

# Individual Project Report

## CMPE 256: Large Scale Analytics, Summer 2019

### Book Recommendation System

Name: Mahesh Reddy Konatham  
SJSU ID: 013823095

#### Abstract:

This project aims at recommending books for users from amazon by their past purchase history using graph theory (network analysis). Searching for a book becomes tedious at times when we have to go through hundreds of books through many pages. This problem can be solved if we have a very good book recommendation system which considers user's purchase history i.e. the books consumers purchased collectively or together in past. Concept of graph theory is used to make books as vertices and edges are formed between those books which were purchased together, and weighting is given based on the similarity.

#### Data collection:

The data for the Book Recommendation system was crawled from Amazon products website for books. This data has fields which includes Id: product id, ASIN (amazon standard identification number) is a ten-digit unique number for every amazon product. Title: Name of the product, sales-rank: overall sales rank of the product, Group: The division of the product like Electronic/book ..etc., similar: ASINs of the products which are purchased together, categories: location of hierarchy in which these product is located and product review information.

```
Id: 15
ASIN: 1559362022
title: Wake Up and Smell the Coffee
group: Book
salesrank: 518927
similar: 5 1559360968 1559361247 1559360828 1559361018 0743214552
categories: 3
|Books[283155]|Subjects[1000]|Literature & Fiction[17]|Drama[2159]|United States[2160]
|Books[283155]|Subjects[1000]|Arts & Photography[1]|Performing Arts[521000]|Theater[2154]
|Books[283155]|Subjects[1000]|Literature & Fiction[17]|Authors, A-Z[70021]|( B ) [70023]
reviews: total: 8 downloaded: 8 avg rating: 4
2002-5-13 customer: A2IGOA66Y6O8TQ rating: 5 votes: 3 helpful: 2
2002-6-17 customer: A2OIN4AUH84KNE rating: 5 votes: 2 helpful: 1
2003-1-2 customer: A2HN382JNT1CIU rating: 1 votes: 6 helpful: 1
2003-6-7 customer: A2FDJ79LDU4O18 rating: 4 votes: 1 helpful: 1
2003-6-27 customer: A39QMV9ZKRJXO5 rating: 4 votes: 1 helpful: 1
2004-2-17 customer: AUUVMSTQ1TXDI rating: 1 votes: 2 helpful: 0
2004-2-24 customer: A2C5K0QTL9UAT rating: 5 votes: 2 helpful: 2
2004-10-13 customer: A5XYF0Z3UH4HB rating: 5 votes: 1 helpful: 1
```

Fig 1. shows the data format before pre-processing

## Data Pre-processing:

Data pre-processing has to be done in order to clean the data before using for recommendation. Some of the fields remain same as id, ASIN, Title but others like group is filtered to only book, categories are stemmed, lemmatized and stop words are removed, purchased\_Together is made from similar ASINs but those which are similar to the product. Total reviews and average rating remain same.

Also, the purchased\_Together data in amazon books file is used to make a Purchased\_Together Graph structure which has books (ASIN) as Nodes, an Edge exists if the books are purchased together in past and also the weight will be determined by similarity. After the Purchased\_Together Graph structure is made Degree- centrality and Clustering co-efficient are also calculated for each Node in the Graph.

Id	ASIN	Title	Categories	Group	SalesRank	TotalReviews	AvgRating
		DegreeCentrality	ClusteringCoeff				
1	0827229534	Patterns of Preaching: A Sermon Sampler	preach subject clergi				
		sermon spiritu book christian religion	Book	396585	2	5.0	8
2	0738700797	Candlemas: Feast of Flames	wicca subject witchcraft spiritu				
		book base earth religion	Book	168596	12	4.5	9
3	0486287785	World War II Allied Fighter Planes Trading Cards	subject				
		home general book hobbi garden craft	Book	1270652	1	5.0	0
4	0842328327	Life Application Bible Commentary: 1 and 2 Timothy and Titus					
		life spiritu general histori commentari subject translat discipleship applic live sacr					
		religion refer bibl studi new book text guid christian testament	Book	631289	1		
		4.0	6	0.79			
5	1577943082	Prayers That Avail Much for Business: Executive	prayerbook busi				
		live subject worship spiritu book christian devot religion	Book	455160	0		
		0.0	4	1.0			
6	0486220125	How the Other Half Lives: Studies Among the Tenements of New York					
		general histori essay nonfict camera urban jewish subject scienc state photo					
		categori unit america social book sociolog art photographi	Book	188784	17		
		4.0	4	0.8			
8	0231118597	Losing Matt Shepard	crimin criminolog u subject general polit				
		book crime gay nonfict lesbian	Book	277409	15	4.5	6
							0.59

**Fig2:** shows the data format after pre-processing

## Methodology:

### Calculating Edge weight:

Edge weight for the edges between nodes of purchased\_Together graph is calculated through two ways

a) Based on content of categories:

Similarity between two books in purchased together data is calculated by taking the number of words common in both categories and dividing it with the total number of words in both categories combined.

**creating purchased together Graph structure using similarities between categories**

```
1 purchased_TogetherGraph = networkx.Graph()
2 for asin,RawData in amazonBooks.items():
3     purchased_TogetherGraph.add_node(asin)
4     for a in RawData['purchased_together'].split():
5         purchased_TogetherGraph.add_node(a.strip())
6         similarity = 0
7         x = set((amazonBooks[asin]['Categories']).split())
8         y = set((amazonBooks[a]['Categories']).split())
9         xIy = x & y
10        xUy = x | y
11        if (len(xUy)) > 0:
12            similarity = round(len(xIy)/len(xUy),2)
13        purchased_TogetherGraph.add_edge(asin, a.strip(), weight=similarity)
```

b) Based on content of Titles:

Similarity between two books in purchased together data is calculated by taking the number of words common in both Titles and dividing it with the total number of words in both Titles combined.

## Creating purchased together Graph structure using similarities between Titles

```
1 purchased_TitleGraph = networkx.Graph()
2 for asin,RawData in amazonBooks.items():
3     purchased_TogetherGraph.add_node(asin)
4     for a in RawData['purchased_together'].split():
5         purchased_TogetherGraph.add_node(a.strip())
6         similarity = 0
7         x = set((amazonBooks[asin]['Title']).split())
8         y = set((amazonBooks[a]['Title']).split())
9         xIy = x & y
10        xUy = x | y
11        if (len(xUy)) > 0:
12            similarity = round(len(xIy)/len(xUy),2)
13        purchased_TitleGraph.add_edge(asin, a.strip(), weight=similarity)
```

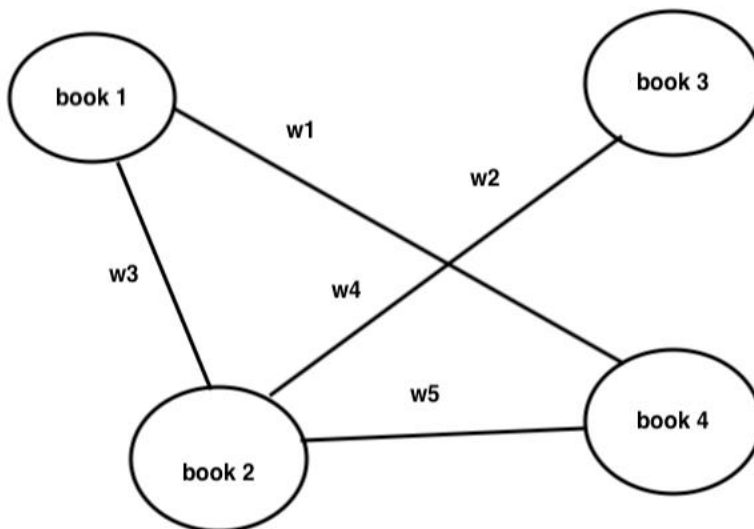


Fig 3: shows the Graph structure of the Purchased\_Together

### Recommendations:

The python function takes the ASIN of the purchased book by user and the threshold to which user want the similarity between the purchased and the recommendations as input and then first locates the direct neighbors of the purchased book and then filters based on Average customer ratings and total number of ratings. Finally, after sorting the results recommendations are made based on threshold values and method of similarity calculation.

Fig 4: shows recommendation for **ASIN: '1577943082'** with similar **CATEGORY** and **Threshold 0.3**

```
[3]: 1 recommend_book('1577943082',0.3)

Recommendations for User Purchasing the Book with:

ASIN = 1577943082
Title = Prayers That Avail Much for Business: Executive
SalesRank = 455160
TotalReviews = 0
AvgRating = 0.0

Recommendation : 1
  Asin = 0892749563
  Title = Prayers That Avail Much for Kids: Short and Simple Prayers Packed With the Power of God's Word
  SalesRank = 298410
  TotalReviews = 1
  AvgRating = 5.0

Recommendation : 2
  Asin = 1577941829
  Title = Prayers That Avail Much for Men
  SalesRank = 115696
  TotalReviews = 0
  AvgRating = 0.0

Recommendation : 3
  Asin = 157794349X
  Title = Prayers That Avail Much for the Workplace: The Business Handbook of Scriptural Prayer
  SalesRank = 221617
  TotalReviews = 0
  AvgRating = 0.0
```

Fig 5: shows recommendation for ASIN: '1577943082' with similar **CATEGORY** and **Threshold 0.4**

```
: 1 recommend_book('1577943082',0.4)

Recommendations for User Purchasing the Book with:

ASIN = 1577943082
Title = Prayers That Avail Much for Business: Executive
SalesRank = 455160
TotalReviews = 0
AvgRating = 0.0

Recommendation : 1
  Asin = 1577941829
  Title = Prayers That Avail Much for Men
  SalesRank = 115696
  TotalReviews = 0
  AvgRating = 0.0

Recommendation : 2
  Asin = 157794349X
  Title = Prayers That Avail Much for the Workplace: The Business Handbook of Scriptural Prayer
  SalesRank = 221617
  TotalReviews = 0
  AvgRating = 0.0
```

---

Fig 6: shows recommendation for ASIN: '1577943082' with similar **CATEGORY** and **Threshold 0.5**

```
1 recommend_book('1577943082',0.5)
```

Recommendations for User Purchasing the Book with:

```
ASIN = 1577943082
Title = Prayers That Avail Much for Business: Executive
SalesRank = 455160
TotalReviews = 0
AvgRating = 0.0
```

```
Recommendation : 1
    Asin = 1577941829
    Title = Prayers That Avail Much for Men
    SalesRank = 115696
    TotalReviews = 0
    AvgRating = 0.0
```

---

Fig 7: shows recommendation for **ASIN: '1577943082'** with similar **TITLE** and **Threshold 0.2(default)**

1	<code>recommend_book_Title('1577943082')</code>
Recommendations for User Purchasing the Book with:	
ASIN = 1577943082 Title = Prayers That Avail Much for Business: Executive SalesRank = 455160 TotalReviews = 0 AvgRating = 0.0	
Recommendation : 1 Asin = 0892749563 Title = Prayers That Avail Much for Kids: Short and Simple Prayers Packed With the Power of God's Word SalesRank = 298410 TotalReviews = 1 AvgRating = 5.0	
Recommendation : 2 Asin = 1577941829 Title = Prayers That Avail Much for Men SalesRank = 115696 TotalReviews = 0 AvgRating = 0.0	
Recommendation : 3 Asin = 157794349X Title = Prayers That Avail Much for the Workplace: The Business Handbook of Scriptural Prayer SalesRank = 221617 TotalReviews = 0 AvgRating = 0.0	

---

Fig 8: shows recommendation for **ASIN: '0805047905'** with similar **CATEGORY** and **Threshold 0.2(default)**

---

```
: 1 recommend_book('0805047905')

Recommendations for User Purchasing the Book with:
ASIN = 0805047905
Title = Brown Bear, Brown Bear, What Do You See?
SalesRank = 171
TotalReviews = 172
AvgRating = 5.0
Recommendation : 1
    Asin = 0694003611
    Title = Goodnight Moon (Board Book)
    SalesRank = 156
    TotalReviews = 339
    AvgRating = 4.5
Recommendation : 2
    Asin = 0399226907
    Title = The Very Hungry Caterpillar board book
    SalesRank = 279
    TotalReviews = 164
    AvgRating = 4.5
Recommendation : 3
    Asin = 068983568X
    Title = Chicka Chicka Boom Boom (Aladdin Picture Books)
    SalesRank = 885
    TotalReviews = 89
    AvgRating = 5.0
Recommendation : 4
    Asin = 0060094273
    Title = Goodnight Moon Board Book & Baby Socks
    SalesRank = 21825
    TotalReviews = 339
    AvgRating = 4.5
Recommendation : 5
    Asin = 0060207051
    Title = Goodnight Moon
    SalesRank = 1614
    TotalReviews = 339
    AvgRating = 4.5
```

---



Fig 9: shows recommendation for **ASIN: '0805047905'** with similar **TITLE** and **Threshold 0.2(default)**

1	<code>recommend_book_Title('0805047905')</code>
Recommendations for User Purchasing the Book with:	
ASIN = 0805047905	
Title = Brown Bear, Brown Bear, What Do You See?	
SalesRank = 171	
TotalReviews = 172	
AvgRating = 5.0	
Recommendation : 1	
Asin = 0805053883	
Title = Polar Bear, Polar Bear, What Do You Hear?	
SalesRank = 2515	
TotalReviews = 29	
AvgRating = 4.5	
Recommendation : 2	
Asin = 0805023461	
Title = Polar Bear, Polar Bear, What Do You Hear? (Henry Holt Big Books)	
SalesRank = 213617	
TotalReviews = 29	
AvgRating = 4.5	
Recommendation : 3	
Asin = 0805017593	
Title = Polar Bear, Polar Bear, What Do You Hear?	
SalesRank = 23211	
TotalReviews = 29	
AvgRating = 4.5	
Recommendation : 4	
Asin = 0805017585	
Title = Panda Bear, Panda Bear, What Do You See?	
SalesRank = 16262	
TotalReviews = 6	
AvgRating = 4.5	
Recommendation : 5	
Asin = 0849976146	
Title = Adam, Adam What Do You See?	
SalesRank = 523223	
TotalReviews = 3	
AvgRating = 5.0	

## **REFERENCES:**

1. Book recommendation system using opinion mining technique : [Shahab Saquib Sohail](#) ; [Jamshed Siddiqui](#) ; [Rashid Ali](#)  
<https://ieeexplore.ieee.org/abstract/document/6637421>
2. Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining : [Anand Shanker Tewari](#) ; [Abhay Kumar](#) ; [Asim Gopal Barman](#)  
<https://ieeexplore.ieee.org/abstract/document/6779375>
3. The state-of-the-art in personalized recommender systems for social networking  
<https://link.springer.com/article/10.1007/s10462-011-9222-1>