

Excerpt from The Book Review Column¹

by William Gasarch

Department of Computer Science
University of Maryland at College Park
College Park, MD, 20742
email: gasarch@cs.umd.edu

Review of²

Who's Bigger? Where Historical Figures *Really* Rank
by Steven Skiena and Charles B. Ward
Cambridge University Press, 2014
379 pages, Hardcover

Review by

Nicholas Mattei
nicholas.mattei@nicta.com.au
Optimization Research Group
NICTA and UNSW

1 Introduction

Steven Skiena and Charles Ward ask an interesting question: can we say whose historical significance through purely statistical means using only public data? The answer is this book and its enjoyable answer, maybe. While at the end of the day this book is more of a playful walk down history the question itself is intriguing. Extremely data driven at its heart, *Who's Bigger* manages to entertain and enlighten throughout. It is an easy, enjoyable read that anyone who loves history and trivia, along with a fondness for arguing over ephemeral measures of “importance,” will never want to put down.

2 Summary

The book is divided into two main parts and several appendices. Using Wikipedia (and Google nGrams for additional analysis) the authors rank order every *person* that is significant enough to have a Wikipedia page. This ranking of over 800,000 individuals serves as a jumping off point for the authors to explore a number of topics related to gravitas, celebrity, and the ways in which we are remembered (or not).

The first several chapters provide an overview of the datasets and ranking methodologies that are used throughout the rest of the book. There are also several different visualizations that are explained in the first chapters. These give the reader an understanding of the sheer volume of data that has gone into what is to follow. From the first chapter, Skiena and Ward jump in front of many of the questions of bias and sufficiency that hard core academics would have about this undertaking. Skiena and Ward do a nice job of answering many of these questions and emphasize that the methods employed can, despite the inherent bias of using the English language Wikipedia to compare the significance of topics as diverse as American

¹© William Gasarch, 2014.

²©2014, Nicholas Mattei

presidents and Chinese artists, shed new light on some interesting questions. The key point being that the content of the open access Wikipedia does capture an idea of *memes* or *mindshare* — the importance that the users of Wikipedia give to the topics. While a historian or literary critic may question the significant conclusions that can be drawn from such a study, the results are still informative and entertaining in their own right.

I found the “Who Belongs in Bonnie’s Textbook?” chapter to be one of the most interesting. The chapter focuses on a 5th grade history textbook and its list of “250 figures highlighted in the text.” Skiena and Ward walk through many of the figures in the textbook asking, “why this figure and not that one.” This comparison is investigated not only their significance scores, but the required nature of some of these figures for different state curricula. They identify several important omissions and make some well reasoned, concrete suggestions for changes to the contents of the history text. This type of analysis, a more objective and critical look at what our children are learning, is one of the big messages that the book can lay claim too.

The later chapters of Part I compare the compiled significance rankings to other measures of historical merit, specifically, the Baseball Hall of Fame and the Hall for Great Americans. As the authors propose these two institutions as qualified arbiters of historical importance and compare the results of their method (and hindsight) to the selection processes for the two halls. I found the discussion (repeated in later sections) of the year over year significance of those included in the respective halls to be of great interest. Seeing the relative significance drop over a number of years, “to clear the backlog of qualified candidates,” was interesting and raised a point I hadn’t thought of before. Namely, it’s hard to select a set number of well qualified individuals for an award that is given annually; awardees may just slip in because someone has too. The rest of discussion in these sections is lively and informative. There is an interesting analysis of the ways in which fame decays, which allows the authors to discuss their “decay” model, which allows them to compare historical figures to more modern ones. These chapters lay the remainder of the foundation for many of the interesting patterns that are observed in Part II.

Part II of the book is, for me, the most entertaining. The authors take us on a journey through a dizzying array of domains, ranking figures in each, with chapters on American political figures, world leaders, science, religion, sports, and the arts. Each section contains a breakdown of the significance scores of key figures in the respective area along with a conversation of why they are significant, and some interesting comparisons to other measures of significance, such as Nobel prizes and Oscars. This section is where the book really shines, I can imagine (and experienced) the book facilitating a number of heated discussions about who was left out of the rankings and why. Some of the most interesting points raised in these sections concern the unique perspective of combining the Wikipedia and Google nGrams data sets. By examining when certain figures were active versus when they were written about, it is possible to visually understand how success can emerge (and be maintained or decline) posthumously. These mixed graphs were among the most interesting for me throughout the book.

The more technical details of the ranking methodologies are saved for the first appendix. There is enough detail here for the curious to understand the general methodologies used and agree or disagree with them. The other appendices included are more playful in nature and contain the most famous people to live or die on a particular date, information about the “Whose Bigger” game and webpage, and brief biographies of the 100 most significant figures in the book.

3 Opinion

This is a well written book that I enjoyed reading. While not technically deep it kept me engaged and reading throughout. Skiena and Ward approach the subject with a mix of humor and rigor that is unique and fun. The book can serve as a springboard for anyone interested in rating and ranking. In addition to being a fun popular science book I could see sections of it being successfully integrated into a statistics or data analytics course.

I can't help but feel that there was an opportunity with this type of analysis that was missed. I would have liked to see a longer version of the text, including a historian, or other domain expert author, to comment on the chapters. Mixing these viewpoints to compare and contrast the rankings would have taken the book from good to great. It would have allowed a dialog to form within the text. The authors attempt to facilitate such a dialogues in the book but I feel the addition of a real counterpoint perspective would have made this book truly phenomenal.

However, this criticism does not stop the book from being fantastic for what is there. I highly recommending the book for anyone interested in history, sports, or getting into good arguments about what's important for being remembered.