Hantz Angrand

Data 620

Homework 4

We will scrap the website Amazon’s Book Review to get the data we need. Book review data will be extracted to create both nodes and edges.

We will use python scraping method to find the User ID, Title, Price and Category. We will establish the following node (users, books) and the following edges:

1. Users Reviewed Book
2. Book Reviewed by User

The book popularity will be measured by degree centrality.

Eigenvector centrality is a measure of how a book importance increases based on how it is connected to other important books.

Betweenness centrality is a count of the number of times a book acts as a bridge the shortest path between two nodes.

We can use t-test or z-score to identify which category produce more sale for the business.