

# Amazon Product Prediction

SNA Project

Group 10

# Objective and Problem Statement

- ❖ The main objective of this project is to do product prediction for Amazon products using graphs and networks as foundational concepts.
- ❖ We have created a simple book recommendation system and a created a graph.

# Review of Literature

- KJ Kim, H Ahn - Hybrid social network analysis and collaborative filtering approach to enhance the performance of recommender systems.
  - Collected a basic information data (gender,age,address) from a university of korea.
  - Collected the top 100 movies references among them and analyse their online relationship and the rating of movies.
- Leskovec, J., Adamic, L. A., & Huberman - The Dynamics of Viral Marketing
  - Program run by large retailer in which the data which they collected is come from user only by giving them some sort of reward.
  - And analyse that data which came from 4m people who give 16m recommendation.
- Types of Recommendation Systems
  - Content Based
  - Collaborative Filtering Based

# 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)

# Approach - Post Mid Semester

## Data Pre-Processing

- Convert the data into an RDBMS structure and filter books
- Format
  - ID
  - ASIN
  - Title
  - Categories
    - Concatenated everything
    - Digits and Stop Words Removed.
  - Co - Purchased
    - Similar filtered to only those ASIN that have metadata associated with it.

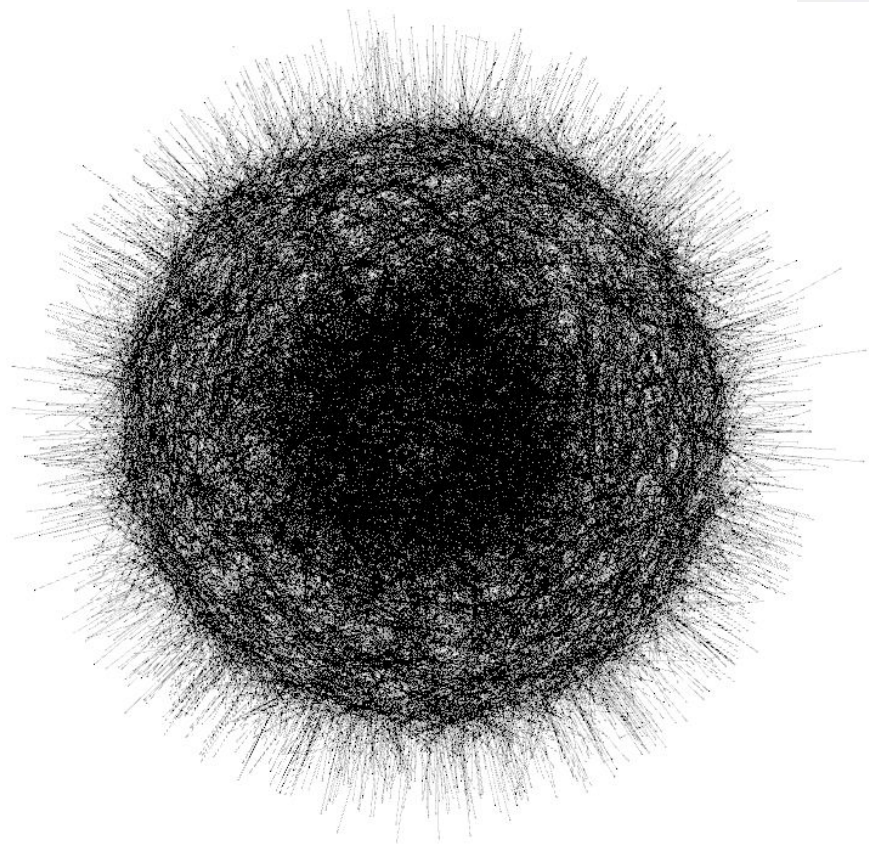
## Co-purchase Graph Structure

- **Node** - ASIN
- **Edge** - connection between the two ASIN.
- **Edge Weight** - Based on Similarity
- **Similarity** - (Common words in Categories)/(Total no. of words in Categories)
- Degree Centrality
- Clustering Coefficient

## Recommendation

- Plotting all the connections
- Then looking at just a single ASIN
- Plotting its EGO Graph
- For a better output, putting the threshold to 0.5
- Sorting the top 5 by Average Rating and then by Total Reviews.
- Then giving the top 5 recommendations for the ASIN.

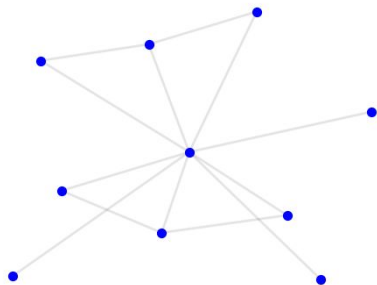
# Results



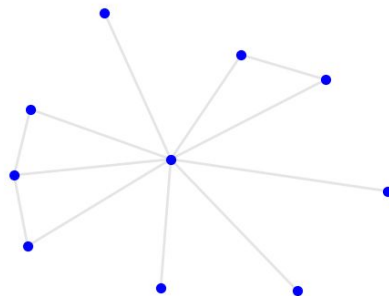
Context ×	—
<b>Nodes:</b>	28787
<b>Edges:</b>	25816
Directed Graph	

## ASIN - 0764565168

Degree-1 Ego Network



Degree-1 Ego Network Trimmed using threshold of 0.5



0471346608,0764565168,0.8  
0764565168,0786885564,0.88  
0764565168,0060937645,0.67  
0764565168,0789205386,0.89  
0764565168,0805069984,0.5  
0764565168,0060394064,0.54  
0764565168,0884195260,0.65  
0764565168,1565302656,0.94  
0764565168,1555612601,0.75

```
-----
ASIN = 0764565168
Title = The Mother of All Pregnancy Books: The Ultimate Guide to Conception, Birth, and Everything In Between (U.S. Edition)
SalesRank = 658
TotalReviews = 90
AvgRating = 4.5
DegreeCentrality = 9
ClusteringCoeff = 0.54
```

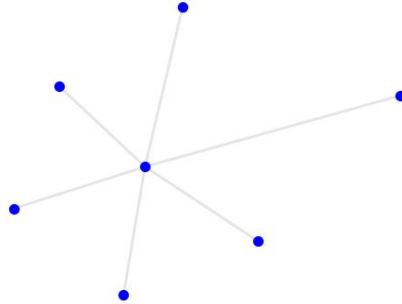
Top 5 Recommendations by AvgRating then by TotalReviews for Users Purchased the book:

```
-----
ASIN    Title    SalesRank    TotalReviews    AvgRating    DegreeCentrality    ClusteringCoeff
('0060937645', 'Taking Charge of Your Fertility: The Definitive Guide to Natural Birth Control, Pregnancy Achievement, and Reproductive Health (Revised Edition)', 265, 701, 5.0, 4, 0.43)
('0060394064', 'Taking Charge of Your Fertility: The Definitive Guide to Natural Birth Control, Pregnancy Achievement, and Reproductive Health (Revised Edition)', 290410, 701, 5.0, 2, 1.0)
('1565302656', '1000 Questions About Your Pregnancy', 307833, 30, 5.0, 2, 1.0)
('0471346608', 'Baby Basics : A Guide for New Parents', 502928, 5, 5.0, 1, 0.0)
('0884195260', 'Healthy Expectations: Preparing a Healthy Body for a Healthy Baby', 903087, 1, 5.0, 1, 0.0)
```

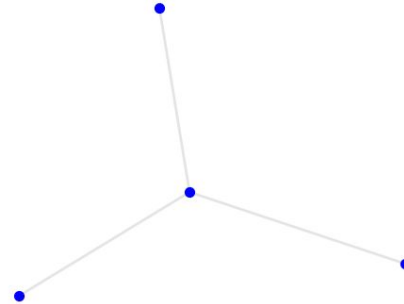


## ASIN - 0395742129

Degree-1 Ego Network



Degree-1 Ego Network Trimmed using threshold of 0.5



```
-----  
ASIN = 0395742129  
Title = A Turkey for Thanksgiving  
SalesRank = 273599  
TotalReviews = 7  
AvgRating = 4.5  
DegreeCentrality = 6  
ClusteringCoeff = 0.0
```

Top 5 Recommendations by AvgRating then by TotalReviews for Users Purchased the book:

```
-----  
ASIN    Title    SalesRank    TotalReviews    AvgRating    DegreeCentrality    ClusteringCoeff  
( '1890817910', 'A Plump and Perky Turkey', 428657, 2, 5.0, 1, 0.0)  
( '0689853181', 'Thanksgiving Day at Our House : Thanksgiving Poems for the Very Young', 1048448, 1, 5.0, 1, 0.0)  
( '080756110X', '1, 2, 3, Thanksgiving', 612658, 0, 0.0, 2, 0.0)  
PS D:\SNA\Social-Network-Analysis> _
```

# Conclusion and Future Work

- This basically gives a simple recommendation system that recommends 5 books based on a certain book.
- In the future:
  - We can look at connections between customers and the products that they reviewed, as in people who like X will they also like Y.
  - Create recommendation network for other groups and products
  - Link Prediction can be done on the product network.
  - We can look at connections between groups (products) and customers.

# References

- Agrawal-Rohit. (n.d.). *Agrawal-Rohit/Amazon-Books-graph-based-recommendation: Amazon Books recommendation using Graph Ego Networks*. GitHub. Retrieved April 30, 2022, from <https://github.com/agrawal-rohit/amazon-books-graph-based-recommendation>
- Ben.k. (2018, June 11). *Book recommendation (python)*. Medium. Retrieved April 30, 2022, from <https://medium.com/@baemaek/amazon-book-recommendation-system-analysis-d9d72b9a7173>
- Sivapanuganti. (n.d.). *Sivapanuganti/social-network-analysis: Book recommendations using social network analysis*. GitHub. Retrieved April 30, 2022, from <https://github.com/sivapanuganti/Social-Network-Analysis>
- Kim, K. J., & Ahn, H. (2012). Hybrid recommender systems using social network analysis. *International Journal of Computer and Information Engineering*, 6(4), 515-518.
- Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1), 5-es