

Who's #1: The Science of Rating and Ranking

Dr. Amy Langville

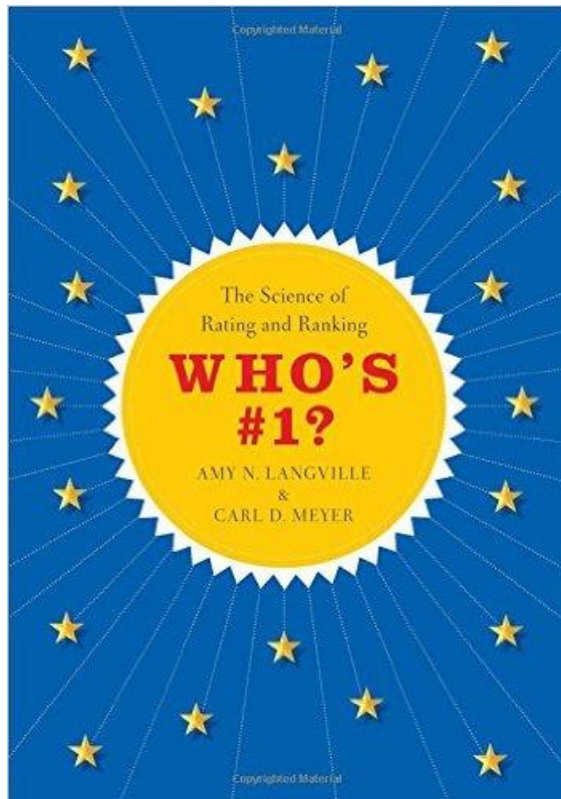
Professor

Mathematics Department

Operations Research Analyst

College of Charleston

In 2012, my book with Carl Meyer, *Who's #1: The Science of Rating and Ranking*¹, appeared. Since then, I regularly receive requests to help companies, law firms, colleagues, or students analyze their data. I'll discuss some of the more interesting recent projects and the corresponding *tools* that fit under this year's Math Awareness Month theme of *The Future of Prediction*.



¹Copyright 2012 Princeton University Press

First, a timely application: March Madness, the annual NCAA college basketball tournament. Millions of fans submit brackets trying to predict the winner of each matchup in this month-long tournament. Along with my colleague, Tim Chartier of Davidson College, we teach our students how to submit brackets based solely on mathematical models. Two such models, the Colley and Massey models, use *linear systems* to rate teams. Another model, the Elo model, uses an *iterative update*. Over the years, the students' models have performed well, some years scoring in the 99th percentile of all brackets submitted. Each year our students ask questions and collect data to add a bit more sophistication to the models. For example, how do we introduce factors like coaching, team cohesion, or tournament experience? How can injuries be accounted for? Can we predict this year's Cinderella team?

Another sports application: My students and I have helped the U.S. Olympic Committee by conducting data analysis on Olympic athletes. In order to optimize resources, the Committee wants predictions of which athletes will medal as the window leading up to the games changes from one year to four years. We used *regression* and *simulation* to answer this question. Other questions had to do with whether or not the country's financial incentive program was working to motivate athletes to win more medals. Getting appropriate data on this second problem was hard. So we expanded to see if the incentive programs of other countries including England worked.

Next, let's discuss recommendation systems such as Amazon's "Customers who bought this also bought ..." system. Again, I team with my colleague Tim Chartier to reply to requests from small startup companies that want to predict which movies to recommend to their customers, or which songs to fill a particular listener's playlist, or which athletic products to recommend to a customer. The common theme is how to use the data the company has collected to make useful predictions that influence customer behavior. We typically use *clustering* and *nearest neighbor classification tools* to solve such problems.

Last year Rootmetrics asked me to improve their current system for rating cell phone carriers. Student Tyler Perini built them a *Markov chain* that elegantly modeled the physical processes of successful connection and transmission of calls and was able to disambiguate the many ties that their current rating system produced.

Two College of Charleston colleagues, a philosophy professor and a psychology professor, asked for help analyzing some text data they had collected for their Humility Project. The goal was to determine a person's level of humility by analyzing a sample of their writing. Student Tyler Perini developed a *text mining tool* that, given a short sample of text, not much longer than a Tweet or Facebook status update, predicts the author to be either humble or not humble. We found that humble writers use inclusive language, such as the inclusive "and" and "we", "all" and "each other" whereas not humble

writers use distancing language such as "they", "people", "them" and the exclusive "or". As a next step, the humanities professors plan to study self-control in the hope of offering behavior training to those children that our tool predicts, based on their speech, to have low self-control.

In another text mining project, graduate students at the College of Charleston are analyzing the text of candidates in this year's presidential campaign. They have found some interesting trends. For example, Donald Trump's dictionary (i.e., his working vocabulary) is about one-third the size of Hillary Clinton's. They will use *similarity measures* and *Markov chains* to predict how voters shift support from one candidate to another as the field narrows.

Finally, a colleague's wife presented me a very difficult problem, the *stable matching problem* of matching medical residents to hospitals after graduation. Medical residents and hospitals both create ranked preference lists of agents on the other side that they prefer. These lists are submitted to a national body that runs a variant of the Gale-Shapley algorithm (of 2012 Nobel Prize fame) to assign residents to hospitals. The local Medical University of South Carolina would like to better advise its own residents as to how to strategically create their preference lists. This problem is particularly challenging given that the data available to the Medical University is incomplete. It only knows its residents' preferences and has no information about its competitors, the other hospitals in the

U.S. We are currently using tools of *linear and integer programming* and *game theory* to analyze this problem.

It is clear from the diversity of problems described above that data are everywhere. Data are collected constantly from sources large and small, from satellites to smart phones. *Here's a prediction that's easy to make: the future will hold exponentially more predictive analytics. All of which makes now a great time to major in mathematics, computer science, data science, or statistics. And even better, some combination of the above.*

References

Wesley Colley. Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained.

<http://www.colleyrankings.com/matrate.pdf>, 2002.

A. E. Elo. The Rating of Chess Players, Past and Present. Batsford, (London), 1978.

Kenneth Massey.

<http://www.masseyratings.com/theory/massey.htm>

The Royal Swedish Academy. The Prize in Economics 2012.

http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2012/popular-economicsciences2012.pdf