Graph based recommendation - Amazon Product Review

Group 2: Charvi Mayank Shastri (cs2095), Dwiti Milind Choksi (dc1846), Hitanshu Raja Oza (hro24), Om Nirav Patel (op187)

Overview :

This project proposes a **PageRank algorithm** for a graph-based recommendation system to identify and rank related products from the **Amazon Product Review Dataset (McAuley et al., UCSD)** [https://amazon-reviews-2023.github.io/]. By modeling products as nodes and their relationships derived from co-purchase and co-view data as edges, PageRank will be used to determine the rank of each product in the graph. Also we will put in an adaptive element which will allow for quick recommendation updates as we add in new reviews or products which in turn will show that our model scales and responds to real world data.

Why :

In large scale e-commerce settings we see that recommendation is a key to user engagement. Nowadays we only see machine learning models used for these recommendations which may be intensive and do not always provide insight into how the model is working. As compared to this, PageRank puts forth a graph based solution for ranking products based on how they are connected. Also it enables efficient generation of recommendations which do not require training data and at the same time is able to identify the large scale relationships between products.

Goals :

- Build a **product–product graph** using co-purchase or co-view relationships from the dataset.

- Implement the **PageRank algorithm** to compute scores for recommendation.

- Design an **adaptive update mechanism** to efficiently recompute PageRank scores when the graph changes i.e when new product/reviews are added.

- **Visualize** the graph structure and highlight top recommended products for any selected node.

Difficulties :

- **Data preprocessing:** Extracting and merging product relationships from large JSON files.

- **Graph scalability:** Managing large-scale graphs with a large number of nodes and edges efficiently.

- **Adaptive recomputation:** Updating PageRank incrementally rather than recomputing from scratch.

- **Evaluation:** Designing a meaningful metric to validate recommendations without ML-based evaluation criteria.

Visualization :

An interactive visualization will be developed using **NetworkX** and **Plotly/D3.js** to display relationships and PageRank results.
 Features include:

- Display of product nodes with size proportional to PageRank score.

- Highlighting of top recommended products for a selected product node.

- Visualization of changes in node ranking after adaptive updates to the graph.

Novelty :

This project we see as a novel take on the use of the PageRank algorithm for real time product recommendation. While PageRank has been a critical part of web search and network analysis, what we propose is to apply it to e-commerce which does not require machine learning and aligns with the project requirements. We put forward an emphasis on efficiency and adaptability which we will demonstrate via implementation of incremental updates to the PageRank which gives accurate results as the data set grows.

Conclusion :

The project showcases the potential for graph-based approaches in recommendation systems, particularly PageRank, to yield good results. In the course of applying PageRank to the Amazon Product Review Dataset, we will demonstrate how product relationships can be leveraged to derive meaningful recommendations without the need for machine learning models. Adding an adaptive mechanism further will make the system responsive to new data, reflecting real-world dynamics.