Social Attributes in Collaboration between Spotify Artists and Comparison with Scientific Collaboration

Ayan Chatterjee

Abstract: Collaboration networks have been widely studied in different social scenarios and these networks provide a broad insight into how people cooperate to achieve certain goals. Historically, inter-disciplinary collaboration has been infrequent in science. Researchers from the same domain of science and humanities tend to work together to contribute towards that domain. The scenario is different in music production. Collaboration between artists from different genres is common and inter-genre collaborations produce numerous popular songs. Social attributes like gender and race have significant influence on musical collaborations. Furthermore, collaboration patterns change over time. Artists mostly collaborate during their early and mid-careers.

Keywords: Collaboration, Network, Spotify, Music, Science, Social Science, Assortative, Temporal patterns

Introduction: Collaboration in various social scenarios enhances the possibility of completion of complex tasks and enables faster execution. Cooperation is an embedded trait in all of us. Collaboration is not only observed modern science and technology, but also in multiple domains in art. With the advent of versatile mathematical analysis techniques and network science, social connections have been easier to visualize and understand, which has provided us with deeper insights into the collaboration patterns in our society.

Various methods have been used over the years to study human collaborations and endorsements, especially in different domains of science. These methods include survey/questionnaire, bibliometrics, and complex network analysis [1]. Easley and Kleinberg [2] have studied several forms of collaboration networks like co-authorship networks, email communication networks, citation networks, and peer-to-peer networks. A work by Lee and Bozeman [3] shows that research collaboration has a positive effect on publishing productivity [3]. Their findings indicate that the number of collaborators is strongly associated with social attributes like age, rank, grant, gender, marital status, family relations, citizenship, job satisfaction, perceived discrimination, and collaboration strategy. Similar sociological study has been conducted in this work for musical collaborations.

Existing literature includes studies on different aspects of creative careers like music. Dynamics and predictability of success has been explored by Yang et al. in [4]. Janosov et al. have quantified luck and the impact of luck in creative careers [5]. In a case study, Danker [6] has analyzed the opportunities and constrains musical artists encounter using a Spotify collaboration network considering the use case of Dutch drum and bass artist Noise. South [7] has studied various network aspects of Spotify collaboration subgraphs and have made conclusions based on network measures like node centrality (e.g., classical artists are found to be the most central to the whole network, while the rap artists are found to be the most central to the popular subgraph).

In this work, multiple hypotheses related to social aspects of musical collaborations have been formulated and tested using network analysis. Broad goal of this work is to understand how collaboration patterns change between domains and the effect of social attributes on collaboration by drawing deeper interpretations from mainstream network analysis. This article is divided in four major sections: description of Spotify data and data cleaning, network visualization and macro level statistics, introduction to collaboration fraction and finally drawing social interpretations from the collaboration network.

Data: Spotify Web API curated data of artists and songs released between 1921-2020 has been used for the analysis. This dataset contains information about more than 160,000 songs. The artist names corresponding to a song available on Spotify are mentioned in list format, which means it contains the info whether multiple artists have collaborated on a song. The dataset also includes features of the songs like acousticness, duration, tempo, genre, loudness, release date, and popularity. The popularity values are normalized and are based on the total number of streams a song has on Spotify. Based on the presence of keywords, the songs with genre information have been classified under twelve musical genres: Metal, Jazz, Pop, Country, Indie, Oldies, EDM, Classical, Rock, Hip-hop, Rap, and R&B. Fetching gender and race for individual artists was not possible using the Spotify Web API. This was achieved using existing Natural Language Processing (NLP) packages [8,9].

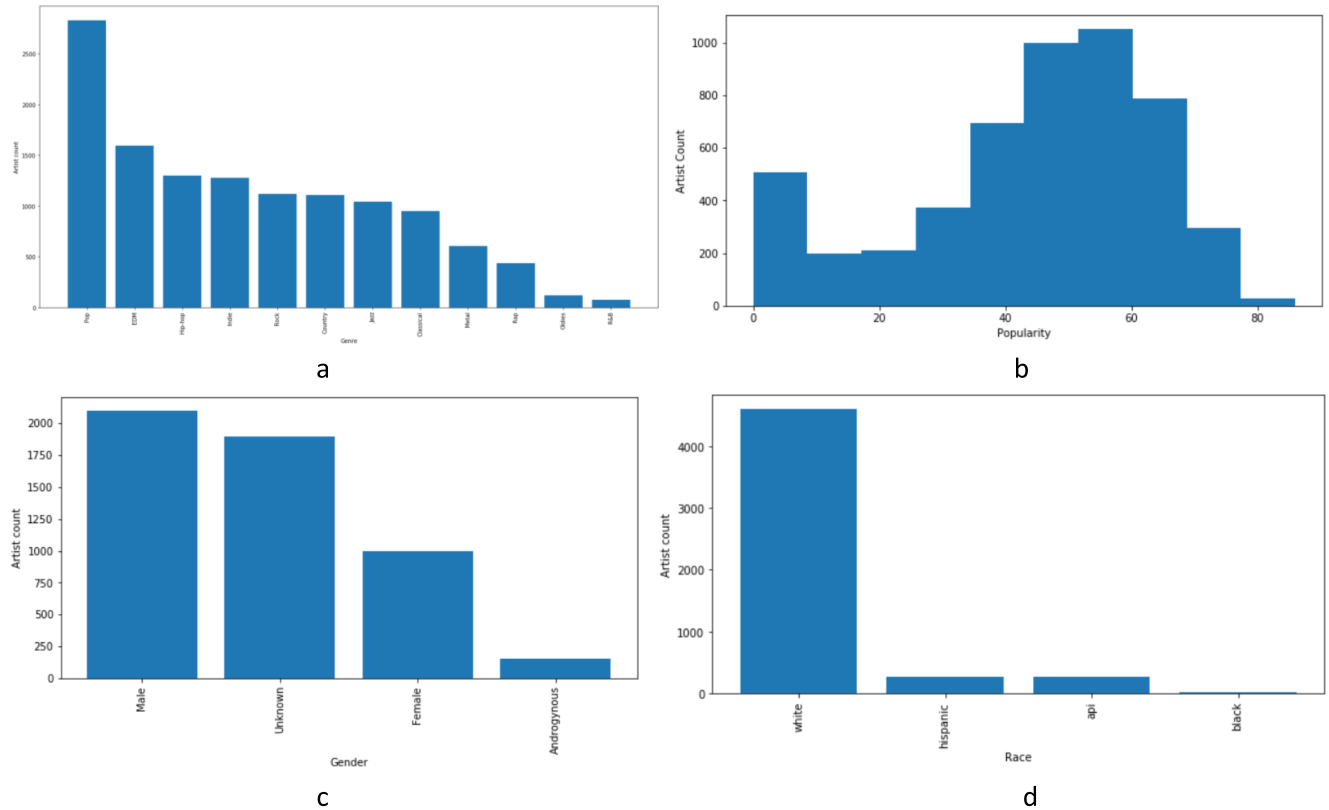


Figure 1: (a) Distrbution of musical genres based on artist count, (b) distribution of normalized populairty in the Spotify collaboration network, (c) distirbution of different genders obtained using NLP methods, and (d) distirbution of different races obtained using different NLP methods

General Questions and Hypotheses:

Collaboration is observed in almost all forms of research and industrial activities. Researchers have explored collaboration patterns to understand underlying social constructs and the social players have manipulated those for better yield. Social attributes have different effects on different forms of collaborations. For example, collaboration between male and female individuals might occur more often in certain types of collaborations (e.g., collaboration in art) compared to others. This pattern provides us a deeper social insight of different fields, like homophily in gender groups or emergence polarization in based on gender. Collaboration between individuals from different cultures provide insight into the commonalities and compatibility of those two cultures. For example, a collaboration between country and rap artists will reach out to people with vast cultural differences. The topic of such collaboration will be at the intersections of both the cultures and will provide deeper insight into the similarities of these cultures. This will also help in understanding other social aspects that are important to both the groups (e.g., movies appealing to both cultures, social movement triggering both groups etc.). This work explores three broad areas of collaborations:

1. Difference between collaborations in various domains (e.g., science and music).
2. Effect of social attributes on collaboration.
3. Changes in collaboration patterns over time.

Keeping the above ideas in mind, specific hypotheses have been formulated and explored in this article:

1. Numerous collaborations do not ensure large popularity in music.

Collaboration patterns differ in academia and music.

2. In academia, productive authors tend to directly coauthor with and closely cite colleagues sharing the same research interests; they do not generally collaborate directly with colleagues having different research topics, but instead directly or indirectly cite them [1].
   Musicians from different genres tend to collaborate more with each other. Inter-genre collaborations tend to be more successful compared to intra-genre music.
3. Highly cited authors do not generally coauthor with each other, but closely cite each other [1].
   In music, popular artists seem to collaborate among themselves more than collaborating with less popular or budding artists.
4. Gender of the artists influence musical collaborations.
5. Race of the artists impacts musical collaboration patterns.
6. Musical collaboration patterns change over time.

Network Analysis and Visualization: The collaboration network includes 5,142 nodes and 12,905 edges. Each node represents an artist. An edge between two artists imply that they have collaborated on one or more than one song that is available on Spotify. Multiple visualizations have been done for this network. Each visualization uses different node attributes for node coloring.

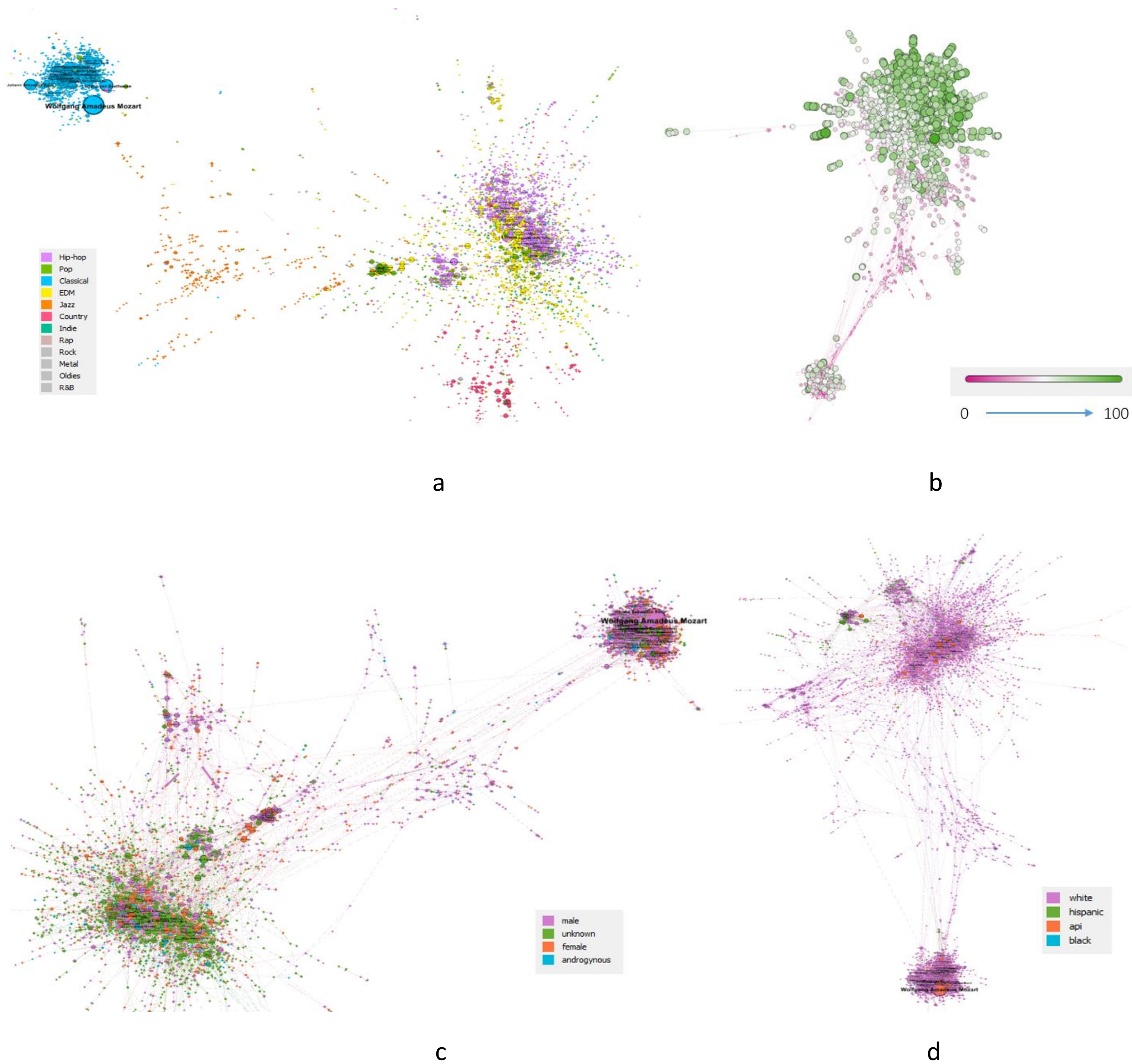
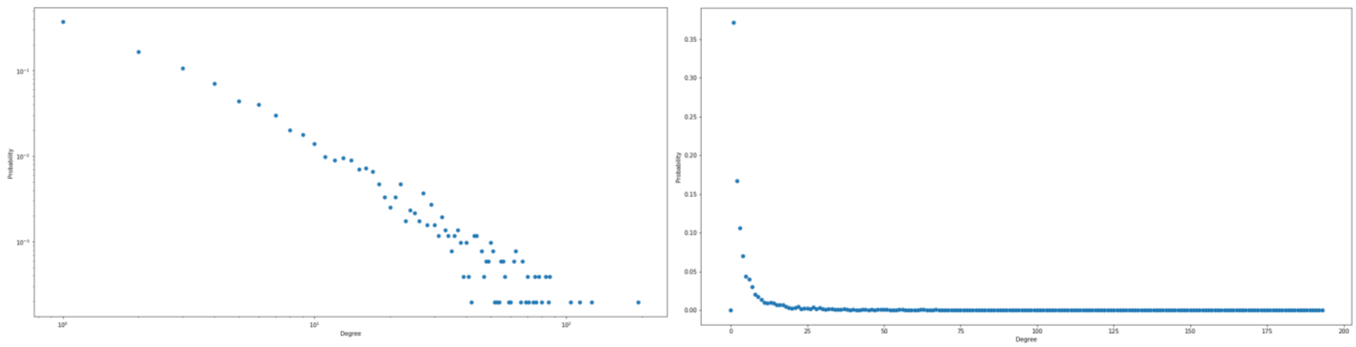


Figure 2: Visualization of Spotify collaboration network using node coloring based on (a) genre, (b) popularity, (c) gender, and (d) race

The collaboration network follows a heavy tailed degree distribution, which follows a linear trend in log-log scale. This implies the existence of hubs in the network, which correspond to the artists who collaborate with other artists often.



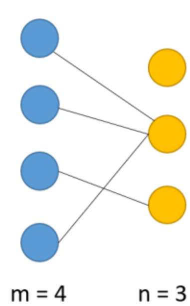
a b

Figure 3: Degree distribution of Spotify collaboration network (a) in log-log scale, (b) linear scale

Common network measures were computed for the network. Average degree of the network is 5.02, which suggests that the artists on an average collaborate with 5 other artists. The network has a diameter of 20, which suggests that two artists are at most 20-hops collaboration apart on Spotify. A modularity value 0.755 suggests that the network contains communities, which we explore in figure (2) based on different node attributes. The network has an average clustering coefficient of 0.425, which suggests transitivity of collaboration in music. If two artists have collaborated with a third artists, it is highly probable that they have collaborated among themselves (maybe all three have collaborated on a song). Average path length of the network is 5.957, which a bit higher than that of an online social network (which is ~3.5). This suggests the small-world property of this collaboration network. Any two artists are 6 hops apart on the collaboration network and 6 is the degrees of separation in this network. Since the network includes artists from various genres, the small-world property suggests that these genres, and broadly the associated cultures are closer to each other than they appear. In the worst-case scenario, two very different cultures are connected via five other cultures. These bridging cultural groups might include elements from its neighboring cultures.

Related Definitions: For studying the collaboration patterns between different groups of artists, a bipartite measure has been proposed. For any two groups of nodes, the maximum possible number of links between them is the product of the size of the vertex sets. The fraction of the possible links present in the network implies a notion of how often artists from the two groups collaborate. This fraction has been defined as 'Collaboration Fraction' and has been used frequently in the forthcoming analysis.

Collaboration Fraction = Existing links between groups / Possible link between the groups



Possible link between the groups = m*n = 12

Existing links between groups = 4

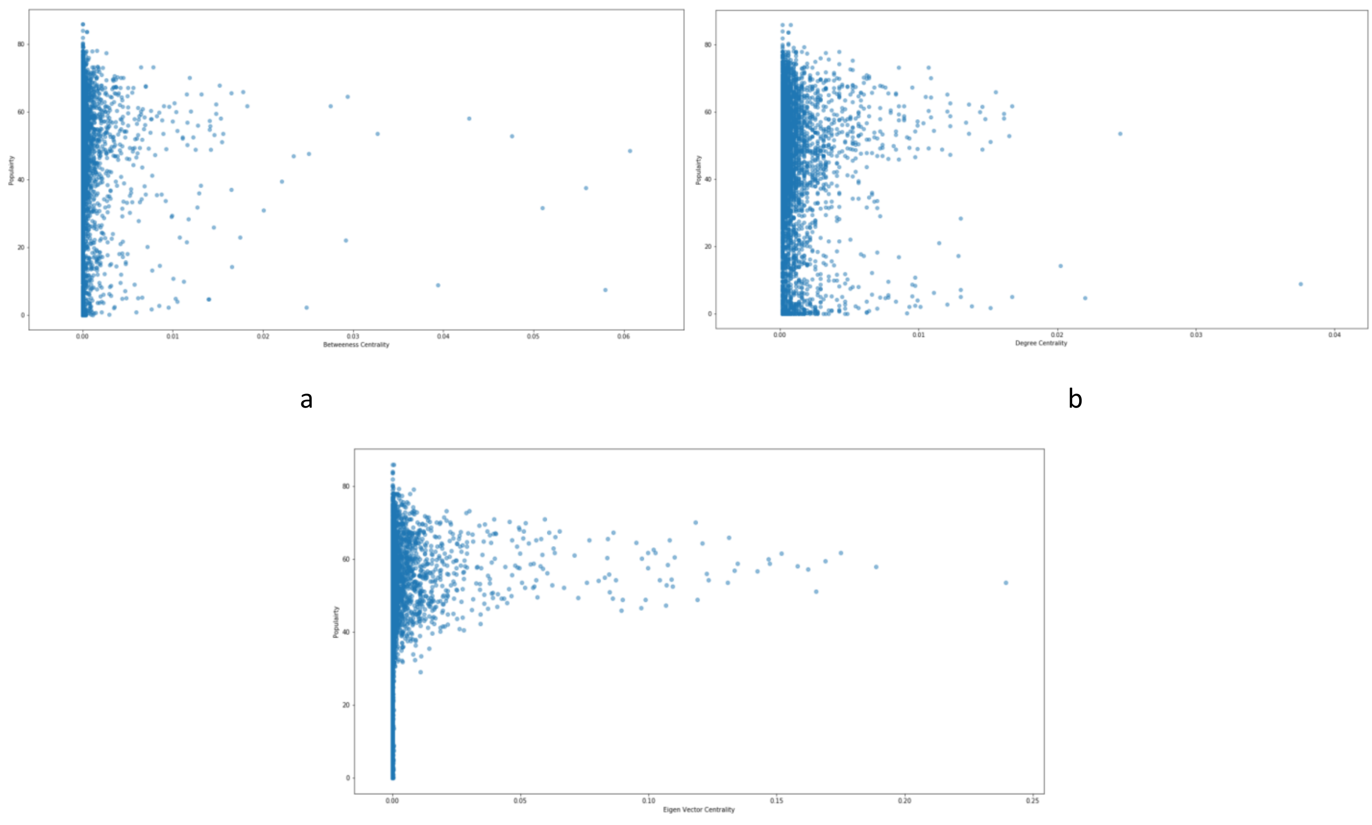
Collaboration Fraction = 4/12 = 0.33

Figure 4: Definition and derivation of *Collaboration Fraction* for two groups of nodes in blue and yellow

This definition has been used to study patterns of collaboration between artists from different genres, genders, and races. Since a small number collaborations ensue compared to all the possible collaborations, the value of Collaboration Fraction is typically low.

Social Analysis: In this section, we draw interpretations from the network analysis and accept or reject the hypotheses mentioned previously.

<u>Hypothesis 1</u> Popularity and centrality of an artist are not correlated, which means that more collaboration does not ensure success in the industry. Similar conclusion can be observed in scientific or other forms of collaboration networks. Popularity of artists have been studied against different types of centralities (degree, betweenness and Eigen Vector) and no significance correlation has been observed (Spearman's correlation coefficient values are *-0.03* in all the three cases). This concludes that popularity is a more important attribute in collaboration scenario than existing number of collaborations everyone has (i.e., the degree in the collaboration network). Though more collaboration helps in expanding social connections, large number of connections do not always provide advantage in different scenarios. This idea maps to the concept of structural holes, where making many connections with low or moderately popular artists has low impact on achieving success.



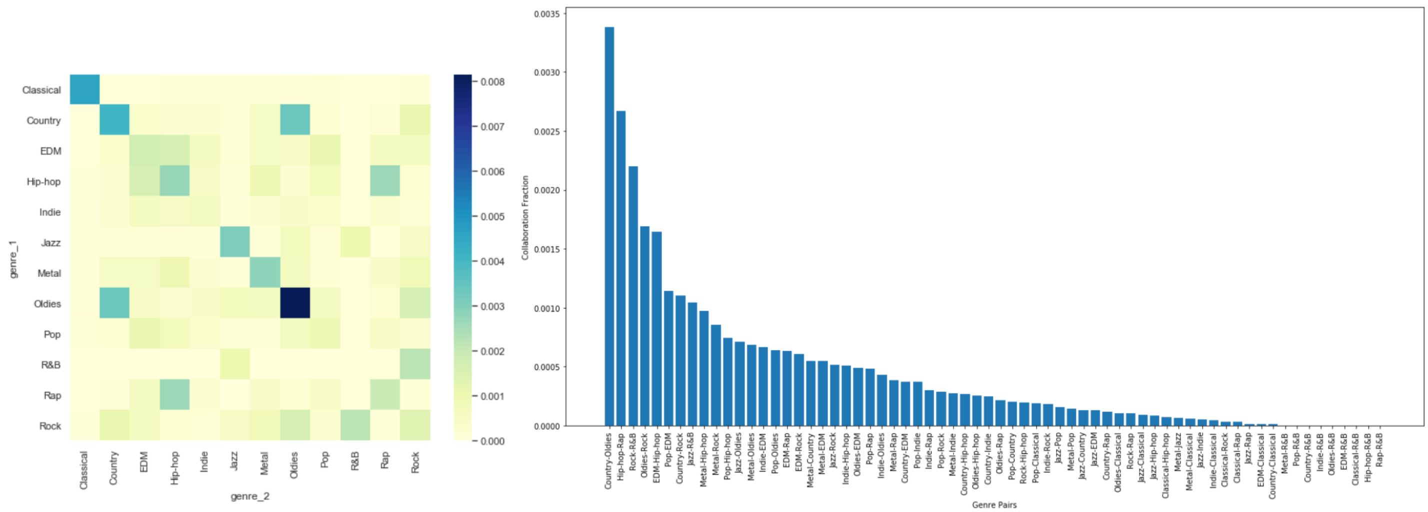
a b

c

Figure 5: Plot of popularity of the artists against (a) degree, (b) betweenness, and (c) Eigen Vector centralities

<u>Hypothesis 2</u> Different aspects of musical collaboration are compared against well-studied scientific collaboration. "In academia, productive authors tend to directly coauthor with and closely cite colleagues sharing the same research interests; they do not generally collaborate directly with colleagues having different research topics, but instead directly or indirectly cite them" [1]. On the other hand, musicians from different genres tend to collaborate more with each other.

Genres using similar style or instruments (e.g., Country and Oldies, Hip-hop and Rap) show more collaboration, since musicians with similar musical knowledge, background and taste contribute to these inter-genre domains. Homophily is observed in all forms of social networks and musical collaboration is not an exception to that. These inter-genre collaborations also reflect the cultural and social intersection between different groups of people. Higher values along the diagonal in figure 6(a) shows that intra-genre collaboration is more common compared to collaboration between different genres.



a b

Figure 6: (a) Heatmap of collaboration fraction between different genres, (b) Inter-genre collaboration

In social networks, weak ties refer to two different social groups having a connection. More of these ties are observed in musical collaboration compared to scientific collaboration. This suggests that despite the existence of homophily, musical collaboration incorporates more heterophily compared to other forms of social interactions. It can be expected that other domains of art (e.g., movies, drama) should show similar nature.

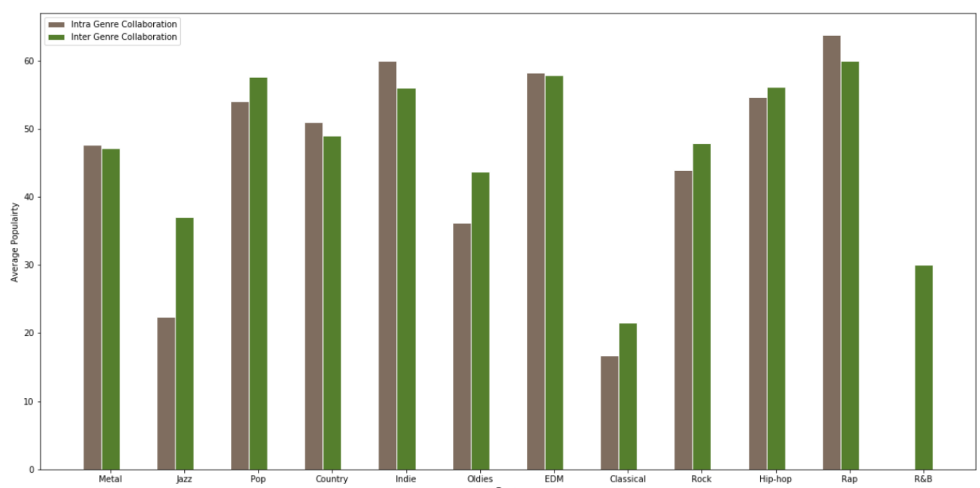
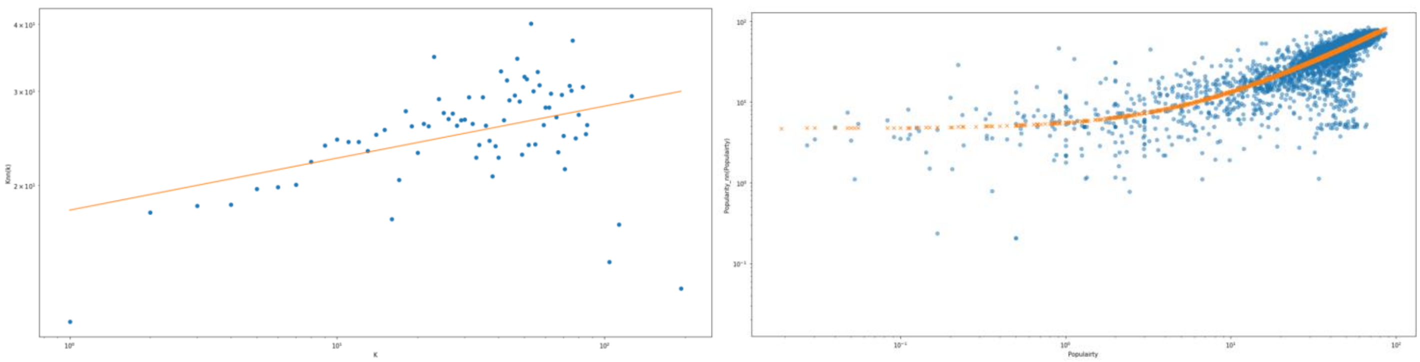


Figure 7: Popularity for intra and inter-genre collaborations

Inter-genre collaborations tend to be more successful compared to intra-genre music for certain genres like classical and pop. Musicians from these genres gain more popularity by producing music together with artists from other genres. Whereas opposite trend is observed for genres like country and indie. These results suggest that success under collaborative and non-collaborative scenarios largely depend on the domain under consideration. Research in interdisciplinary domains like network science demand researchers with different backgrounds to collaborate for better outcomes. Elements from different groups might not always add up to be beneficial. Inter-genre acceptability depends on individual groups and their social open-mindedness.

Hypothesis 3 Another important aspect of scientific collaboration is that highly cited authors generally do not coauthor with each other, but closely cite each other [1]. For musical collaboration, popular artists seem to collaborate among themselves more than collaborating with less popular or budding artists.

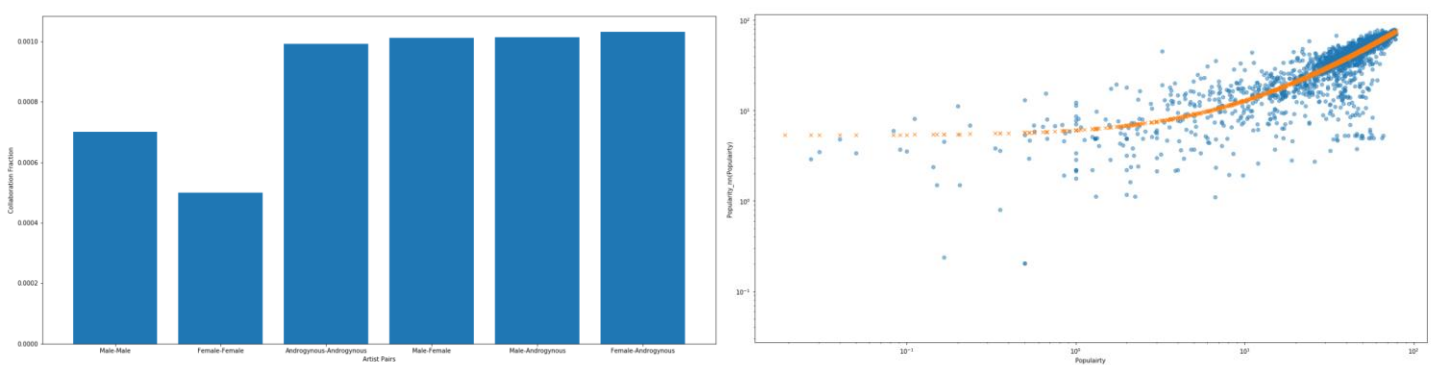


a b

Figure 8: (a) Degree, and (b) popularity assortativity in Spotify collaboration network

Figure 8(a) shows the presence of degree assortativity in musical collaboration network. A linear fit provides the degree correlation exponent of *0.09*. This suggests that the artists who collaborate often, have higher chances of collaborating with each other – which is an intuitive outcome. On the other hand, figure 8(b) shows the popularity correlation in the network. This indicates a stronger assortativity in popularity compared to degree assortativity. Popularity assortativity shows a quadratic trend in log-log scale. This suggests that in certain collaboration scenarios, node attributes contribute more in homophily (or even heterophily). This analysis also provides an insight into the motivation behind collaboration. Depending on the domain, the stimulus for collaboration might change. In general, achieving more success is the prime incentive behind any form of collaboration.

Hypothesis 4 Collaboration patterns among different genders have been studied in this section. All intra and inter-gender collaborations are significant, but more collaboration is observed between dissimilar genders (e.g., male-female, male-androgynous). Collaboration among androgynous artists is higher than male-male or female-female collaborations.



a b

Figure 9: (a) Collaboration fraction between different gender combinations (b) Popularity assortivity among the female artists

In both intra and inter-gender collaborations, high populairty assortativity is observed. This suggests that irrespective of gender, people tend to collaborate with more popular individuals for quick acquirement of success.

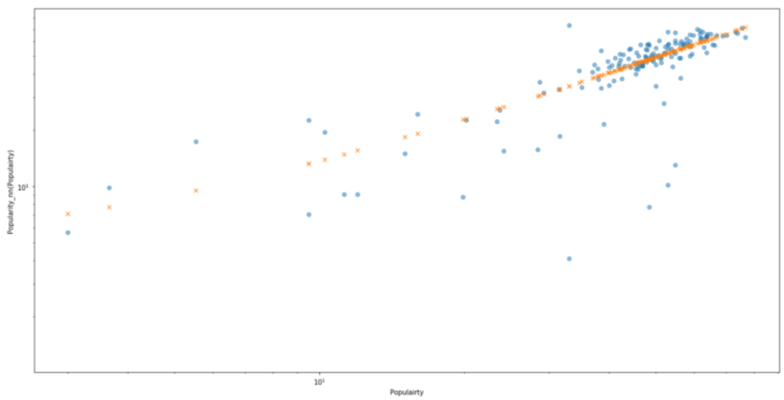
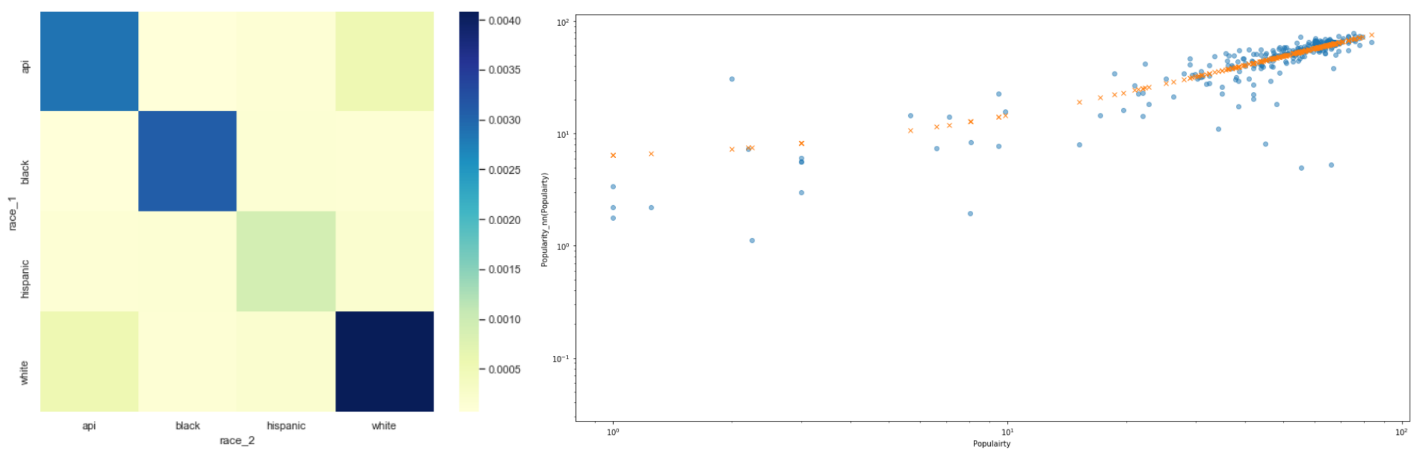


Figure 10: Populairty assortativity among androgynous artists

Interestingly, less popularity correlation is observed among the androgynous artists. This suggests that depending on the gender, the chances of working with less successful individuals might change in certain collaboration scenarios.

<u>Hypothesis 5</u> Collaboration patterns depend on the race of the artists. More intra-race collaborations are observed compared to inter-race collaborations. Black artists seem to have less collaborations with other genres, but this outcome might be highly influenced by the limitations of the NLP algorithm used to determine the genders.

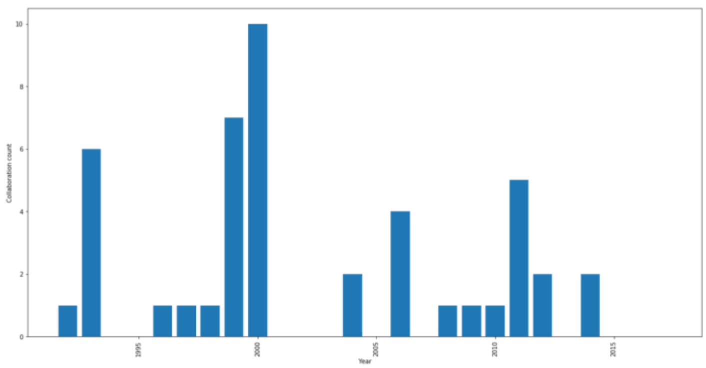


a b

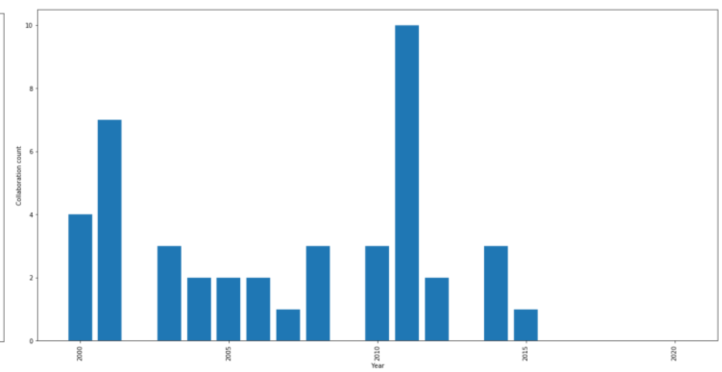
Figure 11: (a) Collaboration fractions across different races, (b) Popularity assortativity between Asian and white artists

Interestingly, less popularity assortativity is observed in inter-race collaboration. This suggests that popular artists of a certain race are more open to collaborate with budding artists of another race. In certain social scenarios, this might be beneficial for certain races in gaining a wider social exposure.

<u>Hypothesis 6</u> Collaboration patterns change over time. Artists seem to be more collaborative during the early stages of their careers. Collaboration increases their chances of succeeding and sharing followers and is an efficient approach of mutually expanding the fan bases. Even for well established artists, most of the collaborations took place during their early or mid-careers. As popularity starts saturating, individuals tend to collaborate less as they already have accomplished a strong floor to release new product on.



a



b

Figure 12: Change of collaboration counts for (a) Snoop Dogg, and (b) Ludacris

Similar patterns might be observed in other collaborative scenarios, as maturity and success in a domain provides independence.

Conclusion: The article has given an insight into the major differences between scientific collaboration and musical collaboration. Although collaboration is observed in various domains of social activities, the fundamental patterns might be different for different domains. These fundamental patterns include individual attributes contributing to the likeliness of others to collaborate with that individual (e.g., popularity in music) and the cultural characteristics associated with different groups (e.g., musical genres). Again, for each domain, large variability of collaboration is seen across social attributes such as gender and race. Cultural and behavioral variabilities are the main contributors for this social dissimilitude. As a conclusion, although the collaboration networks have many similarities across different domains, social constraints affect collaboration to large extents. Thus, drawing conclusions and making predictions based on collaboration network analysis (e.g., predicting success in music) should be done scrupulously by taking all the related social attributes into consideration.

Future Work: Due to lack of time and data, some aspects of musical collaboration network could not be explored in this work. It would be interesting to study the effect of collaboration on musical creativity, productivity, and innovation. Furthermore, common analysis methods might cause collaboration between scientific researchers from different fields. It is intuitive that cultural background influences collaboration between artists from different genres. Understanding the related cultural attributes (e.g., use of similar musical instruments might initiate collaboration) will provide more insight into the mechanism behind collaboration in human society.

References:

[1] Ying Ding, Scientific collaboration, and endorsement: Network analysis of coauthorship and citation networks.

[2] Easley, D., & Kleinberg, J. (2010). Chapter 1: Overview. In Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press., pp. 1–17.

[3] Lee and Bozeman, 2005, The impact of research collaboration on scientific productivity, Social Studies of Science, 35 (2005), pp. 673-702.

[4] Interiano M, Kazemi K, Wang L, Yang J, Yu Z, Komarova NL. 2018 Musical trends and predictability of success in contemporary songs in and out of the top charts.R. Soc. open sci.5: 171274.

[5] Janosov, M., Battiston, F. & Sinatra, R. Success and luck in creative careers. EPJ Data Sci. 9, 9 (2020). https://doi.org/10.1140/epjds/s13688-020-00227-w.

[6] Silvia Donker, Networking data. A network analysis of Spotify's sociotechnical related artist network (2019), Semantic Scholar.

[7] Tobin South, Network Analysis of the Spotify Artist Collaboration Graph (2018), Australian Mathematical Institute.

[8] Predicting gender using NLP: https://github.com/lead-ratings/gender-guesser

[9] Predicting gender using NLP: https://github.com/appeler/ethnicolr

Datasets, Python code and analysis of this project is available here:
https://github.com/ChatterjeeAyan/Spotify_collaboration_network