DATA 620 Summer 2023 Final Project Proposal

Josh Forster & Ben Inbar

Data source(s): https://github.com/wayfair/WANDS/tree/main/dataset

https://www.aboutwayfair.com/careers/tech-blog/wayfair-releases-wands-the-largest-and-richest-publicly-available-dataset-for-e-commerce-product-search-relevance

Proposal

For this project, we aim to apply network analysis on a Wayfair-curated dataset (called WANDS) of queries for products on their website. They annotated and analyzed these queries with labels to classify the search results as “Irrelevant,” “Partial,” or “Exact” match depending on the product(s) returned to the customer, which lends itself well to analysis and evaluation of intra-company search engine effectiveness. Specifically, we are hoping to:

1) Build a bipartite network where nodes are queries and products, and then produce a projection of products, potentially applying a weighting of normalized/factorized label, or average rating, or number of queries.

2) Perform PageRank comparison against the provided labels to evaluate the probabilities of a search given a random customer

3) Perform sentiment analysis on the product descriptions provided by the vendors to assess if sentiment correlates with positioning of search results, potentially for comparison with customer ratings or number of queries.