

# Assignment #3 - Amazon Co-Purchasing History Network Analysis

Group 3 Fireflies

Members: Chunhui Zhu, John Wong and Chunmei Zhu

```

1# Directed graph (each unordered pair of nodes is saved once): Amazon0601.txt
2# Amazon product co-purchasing network from June 01 2003
3# Nodes: 403394 Edges: 3387388
4# FromNodeId ToNodeId
5 0 1
6 0 2
7 0 3
8 0 4
9 0 5
10 0 6
11 0 7
12 0 8
13 0 9
14 0 10
15 1 0
16 1 2
17 1 3
18 1 4
19 1 155
20 1 185
21 1 233
22 1 234
23 1 235
24 1 3943
25 2 0
26 2 1
27 2 3
28 2 4

```

## Co-purchasing data

## Amazon meta

```

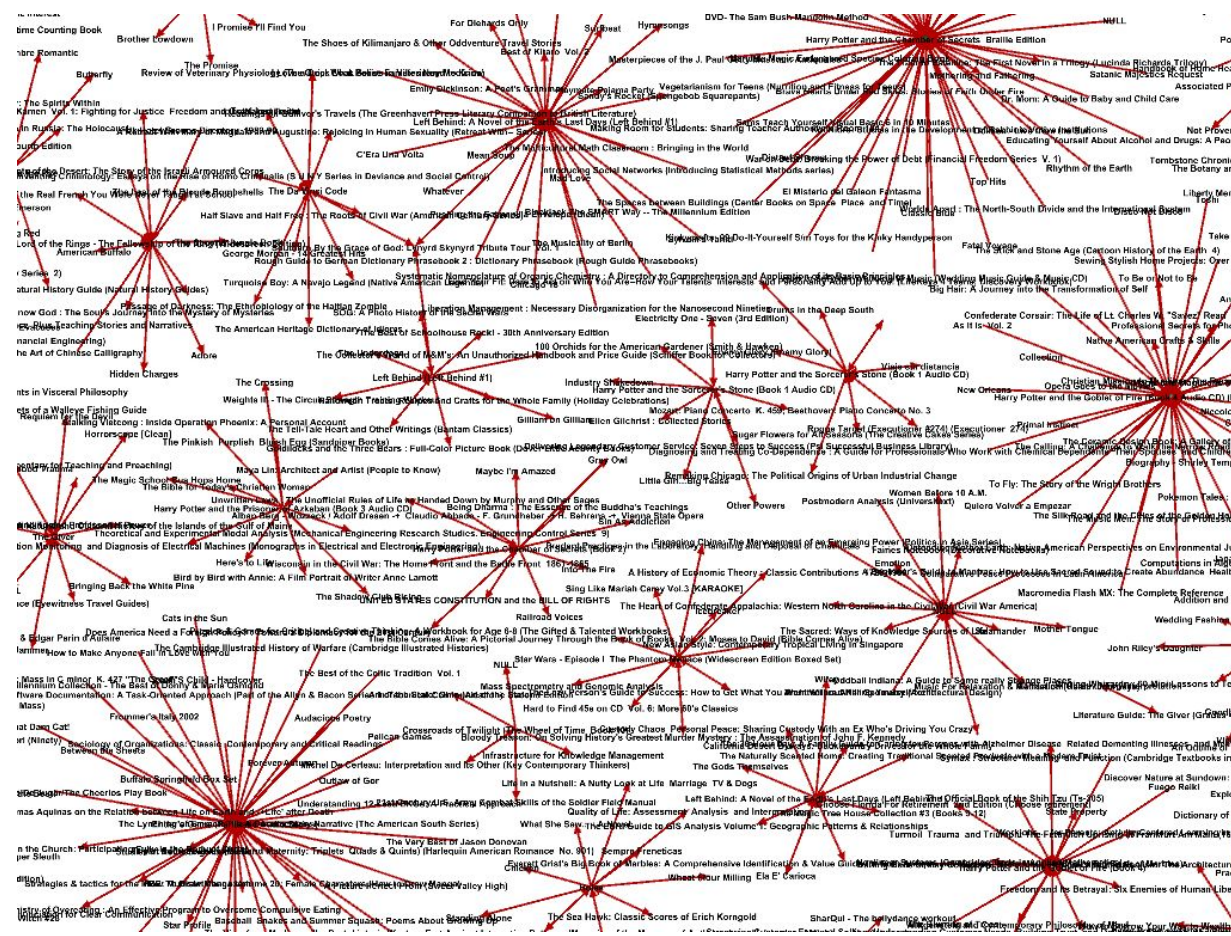
Id: 15
ASIN: 1559362022
title: Wake Up and Smell the Coffee
group: Book
salesrank: 518927
similar: 5 1559360968 1559361247 1559360828 1559361018 07432145
categories: 3
|Books[283155]|Subjects[1000]|Literature & Fiction[17]|Drama[2159]|
|Books[283155]|Subjects[1000]|Arts & Photography[1]|Performing Arts
|Books[283155]|Subjects[1000]|Literature & Fiction[17]|Authors, A-2
reviews: total: 8 downloaded: 8 avg rating: 4
2002-5-13 cutomer: A2IGOA66Y6O8TQ rating: 5 votes: 3 helpful
2002-6-17 cutomer: A2OIN4AUH84KNE rating: 5 votes: 2 helpful
2003-1-2 cutomer: A2HN382JNT1CIU rating: 1 votes: 6 helpful
2003-6-7 cutomer: A2FDJ79LDU4018 rating: 4 votes: 1 helpful
2003-6-27 cutomer: A39QMV9ZKRJXO5 rating: 4 votes: 1 helpful
2004-2-17 cutomer: AUUVVMSTQ1TXDI rating: 1 votes: 2 helpful
2004-2-24 cutomer: A2C5K0QTLL9UAT rating: 5 votes: 2 helpful
2004-10-13 cutomer: A5XYF0Z3UH4HB rating: 5 votes: 1 helpful

```

- raw data: 3.3M rows for co-purchasing relationship among 400k products.
- filtered by top 100 most reviewed products: 900 rows for co-purchasing relationship among 700 products

limited to top 100  
most reviewed  
product in the  
graph:

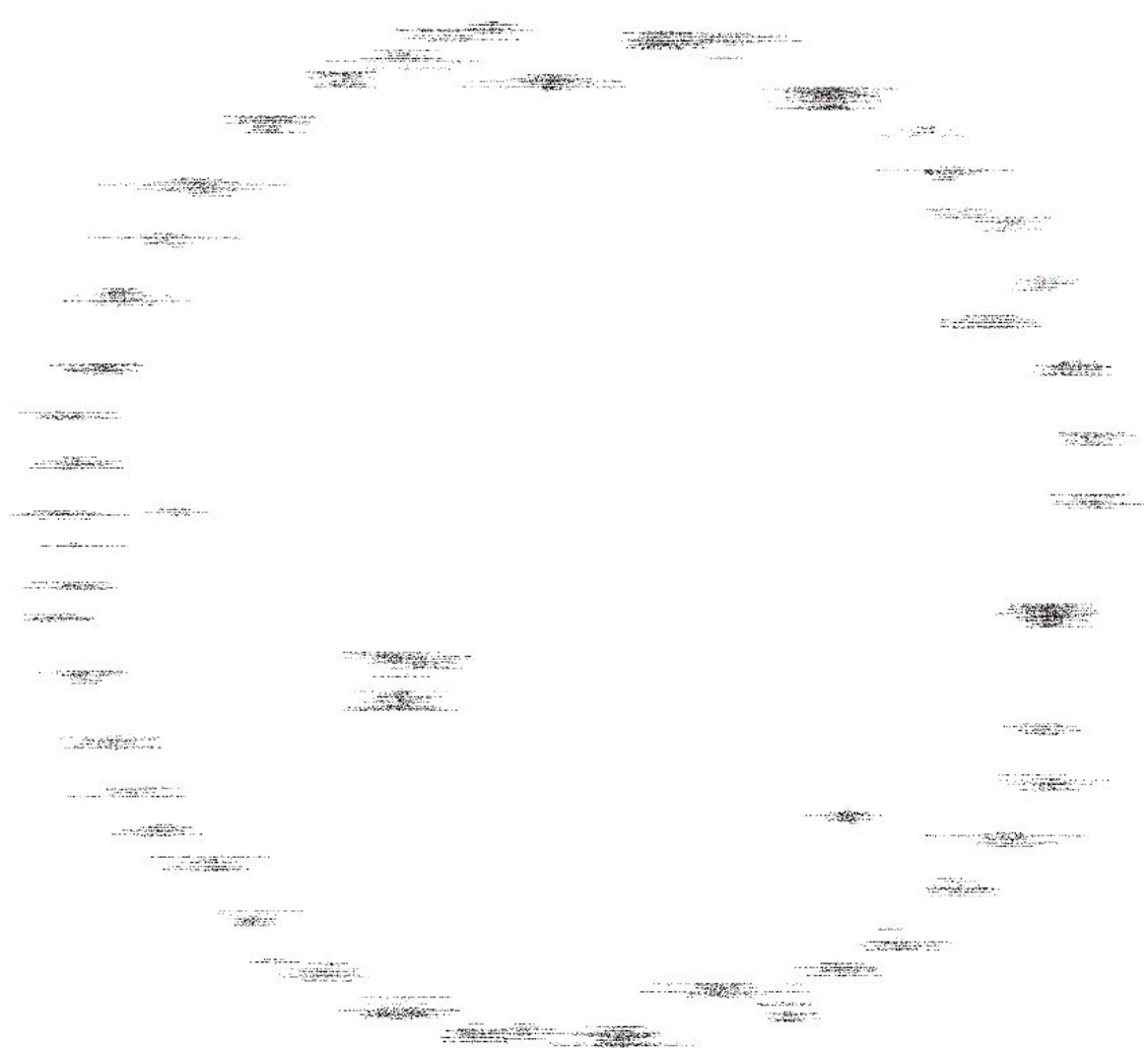
-Gephi



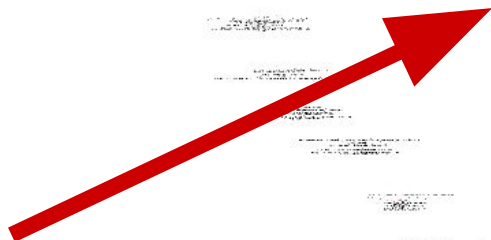
**limited to top 100  
most reviewed  
product in the graph:**

-load as directed graph

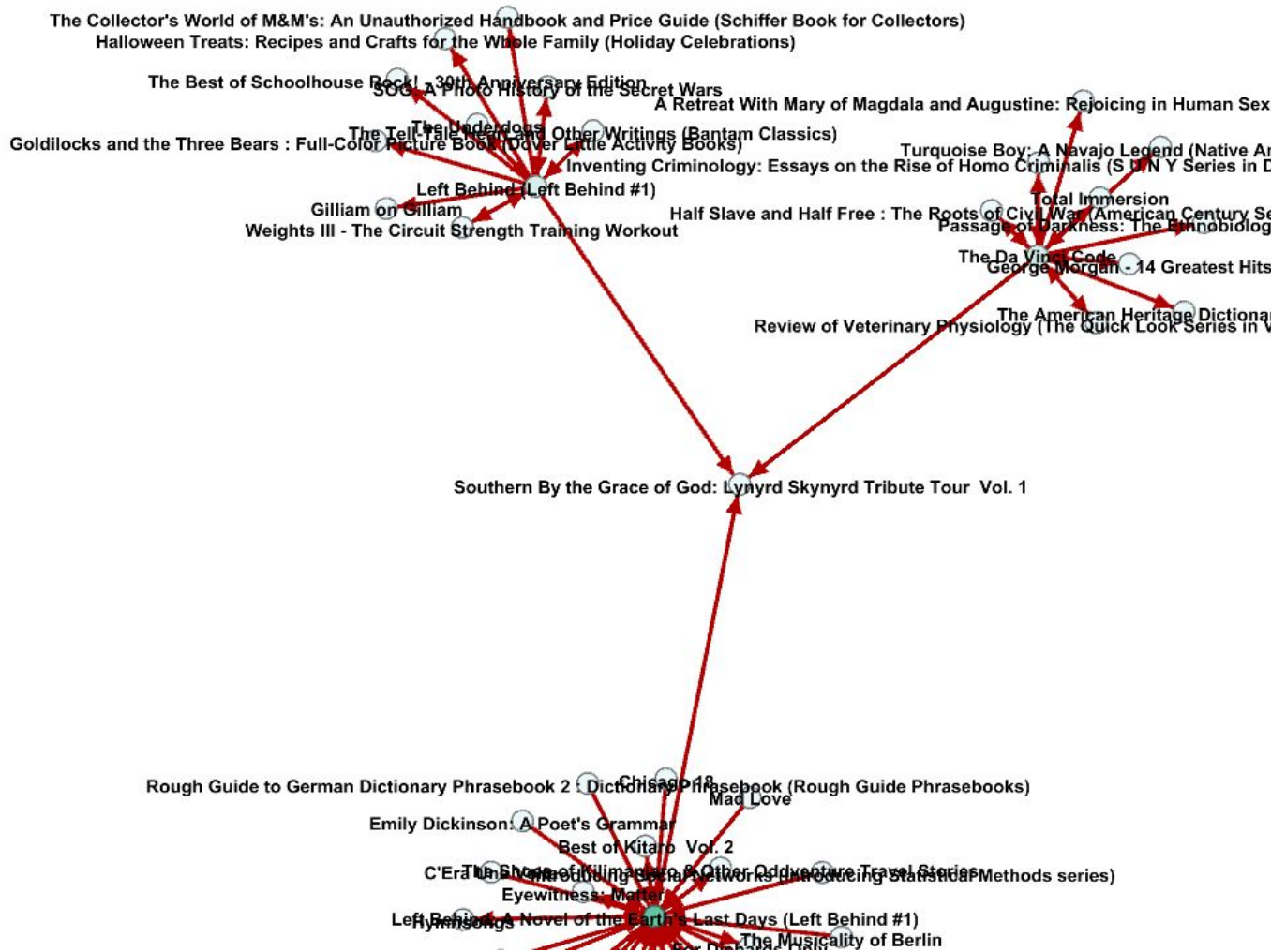
-After we applied layout "YiFan  
Hu". We can see the data are  
clustered or grouped in certain  
way.



# What's in it?



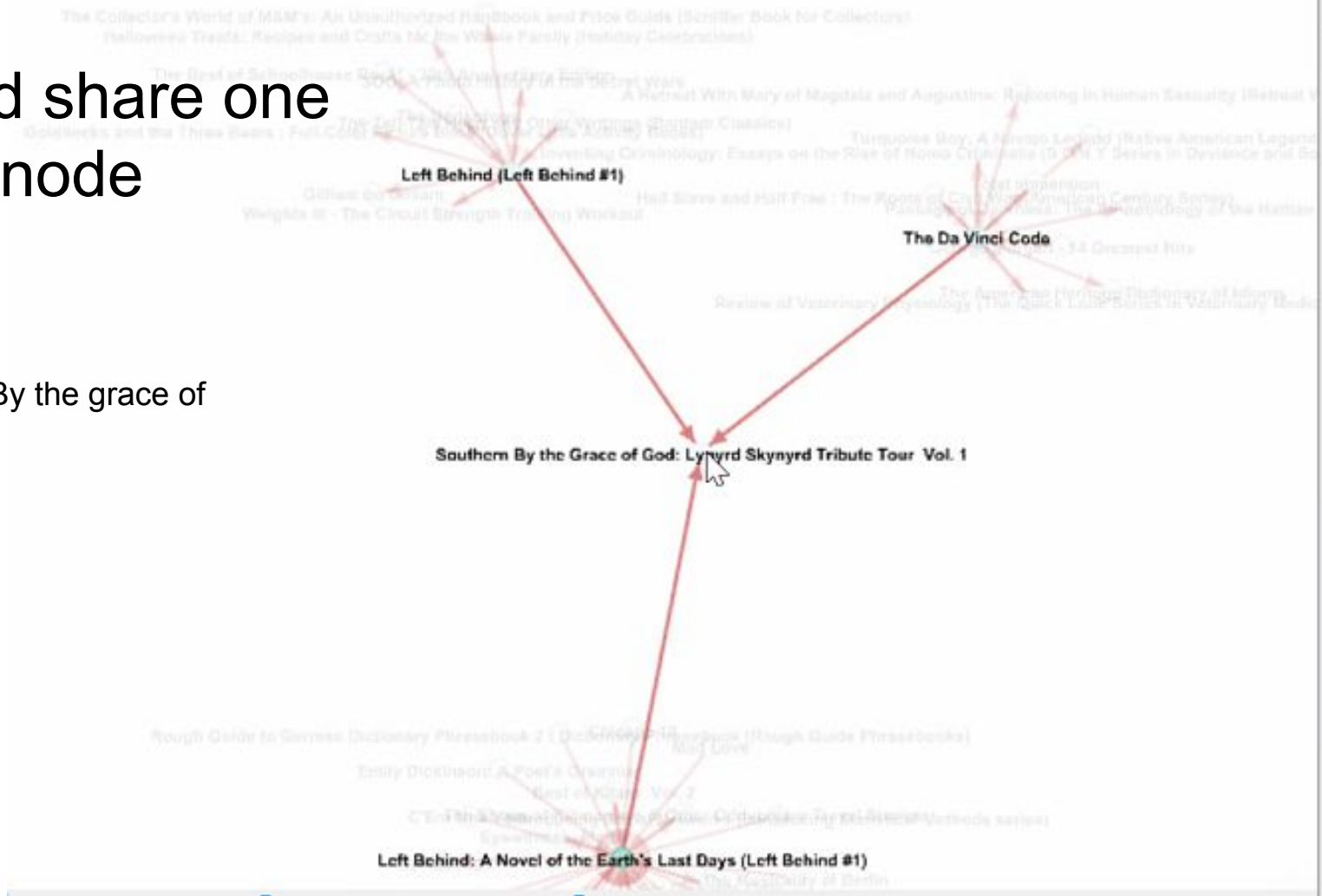
# Zoom In





# 3 grouped share one common node

Book: Southern By the grace of God



# Graph metices

E.g:

Avg Degree =1.3

Network Diameter =2

We will look at In degree and  
out degree

The screenshot shows a software interface for network analysis. At the top, a 'Context' tab is active, displaying 'Nodes: 739' and 'Edges: 968'. Below this, there are tabs for 'Filters' and 'Statistics'. The 'Statistics' tab is selected, showing a 'Settings' section. Under 'Settings', there are three expandable sections: 'Network Overview', 'Node Overview', and 'Edge Overview'. The 'Network Overview' section is expanded, showing a list of metrics with their values and 'Run' buttons. The 'Node Overview' and 'Edge Overview' sections are also expanded, showing their respective metrics. The 'Dynamic' section at the bottom shows '# Nodes' and '# Edges' with 'Run' buttons.

Metric	Value	Action
Average Degree	1.31	Run ⓘ
Avg. Weighted Degree		Run ●
Network Diameter	2	Run ⓘ
Graph Density	0.002	Run ⓘ
HITS		Run ●
Modularity		Run ●
PageRank		Run ⓘ
Connected Components	55	Run ⓘ

Metric	Action
Avg. Clustering Coefficient	Run ●
Eigenvector Centrality	Run ●

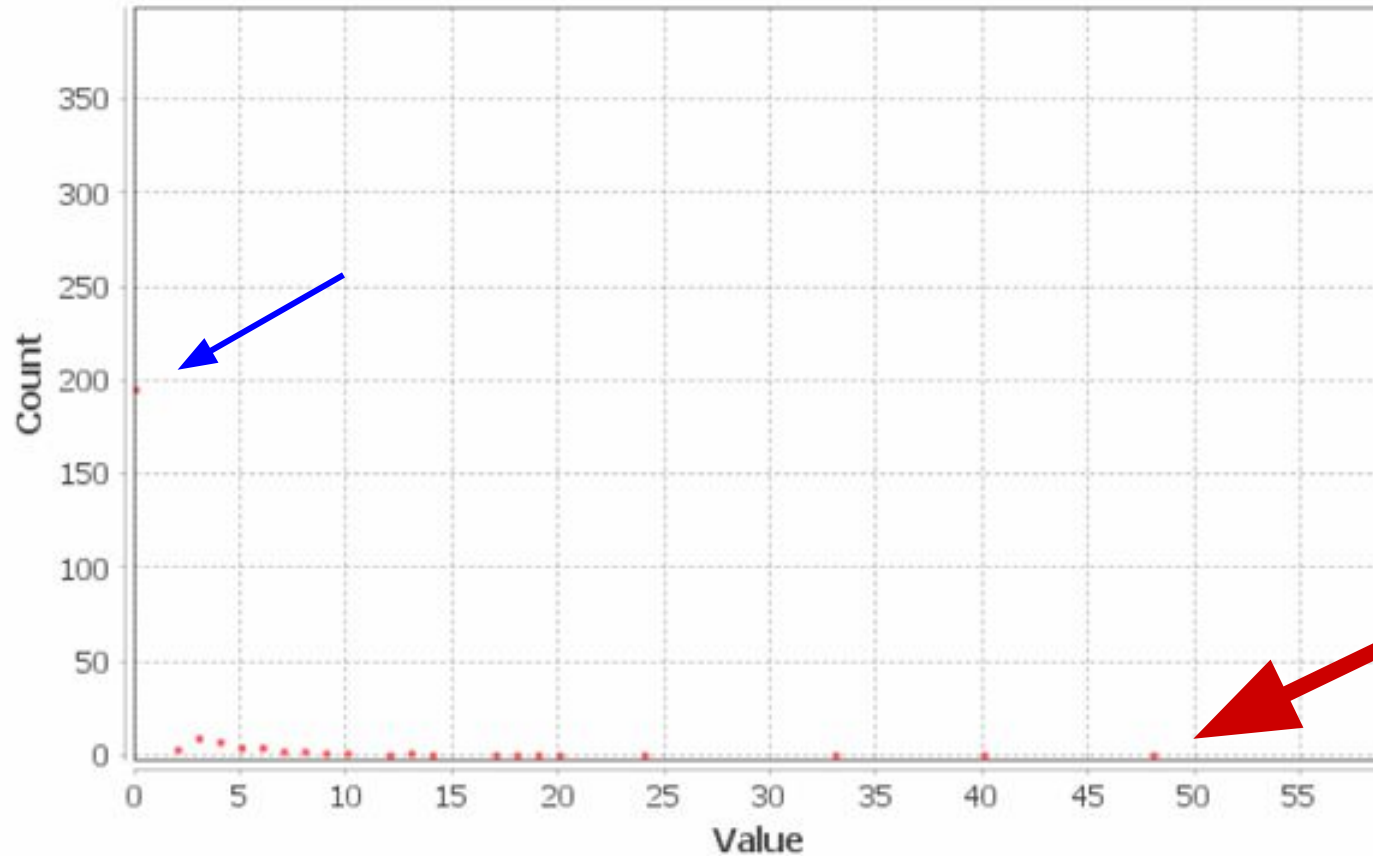
Metric	Value	Action
Avg. Path Length	1.815	Run ⓘ

Metric	Action
# Nodes	Run ●
# Edges	Run ●



## In-Degree Distribution

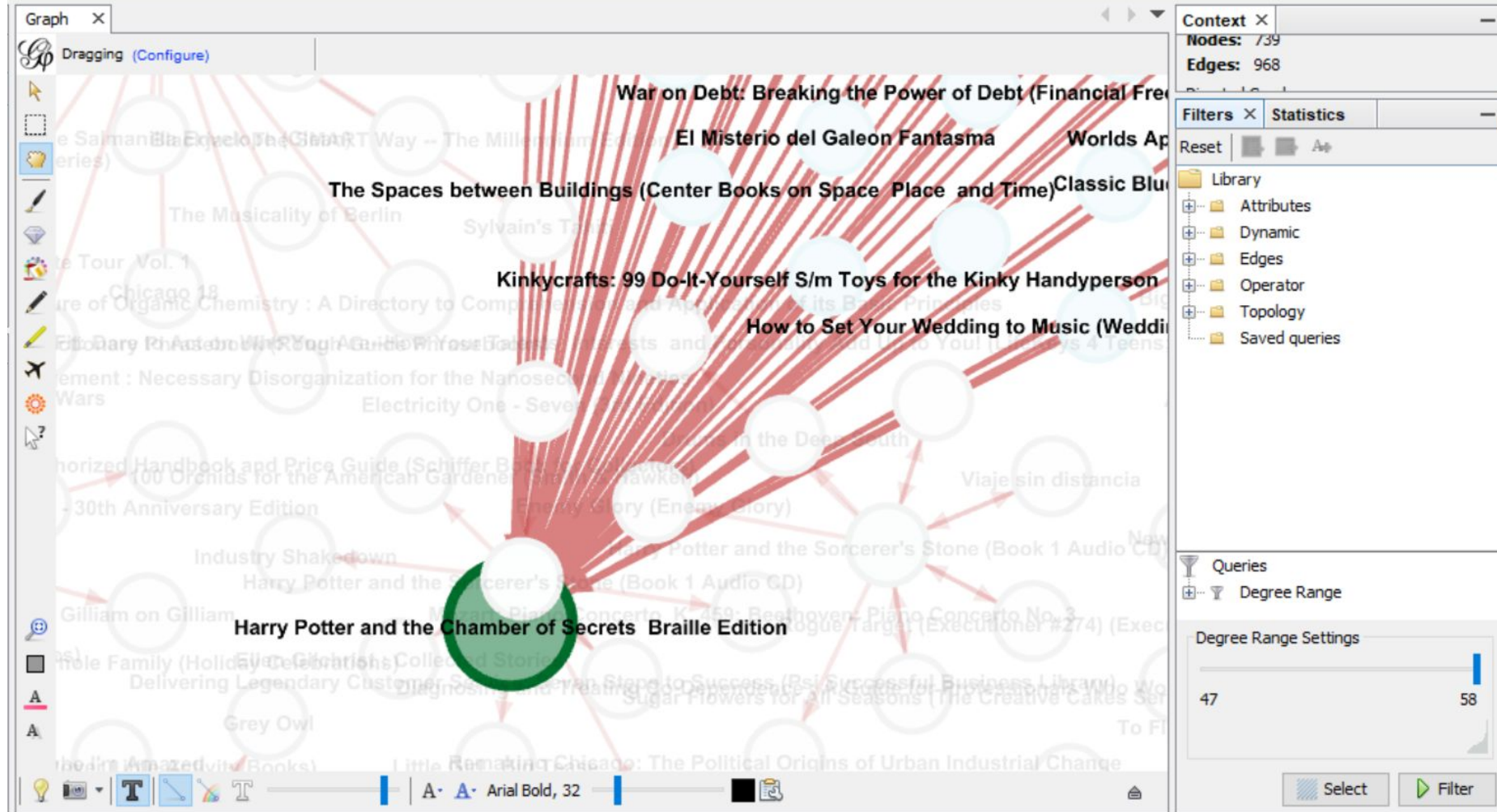


The number of edges goes to one node

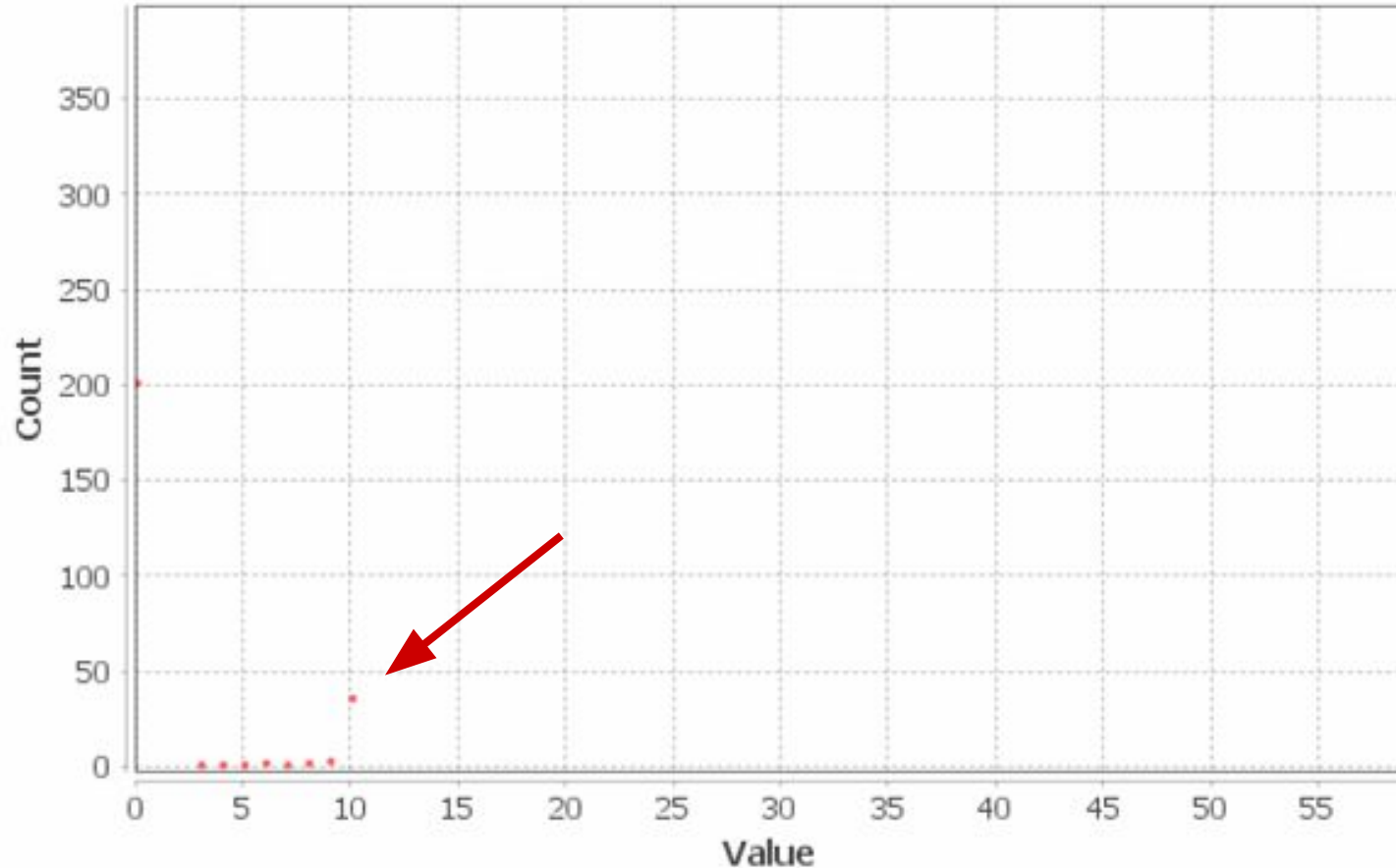
The number of purchased products that relates to one particular product

Blue arrow: about 200 product has only one co-purchasing

Red arrow: next page



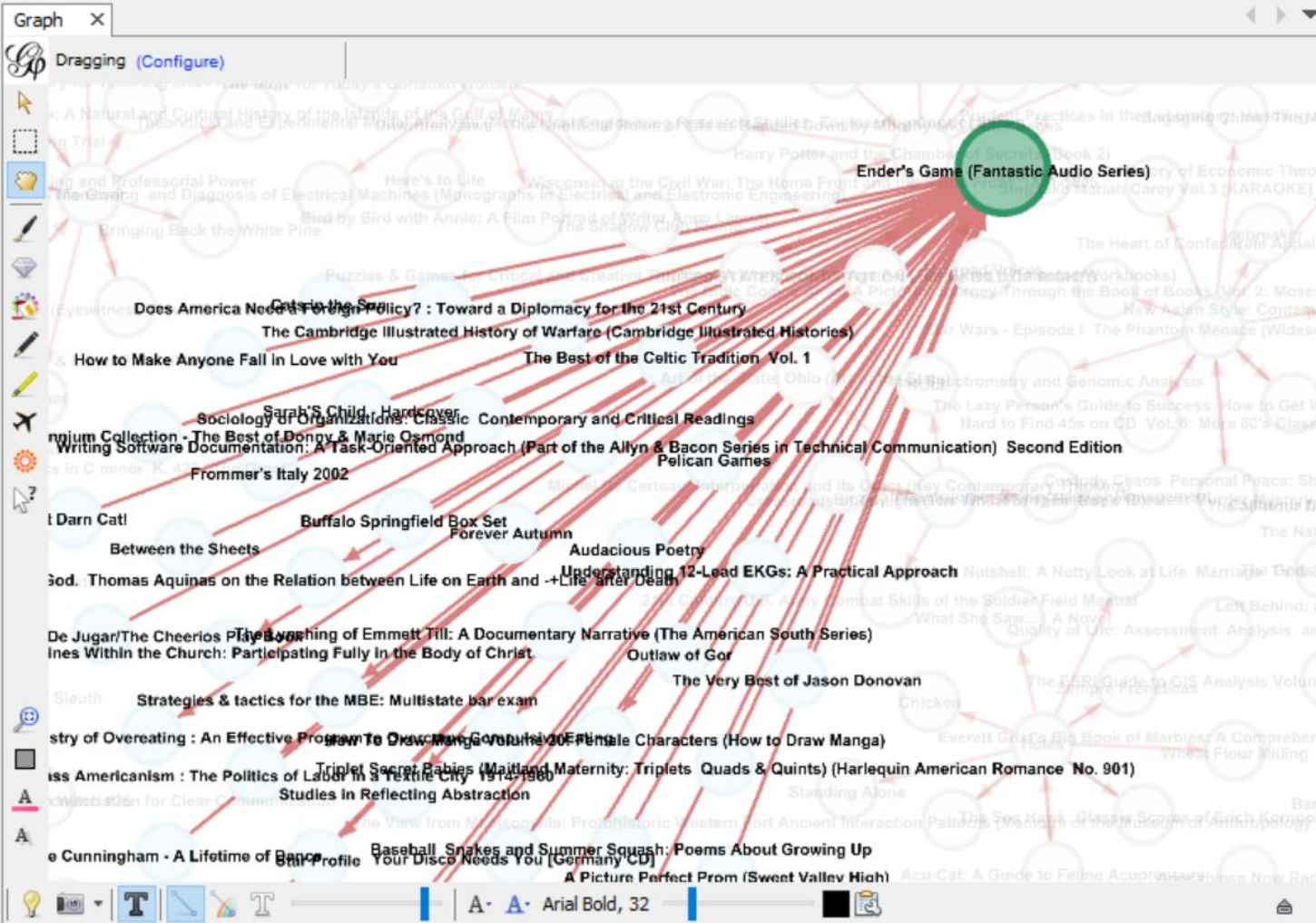
## Out-Degree Distribution



-The number of edges that one nodes link to others

-The number of items people purchased after one particular purchased

-Red arrow: next page



# Conclusion

Purchasing history from customer is a very important asset for Internet Retail company like Amazon. By studying the purchasing relationship, the company can identify some customer's interests by similar purchases and advertise what the customer may purchase.

Thank you