Serena Touqan
SI 507; Intermediate Programming
April 2024

Intermediate Programming Final Project Summary

For my project, I decided to build a network visualization of New York Times Bestselling books. I thought it would be cool to demonstrate the relationship between books that become bestsellers, such as common subjects. As you can see from the graph my program returns, each circle (node) on the graph represents a NYT Bestseller – with each book's size and color depending on their popularity and rating respectively. I was not surprised to see that most of these bestsellers are rated highly (most of the network is green). However, I am surprised that some of these books seem to be that popular based on the data. Obviously they are on this list for a reason (they're bestsellers after all). So, that makes me wonder what you could potentially learn about a book's audience based on their node size (Colleen Hoover books are quite large on this graph, meaning her readers are probably younger/on the internet more…?).

Something I have also noticed is that a good portion of these best selling books tend to be in the contemporary fiction / romance genre, or have "fiction" applied as this catch-all subject. This made it difficult to actually get good recommendations for things like science fiction and fantasy based on subject headings alone. If I could find a dataset that catalogs things like writing tropes, style, pacing, etc (storygraph gathers data on these things from readers but I don't think they have an API yet…) there is a lot of potential to make some good recommendations from a network based on those variables.