Capstone 2 – Book Recommendation System

I've come across a list of 278k users who have implicitly or explicitly reviewed 271k books. I'd like to build a recommendation system to recommend the right books to the right users. I'll attempt both a collaborative approach and a content based approach (although we do not have a lot of dimensional data about the books/users).

My client for this project could be any retailer who sells books or a platform like Good Reads that allows users to follow each other and get recommendations on what book to read next. After this analysis, retailers should be able to offer better products to their customers and increase lifetime value or sales per customer metrics while Good Reads can demand higher advertising revenue by proving they can put publisher's books in front of the right audience under their sponsored content section.

I'll be using the "Book-Crossing" data set win which more info can be found at http://www2.informatik.uni-freiburg.de/~cziegler/BX/.

> *Improving Recommendation Lists Through Topic Diversification*,
>
> *Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, Georg Lausen; Proceedings of the 14th International World Wide Web Conference (WWW '05), May 10-14, 2005, Chiba, Japan.*

Overall there are 1.1M reviews by 278k users across 271k books, most are implicit (not rated or in the case of this data, 0) while some of explicit ratings of 1 – 10. The data also contains other dimensions, such as the user 's physical location, book's publisher, year of publication, and thumbnail to the image.

I'm going to use both collaborative and content based recommendation systems using accuracy from precision and recall. I'll use a graph data structure with networkx in python separating users and books by bipartite.

As a final deliverable set, my code, a paper in pdf form, and a slide deck will be published to my github, in this directory.