1.2
The main methods that I tried to implement for blocking were the first 3 of the ISBN, the first 5 of the title, and the number of authors. Eventually, I stuck with the first 3 characters of the title string and the first 3 of the ISBN (which is the geographical code of the publication)

1.3
In this one, I tried several different approaches and fine tweaks within each one. At first, I attempted to use similarity of the ISBN numbers, but I quickly realized that the data was either not complete or not correct, since after the first 3 characters even the same book can have different ISBN's due to differences in publication location. Eventually I settled upon a combination of the subset of the first 5 characters of the name strings' jaro Winkler similarity (already given to us), then the jaccard similarity on the name string tokens, and finally averaged author jaro Winkler similarity.

2.1

Most of my explanations are in the notebook, but essentially I took preference for the Goodreads version if it was a value that was present, and barnes and nobles if not. For the node name, authors, and publisher I simply concatenated entries to get the end result for each entry.