Amazon Music Products and Genre

Andrew Kos
Ivy Yun
Harry Chen

Abstract

Amazon is an online marketplace with millions of products and a massive amount of reviews. Certainly, within these product and user networks of interaction live a plethora of social network analysis insight and business benefit. In our study, we chose to focus on the Music category of products hoping to find a correlation between how users leave reviews and the genres the products belong to. Our dataset provided review helpfulness scores and ratings which we chose to focus on as they provide a fairly straightforward indication of sentiment about a review. Helpfulness measures how many users found a review “helpful”, indicating the quality of the review. Rating is assigned by the review author and is a quantified measure of their satisfaction with the product they are reviewing.

Our hypothesis was that helpfulness and rating would be correlated to product genres, in a way indicating the sentiment induced by that particular genre of music. Does rap elicit more controversy? Are there genres of rock that more negative users gravitate towards, causing the genre to have more negative reviews and lower helpfulness ratings? We explored these questions and our answers came up with some interesting insights, but are ultimately inconclusive, though later studies could employ some of these methods to build on our findings.

Data Preparation

Our analysis focuses on a subset of products and users in the Music category which is composed of 64,706 reviews spanning across 5,541 unique users and 3,568 unique products, representing a pre-filtered set of 5-core entities, or products that have received at least 5 reviews and users that have reviewed at least 5 products. To get the dataset into a format we could use for analysis, some transformations were performed. The 5-core data is provided as a JSON list of reviews from which we extracted unique products and users into nodes on a directed graph, the review representing a directed path from a user to a product. Review paths in the graph include the following attributes we used for analysis:

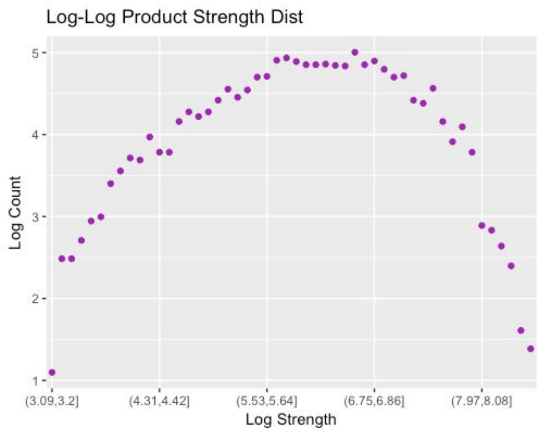
Name	**Range**	**Description**
helpful_1	[0, ∞)	Number of users who responded “Yes” to “Was this review helpful to you?” for this particular review.
helpful_2	[0, ∞)	Number of users to responded “Yes” *or* “No” to “Was this review helpful to you?” for this particular review.
overall	[0, 5]	Overall score user gave to the product being reviewed.

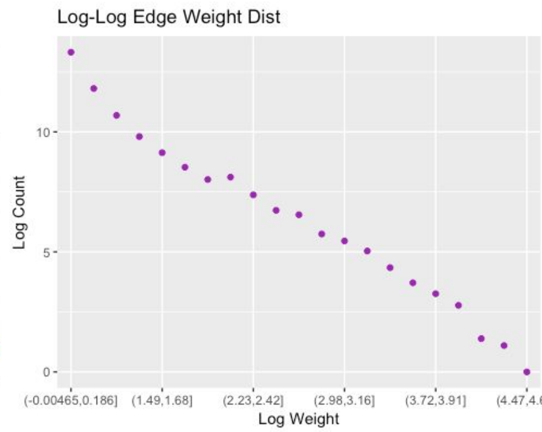
We also added titles to the products and extracted product categories into a hierarchy relating categories to each other and categories to products; some examples of categories we extracted: *Indie Rock, Gangsta & Hardcore, Singer-Songwriters, Soul*.

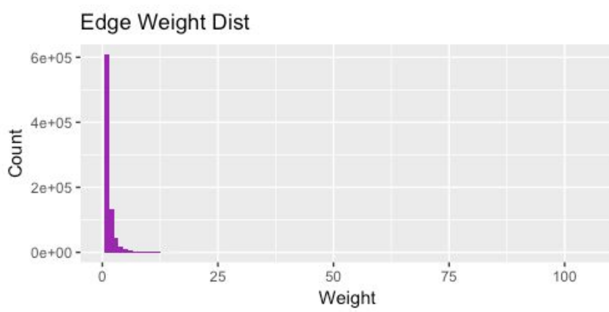
The following visualization is the full graph with high degree categories and products shown. It is interesting to note that Rock and Rap end up on different ends of the network with genre mixing products in the middle.

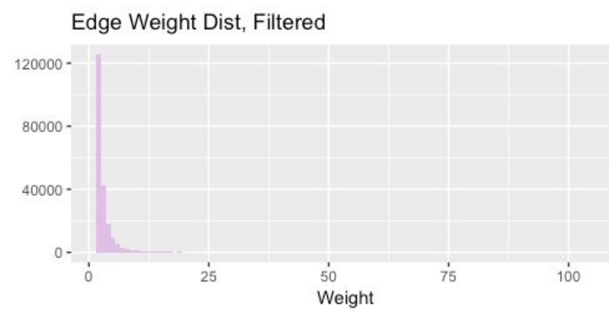
Digital Music
Country Rock
Progressive
Album-Oriented Rock (AOR)
Folk Rock
Progressive Rock
Machine Head
Soft Rock
Hard Rock
Contemporary Folk
Blood on the Tracks
Folk
Metal
Oldies
Singer-Songwriters
Tapestry
Pet Sounds
Van Halen
New Wave
1984
New Wave & Post-Punk
Vocal Pop
Adult Contemporary
Are You Experienced
Indie & Lo-Fi
Indie Rock
Pop Metal
Adult Alternative
Grace
Alternative Rock
What's Going on
Like a Prayer
Afterglow
Little Earthquakes
Punk
Come Away with Me
Bad
Feels Like Home
Californication
Elephant
Ray of Light
No Angel
Weezer (Blue Album)
Thriller
Off the Wall
Dance Pop
Electronica
House
Toxicity
Soul
Dance & Electronic
Blues
Funk
Chocolate Starfish and the Hot Dog Flavored Water
Break the Cycle
Meteora
The Diary of Alicia Keys
Contemporary R&B
Who Is Jill Scott? Words and Sounds, Vol. 1
Speakerboxxx/ The Love Below
The Eminem Show [Limited Edition w/ Bonus DVD]
The Marshall Mathers LP
Encore (Deluxe Edition)
Get Rich Or Die Tryin
2Pac Greatest Hits
Ready to Die [Vinyl]
The Massacre
All Eyez on Me
Chronic 2001
Pop Rap
Life After Death
Blueprint
Beg for Mercy
West Coast
East Coast
The Documentary
Me Against the World
God's Son (Clean)
Gangsta & Hardcore
The Don Killuminati (The 7 Day Theory)
Hunger for More (Bonus DVD)
Southern Rap

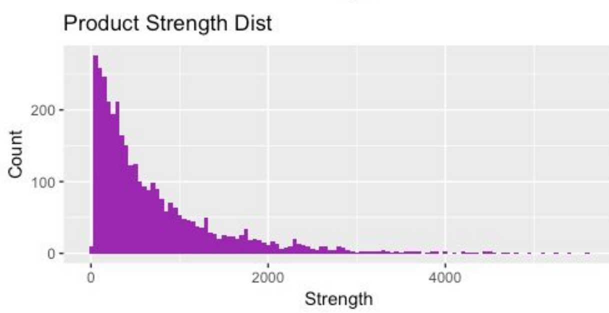
Once this bipartite graph of users and products was put together, we created a product-product projection to aid in investigating this network. This projection ends up being large, 829,575 edges, so to make these projections more manageable, we filtered out edges of weight 1 then removed nodes of strength less than 50.

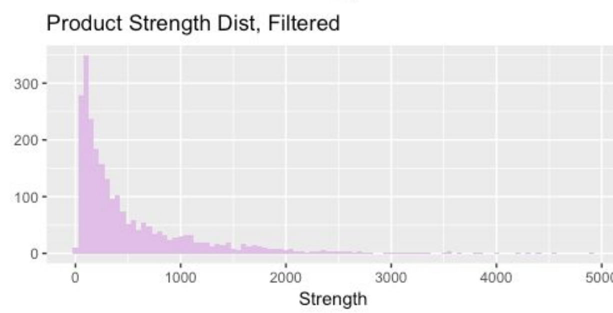












Using this filtered projection we investigated the hypothesis that music genres affect the way users interact with products and each other, specifically by looking at communities of products through the lenses of helpfulness scores and ratings.

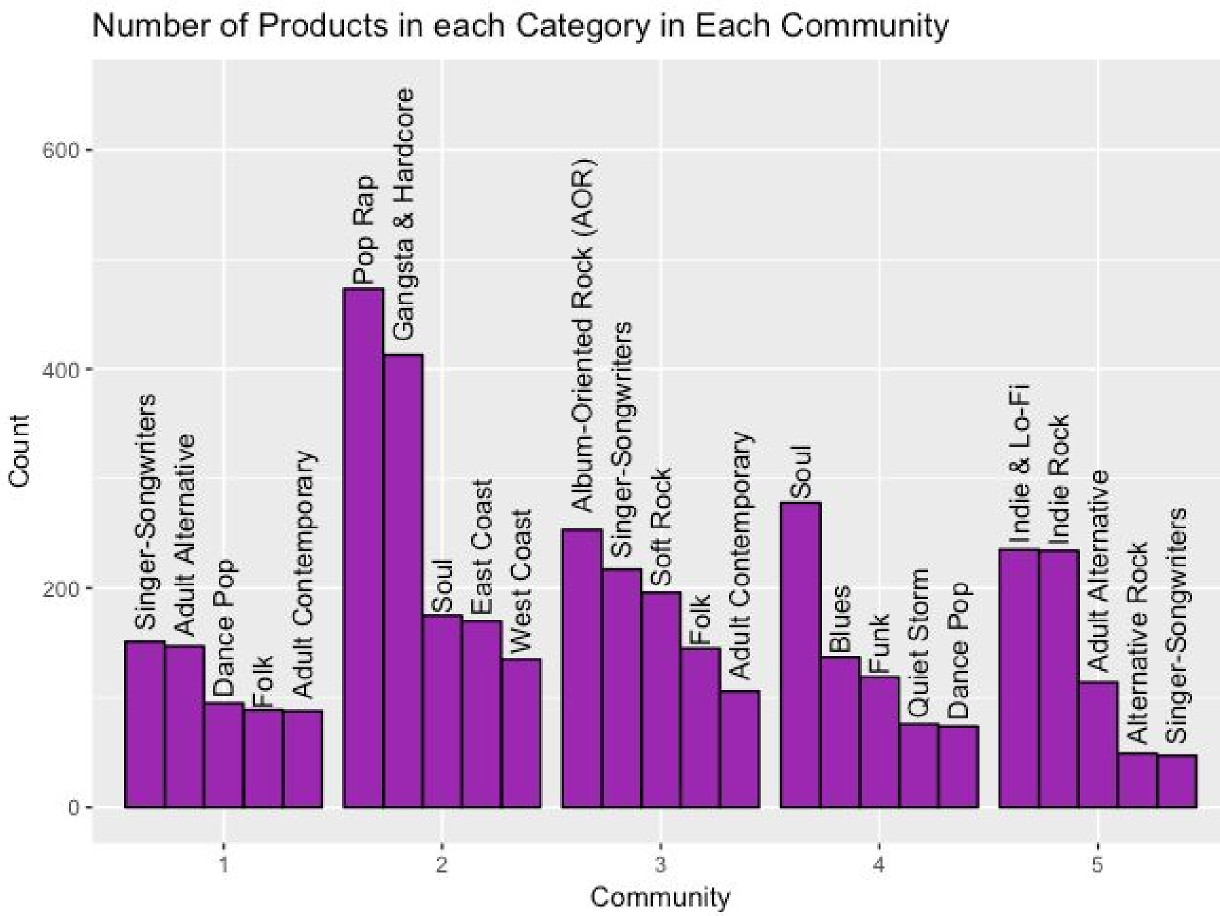
Discovering Genres

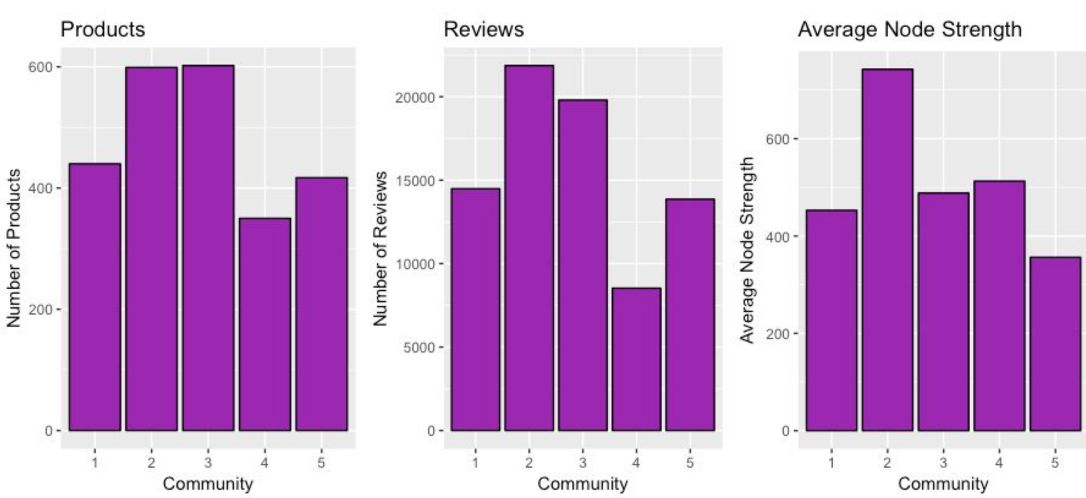
Before analysis on “genres” could begin, we had to construct them using a clustering algorithm. Running the product-product projection through the basic list of clustering algorithms produced the following results:

Algorithm	**Clusters**	**Modularity**
Walktrap (4 steps)	14	0.4142
Leading Eigenvector	7	0.3990
Walktrap (10 steps)	9	0.3906
Fast Greedy	5	0.3815

We also attempted betweenness clustering, but the graph is too large to calculate clusters with this algorithm in a reasonable time with the computers we had available to us. Walktrap with 4 steps produces a high modularity compared to the other algorithms, but there are many clusters that contain only a few products primarily in the *Digital Music* category--which represents digital singles that are missing some data and not really directly comparable to the rest of the products which are albums. In a similar manner, walktrap with 10 steps produces too many communities. Fast greedy produces a meaningful number of communities but has a lower modularity than leading eigenvector clustering. The latter clustering algorithm also successfully isolates products in *Digital Music* and a couple miscategorized products. Given these distinctions, we elected to use the communities produced by leading eigenvector clustering for the rest of the analysis.

To determine the “genre” of each community, we iterated through the list of products in each community and compiled a count of the categories each was connected to, then took the top 5 as representative of that community’s genre. Before running this analysis we removed the two highest degree categories *Pop* (degree of 1486) and *Rock* (degree of 869) because they cover too broad of a set of products to be meaningful: *Pop* connects almost every other category and *Rock* is better described by its sub-categories. The following shows the top 5 categories in each genre, note that multiple categories can be assigned to a single product:

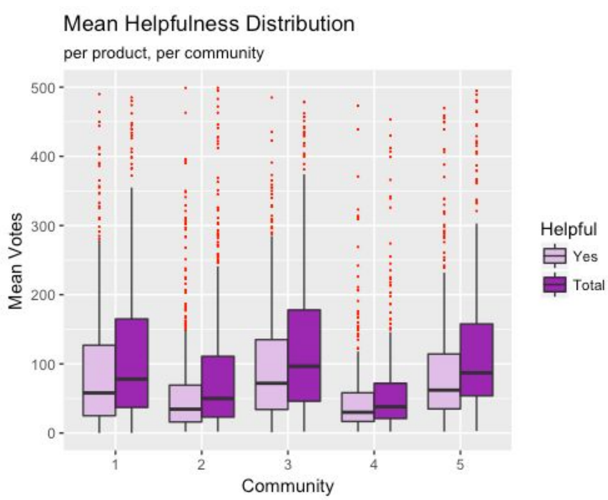


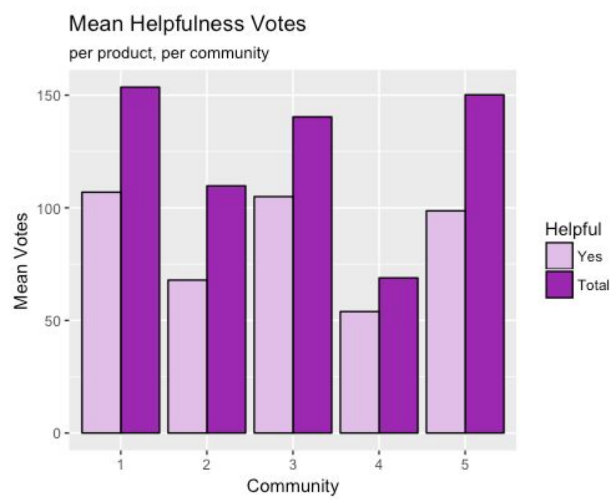


Helpfulness and Genre

Given our genres, we first explored their relationship to helpfulness scores. Helpfulness is based on a user responding “Yes” or “No” to the question “Was this review helpful?” which is presented under a review. Provided for each review in the dataset is count of users who responded “Yes” (H_Y) and the total number of respondents to this question (H_T); thus, the count of users who responded “No” would be $H_T - H_Y = H_N$. To facilitate some comparisons between products we also calculated totals for these helpfulness metrics across the reviews for a product ($H_T(p) = \Sigma_R H_T$ and $H_Y(p) = \Sigma_R H_Y$).

At first glance, it may be Our hypothesis was that certain genres of music might provoke a higher count of H_Y relative to other genres, perhaps based on musical qualities or some other underlying attribute of the genre. Given the $H_T(p)$ and $H_Y(p)$ across products, we compiled those into totals for each genre:



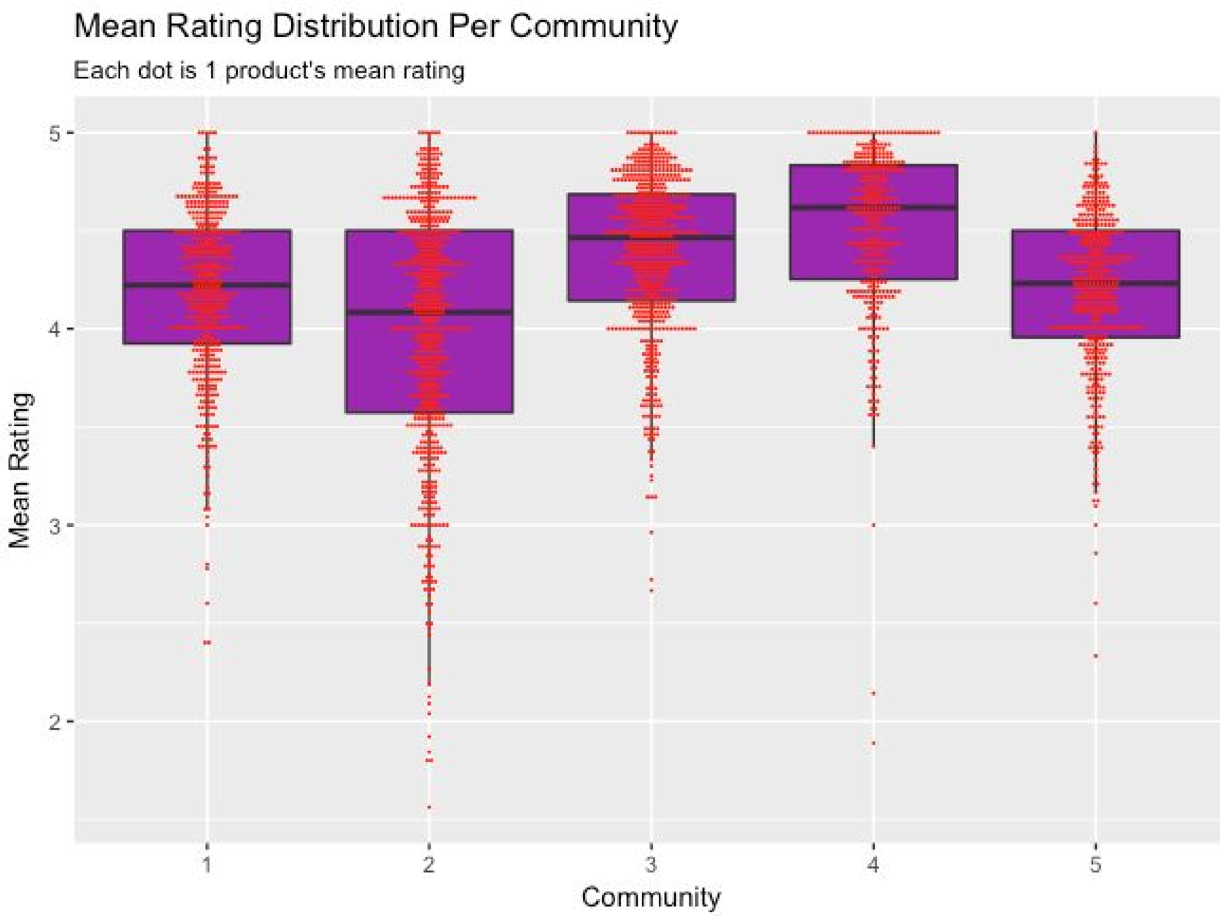


It is interesting to note that genre 2 (*Pop Rap, Gangsta & Hardcore, Soul, East Coast, West Coast*) has a low response rate to the helpfulness of reviews compared to other genres even though it has the most products and reviews; this pattern is also apparent in the box plot. We hypothesize this could be related to genre 2 (*Pop Rap, Gangsta & Hardcore, Soul, East Coast, West Coast*) being more divisive than other genres.

As a last test of helpfulness ratings and genre, we tested the assortativity of these attributes. Binning these ratings into 10 discrete bins and testing assortativity gives a score of 0.02024 for H_T across all products and a score of 0.02724 for H_Y across all products. We ran a CUG test and QAP test on these attributes to see what kind of results we would get; however, given that the assortativity of these attributes is so near 0, CUG and QAP does not verify or deny much as all the random nets produced have similar assortativity near 0. It appears through our analysis that there is little correlation between genre and helpfulness.

Rating and Genre

Another hypothesis we considered was if different genres provoked different overall review ratings. When a user makes a review on Amazon, they can leave an overall score for the product at 0.5 intervals from 0 to 5, inclusive, presumably summarizing the sentiment of their review. Taking the mean of the overall rating across all reviews for a product (*O(p)*) and visualizing the distribution per genre yields the following:

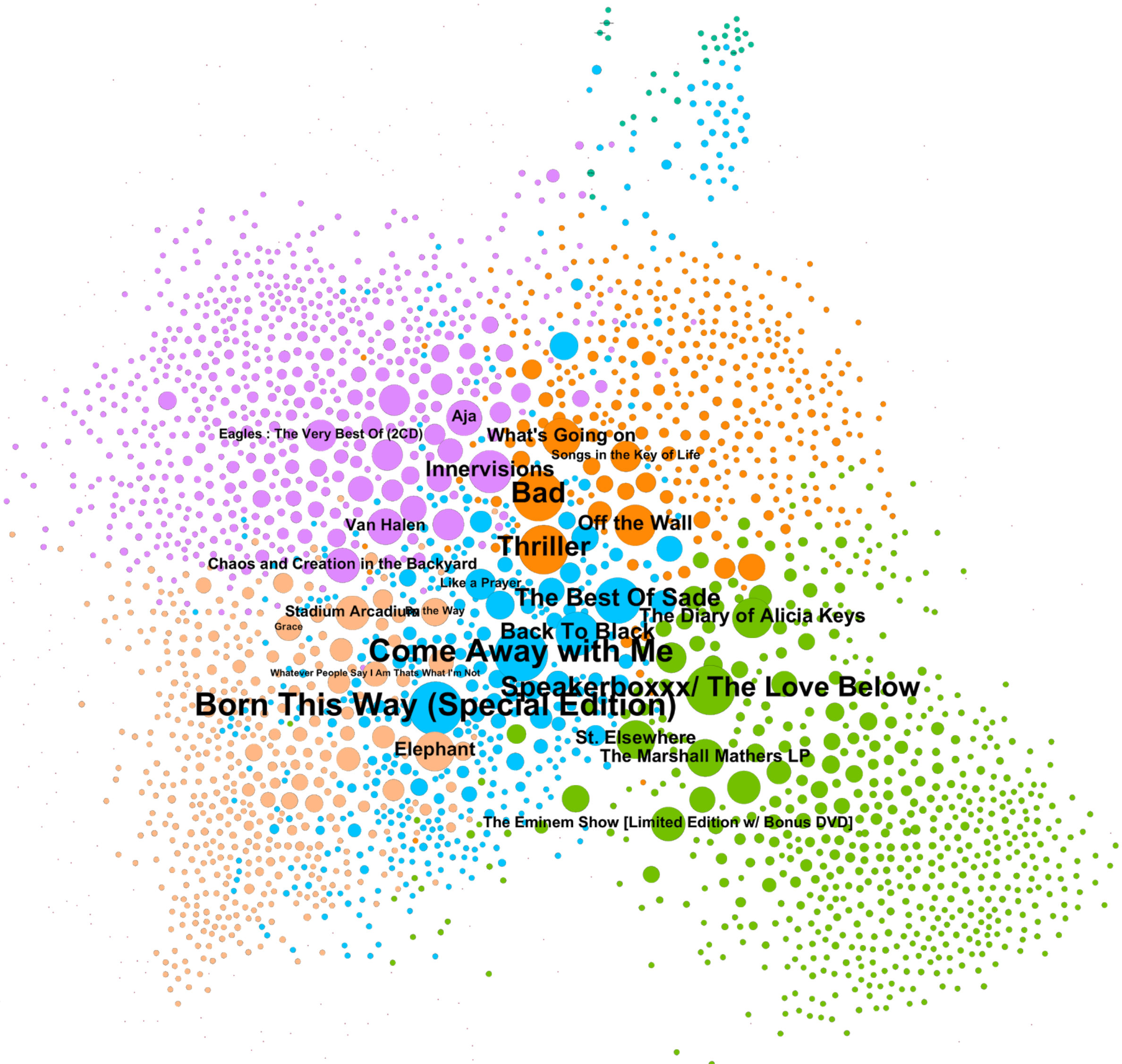


Interestingly, genre 2 (*Pop Rap, Gangsta & Hardcore, Soul, East Coast, West Coast*) has the widest distribution of ratings; a possible hypothesis is that this genre is more controversial than others. Another note is genre 4 (*Soul, Blues, Funk, Quiet Storm, Dance Pop*) appears to have a rating distribution highly skewed towards 5; a hypothesis here is that users reviewing this genre generally view this music positively perhaps lending credit to our original hypothesis about ratings and genre. Overall, most users appear to be leaving reviews rating products 4 and above which could mean either 5-core users are more apt to leave positive reviews, or users generally are leaving positive reviews; this could be indicative of the user norm problem.

Unfortunately, checking the assortativity of review means, binned into 10 bins, gives a value of 0.04554, which suggests that review means have little to no relation to these genres. We also ran CUG and QAP tests for this metric, see appendix.

High Betweenness Products

Being a measure of how information flows through a network, in the context of this product review network, we hypothesize high betweenness correlates with products that are mixing genres. At a glance, the high betweenness products sit near the center of a Force Atlas 2 layout of the product-product projection which, at least, suggests that these products share more user reviews with people outside the genre than other products in the genre the products belong to. To really dive into this hypothesis we assume we would need a much deeper analysis than provided here of the musical properties of these products and their correlation with betweenness.



Conclusions

Helpfulness scores and ratings appear to have little to no correlation with the properties of the genres we have discovered. There may be other underlying properties of the network that correlate with these statistics, but genres as we have them appear not to be related. Further investigation could be done into why genre 2 (*Pop Rap, Gangsta & Hardcore, Soul, East Coast, West Coast*) has such a wide range compared to the other genres and why genre 4 (*Soul,* *Blues, Funk, Quiet Storm, Dance Pop*) has ratings skewed highly towards 5. In general, more work could be done to dig into the individual genres themselves to see what types of sub-communities there are; especially within genre 2 (*Pop Rap, Gangsta & Hardcore, Soul, East Coast, West Coast*) and within the *Rock* category.

More broadly, having insight into how users interact with products through genre could benefit Amazon in a number of ways. For instance, if a user is reviewing products in a genre that has a low amount of helpfulness votes, Amazon could encourage the user to leave a helpfulness vote, improving the credibility of reviews in less popular genres. Using mean review scores, Amazon could target advertisements for high score products in a particular genre at users who rate many products in that same genre. Assuming our hypothesis about high betweenness products being genre mixers, users who prefer one genre over another could be given a bridge from their preferred genre to others by encouraging them to purchase high betweenness products.

Given more time and data, there are some avenues of analysis we would pursue. The dataset included time stamps for reviews which could be run through a dynamic network analysis to parse out trends in the way products are reviewed: perhaps there are influential users who, after leaving a review on a product, attract other users to review that product. With additional computing power, we could run this network through a link prediction model to see what genres a user might be interested in based on which products they've reviewed in the past. User review data could also be used with a technique like correlation-based similarity to find users who have reviewed similar products then recommend new products to purchase and review.

Amazon has an abundance of data and this analysis barely scratches the surface of the insights hiding within the bits. Deeper exploration of this data would be beneficial to both social network analysts and business alike.

Citations

McAuley, Julian. "Amazon Product Data." *Amazon Review Data*. Web. 02 May 2017. http://jmcauley.ucsd.edu/data/amazon/links.html

Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering

R. He, J. McAuley

WWW, 2016

pdf

Image-based recommendations on styles and substitutes

J. McAuley, C. Targett, J. Shi, A. van den Hengel

SIGIR, 2015

pdf