# Amazon Products Recommendation with Graph based models

**Xueyuan Li**
Vanderbilt University
xueyuan.li@vanderbilt.edu

**Ruiqui Zhou**
Vanderbilt University
ruiqui.zhou@vanderbilt.edu

### Abstract

This paper introduces a novel recommendation system for Amazon's software products utilizing Graph Neural Networks (GNN) and NetworkX. Leveraging relational data structures, the system creates personalized recommendations to enhance customer satisfaction and drive sales. Using Amazon's 2018 software purchases dataset, the proposed approach integrates user-item interaction data with advanced graph analytics to predict user ratings and identify key product relationships. The GNN framework captures essential user and product characteristics through embedding techniques, predicting potential ratings for unrated products and using cosine similarity measures to find related products. NetworkX is employed to explore and visualize connections between users and central products, enhancing the social dimensions of recommendations. The evaluation on standard metrics such as RMSE and MAE demonstrates improved accuracy and efficiency. The system is further showcased through an interactive web application, providing insights into user behavior and product dynamics, thus offering a scalable, data-driven solution for e-commerce platforms

## Introdution

In this paper, we present a novel recommendation system for Amazon's software products using Graph Neural Networks (GNN) and NetworkX. We leverage relational data structures to create personalized product recommendations (0), aiming to enhance customer satisfaction and increase sales. Our approach integrates user-item interaction data with advanced graph analytics to predict user ratings and identify key product relationships (0) . We evaluate our system using the Amazon 2018 software purchases dataset, demonstrating improvements in prediction accuracy and computational efficiency. Our findings offer insights into user behavior and product dynamics within e-commerce platforms.

## Problem Statement

Our project leverages Graph Neural Networks (GNN) and NetworkX to enhance the personalization and accuracy of Amazon's software product category recommendations. Specifically, our system not only predicts user ratings for products to generate top recommendations but also uses complex relationships between products and users to increase the relevance and satisfaction of these recommendations.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

### Two Core Problem

1. How can the system utilize GNN to predict user ratings for products and provide recommendations based on product similarities? Furthermore
2. How can NetworkX explore potential connections between users and identify as well as visualize the most centrally connected products and users to enrich the social dimensions and depth of the recommendation system?

## Challenges

Accuracy of Rating Predictions: Predicting user ratings for products with GNN involves dealing with sparse user-product interaction data, requiring the model to accurately capture user preferences.

Efficiency in Calculating Similarities: Calculating cosine similarities between items in a large product database demands efficient algorithms and sufficient computational resources to update recommendation lists in real-time.

Complexity in Exploring User Connections: Exploring potential connections between users with NetworkX involves processing and analyzing large volumes of social graph data, presenting challenges in data handling and model training.

Identification and Visualization of Centrally Connected Products and Users: Identifying and visualizing the most densely connected products and users not only requires sophisticated graph analysis techniques but also intuitive visualization tools to help users understand the logic behind recommendations.

Scalability and Flexibility of the System: As the user base and number of products increase, maintaining system performance and responsiveness becomes crucial.

## Methodology

### Graph Neural Network (GNN)

In the methodology section of our project on enhancing Amazon product recommendations using graph-based models, we focused on the application of Graph Neural Networks (GNN). Initially, we prepared the data by utilizing user interaction data, which included ratings and reviews, to form a comprehensive user-item interaction graph. This setup allowed us to map the complex relationships and interactions between users and products effectively.

Following the data preparation, we implemented a GNN architecture designed to learn low-dimensional embeddings for both users and products. These embeddings are crucial as they capture the essential characteristics and preferences of users, as well as the features of the products, based on historical interactions. The learned embeddings facilitated the core functionality of our recommendation system.

For rating prediction and recommendation generation, we utilized these embeddings to predict potential ratings for products that a user has not yet rated. By predicting these ratings, we could rank products according to their likelihood of appealing to a specific user, thereby personalizing the recommendations. Additionally, we employed cosine similarity measures on the product embeddings to identify the top similar products. This approach not only suggested alternative products that users might find appealing but also helped in understanding product relationships within the catalog, enhancing the recommendation system's accuracy and relevance.

### NetworkX

In the methodology section focused on NetworkX for graph-based recommendations, our approach began with the construction of a bipartite graph. This graph structurally represented the users and products as two distinct sets of nodes, connected by edges that signify interactions such as purchases or ratings. This model effectively captured the complex web of relationships between users and the products they interact with.

Following the construction of the graph, we engaged in link prediction and centrality analysis to enhance the system's recommendation capabilities. By predicting potential links between users based on shared behaviors and product interactions, we could suggest new connections and discover influential users and products. Centrality analysis further assisted in identifying the most connected nodes within the graph, which often represent the most influential products or active users, providing valuable insights for targeted marketing and product placement strategies.

Lastly, we implemented a visualization component to make the intricate relationships within the graph more comprehensible. This visualization highlighted the key network components and their interconnectivities, offering a clearer understanding of the underlying structure of the user-product relationships. This not only served as a tool for analyzing the network's dynamics but also as a compelling visual aid in presentations and strategic discussions, making the complex data accessible and actionable.

## Experiments

### Data Source and Selection

The data originates from the Amazon 2018 dataset, which was compiled and released by Jianmo Ni, Jiacheng Li, and Julian McAuley in their 2019 research. This dataset is commonly used for analysis and model training, especially in the fields of recommendation systems and user behavior analysis.

To handle the large scale of the dataset and limitations in computational resources, the team utilized a method known as ”5-kcore selection” to extract a subset of the dataset(same method as reference paper). This method ensures that each selected data point is connected to at least five other points, thereby maintaining a degree of richness and connectivity in the graph structure. This helps to improve the learning efficiency and prediction accuracy of the graph neural network models.

Additionally, the data is divided into user nodes and product nodes, and we built relationships between them are established, such as ”Reviews” edges (connecting user nodes to product nodes, with attributes indicating the user's rating), ”Also Bought” edges (connecting product nodes to represent products frequently purchased together), ”Descriptions” edges (basic information about each product, including features and specifications), ”Category” ( the specific category to which each product belongs) and ”Ratings” (numerical ratings given to products by users, used for evaluating product popularity and user satisfaction).

Our dataset includes two part, ”Reviews” (12,805 entries capturing user feedback and ratings), ”Metadata” (26,790 entries detailing product information), which contains graph elements comprising 21,639 product nodes, 1,826 user nodes, and various edges representing relationships such as purchases, textual comments, and product features.

### Performance Metrics

We evaluate the models based on four key metrics: RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), Precision, and Recall.

**RMSE and MAE**: These metrics quantify the average magnitude of the errors between predicted and actual ratings. RMSE penalizes larger errors more heavily than MAE, providing a measure of the model's accuracy in rating prediction.

**Precision**: Precision measures the proportion of relevant items among the recommended items. It reflects the ability of the model to recommend items that users are likely to rate positively.

**Recall**: Recall measures the proportion of relevant items that have been recommended out of all the relevant items. It indicates the model's ability to capture all relevant items in its recommendations.

## Results

### Evaluation Results

In this section, we present the results of the evaluation of our Amazon product recommendation system using Graph Neural Network (GNN) and Graph-based Network implemented with NetworkX. We analyze the performance of the models based on key metrics including RMSE, MAE, Precision, and Recall.

The evaluation results for both the training and test phases are summarized in Table 1 below:

| Data | Model | RMSE | MAE | Precision | Recall |
|---|---|---|---|---|---|
| Train | GNN | 0.2181 | 0.1695 | 0.8738 | 0.9797 |
| | NetworkX | 1.1165 | 0.8223 | 0.8898 | 0.6889 |
| Test | GNN | 0.7765 | 0.6567 | 0.8353 | 0.7167 |
| | NetworkX | 1.2561 | 0.9198 | 0.8317 | 0.6523 |

Table 1: Summary of evaluation results for both training and testing datasets.

Table 1 presents the evaluation results for both the training and test phases. In both phases, the GNN outperforms the NetworkX-based model across all metrics. The GNN demonstrates lower RMSE and MAE, indicating higher accuracy in rating prediction, as well as higher Precision and Recall, suggesting its effectiveness in recommending relevant items to users based on their preferences.

## Web App Demo

We utilized Streamlit, a Python library for creating web applications, to develop an interactive platform for our Amazon product recommendation system. The web app allows users to input their user ID, log in, and view relevant information tailored to their preferences.

Upon entering their user ID and logging in, the user is presented with three key components:

Figure 1 shows the user's purchase history, providing insights into their past interactions with the platform. Figure 2 showcases the top 10 recommended products for the user, based on our recommendation algorithm's predictions. Figure 3 presents a node plot depicting the user, top recommended products (indicated by red nodes), and top similar products for these recommended items (indicated by green nodes).

This interactive web app enhances user engagement and facilitates a personalized shopping experience by leveraging the power of our recommendation system.

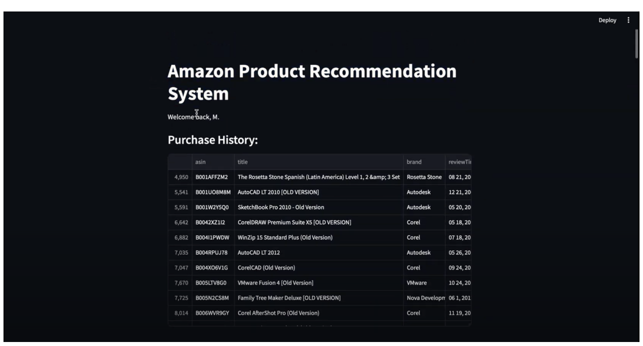


Figure 1: User's purchase history.

# Discussion

In the discussion section of our project on enhancing Amazon product recommendations using graph-based models,

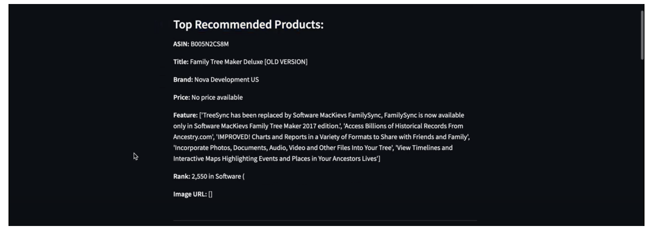


Figure 2: Top 10 recommended products for the user.

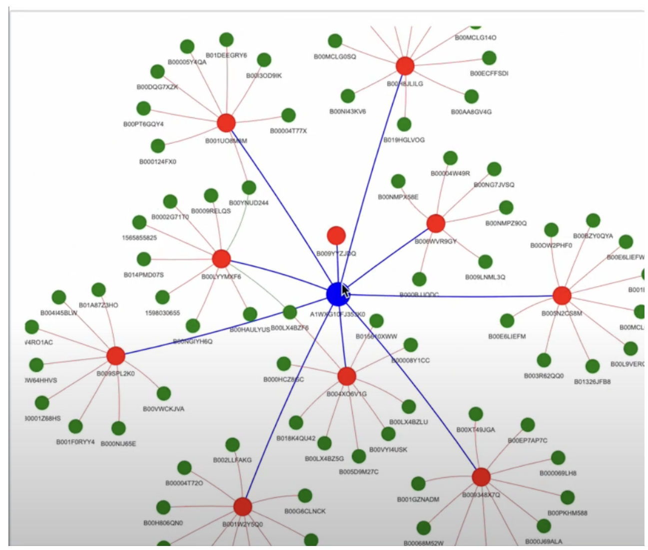

Figure 3: Node plot depicting user, recommended products, and similar products.

we delve into the implications, strengths, and limitations of the methodologies employed, namely Graph Neural Networks (GNN) and NetworkX. This analysis not only sheds light on the effectiveness of these technologies in the context of e-commerce but also opens avenues for further research and application improvements. The application of GNN in our project has demonstrated significant potential in accurately modeling user preferences and product features through sophisticated embedding techniques. The ability of GNN to learn from the rich, interconnected data provided by user-item interactions allows for highly personalized recommendations. These personalized recommendations have the potential to increase customer satisfaction and retention, as users receive suggestions that are closely aligned with their interests.

Similarly, the use of NetworkX to create and analyze a bipartite graph of users and products has proven instrumental in understanding the complex network dynamics within large datasets. Through centrality analysis and link prediction, we could identify key influencers and central products, which are critical for strategic marketing and inventory management. The visualization of network data further aids in comprehending the intricate relationships and flow of influence within the marketplace, making it a valuable tool for decision-makers.

## Conclusion

Our project successfully integrates Graph Neural Networks (GNN) and NetworkX to enhance recommendation systems for Amazon's software product category. By predicting user ratings with GNN, we provide personalized top 10 product recommendations and identify top 10 similar products based on cosine similarity. Using NetworkX, we explore potential connections between users, identifying and visualizing products and users with the highest centrality in the network. This approach has allowed us to create a more connected, personalized, and context-aware recommendation system that leverages both user-item interactions and the relational structure of the data

## Future Work

1, Model Optimization: We will continue to refine our algorithms to improve the accuracy of rating predictions and the relevance of product similarities. This includes tuning model parameters and exploring newer GNN architectures.

2, Enhanced User Insights: By delving deeper into graph analytics, we aim to uncover more nuanced user behaviors and preferences, enhancing the complexity of our recommendation logic.

3, Scalability Improvements: As we scale our system to contain more users and products, optimizing data processing and model performance will be importrant . We plan to implement distributed computing techniques to handle this growth.

4, Broader Product Categories: Expanding our recommendation system to include more of Amazon's product categories, adapting our models to diverse data characteristics and improve user interaction patterns.

## References


Yuliana Lavryk and Yurii Kryvenchuk. Product recommendation system using graph neural network. *Lviv Polytechnic National University, Stepana Bandery Street 12, Lviv, 79013, Ukraine*, 2023.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983. ACM, 2018.