# Discussion11: Recommender Systems

Rachel Greenlee

11/2/2020

## Introduction

For this discussion I'll take a look at Goodreads, an incredibly popular "Facebook of Books" that also has a recommended system for users to discover new books. I use this website myself but with many disappointments, so it will be interesting to analyze this further.

## Scenario Design Analysis

### Who are Goodreads' target users?

Goodreads was created in the 1990s by a (now-married) couple that realized one of the best ways to discover new books was by seeing what was on your friends' bookshelves when you were in their house. Clearly others liked this approach as the website has grown immensely in the past 20 years (despite the dated interface) and was aquired by Amazon in 2013.

With this history in mind, originally I'd say Goodreads' target user was anyone looking for book reccomendations based on their reading history and their friends' history, that user also inputting their reading habits would also expand their dataset for further recommendations to users.

Once acquired by Amazon, I would imagine Amazon also hoped to target users who would use the on-page links to buy the physical/ebook copies on their website. Further, a lot of Amazon ebook self-published authors are on Goodreads, giving those books essentially free marketing - particularly when Goodreads (Amazon) gives popular/influential reviewers free copies of the books to review.

### What are Goodreads' key goals?

Goodreads wants engaged users on the website adding to their datapoints and using the website regularly. After being acquired by Amazon, I'd assume there is a goal to increase ebook and physical book sales on Amazon as well.

### How can data science/recommender systems help accomplish these goals?

If Amazon just wanted to increase book sales, they could put more effort into Amazon.com's book recommended system - but the supposed secret sauce of Goodreads is the *social* aspect of book recommendations. The recommender system gives Amazon/Goodreads a ton of data on users, as many link it to their Amazon account, and a ton of datapoints on what books to recommend to who, driving up website engagement and book sales. Further, there are a mind-boggling number of books available (more than say restaurants in a city or movies on Netflix) and it can be daunting to pick a book to read, and then time-consuming if you don't end up liking it all that much. If a recommender system can have a good track record (in the user's eye) of finding them books they love, they will use the website more and presumably buy more books.

## Reverse Engineer the Recommender System

Any actual code or insight into the recommender system algorithm seems to be out of public sight, the announcement linked below dates 2011 and states:

*"Earlier this year, Goodreads purchased a company that had built a very sophisticated book recommendation system... The Goodreads Recommendation Engine combines multiple proprietary algorithms which analyze 20 billion data points to better predict which books people will want to read next. It maps out the connections between books by looking at how often they appear on the same bookshelves and whether they were enjoyed by the same people. On average, Goodreads members have 140 books on their shelves. With this information, the engine learns how your tastes are similar to or different from the tastes of other Goodreads members. So, a big part of the secret sauce is... you, the Goodreads community. The Goodreads community is almost six million members strong, and you've added a combined total of 190 million books to your shelves!"*

https://www.goodreads.com/blog/show/303-announcing-goodreads-personalized-recommendations

So from this snippet it sounds like a standard recommendation system in which Goodreads recommends books that "other users like you" also liked. From my personal experience as a user, you do set up explicitly what genres you are interested in, which can help clean up your recommendation if you know you really like science-fiction but are very uninterested in fantasy, two related genres that often recommender systems lump together.

## Recommendations

Much of the reading community has been underwhelmed with Goodreads for years, particularly considering how much data is behind the website and it's acquisition by Amazon. One issue is that popular books with really vocal reviewers are not always good books. There some users on Goodreads who must be spending 40+ hours a week reading and crafting very long, gif-heavy reviews. This handful of power users dominates ratings and reviews in their specific genres - and I don't feel the algorithm takes that into account. Ideally I'd like to see a 5-star rating from someone who only gives 5-stars 10% of the time be worth more than a 5-star rating from someone who rates 40% of their 300 books a year at that level.

Another recommendation, is that when you hover over your recommendation it will show *why* the book was recommended to you, and it's usually just based on one other book you've read. I wish there was a way to filter by book recommendations that are related to 2, 5, 10 other books you've read and loved - wouldn't that be a stronger connection?

Finally, such a small thing, there should be the option to exclude young adult books if desired. It's a really strong genre/community on Goodreads but adults either love/hate young adult books, and it'd be such an easy fix and would save us non-young adult readers a lot of time reading a book summary only to see it tagged at YA and have to start again.

In my research I found this Medium article which collects a lot of the disappointments readers have about Goodreads that echo my thoughts, one snippet captures the issue with the young adult genre dominating the algorithm:

*"For some reason, Goodreads seems to attract an audience of people with insanely bland and entry-level taste," Martin says. He points to the site's Best Books Ever list, which includes Harry Potter, high school curriculum novels, and copious YA. "That would be fine if it didn't seem to poison the site's recommendation algorithm, which in my experience is entirely useless."*

https://onezero.medium.com/almost-everything-about-goodreads-is-broken-662e424244d5

## Conclusion

Goodreads has so much data and so much potential, but with no real competitor there isn't much incentive for Amazon to invest in improving their reccomender system to the standards users have come to expect.

Any competitor has a huge uphill battle as they'd need to quickly attract a huge base of users to generate the data points in order to make their recommender system operate better than Goodreads.