankings are plausible, and in particular, most of the comparative rankings of people in the same field are plausible. Jesus[1], Napoleon [2], Muhammed [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 to work very well, on the whole. (Among intellectuals, in fact, it seems to me that they are *over*-compensating for time, and ranking 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 these kinds of comparison are meaningful at all, it is simply wrong to say that two of the top 20 most important people 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], Cheybshev [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; and so on.

The problems are compounded by the fact that some figures that 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] is, absurdly, listed as the third greatest novelist rather than, less implausibly, as the second greatest playwright.

As of the time of writing, the web site is full of bugs. About one-fifth of the pages do not display the statistics correctly. The web page for Queen Victoria strangely compares her ranking to New York, Toronto, San Francisco, and so on. The web site includes 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 one would expect from a collection based in English-language Wikipedia, there are biases in favor of English-speakers, against women, and, in descending order, in favor of US, UK, Western Europe, classical Greece and Rome, Eastern Europe, Middle East, Far East, rest of the world.

There are also marked biases in the categories. Among the top 200 figures, there are ten classical composers and five artists but, ironically, there is only one historian (Herodotus [123]) (i.e. person primarily known as a historian). In the top 1000, there are only eleven 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 itself 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 this kind of project. To the extent that the rankings correspond to the conventional wisdom (Jesus, Napoleon, Muhammed), 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 can argue this in the exact reverse: To the extent that they correspond, that validates the rankings; to the extent that they differ, that offers us new insights. The problem, though, is that the new insights — i.e. the people that are more highly ranked than one would expect — do not seem particularly interesting or insightful; they are just people (Queen Victoria, Jules Verne, Ada Lovelace, etc.) who, for one reason and another, are much better known than their actual accomplishments would warrant.

Skiena and Ward suggest a number of uses that these rankings might serve. One is for vetting history textbooks; Skiena has a lengthy discussion of his daughter's fifth-grade history textbook which includes some very obscure people. He proposes to substitute other people, judged more important in his ranking. His suggestions mostly seem sensible; however, precisely because they are self-evidently sensible, it is not clear why you would need the rankings to arrive at them (except to intimidate reluctant educationalists with numbers). They also suggest that the lower rankings of women as a whole 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.

What is the harm in it? Against these uses, one has to weigh the harm that this kind of book does 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 twenty-five word summary. We are awash in lists of all kinds; the last thing we need is an exhaustive list of everyone on one single criterion, supported by pretences to objectivity.

Bottom line: All in all, the book seems to me *bloated*, both in its claims and in its length. The claim that this constitutes any kind of contribution to our actual understanding of what figures are historically significant seems to me entirely baseless. And the book is about 10 times too long. There are all kinds 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 is, as I said, sensible, but the point could have been made in about one page rather than thirty. There is a long, entirely uninteresting, 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. If this 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.