

Analysis of Amazon-Co-Purchasing-Network

By Imaz, Suchir, Sathvik, Srinath, Phalgun

Analysis of Amazon-Co-Purchasing-Network

Our project involves the analysis of an Amazon Product Co-Purchasing Dataset retrieved from SNAP. We are making use of the igraph library in python to analyze the dataset.

While SNAP provides us with the actual list of edges in its dataset, it also provides us with a metadata file. This metadata file contains the actual info about the products referenced by the node IDs in the dataset. The metadata file contains information about 548,552 different products (Books, music CDs, DVDs, VHS video tapes, Toys, Video Games, etc.).

For each product the following information is available in the metadata file:

- ▶ Title
- ▶ Group (Books, DVDs, etc.)
- ▶ Salesrank
- ▶ List of similar products (that get co-purchased with the current product)
- ▶ Detailed product categorization
- ▶ Product reviews: time, customer, rating, number of votes, number of people that found the review helpful

This Salesrank bit in particular is quite interesting. It is Amazons metric for determining how well a product is currently selling on their website. The lower the Salesrank the better the product is selling.

- ▶ "In graph theory, eigenvector centrality is a measure of the influence of a node in a network."
- ▶ "based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes."

It seems like products with higher eigenvector centralities would be considered the influential nodes in the network. If we were to check these products Salesrank values I feel like we could see if there is a connection between best selling products and the most influencial nodes in the network.

Now this is what the metadata file actually looks like

```

Id: 19
ASIN: 078510870X
title: Ultimate Marvel Team-Up
group: Book
salesrank: 612475
similar: 5 0785114572 078511078X 0785114033 0785114041 0785113126
categories: 4
|Books[283155]|Subjects[1000]|Children's Books[4]|Literature[2966]|Science Fiction, Fantasy, Mystery & Horror[3013]|Comics & Graphic Novels[3014]
|Books[283155]|Subjects[1000]|Comics & Graphic Novels[4366]|Publishers[4396]|Marvel[4400]
|Books[283155]|Subjects[1000]|Teens[28]|Science Fiction & Fantasy[17466]|Fantasy[17468]
|Books[283155]|Subjects[1000]|Comics & Graphic Novels[4366]|Graphic Novels[4390]|Superheroes[3825121]
reviews: total: 8 downloaded: 8 avg rating: 3.5
2002-3-23 cutomer: A1RLXSCKKM14KF rating: 3 votes: 9 helpful: 5
2002-9-29 cutomer: ATS7LYGLUX34L rating: 5 votes: 3 helpful: 3
2002-10-3 cutomer: ATS7LYGLUX34L rating: 5 votes: 2 helpful: 2
2002-11-7 cutomer: A3DQ6L24TOMXAZ rating: 4 votes: 10 helpful: 10
2003-3-17 cutomer: A3M6J4CTJ13QBX rating: 2 votes: 11 helpful: 4
2003-12-3 cutomer: A4JF4T1F95DOA rating: 2 votes: 0 helpful: 0
2004-3-14 cutomer: A2S166WSCF1FP5 rating: 4 votes: 5 helpful: 2
2004-10-20 cutomer: ADC571K8L3ZLN rating: 4 votes: 6 helpful: 6

```

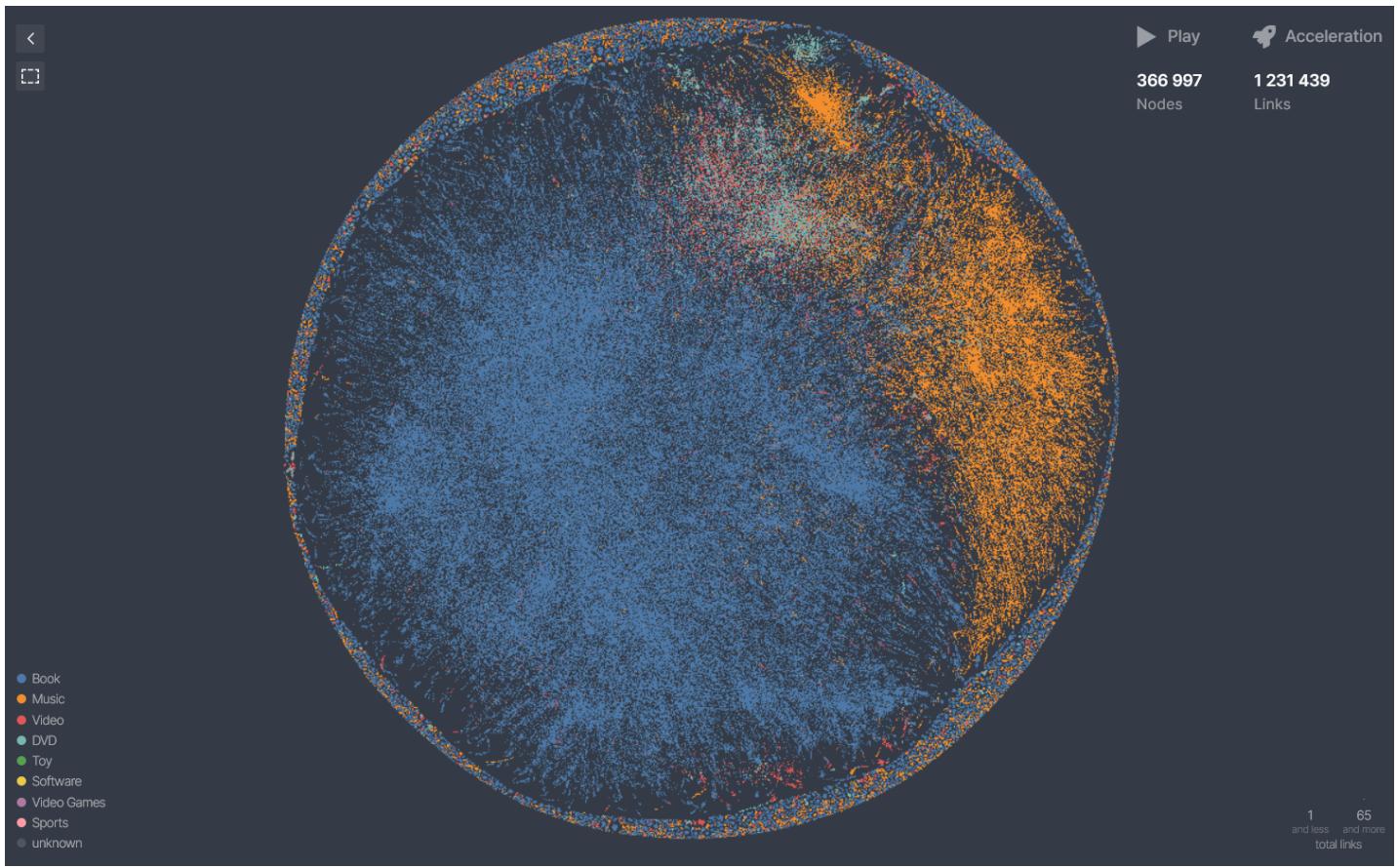
We have code written to parse all this information and neatly place it into a Pandas DataFrame so that we can use the Salesrank information and also other information like product categories and average rating of the product in our network analysis.

Network Visualisation

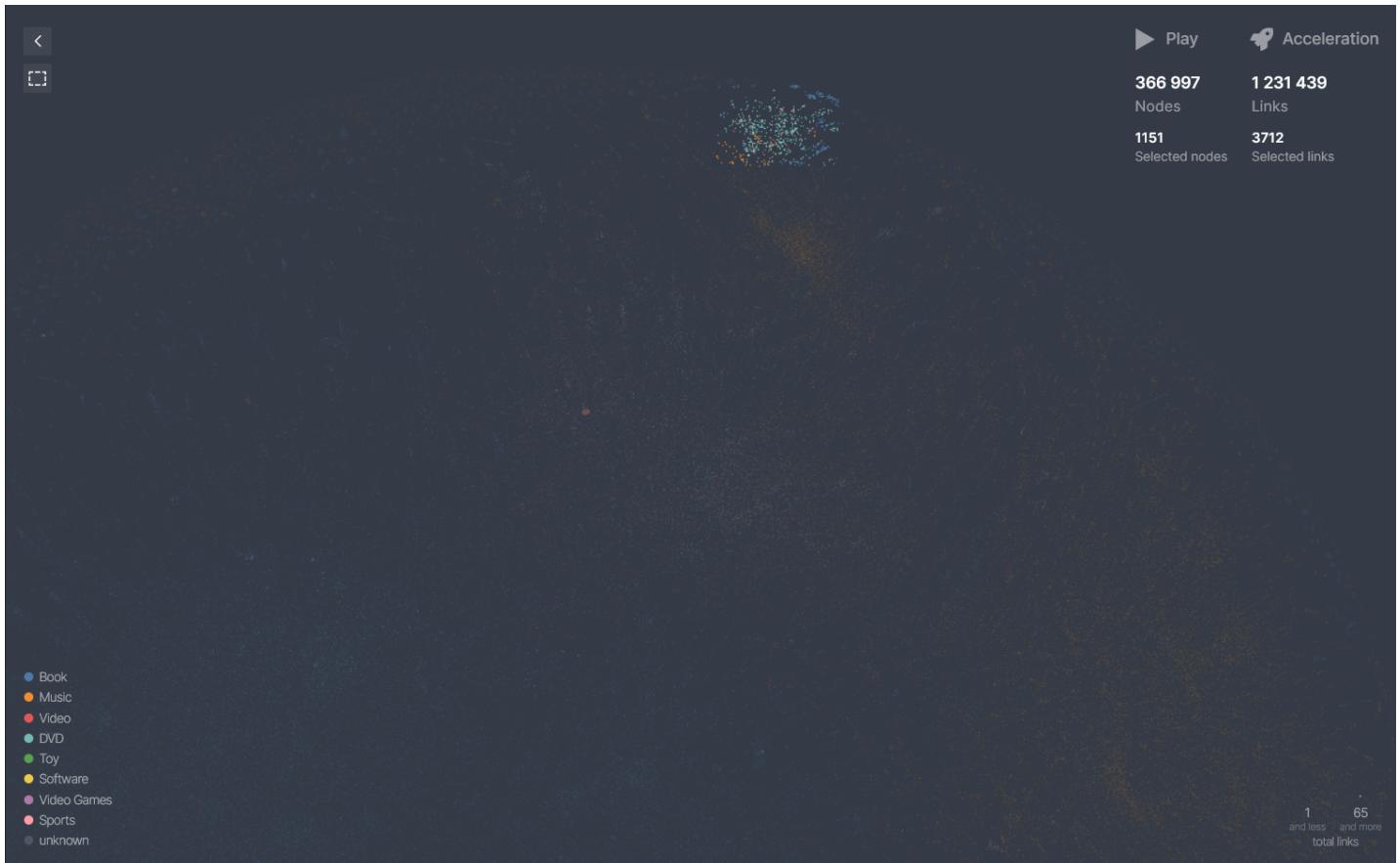
We used an online app called Cosmograph to visualize the network. It can visualize networks with a million nodes very quickly. It also uses Force layout so that all the nodes are clearly visible. Useful for us since our network contains around 400K nodes. The app is also slightly interactive letting you click on a node to highlight it and its immediate neighbours.

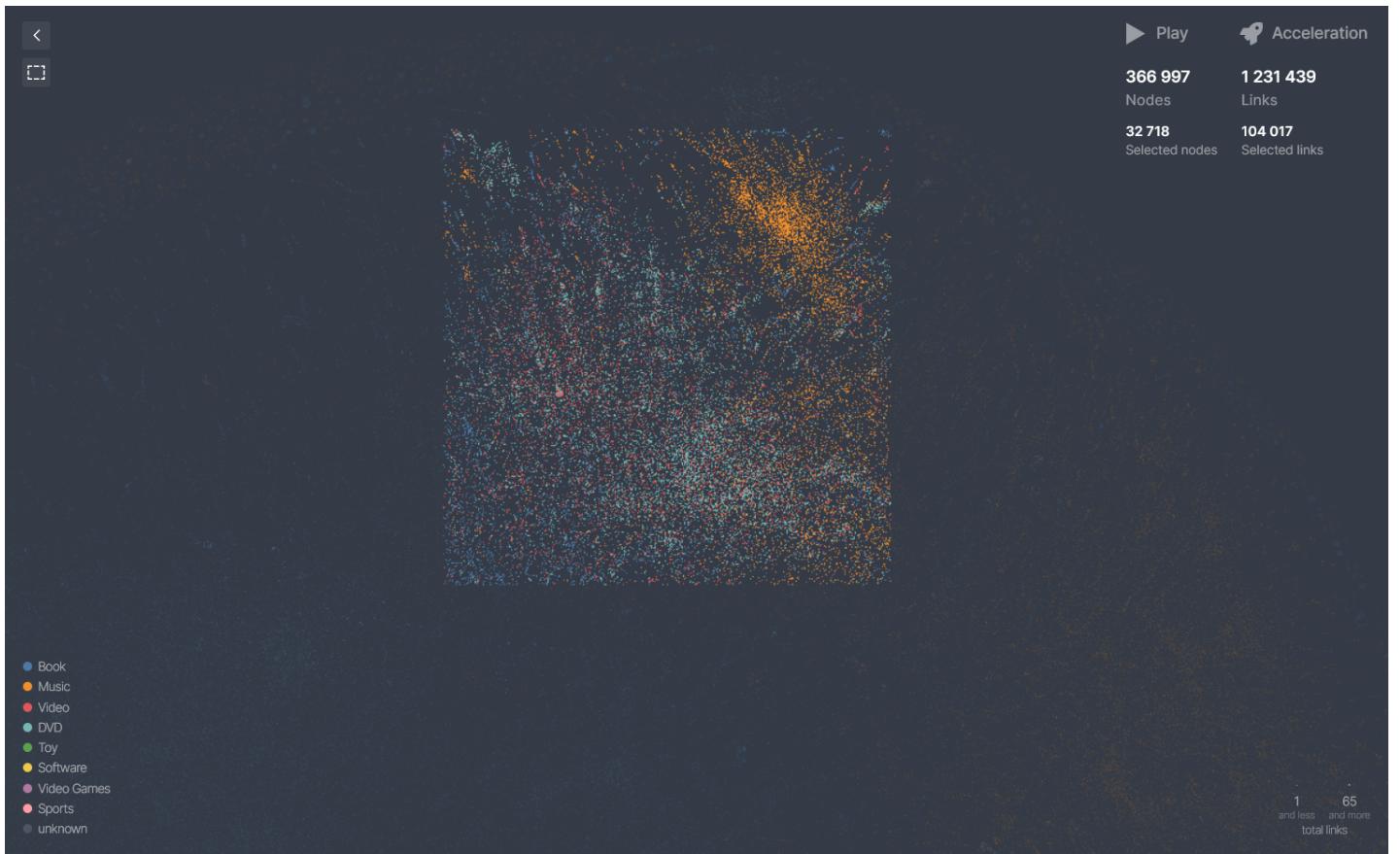
Ofcourse the nodes represent products, and the edges the copurchasing relationships. To make the visualization more informative, we sized the nodes based on their out-degree, with larger nodes representing products which are purchased along with many more products. Additionally, we assigned colors to the nodes based on their category, such as DVD, music, book, video, toy, software, video games, and sports.

The resulting network provided a clear visual representation of the copurchasing patterns in the Amazon network. The colors of the nodes allowed us to quickly identify different product categories. This visualization approach was effective in analyzing the copurchasing network and gaining insights into the relationships between products and their categories.



We noticed an interesting cluster in the copurchasing network visualization. It was a cluster mainly made up of DVDs but it was separated from the main, much larger, DVD and Video VHS cluster.



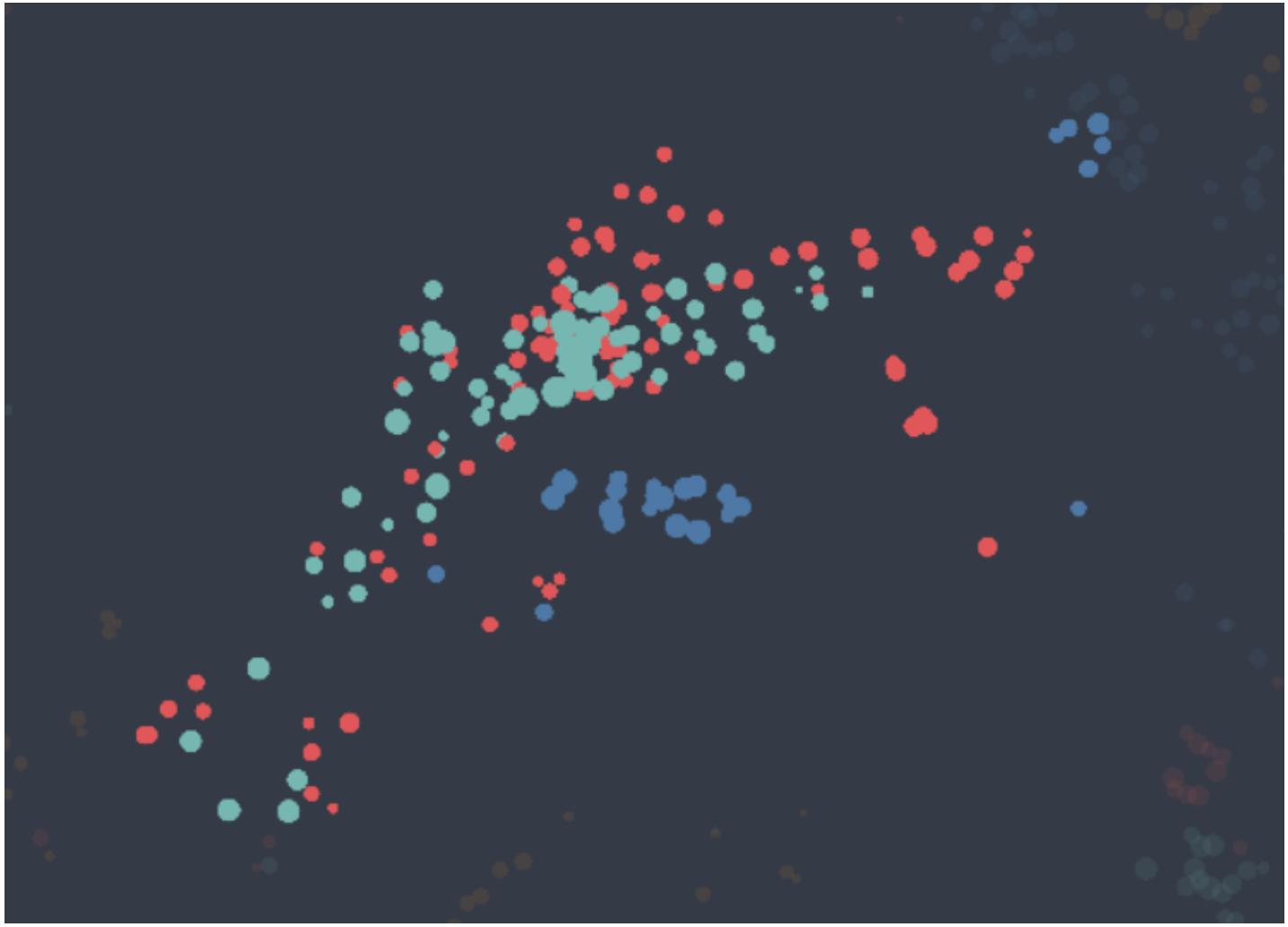


Upon further inspection, we noticed that this cluster consisted of mostly DVDs related to Anime(Japanese Cartoons). And as for the reason this DVD cluster was separated from the main DVD cluster, if you examine the area around the DVD cluster, you will notice that all the non DVD products in this area are also Japan related, mostly consisting of Manga(Japanese Comic Books), Japanese Video Game Strategy guides, etc. We think that since these Anime DVDs were probably co-purchased with other Japanese products like Manga and Japanese Video Games, they formed their own little Japan cluster away from the main cluster.

id	title	group	↓	salesrank	similar
B00006IUHQ	Robotech - The Macross Saga - Complete Collection	DVD		8118	['B00006RCKO', 'B00007AJD9', 'B000051YJY', 'B0000639E0', 'B0001DCYCS']
B000096IAJ	R.O.D. - Read or Die	DVD		16311	['B00020HBX2', 'B0000066IC', 'B0002IQHJI', 'B0002VYPKS', 'B0006FFRIE']
B000055ZL5	Generator Gawl - Secrets and Lies (Vol. 3)	DVD		45234	['B000053UYX', 'B00004W5XG', 'B00008AOSO', 'B0000A2ZOO', 'B00008G8PK']
6305951667	Sword For Truth	DVD		48132	['6305131082', 'B00009MGFR', 'B0000CABH8', 'B00005UW7E', 'B00005V1XW']
B00008L3MH	Hellsing - Complete Collection	DVD		2187	['B00005OW09', 'B00005QCW4', 'B0001DCYCS', 'B0000639E0', 'B00005NX1N']
B00005YUPH	The Soultaker - Flickering Faith (Vol. 2)	DVD		40516	['B00005UM46', 'B0001WTW2E', 'B00079I0BO', 'B000063K2C', 'B0002VEVR0']
B000060MWI	GTO - Outcasts (Vol. 3)	DVD		36555	['B00007AJG0', 'B0000844KU', 'B0000844KV', 'B0000A2ZOH', 'B00009AV86']
B000060MWG	Vampire Princess Miyu - The Last Shinma (TV Vol. 6)	DVD		41796	['B00005UQ9N', 'B00005RIYC', 'B00005OSLE', 'B00005M1Z9', 'B00005B8U8']
B000060MWF	Vampire Princess Miyu - Dark Love (TV Vol. 5)	DVD		22306	['B00005OSLE', 'B00005M1Z9', 'B00005B8U6', 'B0000844KV', 'B00005B8U8']
B000060MW8	Chet Atkins: Rare Performances 1976-1995	DVD		27099	['B00005NC53', 'B00006AUJP', 'B0000JLLC2', 'B00005KH20', 'B000096IAQ']
B00008WT62	Mobile Fighter G Gundam Boxed Set - Rounds 10-12	DVD		36379	['B0001BMMB8', 'B0001BMMAY', 'B0003JAHOG', 'B0000CBL7G', 'B00006HAW6']
B00008WT61	Crest of the Stars - Complete Series Set	DVD		31854	['B0002MHDZG', 'B0002VEVOI', 'B00007G1U6', 'B00008DDIN', 'B00005NX1N']
B00007G1W8	Vandread - Second Stage - Revelations (Vol. 3)	DVD		28114	['B0000639GY', 'B00005UW8F', 'B00005RZPV', 'B0002IQNWY', 'B00020X9A6']

Similar to this Japan cluster we also saw a few other clusters formed based on the Genre of the product although on a much smaller scale. For example, there was a cluster full of Clint Eastwood Western films

and one full of WWE Books, DVDs and Video VHS tapes.



id	title	group	salesrank	similar
6305276854	WWE WrestleMania 2 - What The World Is Coming To	Video	16560	['6305276838', '6305276862', '6305276889', '6305276900', '630527696X']
0671525751	Someone to Watch Over Me : A Novel	Book	266553	['0671737600', '0671525743', '0671880608', '0671795554', '0671776800']
6305565724	WWE Fully Loaded 1999	Video	3672	['B00001W9GL', '6305710341', '6304826214', 'B00004YRWV', '6305710481']
0425176347	Duel Identity (Tom Clancy's Net Force; Young Adults)	Book	598943	['0425173674', '0425181502', '0425171132', '0425171914', '042516991X']
B00001MXWK	WWE - Hell Yeah: Stone Cold's Saga Continues	DVD	19281	['B00019GHTI', 'B00005RIYR', 'B00005RIYQ', 'B00006JU8M', 'B00005RIYZ']
B00001MXWJ	WWE WrestleMania XV - The Ragin' Climax	DVD	37224	['6305994781', 'B00005KH1M', 'B00004STEE', 'B00005RIYO', 'B00005QAQJ']
6304328605	Gullah Gullah Island: Sing Along with Binyah Binyah	Video	12029	['630507206X', '6304328613', '6304328583', '630432863X', 'B00000HF1B']
B00005O5DO	WWE - Invasion 2001	DVD	7331	['B00005QAQJ', 'B00005Q4EU', 'B00005Q4ER', 'B00005RIYI', 'B00005Q4EV']
0786251522	Someone to Watch over Me	Book	1112015	['0671737600', '0671525743', '0671880608', '0671795554', '0671776800']
B00005MMA6	WWE - Best of Raw, Vol. 3	Video	9089	['B00005AADX', 'B00005RIZ0', 'B00005O5DO', 'B00005KHKG', 'B00005KHKF']
6305277001	WWE WrestleMania XI	Video	1896	['6305277036', '6305276994', '6305277044', '630527696X', '6305276986']
6305454302	WWE The Rock - Know Your Role	Video	29938	['B00004STEF', 'B00006JU8M', 'B00001MXWK', 'B00005RIZ2', 'B00005RIYS']
B00008HCA1	WWE No Way Out 2003	Video	18320	['B00009AV93', 'B00008HCA6', 'B00008HC9Y', 'B00009W0UB', 'B00009W0U9']

As Google Once Said "PageRank", And Rank The Pages We Shall

In order to find the most influential node we found the nodes with the highest pagerank scores. When

computed for the entire graph, the following results were found :-

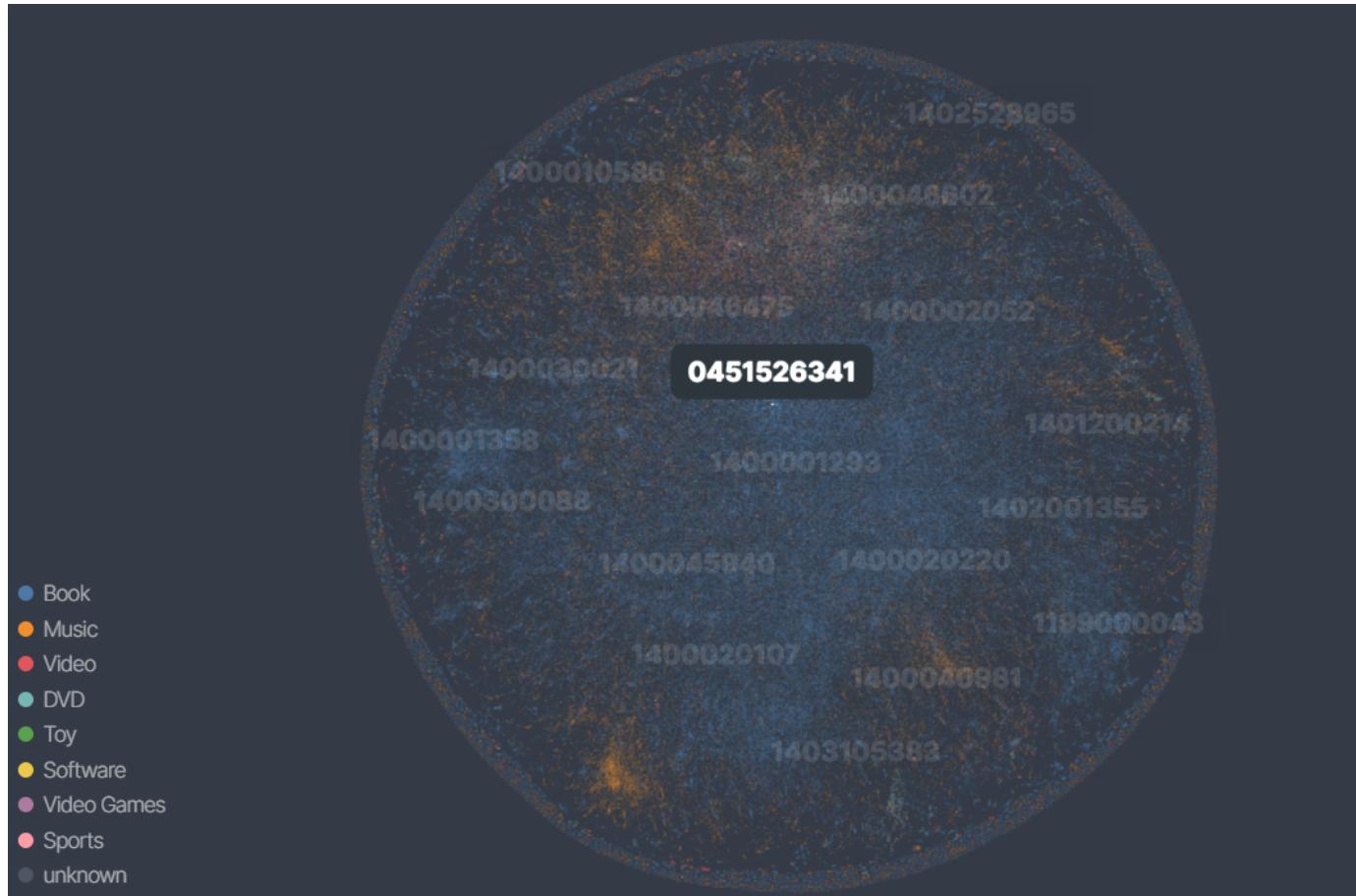
- ▶ The most influential node was found to be

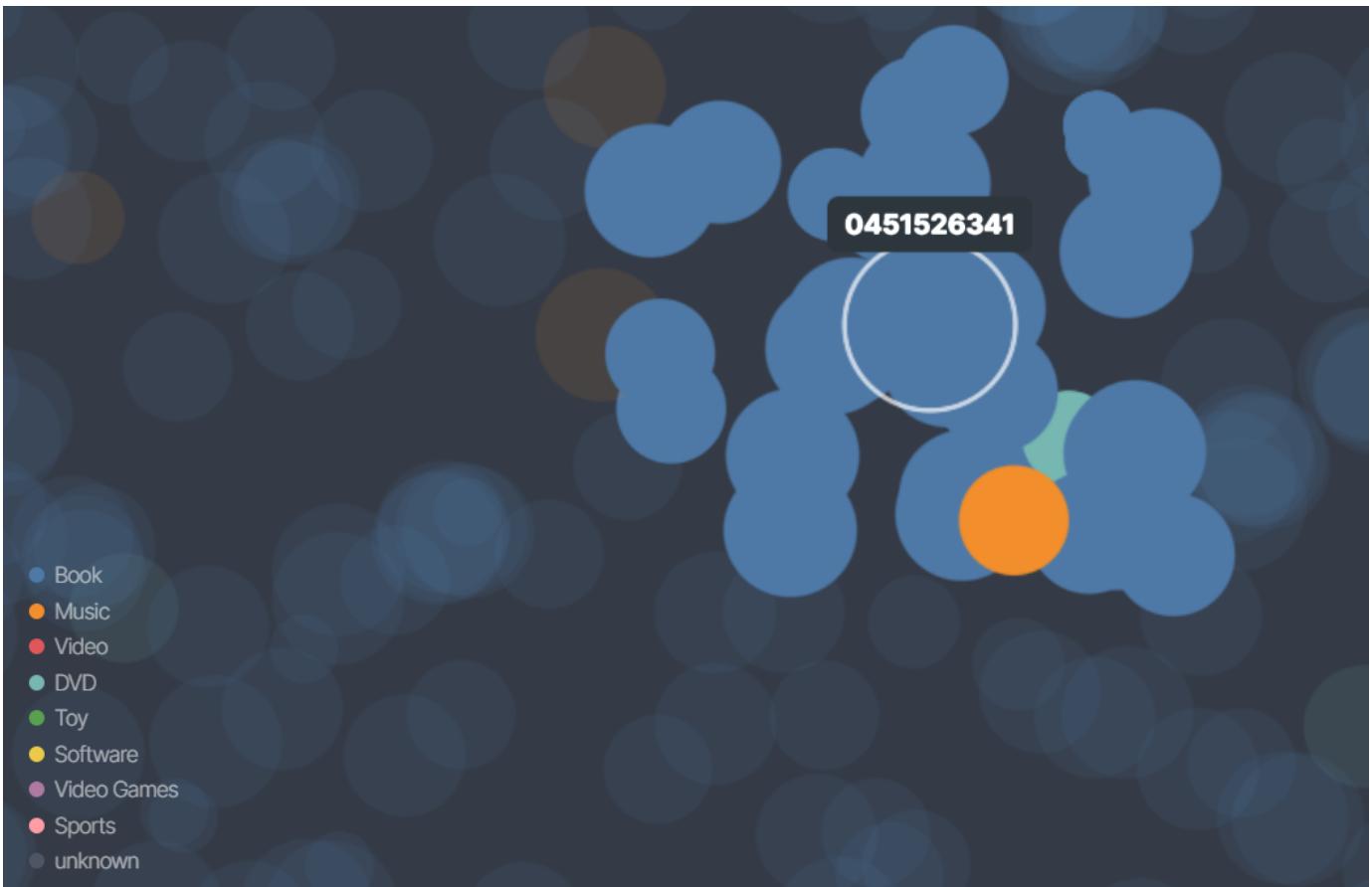
ASIN: 0451526341

title: Animal Farm

group: Book

salesrank: 275





The node has a pagerank score of 0.000915585279743769.

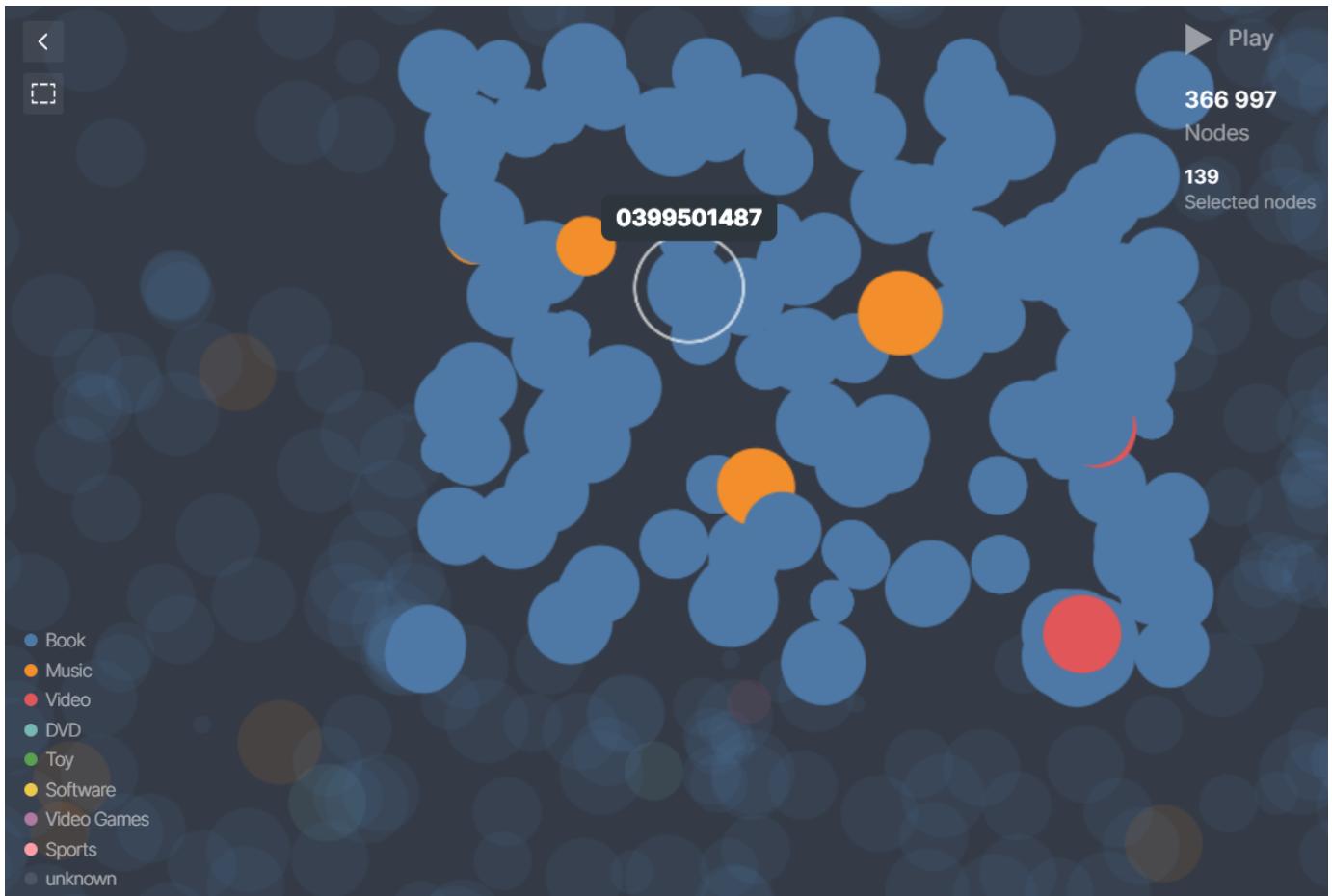
The node also has a betweenness centrality of 0.32.

The kids did not just stop at Animal Crackers, it appears they have gone for the entire farm as well.

However, we notice a discrepancy between the pagerank score and the salesrank score. According to our discussion before, the lower the salesrank score is the more the product is being sold, that is, it's a best selling product. However, we see that the product with the highest pagerank score does not actually have the lowest salesrank score!

This can be seen by the product with the 5th highest pagerank :-

- ▶ ASIN: 0399501487
- title: Lord of the Flies
- group: Book
- salesrank: 143



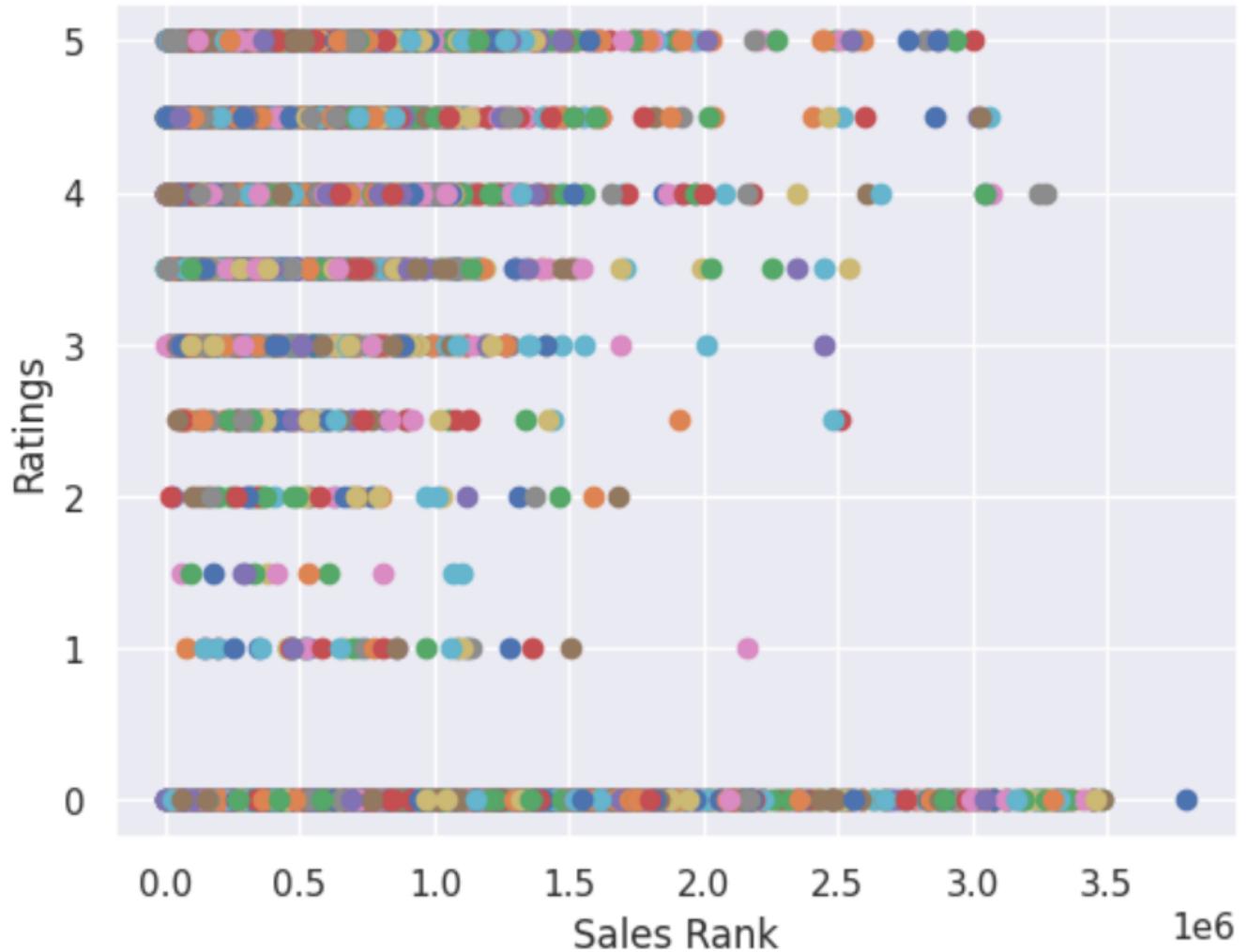
The pagerank of this product is 0.0007858643024663072 and has a betweenness centrality of 0.463!

To analyze salesrank a bit we plotted a graph between salesrank and avg rating.

Logically speaking, you would expect salesrank to be low (Better selling) for a product with a higher rating.

However, the graph we achieved can be seen as below :-

Sales Rank vs Ratings for Books



While majority of the products are plotted as expected,(products with ratings 3-5 are plotted with lower sales rank), the products with low ratings (close to 0) are all over the place, as well as some products with higher rating have high sales rank! (not being sold well even though they have high ratings)

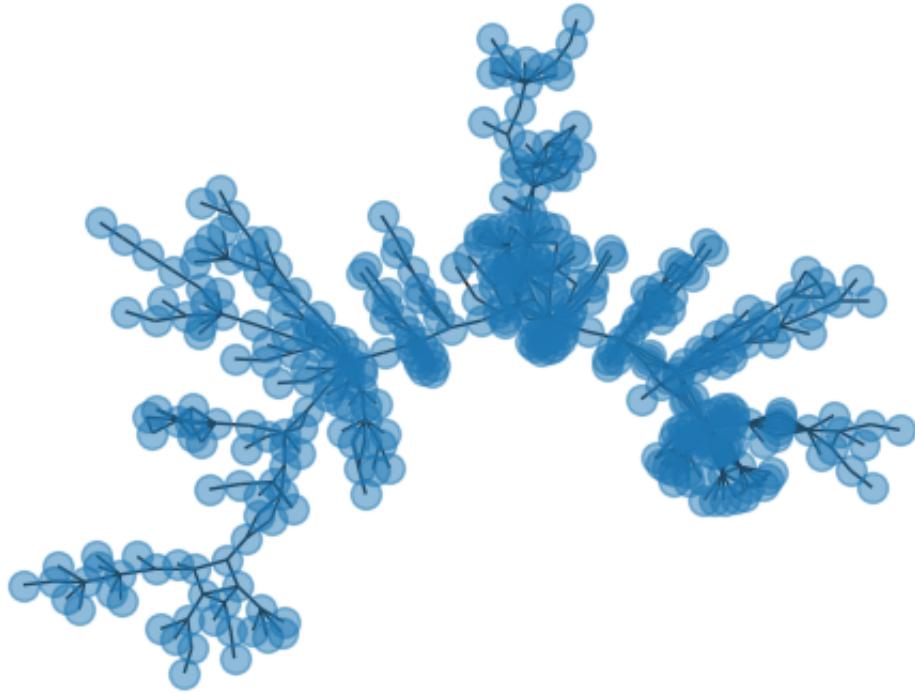
The dataset being from 2003, we can assume that most people just were not writing reviews at the time. Good chance these products had high reviews among their peers and in the newspapers and people just used Amazon as a convenient place to purchase it. These days most things are bought online so you are more likely to find their reviews online as well.

Everyone Loves Graphs

Especially the ones that look like cute little nervous systems

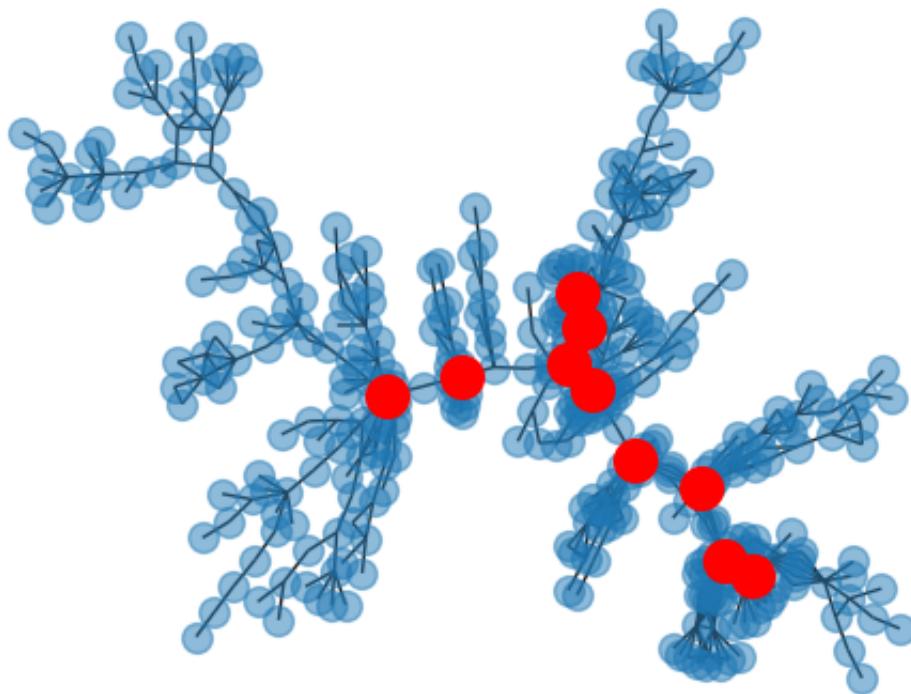
Time for everyones favourite part of the day.

We found the largest connected component in the graph and it looks like this :-



Intuitively looking at this graph, we want to find the top 10 nodes with the highest degree centrality. We should expect the nodes to be somewhere in the areas with high density!

The top 10 nodes with highest degree centrality :-



Now that we found the top 10 nodes with the highest degree centrality, we need to analyse the related data.

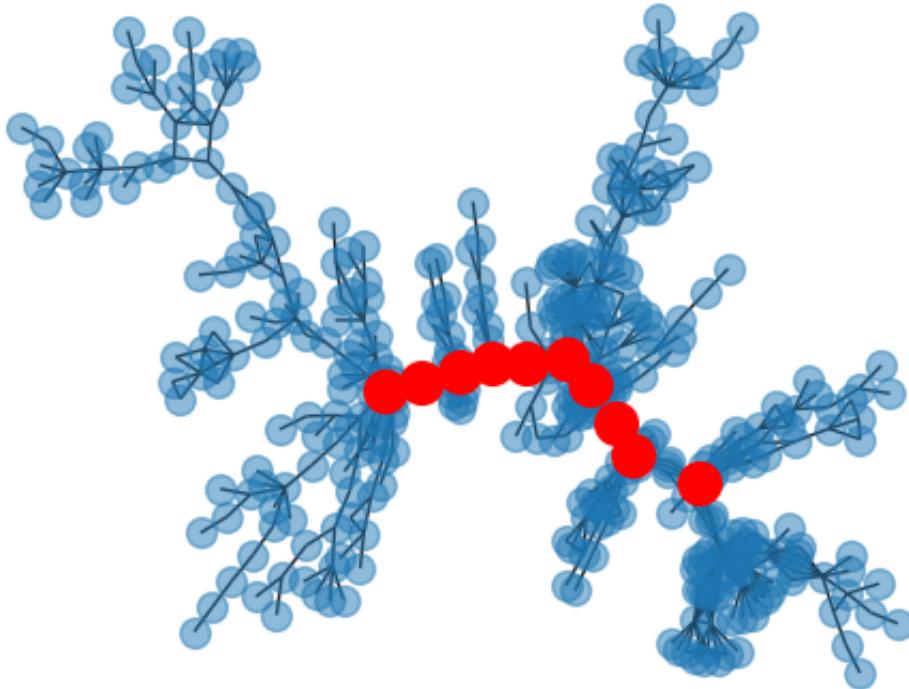
On analysing these nodes, we find that majority of the products are either books or music. The ratings of most of the products are around 5 or 4 however we do find some products with rating as 0. These could be

newly added products which havent been rated by the users yet. It is also worth mentioning that these are the products which havent received any reviews!

level_0	index	id	title	group	salesrank	categories	total_reviews	avg_rating
151	B00000IC82	B00000IC82	Laurel & Hardy - Flying Deuces/Utopia	DVD	44671	#Titles#Deals Under \$15#All Deals#Caprioli, Vittorio#Whelan, Tim#Laurel, Stan#Genres#Categories#Delair, Suzy#(L)#(C)#Classic Comedies#(F)#Amazon.com Stores#(D)#General#(B)#(H)#Hardy, Oliver#Frees, Paul#Today's Deals in DVD#Special Features#Directors#(W)#DVD Outlet#DVD#Actors & Actresses#Comedy#Berry, John#Amazon.com Outlet	13	3.0
170	0471394327	0471394327	Lafayette	Book	46461	#Books#General#History#Presidents & Heads of State#Historical Study#Revolutionary#Leaders & Notable People#Subjects#Biographies & Memoirs#Historical	8	5.0
932	0971024502	0971024502	Cabin III: The Unlawful Assembly at Winding Ridge (Cabin)	Book	498145	#Books#Literature & Fiction#General#Subjects#Contemporary	6	5.0
933	1878075063	1878075063	Beeswax: Production, Harvesting, Processing and Products	Book	656511	Business & Investing##Business & Investing Books#Books#General#Subjects#Home & Office#Amazon.com Stores	0	0.0
1224	0268000069	0268000069	An Anthology of Beowulf Criticism.	Book	215712	#Criticism & Theory#Books#Poetry#Literature & Fiction#General#History & Criticism#Epic#Subjects	0	0.0
2460	B00006L50C	B00006L50C	Sharanam (Refuge)	Music	287813	International##General#New Age#Indie Music#Specialty Stores#Music#Styles	3	5.0
2571	B000003C0B	B000003C0B	Firing Squad	Music	59068	#General#Indie Music#East Coast#Specialty Stores#Rap & Hip-Hop#Music#Styles#Gangsta & Hardcore	11	4.5
4722	B000002TNS	B000002TNS	Sings the Songs of Andrew Lloyd Webber	Music	67071	#Vocal Pop#General#Traditional Vocal Pop#Broadway & Vocalists#Easy Listening#Music#Styles#Pop	9	4.0
5709	1589131622	1589131622	0	Music	0		0	0.0
9426	B00005NORT	B00005NORT	A Scandal in Bohemia	Music	124702	#British Alternative#General#Indie Music#Rock#Specialty Stores#Alternative Rock#Music#Styles	4	5.0

We also did the same thing as above for betweenness centrality!

The results look quiet similar :

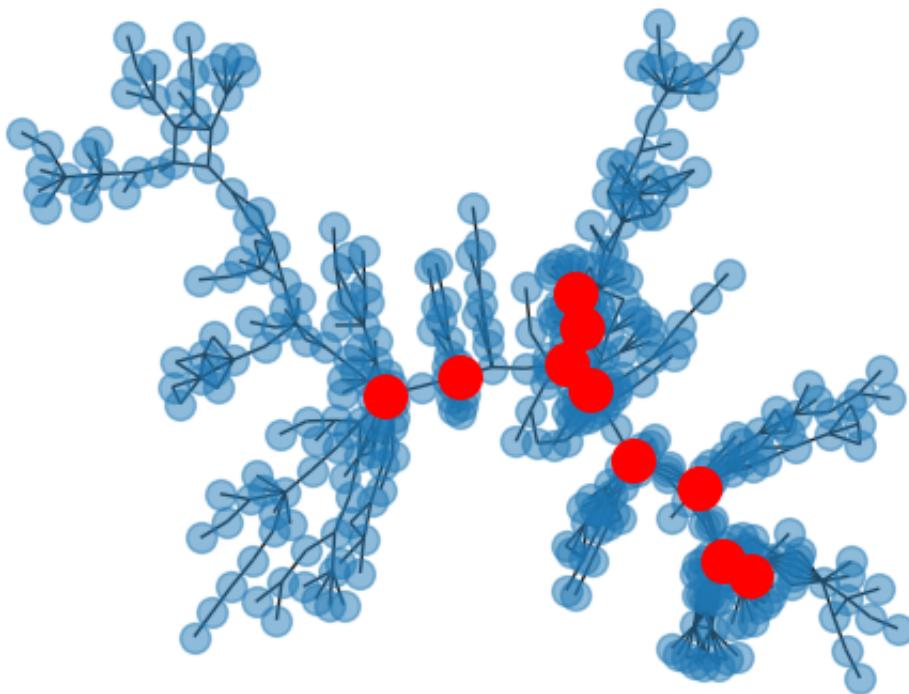


The top 10 nodes also belong to the similar categories, being books music or videos.

level_0	index	id	title	group	salesrank	categories	total_reviews	avg_rating
932	0971024502	0971024502	Cabin III: The Unlawful Assembly at Winding Ridge (Cabin)	Book	498145	#Books#Literature & Fiction#General#Subjects#Contemporary	6	5.0
1224	0268000069	0268000069	An Anthology of Beowulf Criticism.	Book	215712	#Criticism & Theory#Books#Poetry#Literature & Fiction#General#History & Criticism#Epic#Subjects	0	0.0
2460	B00006L50C	B00006L50C	Sharanam (Refuge)	Music	287813	International##General#New Age#Indie Music#Specialty Stores#Music#Styles	3	5.0
2571	B000003C0B	B000003C0B	Firing Squad	Music	59068	#General#Indie Music#East Coast#Specialty Stores#Rap & Hip-Hop#Music#Styles#Gangsta & Hardcore	11	4.5
4722	B000002TNS	B000002TNS	Sings the Songs of Andrew Lloyd Webber	Music	67071	#Vocal Pop#General#Traditional Vocal Pop#Broadway & Vocalists#Easy Listening#Music#Styles#Pop	9	4.0
6333	0697202097	0697202097	Introductory Biological Statistics	Book	1045813	Biology##Books#Professional & Technical#General#Research#Professional Science#Biological Sciences#Medicine#Subjects#Biostatistics#Science	1	5.0
9426	B00005NORT	B00005NORT	A Scandal in Bohemia	Music	124702	#British Alternative#General#Indie Music#Rock#Specialty Stores#Alternative Rock#Music#Styles	4	5.0
10676	1583480056	1583480056	The Dynamic Decision Maker: Five Decision Styles for Executive and Business Success	Book	306131	Business & Investing##Business & Investing Books#Books#General#Leadership#Management#Subjects#Home & Office#Amazon.com Stores#Management & Leadership#Reference	0	0.0
12262	6303046495	6303046495	The Leech Woman	Video	4476	#Genres#Dein, Edward#(A)#Categories#Classic Horror & Monsters#(G)#(T)#Gray, Coleen#Hamilton, Kim#Things That Go Bump#Video Outlet#Williams, Grant#Amazon.com Stores#Occult#(D)#Today's Deals in Video#General#VHS#(H)#Science Fiction & Fantasy#Horror#Special Features#Directors#(W)#Actors & Actresses#Goodwin, Harold#Alper, Murray#Talbott, Gloria#Amazon.com Outlet	9	2.5
26278	0862416310	0862416310	The Harlem Cycle: The Big Gold Dream; All Shot Up; The Heat's on	Book	1437012	#African American#Books#World Literature#Literature & Fiction#United States#Subjects#(H)#Himes, Chester#Authors, A-Z	0	0.0

Comparing the top 10 nodes for degree centrality and betweenness centrality, we can see that some of these products are present in both!

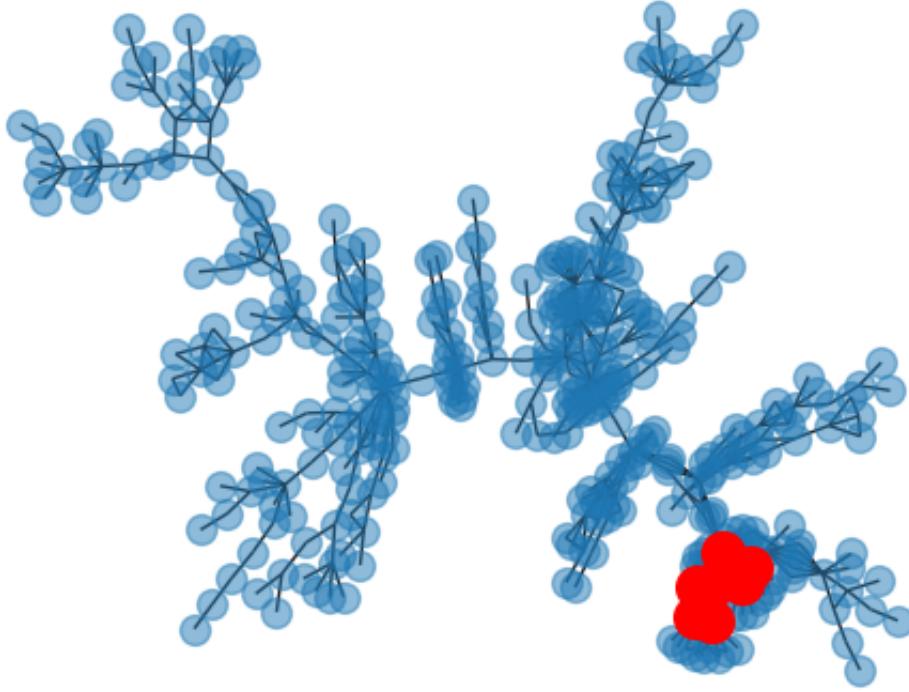
Doing the same as above for pagerank, we get the following graph:-



On analyzing these nodes, once again we find products which can be seen in the tables of degree and betweenness centrality!

level_0	index	id	title	group	salesrank	categories	total_reviews	avg_rating
151	B00000IC82	B00000IC82	Laurel & Hardy - Flying Deuces/Utopia	DVD	44671	#Titles#Deals Under \$15#All Deals#Caprioli, Vittorio#Whelan, Tim#Laurel, Stan#Genres#Categories#Delair, Suzy#(L)#(C)#Classic Comedies#(F)#Amazon.com Stores#(D)#General#(B)#(H)#Hardy, Oliver#Frees, Paul#Today's Deals in DVD#Special Features#Directors#(W)#DVD Outlet#DVD#Actors & Actresses#Comedy#Berry, John#Amazon.com Outlet	13	3.0
170	0471394327	0471394327	Lafayette	Book	46461	#Books#General#History#Presidents & Heads of State#Historical Study#Revolutionary#Leaders & Notable People#Subjects#Biographies & Memoirs#Historical	8	5.0
932	0971024502	0971024502	Cabin III: The Unlawful Assembly at Winding Ridge (Cabin)	Book	498145	#Books#Literature & Fiction#General#Subjects#Contemporary	6	5.0
933	1878075063	1878075063	Beeswax: Production, Harvesting, Processing and Products	Book	656511	Business & Investing##Business & Investing Books#Books#General#Subjects#Home & Office#Amazon.com Stores	0	0.0
1224	0268000069	0268000069	An Anthology of Beowulf Criticism.	Book	215712	#Criticism & Theory#Books#Poetry#Literature & Fiction#General#History & Criticism#Epic#Subjects	0	0.0
2460	B00006L50C	B00006L50C	Sharanam (Refuge)	Music	287813	International##General#New Age#Indie Music#Specialty Stores#Music#Styles	3	5.0
2571	B000003C0B	B000003C0B	Firing Squad	Music	59068	#General#Indie Music#East Coast#Specialty Stores#Rap & Hip-Hop#Music#Styles#Gangsta & Hardcore	11	4.5
4722	B000002TNS	B000002TNS	Sings the Songs of Andrew Lloyd Webber	Music	67071	#Vocal Pop#General#Traditional Vocal Pop#Broadway & Vocalists#Easy Listening#Music#Styles#Pop	9	4.0
5709	1589131622	1589131622			0		0	0.0
9426	B00005NORT	B00005NORT	A Scandal in Bohemia	Music	124702	#British Alternative#General#Indie Music#Rock#Specialty Stores#Alternative Rock#Music#Styles	4	5.0

Finding the eigenvector centralities, the following graph is seen for the top 10 nodes:-

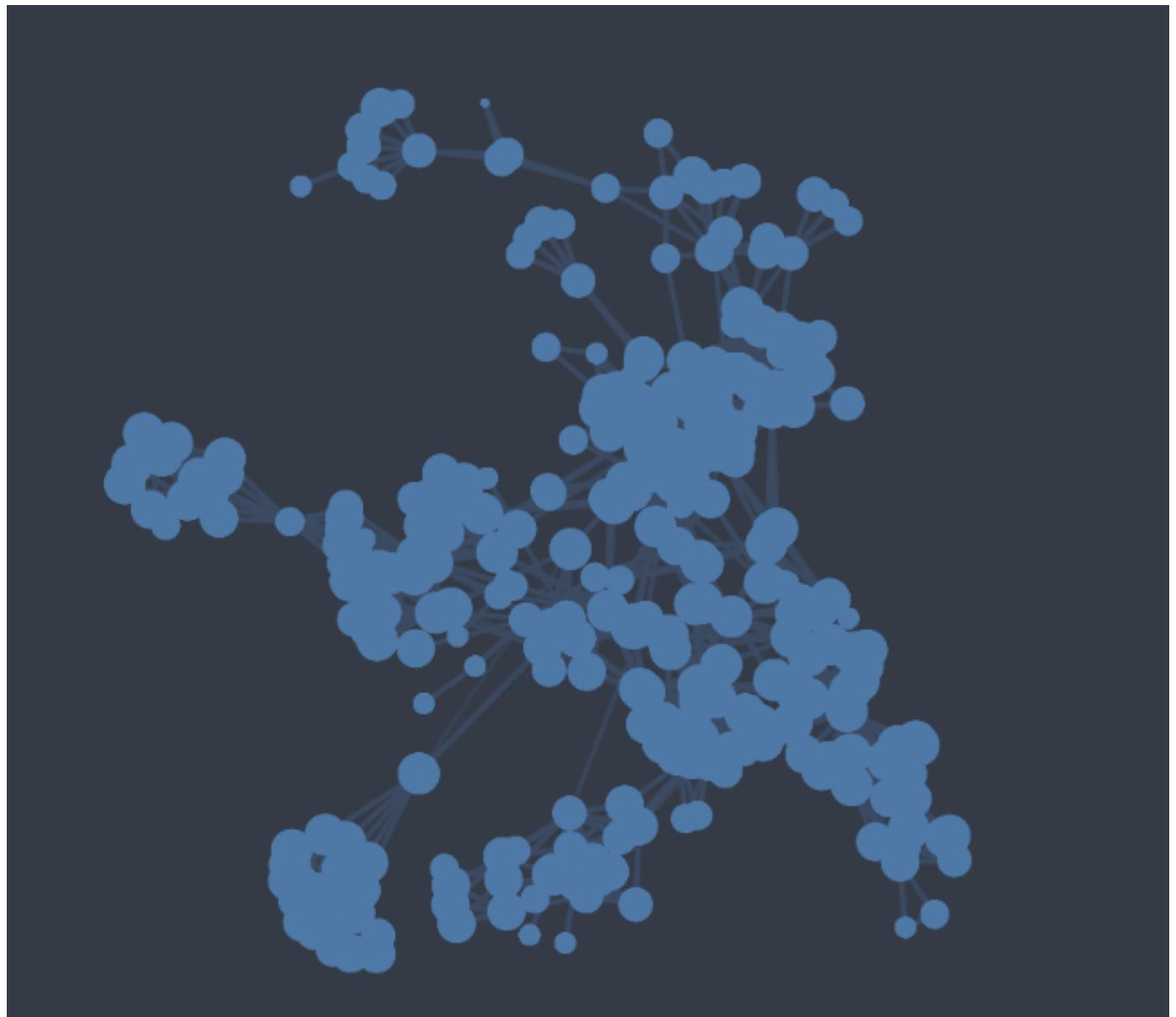


It is interesting to see that all the nodes are around one cluster and not spread out unlike the results from other metrics!

However, on analyzing these nodes we do find some similar products! Once again the majority products are books or music.

	index	id	title	group	salesrank	categories	total_reviews	avg_rating
932	0971024502	0971024502	Cabin III: The Unlawful Assembly at Winding Ri...	Book	498145	#Books#Literature & Fiction#General#Subjects#C...	6	5.0
933	1878075063	1878075063	Beeswax: Production, Harvesting, Processing an...	Book	656511	Business & Investing##Business & Investing Boo...	0	0.0
4122	0658016903	0658016903	Way-Cool French Phrase Book : The French That ...	Book	193304	Instruction##Books#French#General#People & Pla...	0	0.0
6515	0471252050	0471252050	Radar Principles	Book	424679	#Books#Electrical & Optical#Engineering#Subjec...	2	3.0
9726	0911910131	0911910131	Merck Index: An Encyclopedia of Chemicals, Dru...	Book	4237	#Science#Health, Mind & Body#Books#Professiona...	8	4.0
9727	0867094753	0867094753	The English Teacher's Companion: A Complete Gu...	Book	249814	Words & Language##Books#Nonfiction#General#Ped...	14	5.0
16453	B00004W40Y	B00004W40Y	Strangers in Paradise	Music	517059	#General#Indie Music#Rock#Specialty Stores#Cou...	5	4.5
21013	B000007WGY	B000007WGY	H	Music	421301	International##Imports#General#Rock#Specialty ...	0	0.0
24626	B00000DPQS	B00000DPQS	Dubliners Irish Favorites	Music	377372	#Travel#Bargains#Amazon.com Outlet#Music#Brita...	0	0.0
30742	0325000964	0325000964	Green Land, Brown Land, Black Land: An Environ...	Book	705517	#Books#Africa#Professional & Technical#History...	2	3.5

We found the communities in the entire network which amounted upto 6249 communities.
 We then found the largest community among these:-



It is worth mentioning that the community comprises entirely of books. This shows that on an average the product which is most co-purchased together in the dataset is books.

Analysing this community shows that the product with the highest pagerank is none other than the book 'Sleeping Beauty'!

Everyone Loves Graphs?

Not these ones man

Time for everyones least favourite part of the day.

Matplotlib 😞

It's okay, reading graphs is an important skill to have for writing the GRE.

Harnessing the power of placing things on the x and y axes and *matplotlib* 😞 we have for you a great selection of graphs that I will proceed to explain.

The intuition behind these graphs is one simple question:

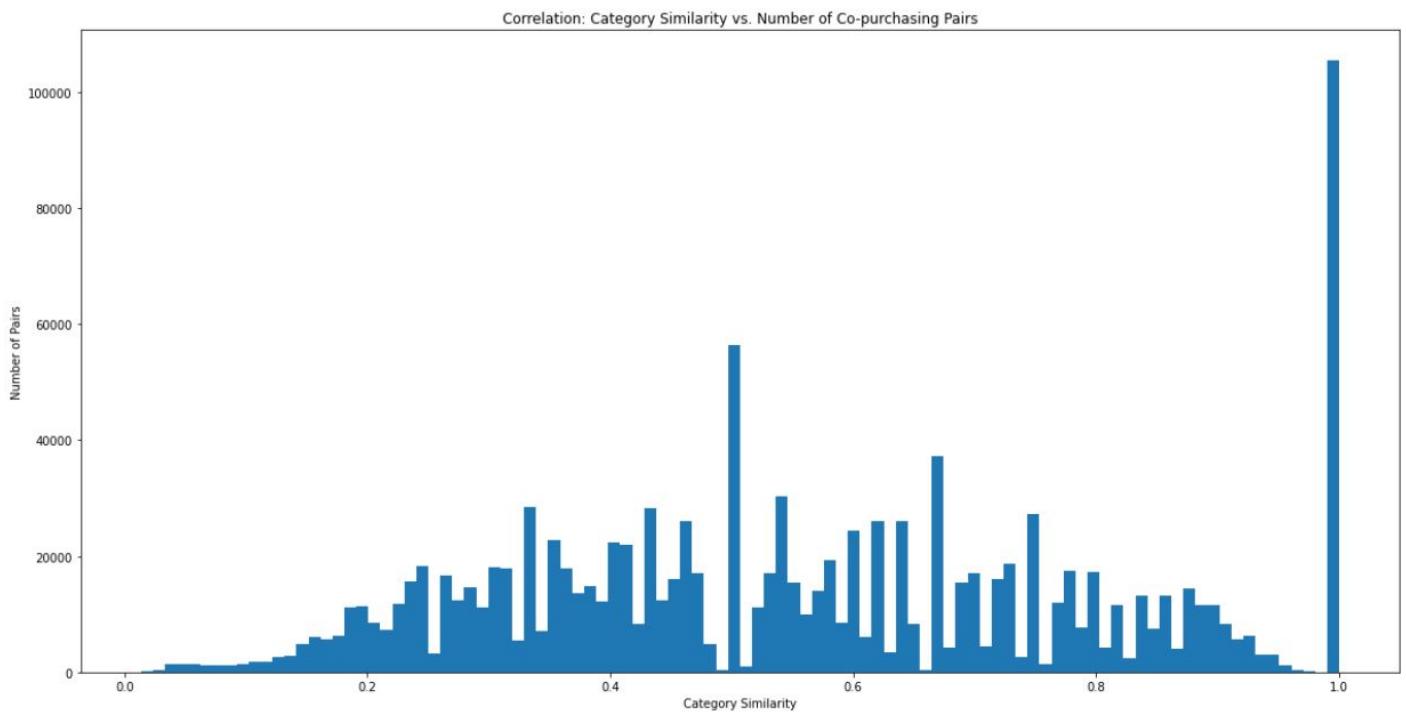
"Is there information about this product that will lead to the purchase of another product?"

First up,

1. Category

Any product that you buy can be placed into a category based on its contents or function. For example, if you were to buy a book, say Sherlock Holmes, it would probably be in a category like Mystery.

So what we want to see is if two products with similar categories are more likely to be co-purchased.



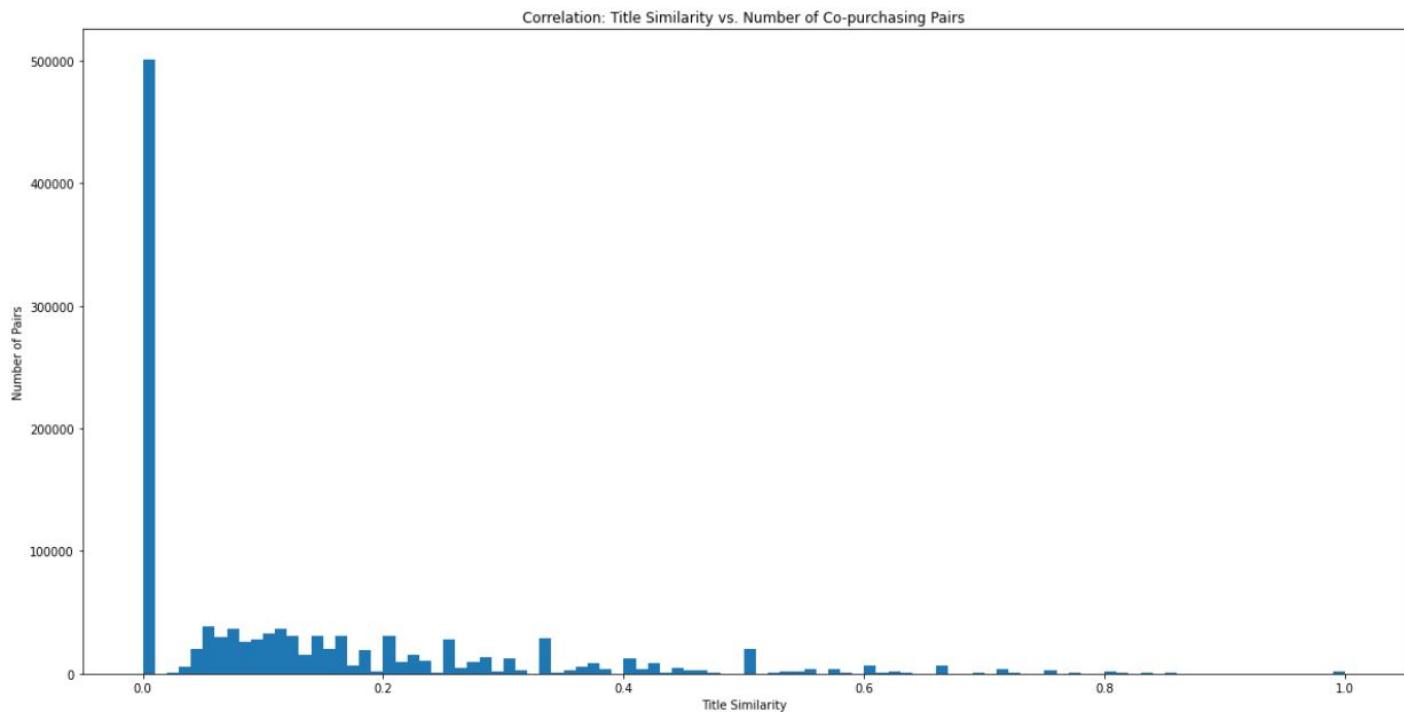
So as we can see in this graph it looks like the influence of category similarity on whether two products are co-purchased or not increases until the halfway point. At this point it keeps decreasing until it suddenly spikes to the highest value for full category similarity.

What this is telling us is that having the same categories makes the co-purchase of two products extremely likely. It also looks like the likelihood of two products being co-purchased increases with higher category

similarity as we can see with the graph going up until 0.5 similarity. I imagine that the reason co-purchased products starts to go down after that point is that having a 0.8 or 0.6 similarity just is not very likely, so there are not enough samples to show a linear increase.

2. Title

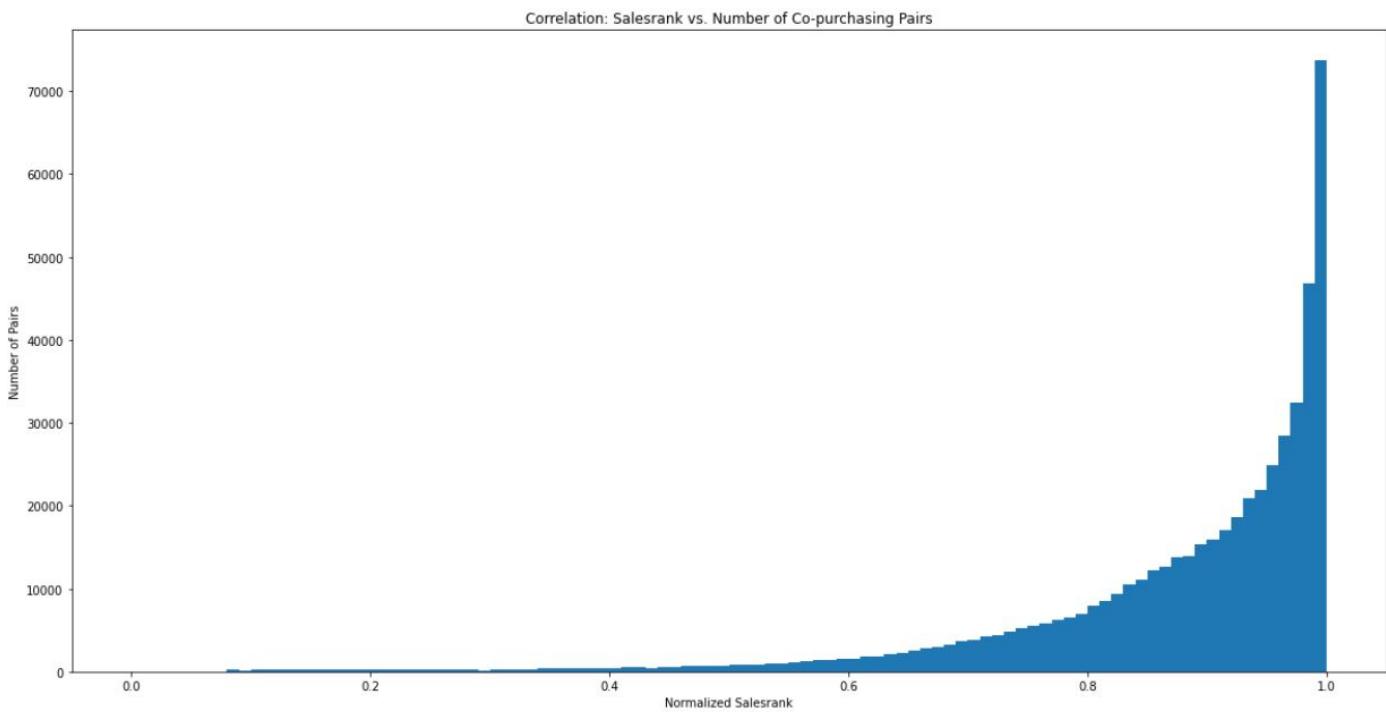
Now, we are checking if two products with similar names are more likely to be co-purchased.



Clearly not. However, this might just be because most products just have completely different names. If we were looking at a more specific dataset like a dataset of smartphones we might start to see a positive correlation, owing to phone brands having product lines which will have similar phone names like Galaxy M series and A series and also many phone companies using blanket terms like Pro and Pro Max.

3. Salesrank

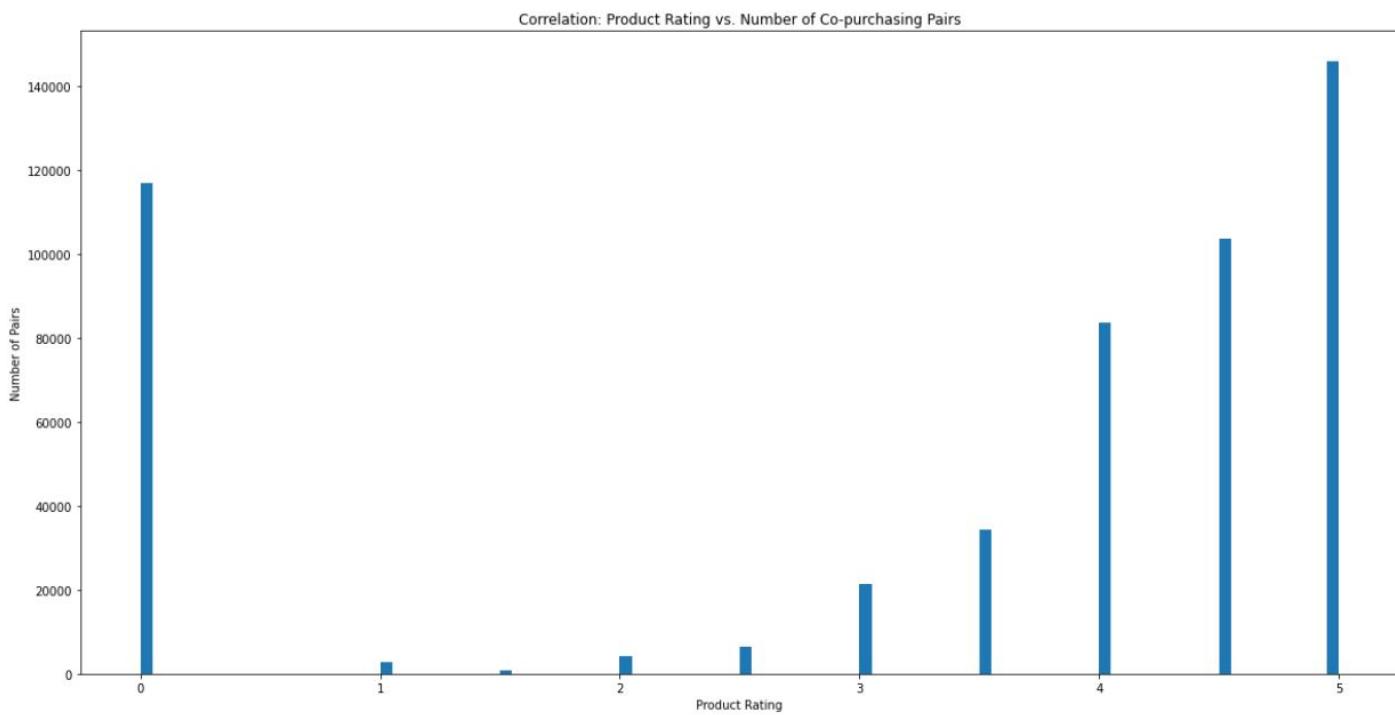
Here's a straightforward one. Are best selling products more likely to be purchased with other bestsellers?



This kind of question is what we call a no-brainer. We also call it a rhetorical question (Note that one down too GRE preppers, you'll need it for the writing section).

4. Product Rating

This one is also a no-brainer. It makes a lot of sense that better rated products will be purchased together.



And yeah it's true. You'll also notice that the number of co-purchasing pairs increases as the product rating gets higher. Which again makes a lot of sense, since people aren't gonna buy many poorly rated products. Although you will notice that a whole lot products with no rating are also purchased together. But as I explained before this discrepancy probably has a lot to do with the landscape of e-commerce in the early

2000s.

Conclusion

And there you have it, our analysis of the Amazon Co-Purchasing Network. Of course we don't have a rigorous analysis with a ton of numbers and spreadsheets, but we hope that this lighthearted look into the early days of e-commerce on Amazon was insightful for you.

The type of products people used to purchase, what was popular at the time, etc.

We saw how online shopping was primarily used for books at the time, and interestingly music as well which no one purchases these days due to the existence of Spotify.

We saw the different communities that can form around certain types genres. Like who knew so many people were buying Anime VHSs on Amazon considering the low popularity of Anime at the time. Most people just watched Pokemon on TV, yet here we're seeing tons of obscure shows being bought in that tiny little clusters.

And finally we encourage all of you to use Cosmograph as well to take a look around the graph yourselves and see how popular the things you like now were back in the day. Of course you would need to pre-process the metadata first, but we can provide with the pre-processed data if you'd like.

Thank you all.

Meet The Team



Trust me he was in the picture.

Now why would I lie to you.

He was just sick, that's why he looked so green.