UE19CS345 – Network Analysis and Mining
Course Project

Project Title : Graph Based Recommendation system for amazon products

Team member 1: Bhavikha Chopra
PES1UG19CS111
Team member 2: Gitika Jain
PES1UG19CS169
Team member 3: N. Mahendra
PES1UG19CS285
Team member 4: Snigdha Sinha
PES1UG19CS491

Topic and its uniqueness

- Amazon Recommendation with Social Network Analysis based on co-purchase data: this method works well for all sorts of item groups in the dataset eg. Books, DVDs, Music CDs, etc
- The advantage of this approach is that the graph needs to be built only once. After that, it’s very easy to add products to it by just adding the node and edges to other products. Also the search algorithm is the online feature and is fast in returning recommendations.

Dataset

The dataset contains product metadata and review information for about 548,552 different products. For each product the following information is available:

- Product ID: numeric values (0,1,2….,548,551)
- ASIN: Stands for Amazon Standard Identification Number is a 10-character alphanumeric UID assigned by Amazon for the identification of product.
- Title: Name of the Product.
- Group: Product type, could be books, music CDs, DVDs and VHS video tapes.

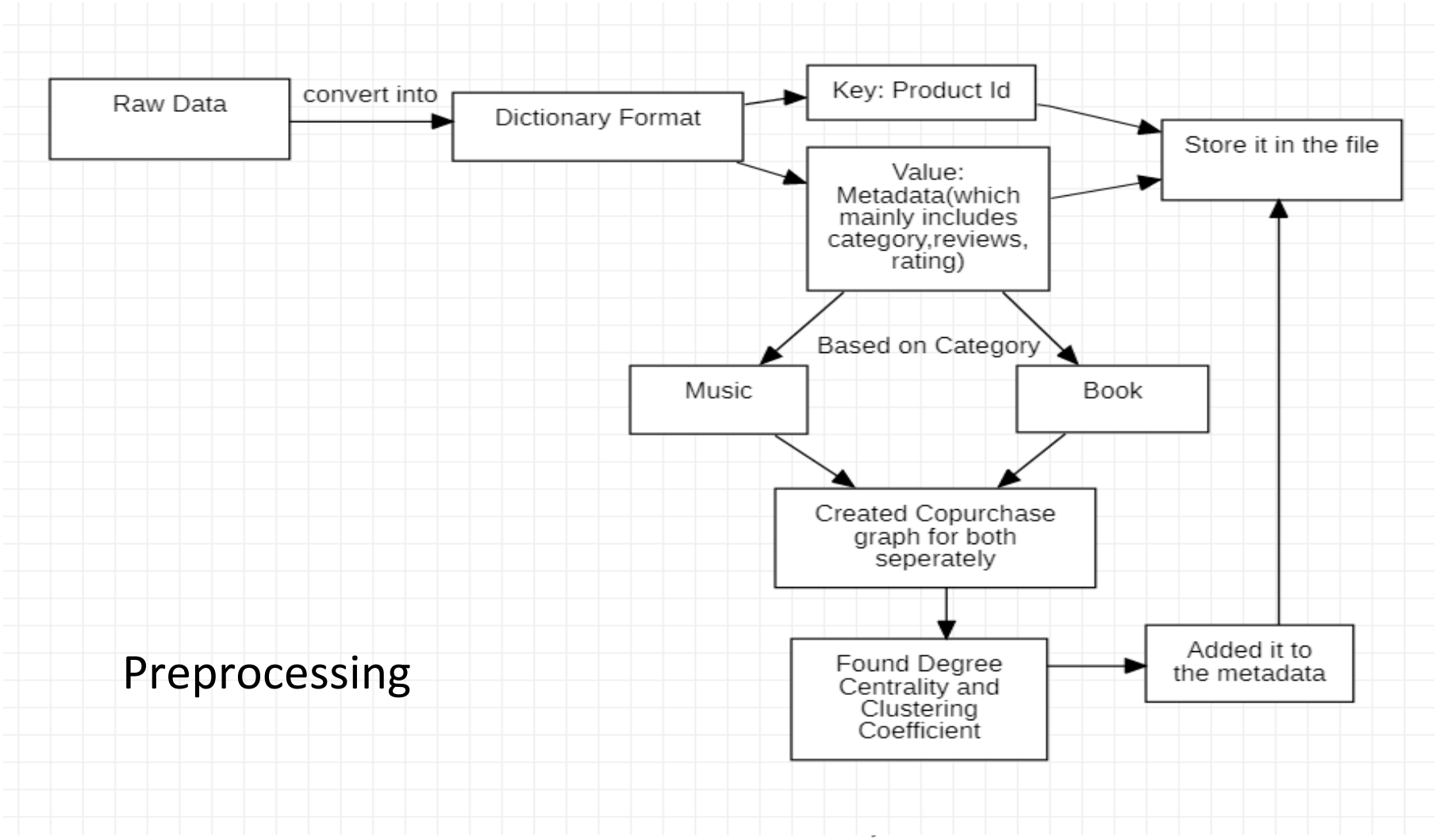
Dataset

- SalesRank: Representation of the sales of that product compared to the others in its category.
- Similar: ASINs of co-purchased products i.e., people who buy A also buy .
- Categories: Gives the specification of the product's category hierarchy, e.g., genre etc. (separated by |, category id in [ ]).
- Reviews: Product's review information will have Total Number of Reviews,Average Rating and individual customer review with time, user id, rating, total number of votes on the review and the number of people who found the review helpful.

Overall design or approach in a free hand diagram

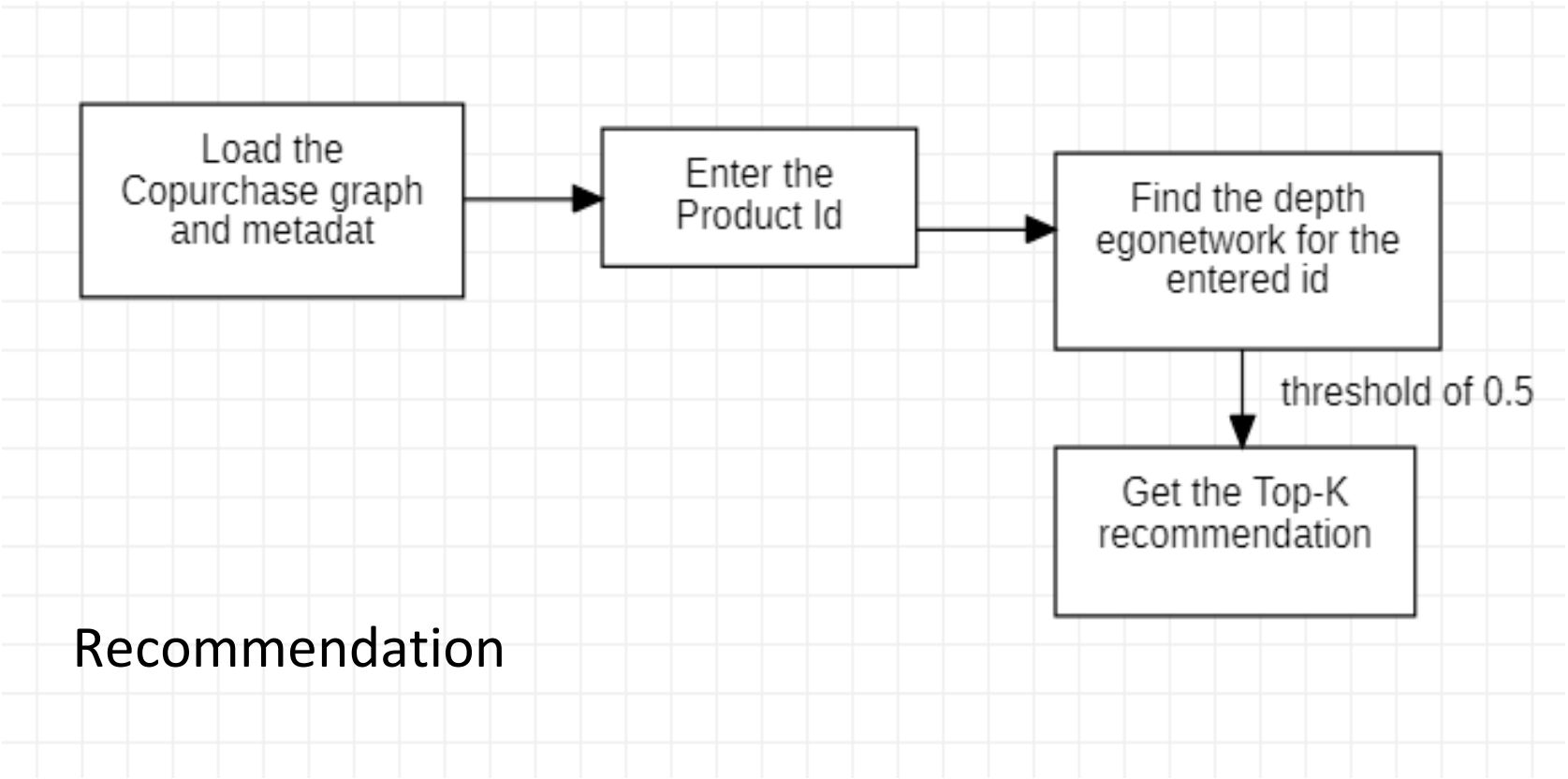
- Our Design consists of two parts:
  - Preprocessing
  - Recommendation

Overall design or approach in a free hand diagram



Preprocessing

Overall design or approach in a free hand diagram



Recommendation

Final results

Output for book recommendation:

```
Top  Recommendation by AvgRating then by TotalReviews for Users Purchased the book:

-------------------------------------------------------------------------------
ASIN     Title   SalesRank       TotalReviews    AvgRating       DegreeCentrality        ClusteringCoeff
('1559361247', 'Sex, Drugs, Rock & Roll', 444861, 3, 5.0, 7, 0.9)
('1559361018', 'Suburbia', 418461, 9, 4.5, 12, 0.66)
('1559360968', 'Pounding Nails in the Floor With My Forehead', 473331, 6, 4.5, 7, 0.9)
('1559360828', 'The Essential Bogosian: Talk Radio, Drinking in America, Funhouse & Men Inside', 384133, 1, 4.0, 7, 0.9)
```

Final results

Output for music recommendation:

```
Top  Recommendation by AvgRating then by TotalReviews for Users listening the music:

-------------------------------------------------------------------------------
ASIN      Title    SalesRank        TotalReviews     AvgRating        DegreeCentrality        ClusteringCoeff
('B00000JQIE', 'The Modern World', 29836, 5, 5.0, 2, 0.0)
```

Our final result is a recommendation by combining 2 similarity metrics namely AvgRating and TotalReviews.

What are the remaining portions in this project ?

- Given more time to work on the project, we had a few ideas to further implement. One of them is to perform NLP on the review text to extract sentiment from each review. We would attribute a score to how positive or negative a review is, which would help us in recommending better products.
- We provided recommendations for amazon books and music and couldn’t do for other products because of lack of data.

Quantity and quality of work

no	Code functionality	% Complete	Runs without problem (Y/N)	If there are minor issues, indicate
1	Separation of books from co-purchased list	100	Y	NO
2	Plotted co-purchase graph for analysis.	100	Y	NO
3	Adding degree centrality and clustering coefficients of each ASIN.	100	Y	NO
4	To perform NLP on the review text to extract sentiment from each review.	25	N	The work is yet to be completed.
5	Printing the top 5 recommendations using similarity metrics like avg rating and total reviews	100	Y	NO

Top few learning

Serial No	Top learning in this project
1	Graph Based Recommendation
2	Use of ego networks for analysis

Top unresolved challenges

Serial No	Brief description of unresolved challenges	Type of challenge (scope/data/design/implementation / others)
1	Performing NLP on the review text to extract sentiment from each review. And having a score to how positive or negative a review is to recommend better products.	Scope
2	Use of other similarity metrics for recommendation	implementation
3	Providing recommendation for other amazon products.	Others - Our dataset had very less index on products like DVD because of which we couldn't do recommendations on these products.

Reference papers, if any

No	Paper Title	Authors
1	Graph-Based Recommendation of Amazon Products	Aaron Effron, Kelly Shen, Ryan Mui
2	A graph-based recommender system for digital library	Zan Huang,Wingyan Chung,Hsiu-chin Chen
3	Graph Based Web Recommendation System to Improve Time Efficiency	Shah Ankur , Varachhia Hemraj
4	Graph based recommendation engine for Amazon products	Mohtadi Ben Fraj

PES UNIVERSITY

THANK YOU