Amazon Product Prediction

SNA Project

Group 10

Dataset

- The dataset is taken from SNAP and titled as *"Amazon product co-purchasing network metadata"*
- The data was collected by crawling Amazon website and contains product metadata and review information about 548,552 different products (Books, music CDs, DVDs and VHS video tapes).
-

Dataset statistics	
Products	548,552
Product-Project Edges	1,788,725
Reviews	7,781,990
Product category memberships	2,509,699
Products by product group	
Books	393561
DVDs	19828
Music CDs	103144
Videos	26132

Data Format

1. **ID**: Product id (number 0, ..., 548551)
2. **ASIN**: Amazon Standard Identification Number
3. **Title**: Name/title of the product
4. **Group**: Product group (Book, DVD, Video or Music)
5. **Salesrank**: Amazon Sales Rank
6. **Similar**: ASINs of co-purchased products (people who buy X also buy Y)
7. **Categories**: Location in product category hierarchy to which the product belongs (separated by |, category id in [])
8. **Reviews**: Product review information: *time, user id, rating, total number of votes on the review, total number of helpfulness votes* (how many people found the review to be helpful)

Objective and Approach

- The main objective of this project is to do product prediction for Amazon products using graphs and networks as foundational concepts.
- Through the dataset we have found we wish to explore If one person brought X will they buy Y? We will also explore the likeness of a product through review's, as in people who like X will they also like Y?
- We will make Bipartite graph between different attributes such as title, sales rank and different categories to make predictions.
- We will also make unipartite graph between different attributes, as we move ahead, we will explore more and will be able to find more specific questions to answer.

Review of Literature

1. Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. ACM Transactions on the Web (TWEB), 1(1), 5-es.
2. Bag, S., Tiwari, M. K., & Chan, F. T. (2019). Predicting the consumer's purchase intention of durable goods: An attribute-level analysis. Journal of Business Research, 94, 408-419.
3. Pham, M. C., Cao, Y., Klamma, R., & Jarke, M. (2011). A clustering approach for collaborative filtering recommendation using social network analysis. J. Univers. Comput. Sci., 17(4), 583-604.
4. link prediction on amazon product network (Existing Project on SNAP)
5. Mohan, A., Venkatesan, R., & Pramod, K. V. (2017). A scalable method for link prediction in large real world networks. Journal of Parallel and Distributed Computing, 109, 89-101.

Group Members

Sr No	Name	Enrollment Number
1	Gagan Sankhla	AU1940115
2	Kushal Gandhi	AU1940147
3	Aatman Vaidya	AU1940108