PES UNIVERSITY

# UE20CS344 - Network Analysis and Mining Course Project

**Project Title** : Spotify Artist Feature Collaboration Network Analysis

Team members : PES2UG20CS479 V NIMISH BHASU- Section H
PES2UG20CS448 POOJA H - Section G
PES2UG20CS453 RIA TREZA SERRAO -Section G
PES2UG20CS399 YOGITHA C SHEKHAR - Section G

# Topic

Exploring Collaborative Relationships between Artists on Spotify-
An Analysis and Visualization of Feature Collaborations through Network Representation.

FUNCTIONALITIES:

1. Graph generation and centrality measure visualization
2. Artists' gender prediction and assortativity
3. Popularity analysis wrt genre
4. Temporal analysis

## Uniqueness

- Collaboration networks have been extensively examined in a variety of social contexts and these networks offer a thorough understanding of how people work together to accomplish specific goals.
- In the past, science has not done interdisciplinary collaboration very often.
- Researchers from related fields of science and the humanities frequently collaborate to advance those fields.
- In the case of music production, things are different.
- Collaborations between musicians from various genres are frequent, and these partnerships often result in hit songs.
- Social factors like gender have a big impact on musical partnerships.
- Additionally, collaboration styles evolve over time.

# Dataset

## Collaboration links with genre

```
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   source        14860 non-null  object
 1   target        14860 non-null  object
 2   release_date  14860 non-null  int64
 3   populairty    14860 non-null  int64
```

# Dataset

## Collaboration nodes with genre

```
 8   Id                  5142 non-null   object
 9   acousticness        5142 non-null   float64
 10  danceability        5142 non-null   float64
 11  duration_ms         5142 non-null   float64
 12  energy              5142 non-null   float64
 13  instrumentalness    5142 non-null   float64
 14  liveness            5142 non-null   float64
 15  loudness            5142 non-null   float64
 16  speechiness         5142 non-null   float64
 17  tempo               5142 non-null   float64
 18  valence             5142 non-null   float64
 19  key                 5142 non-null   int64
...
 29  white               5142 non-null   float64
 30  populairty          5142 non-null   float64
```

PES UNIVERSITY

# Dataset

## Artists with genre

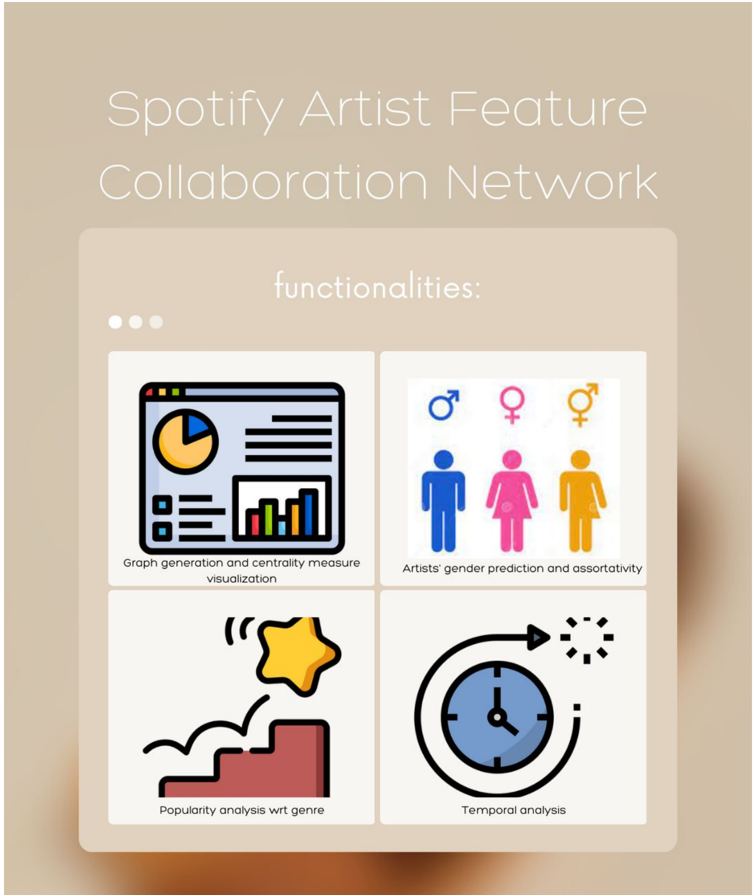
| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| --- | ------ | -------------- | ----- |
| 0 | artists | 12454 non-null | object |
| 1 | acousticness | 12454 non-null | float64 |
| 2 | danceability | 12454 non-null | float64 |
| 3 | duration_ms | 12454 non-null | float64 |
| 4 | energy | 12454 non-null | float64 |
| 5 | instrumentalness | 12454 non-null | float64 |
| 6 | liveness | 12454 non-null | float64 |
| 7 | loudness | 12454 non-null | float64 |
| 8 | speechiness | 12454 non-null | float64 |
| 9 | tempo | 12454 non-null | float64 |
| 10 | valence | 12454 non-null | float64 |
| 11 | popularity | 12454 non-null | float64 |
| 12 | key | 12454 non-null | int64 |
| 13 | mode | 12454 non-null | int64 |
| 14 | count | 12454 non-null | int64 |
| 15 | genres | 12454 non-null | object |
| 16 | filtered_genre | 12454 non-null | object |

# Dataset

## Data with genre

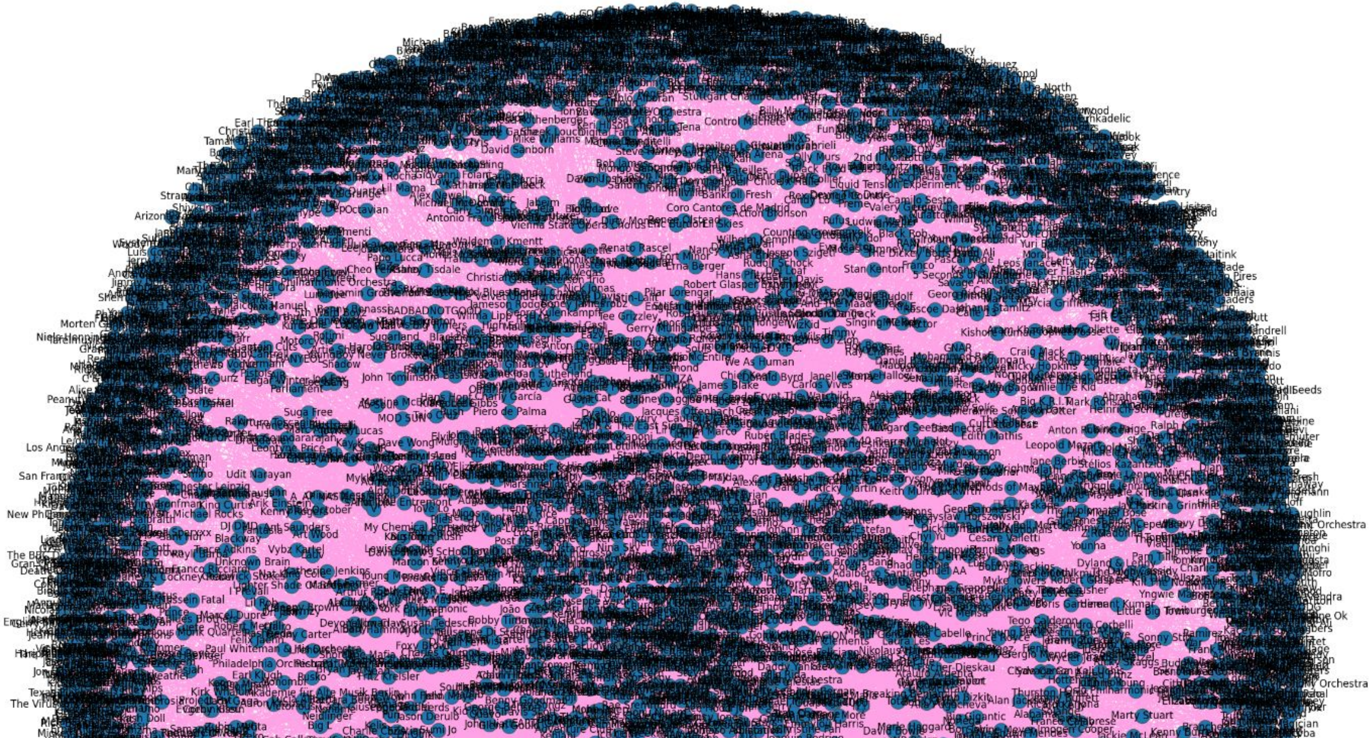
```
1    acousticness       27621 non-null  float64
2    danceability       27621 non-null  float64
3    duration_ms        27621 non-null  float64
4    energy             27621 non-null  float64
5    instrumentalness   27621 non-null  float64
6    liveness           27621 non-null  float64
7    loudness           27621 non-null  float64
8    speechiness        27621 non-null  float64
9    tempo              27621 non-null  float64
10   valence            27621 non-null  float64
11   popularity         27621 non-null  float64
12   key                27621 non-null  int64
13   mode               27621 non-null  int64
14   count              27621 non-null  int64
15   genres             27621 non-null  object
```

# Design diagram



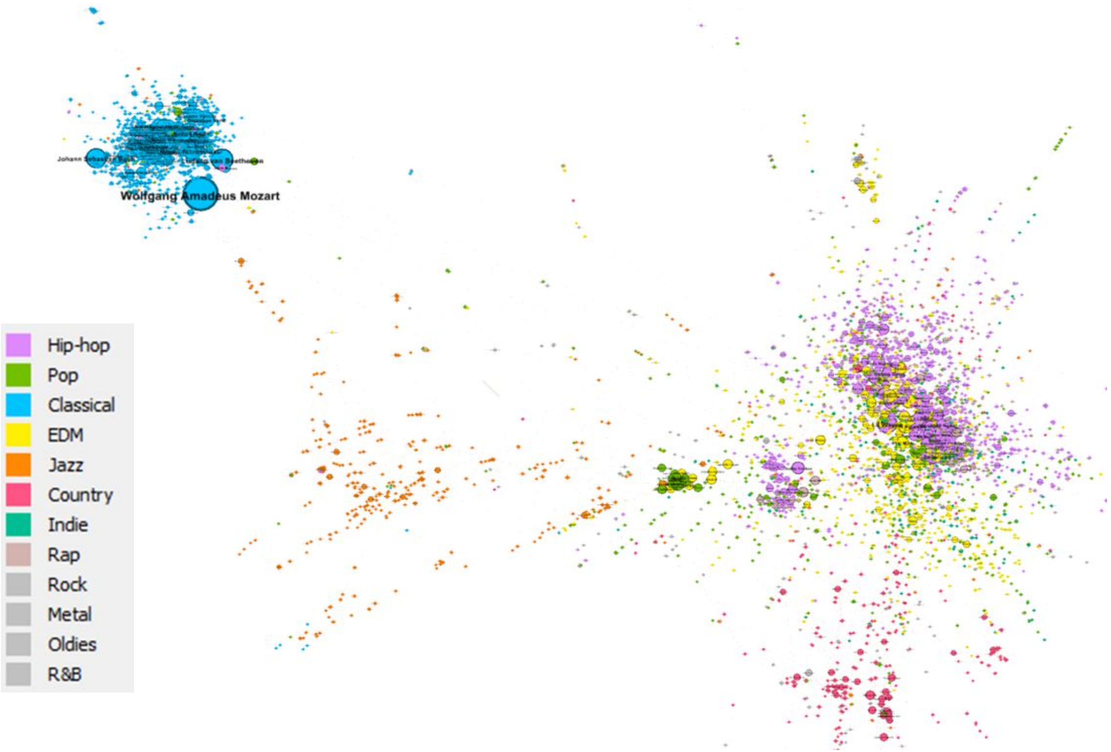
# Graph generation

Graph generation: A directed graph representing a collaboration network and sets node attributes based on given data. It also computes node positions for potential visualization of the graph.



# Graph generation

## GENRE

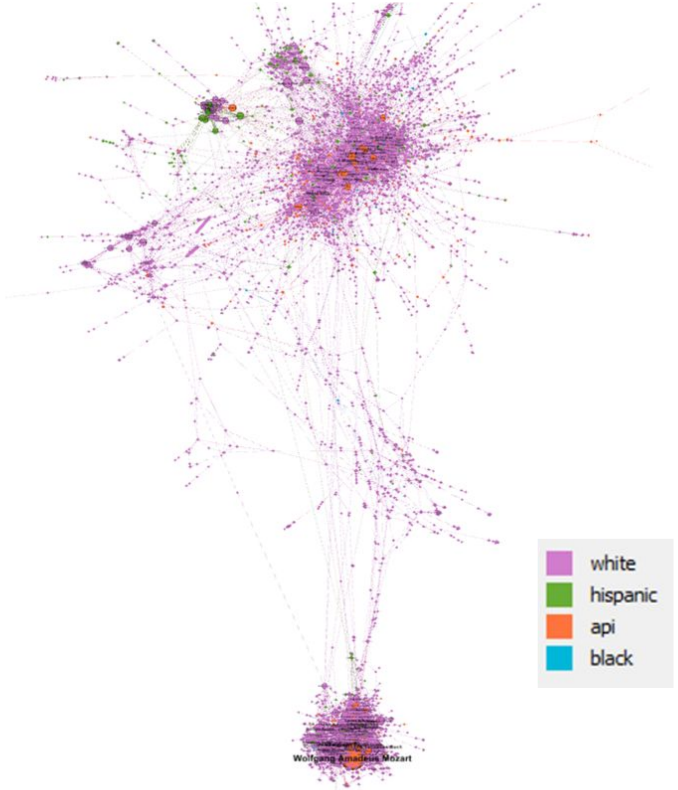



## POPULARITY

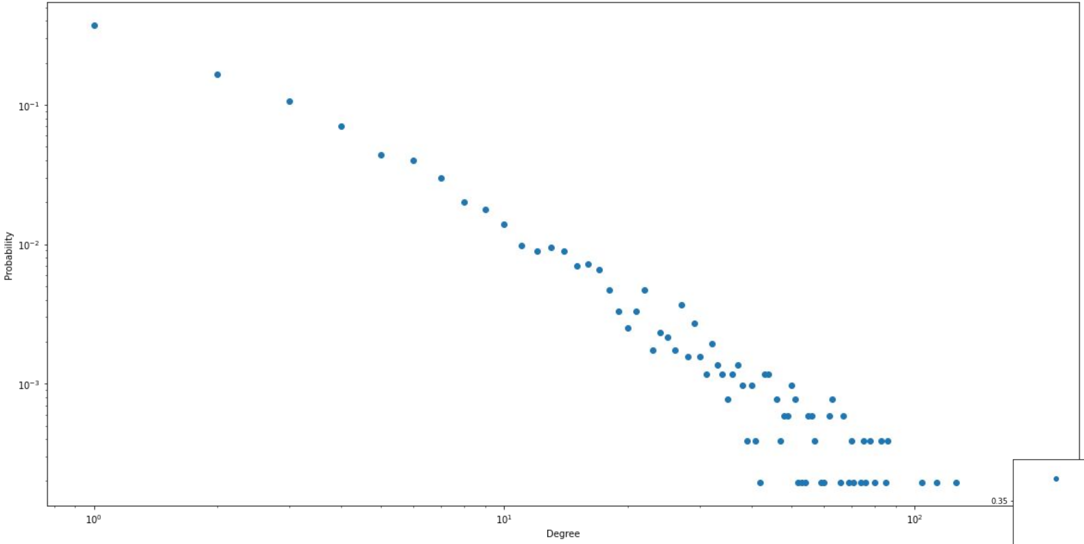
# Graph generation

## GENDER


## RACE

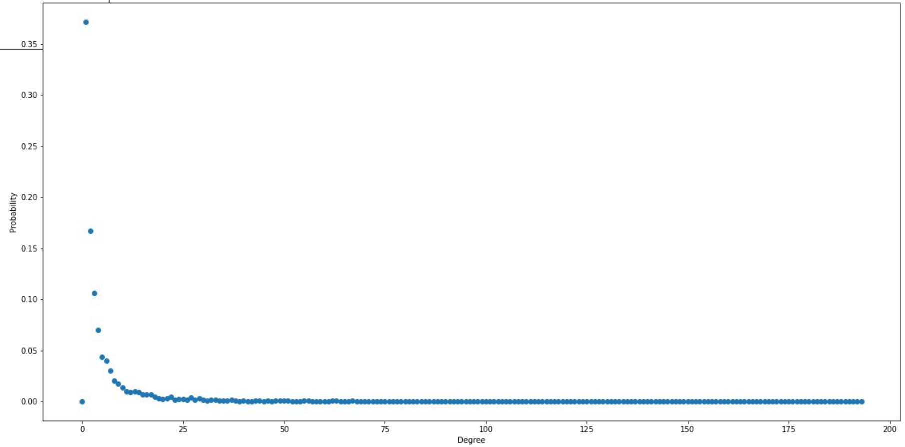


# Centrality measure visualization



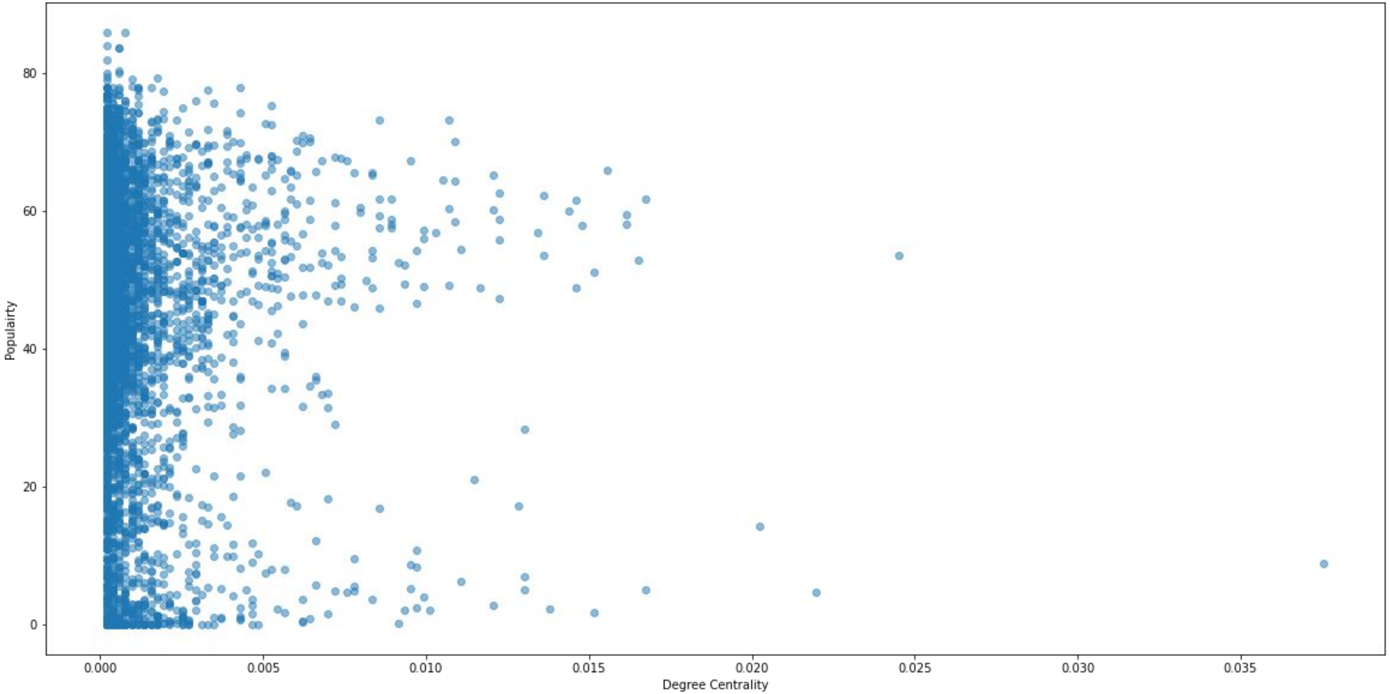

Logarithmic plot showing the relationship between the degrees of nodes in a graph and their corresponding normalized frequency

Showing the relationship between the degrees of nodes in a graph and their corresponding probability



## Centrality measure visualization

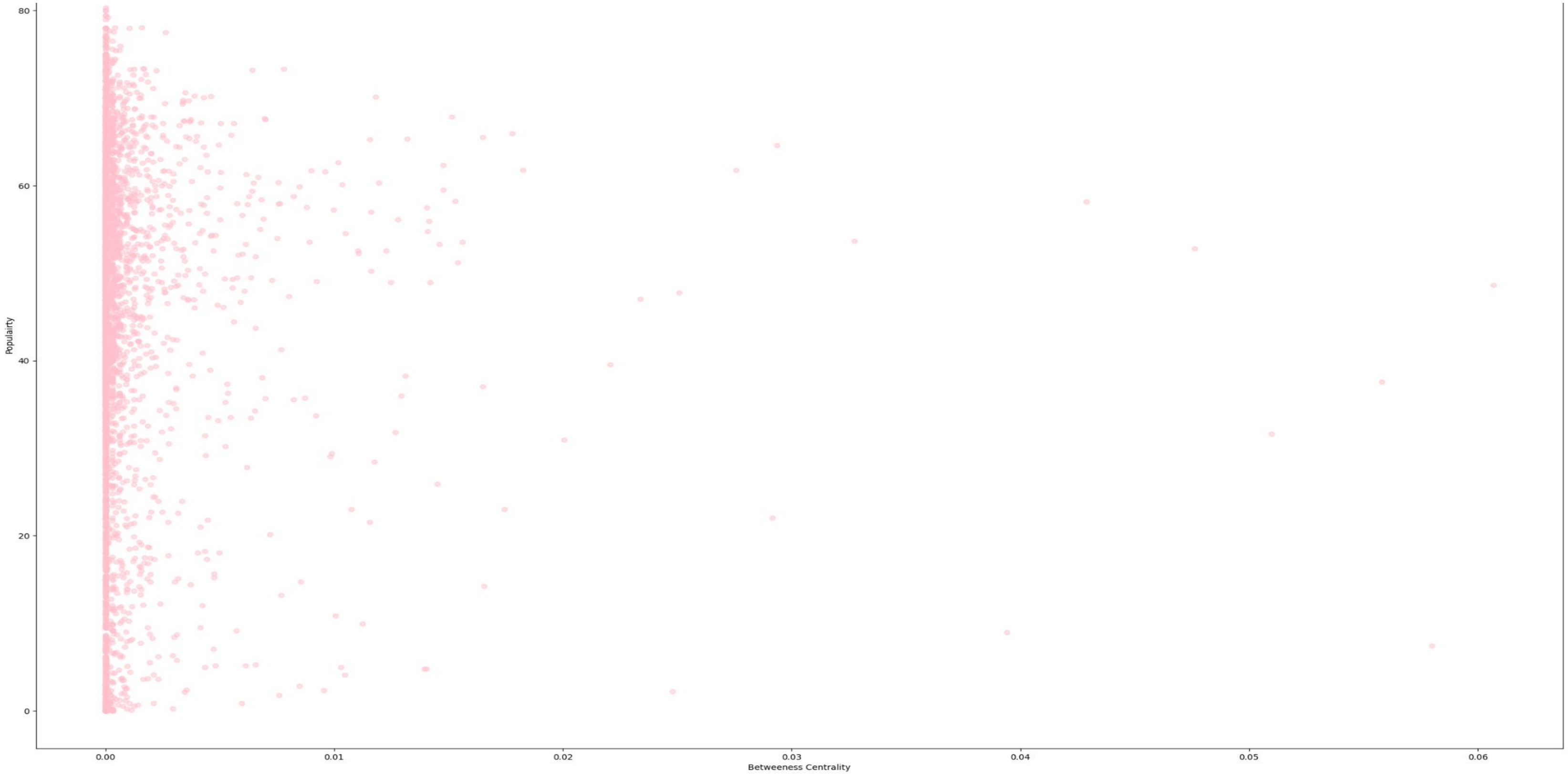
### Degree Centrality




Spearmanr Result(correlation=-0.031566179716198885, pvalue=0.023601613249444233)

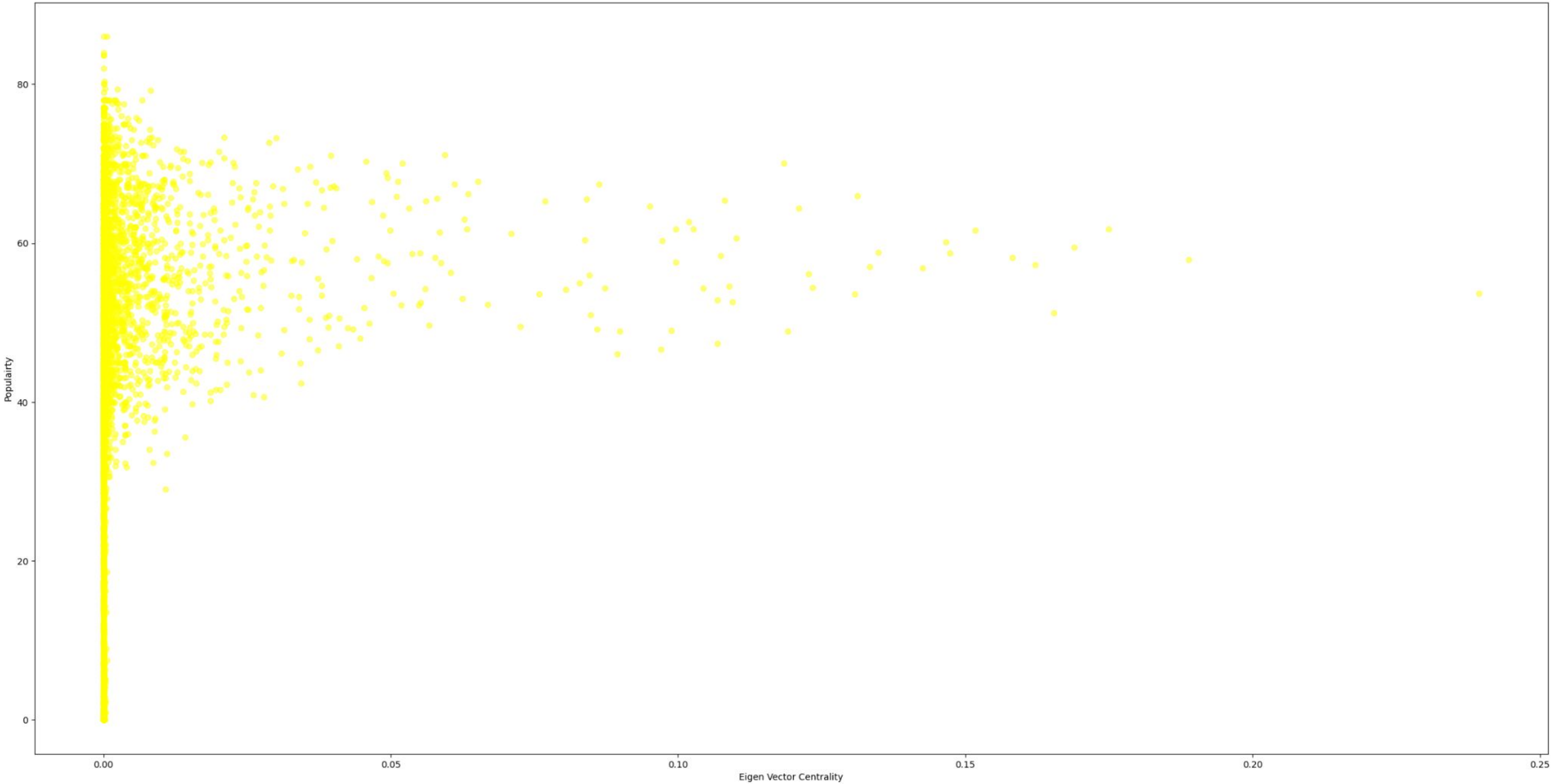
Centrality measure visualization

**Betweeness Centrality**

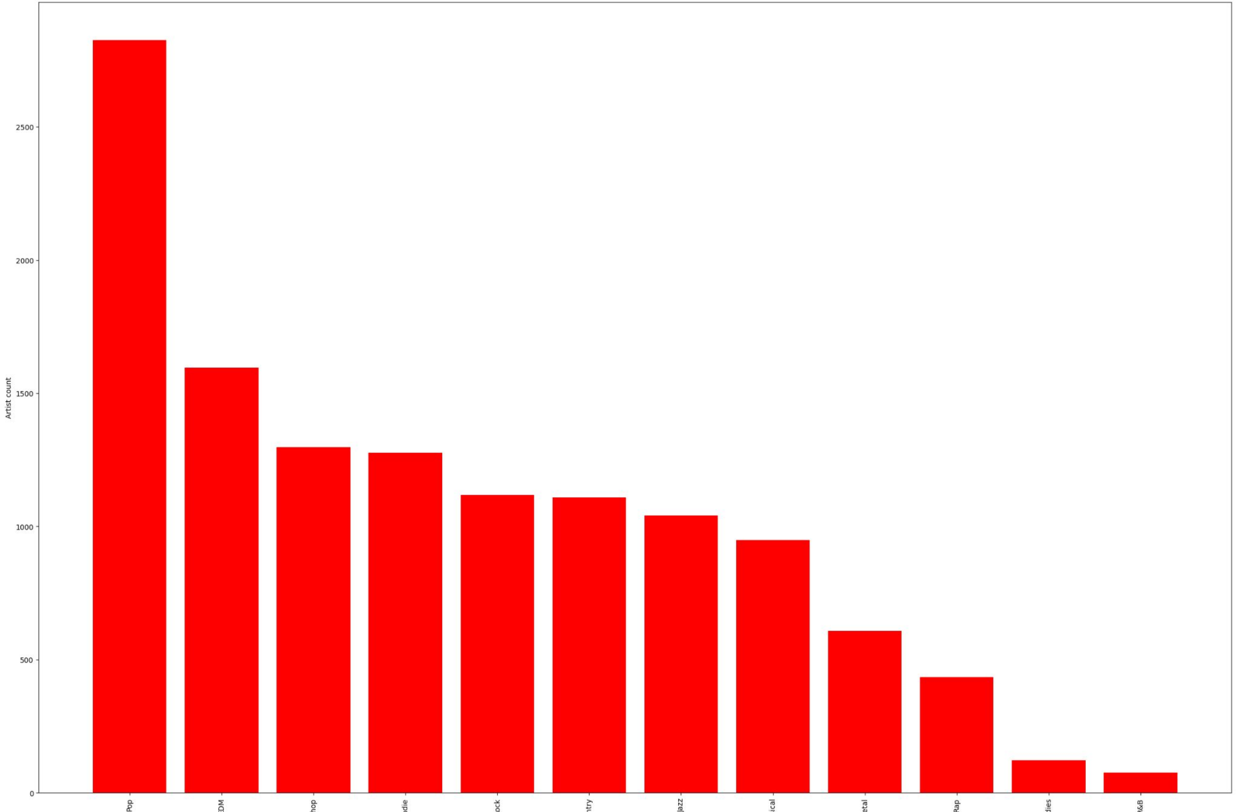


# Centrality measure visualization

## Eigen Vector Centrality



**Bar plot for Artist count vs Genre:** Many artists tend to make music of Pop genre and fewer artists make RnB music

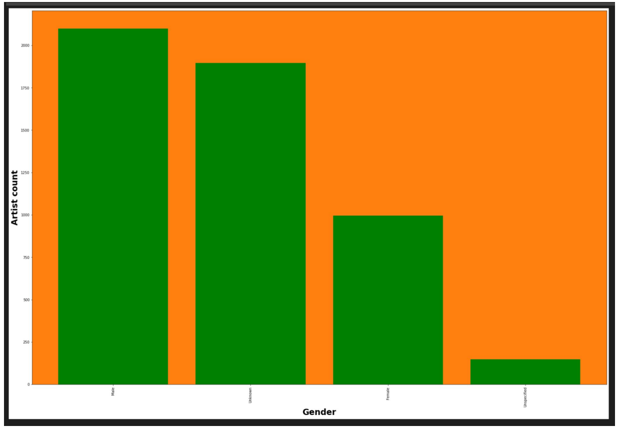


# Artists' gender prediction and Assortativity

1. Guessing the gender based on first name using gender_guesser Python library.

```
male artist counts:  2100
female artist counts:  996
andy artist counts:  149
unknown artist counts:  1897
```

2. Bar graph showing the number of artists for each gender label.

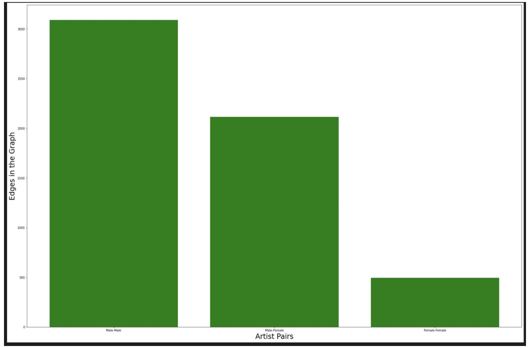


# Artists' gender prediction and assortativity

3. Obtaining the edges that connect the two groups and returning the length of the resulting list of edges.

```
Number of edges between male artists:  3089
Number of edges between male and female artists:  2114
Number of edges between female artists:  496
```

4. Bar chart to visualize the number of edges in the graph that connect different groups of artists based on their gender label.

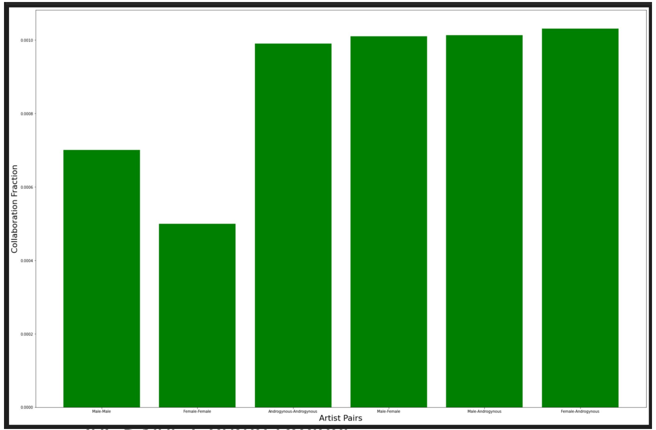


# Artists' gender prediction and assortativity

5. Calculating the fraction of edges in the graph.

```
Fraction of edges between male artists:  0.0007004535147392291
Fraction of edges between male and female artists:  0.0010107095046854083
Fraction of edges between female artists:  0.0004999919356139417
Fraction of edges between androgynous artists:  0.0009909463537678483
Fraction of edges between male and androgynous artists:  0.0010131032278683286
Fraction of edges between female and androgynous artists:  0.0010309695156464785
```

6. Bar graph to visualize the fraction of collaborations between different pairs of artists' gender groups.



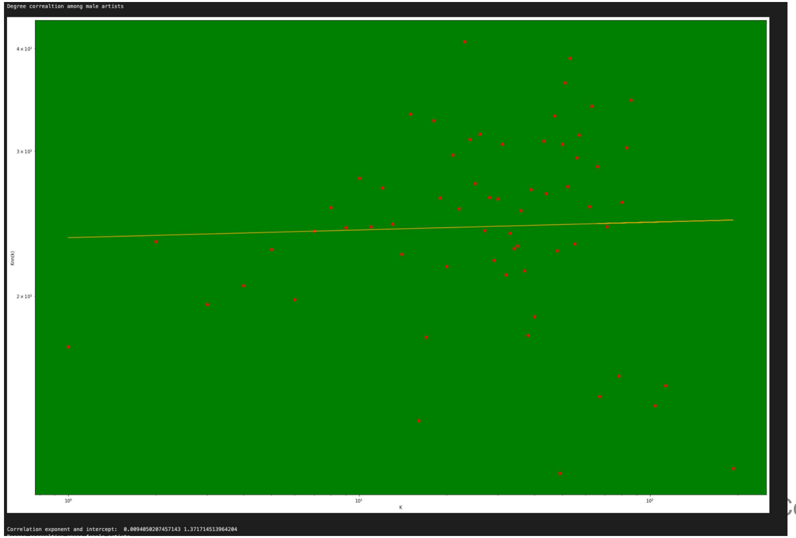
# Artists' gender prediction and assortativity

## 7. Calculating the assortativity coefficient between two groups of artists in a collaboration network.

Degree correaltion among male artists

Correlation exponent and intercept:

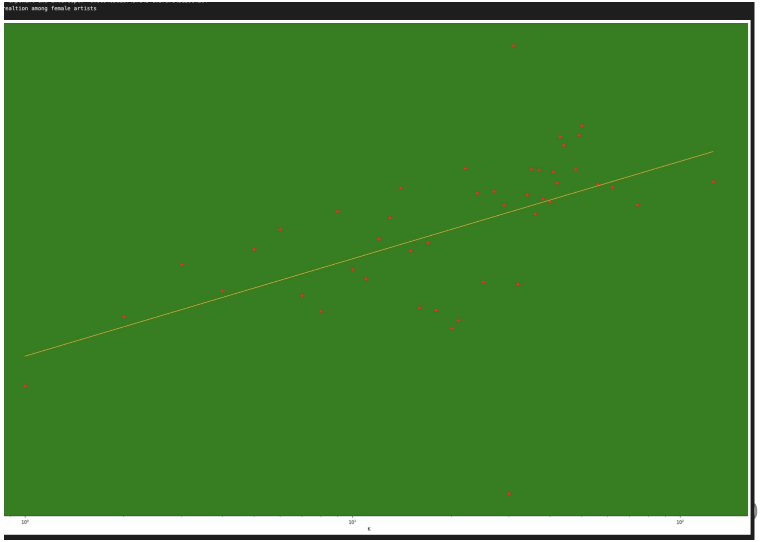
0.0094050207457143, 1.371714513964204




Degree correaltion among female artists

Correlation exponent and intercept:

0.3084186941703006, 0.9112585201787188

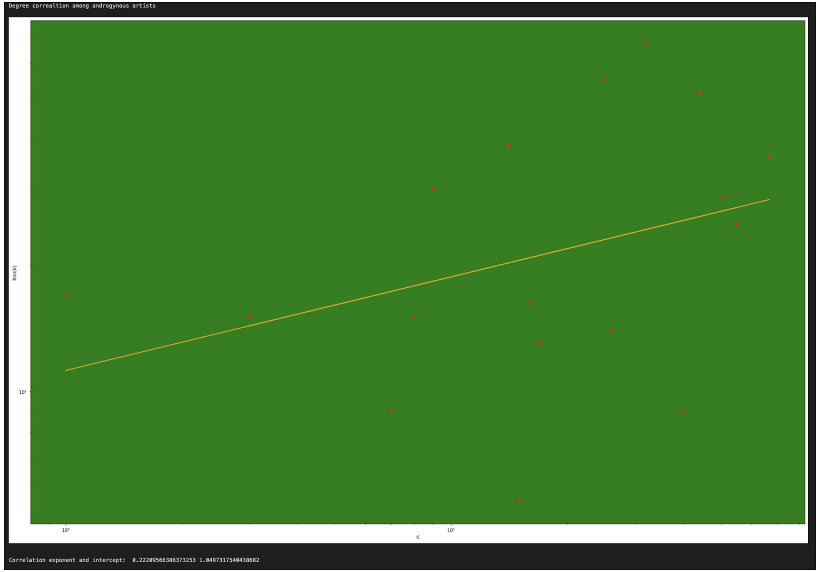


# Artists' gender prediction and assortativity

Degree correaltion among androgynous artists

Correlation exponent and intercept:

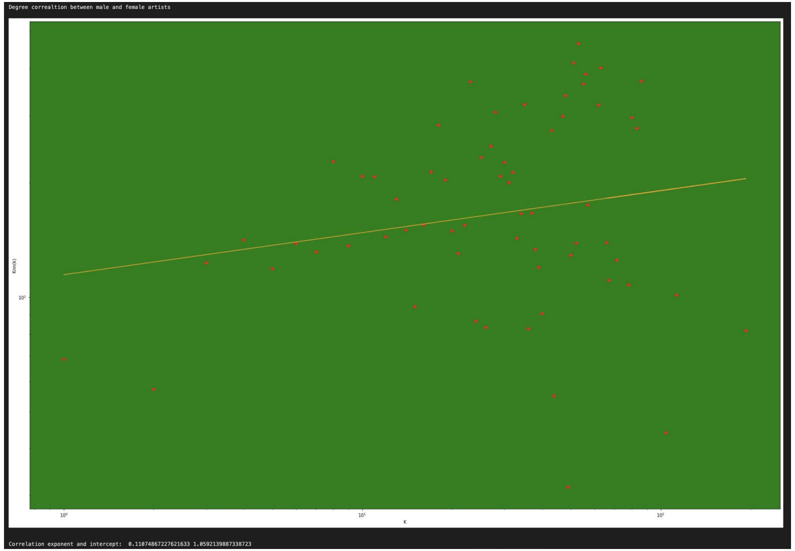
0.22209566386373253, 1.0497317540438682




Degree correaltion between male and female artists

Correlation exponent and intercept:

0.11074867227621633, 1.0592139887338723



# Artists' gender prediction and assortativity

Degree correaltion between male and androgynous artists

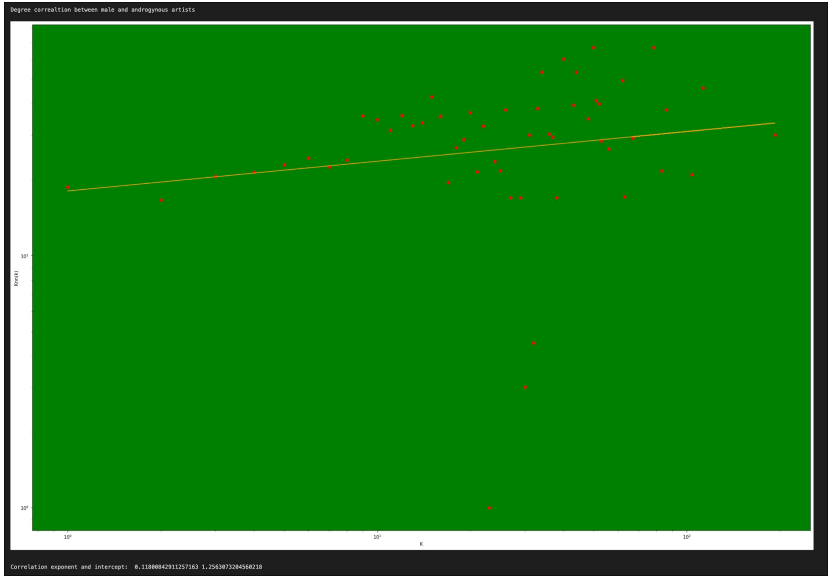
Correlation exponent and intercept:

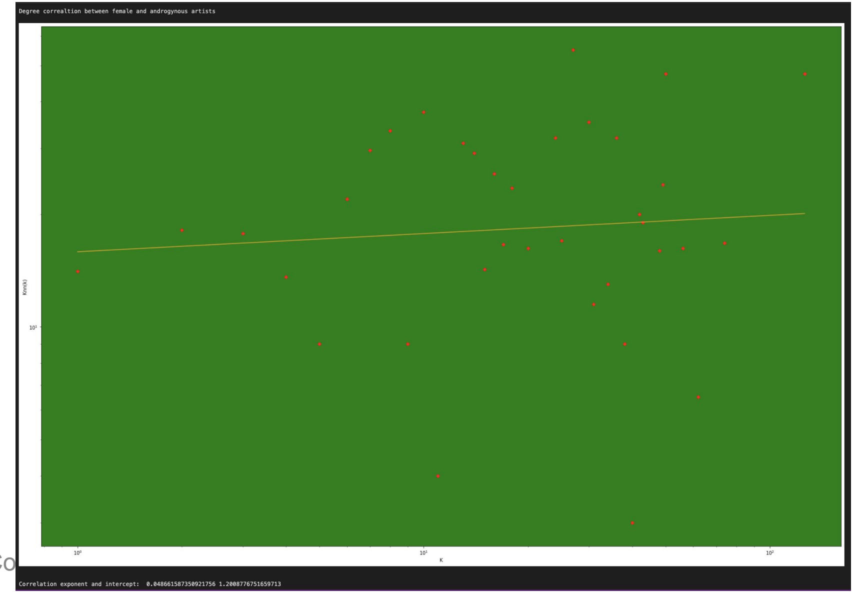
0.11800842911257163, 1.2563073204560218

Degree correaltion among female and androgynous artists

Correlation exponent and intercept:

0.048661587350921756, 1.2008776751659713



## Popularity analysis wrt genre

- Steps:

- The CSV file that contains information about collaborations between artists, including their respective genres and the individual artist is loaded
- Dictionary of artist and genre along with artist link dataframe is formed

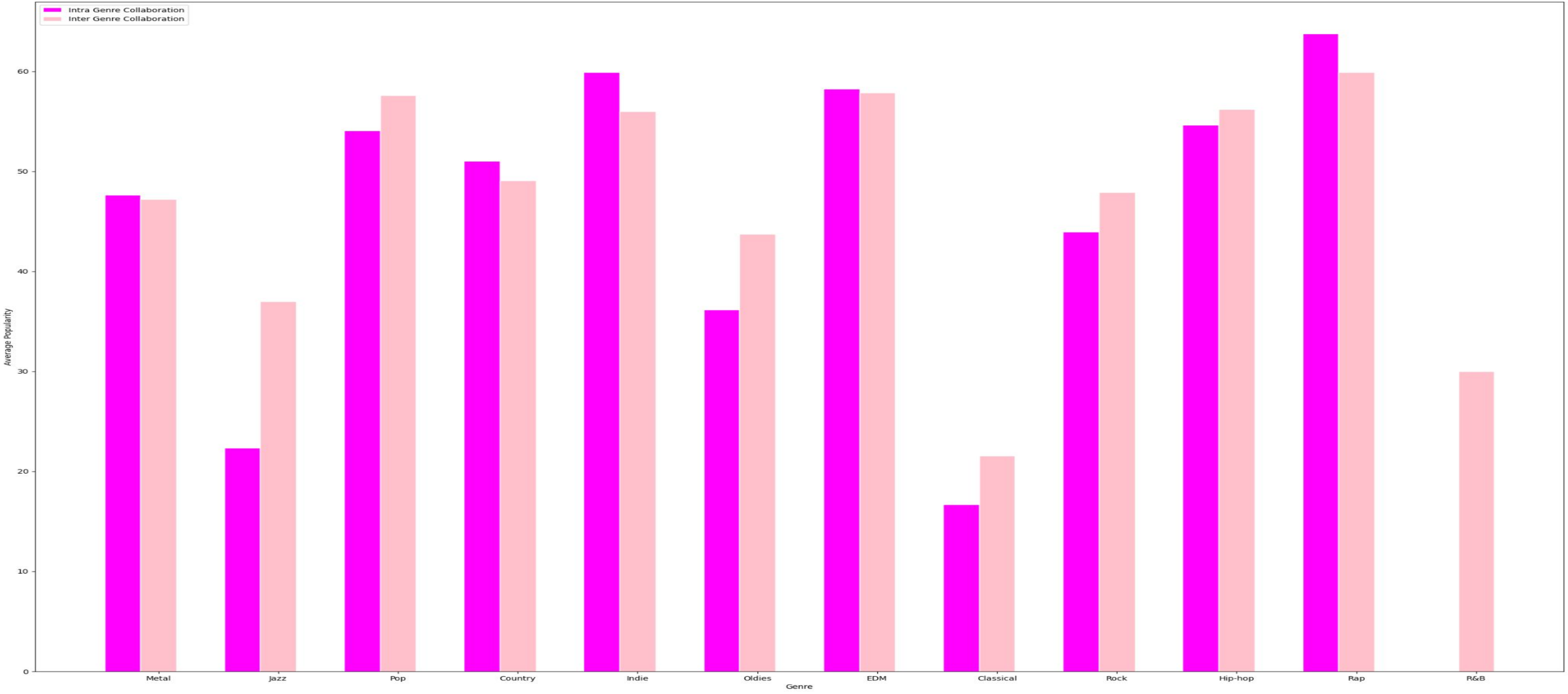
```
'Alfonso X El Sabio': 'Classical',
'Alfred Brendel': 'Classical',
'Alfred Deller': 'Classical',
'Alfred Prinz': 'Classical',
'Alfredo Olivas': 'Pop',
'Ali': 'Pop',
'Ali & Gipp': 'Hip-hop',
'Ali Farka Touré': 'Pop',
```

| source | target | release_date | popualirty | source_genre | target_genre |
|---|---|---|---|---|---|
| Robert Schumann | Vladimir Horowitz | 1928 | 0 | Classical | Classical |
| Frédéric Chopin | Vladimir Horowitz | 1928 | 1 | Classical | Classical |
| Felix Mendelssohn | Vladimir Horowitz | 1928 | 0 | Classical | Classical |
| Franz Liszt | Vladimir Horowitz | 1928 | 0 | Classical | Classical |

# Popularity analysis wrt genre

PES UNIVERSITY

- Intralist and inter list is created that has the mean popularity score of these(inter and intra) collaborations, using which the following bar graph is plotted



## Popularity analysis wrt genre

- A graph is generated using  artist link dataframe
- 2 dictionaries are created

1. maps artist IDs to their popularity scores
2. Maps the artist's popularity and the average popularity of its neighbors in the graph

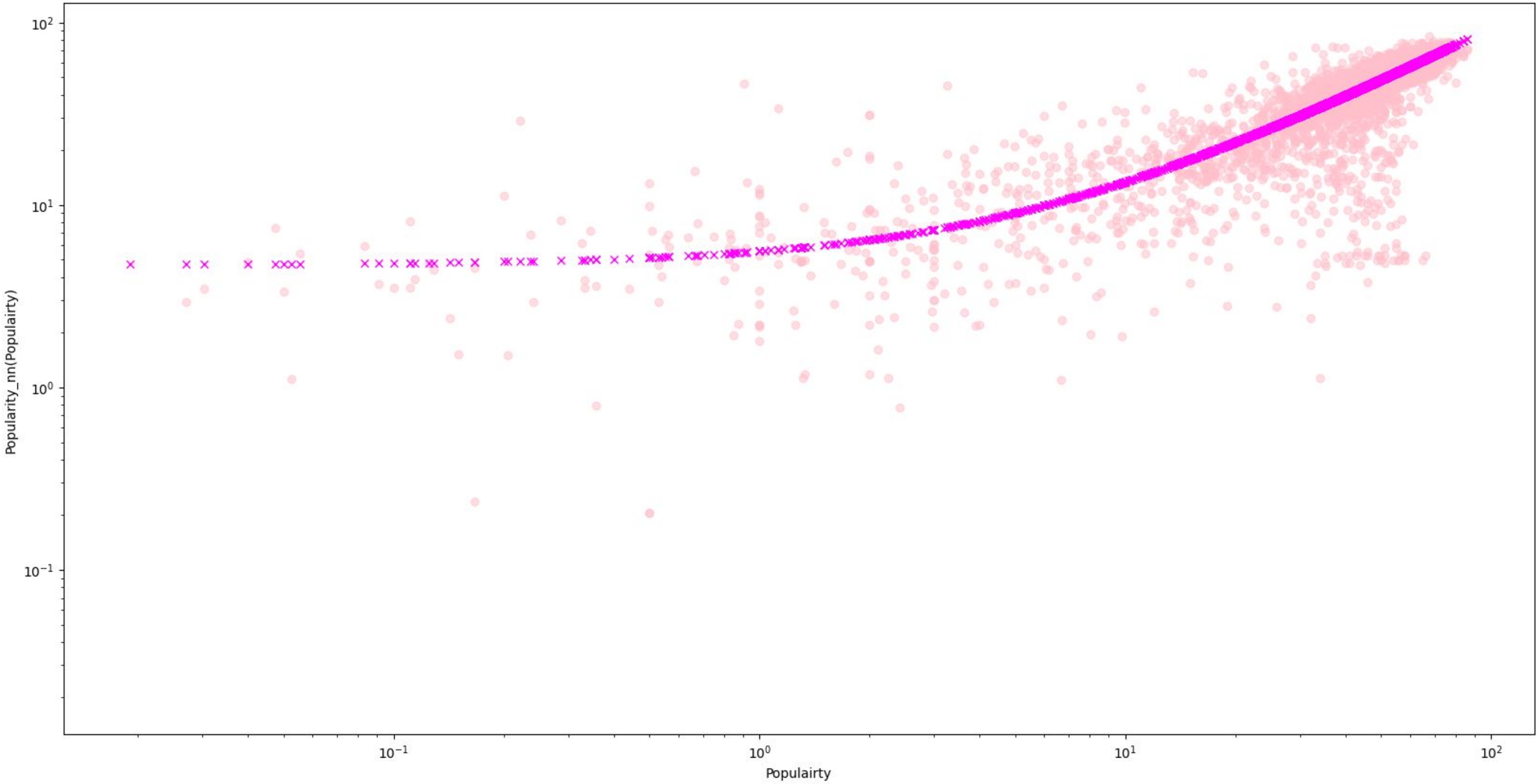
- Using which we print the log log plot in the next slide,there is a positive correlation between artist popularity and the average popularity of their neighbors, with a correlation exponent of 0.00032. The intercept 0.86 indicates that the curve starts at a relatively high value of popularity_nn when popularity is zero. However, the correlation exponent is quite small, which means that the curve is relatively flat and the correlation is not very strong.


```
'Ciara': 56.28571429,
'Cisco Adler': 50.8,
'City Girls': 67.54545455,
'City High': 51.33333333,
'City Of Birmingham Symphony Orchestra': 44.5,
'City of London Sinfonia': 42.5,
```

```
'Clara Haskil': {'popularity': 0.0, 'popularity_nn': 6.8680170574999995},
'Clark Terry': {'popularity': 17.4893617, 'popularity_nn': 9.6379197995625},
'Classixx': {'popularity': 50.0, 'popularity_nn': 50.0},
'Claud': {'popularity': 64.0, 'popularity_nn': 67.9166666500001},
'Claude Bolling': {'popularity': 19.6, 'popularity_nn': 9.111111111},
'Claude Debussy': {'popularity': 4.966153846,
```

# Popularity analysis wrt genre



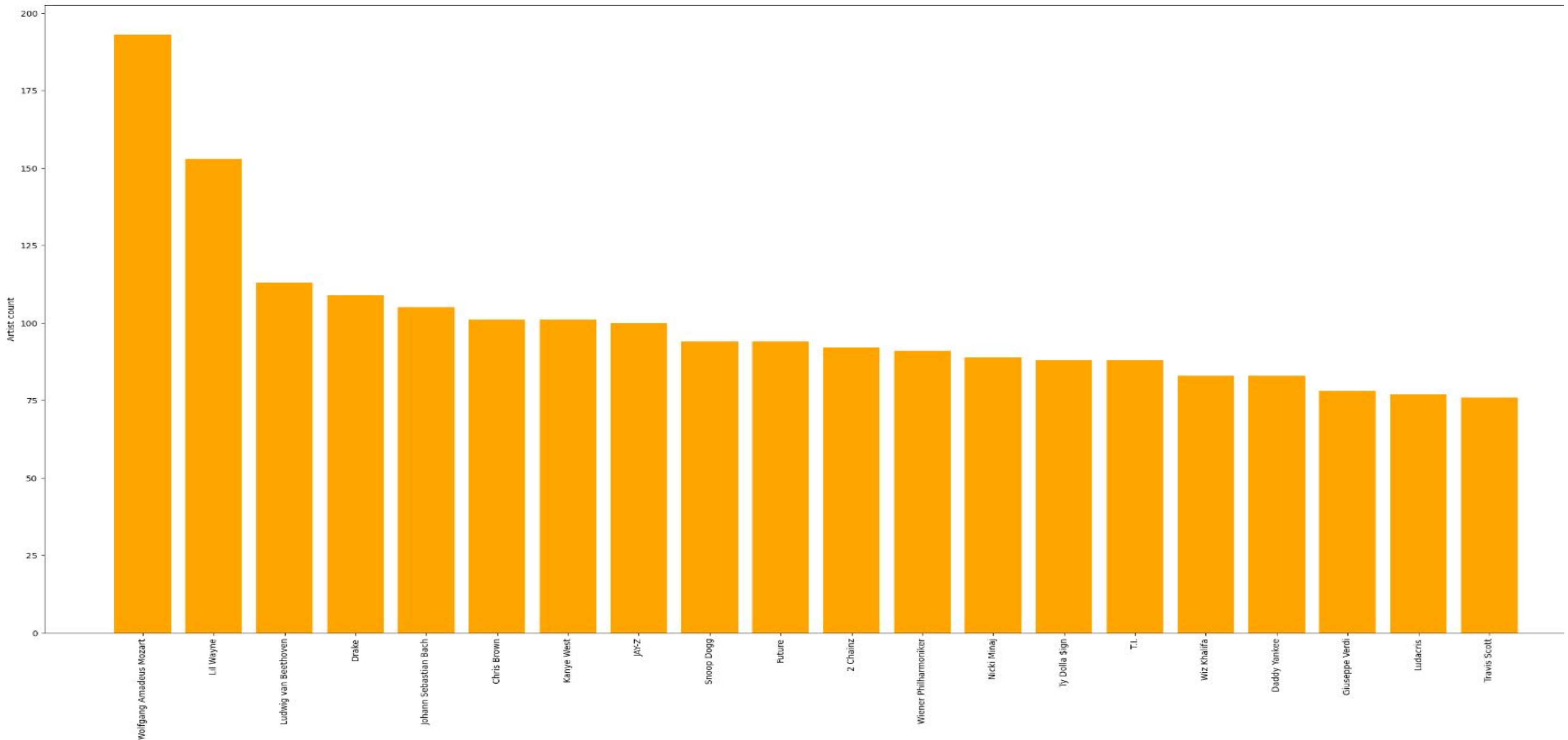


# Popularity analysis wrt genre

- Conclusion:
- 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.
- Elements from different groups might not always add up to be beneficial. Inter-genre acceptability depends on individual groups and their social open-mindedness.
- This suggests that despite the existence of homophily, musical collaboration incorporates more heterophily compared to other forms of social interactions
- 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.

# Temporal analysis

Temporal analysis is done to visualize the number of active years for the top 20 artists with the highest number of active years.

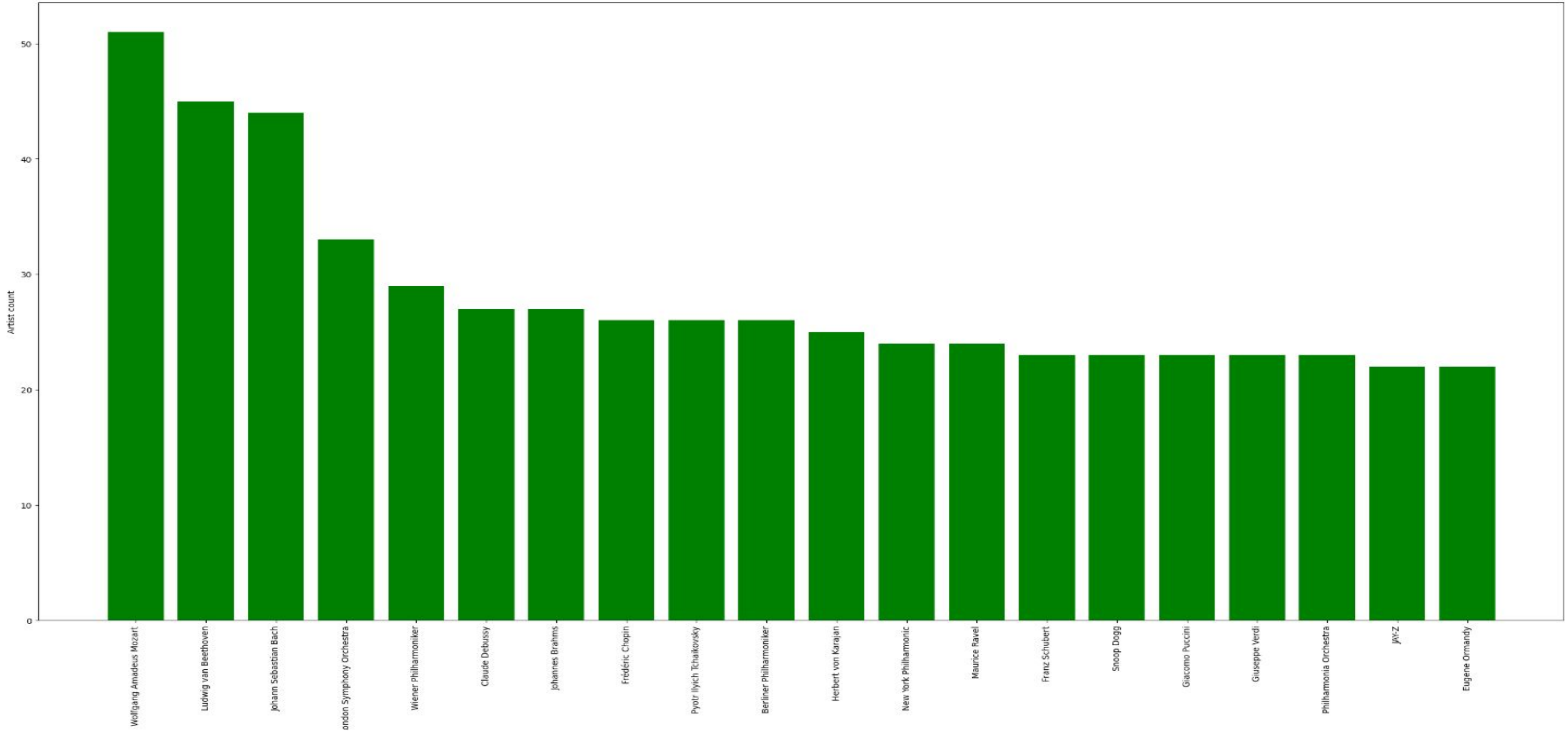


# Temporal analysis

Temporal analysis is done to visualize the number of active years for the top 20 artists with the highest number of active years.

- A dictionary(artistyear_dict) that maps each artist to a list of years in which they released music.
- An empty list years_list to store the years in which the artist released music.
- **artist_link** is a DataFrame that contains information about collaborations between artists, including the release date of each collaboration.
- **artist** is a string representing the name of the artist for whom we want to visualize their collaborations over time.

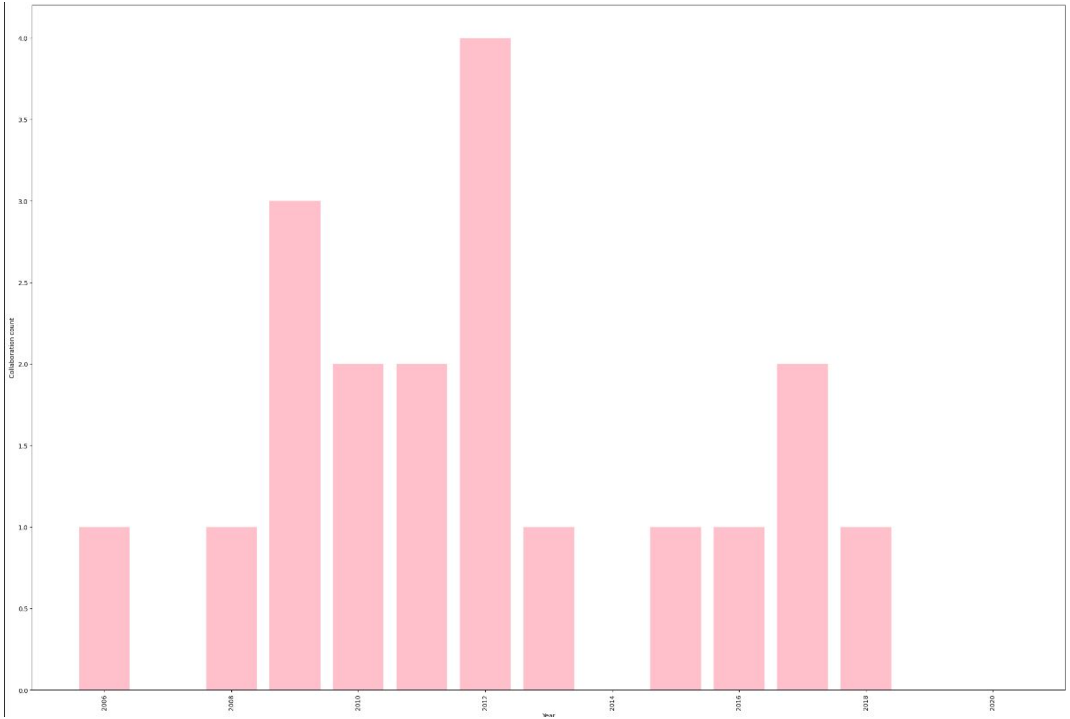
A bar chart of the top 20 artists by the number of distinct years in which they released music.



## Temporal analysis

- Based on the output of the **temporal_collab()** function with '**Rihanna**' as the input for the artist parameter, Rihanna has been releasing music since 2006 and has released music in every year until 2020.
- The bar chart shows the number of collaborations that Rihanna participated in each year. There are noticeable peaks in 2007, 2010, and 2011, where she collaborated on 11, 13, and 10 songs, respectively.
- The chart also shows a gradual increase in collaborations throughout her career, with the highest number of collaborations occurring in the last few years. Overall, the chart suggests that Rihanna has been highly active in collaborating with other artists throughout her career.

# Temporal analysis



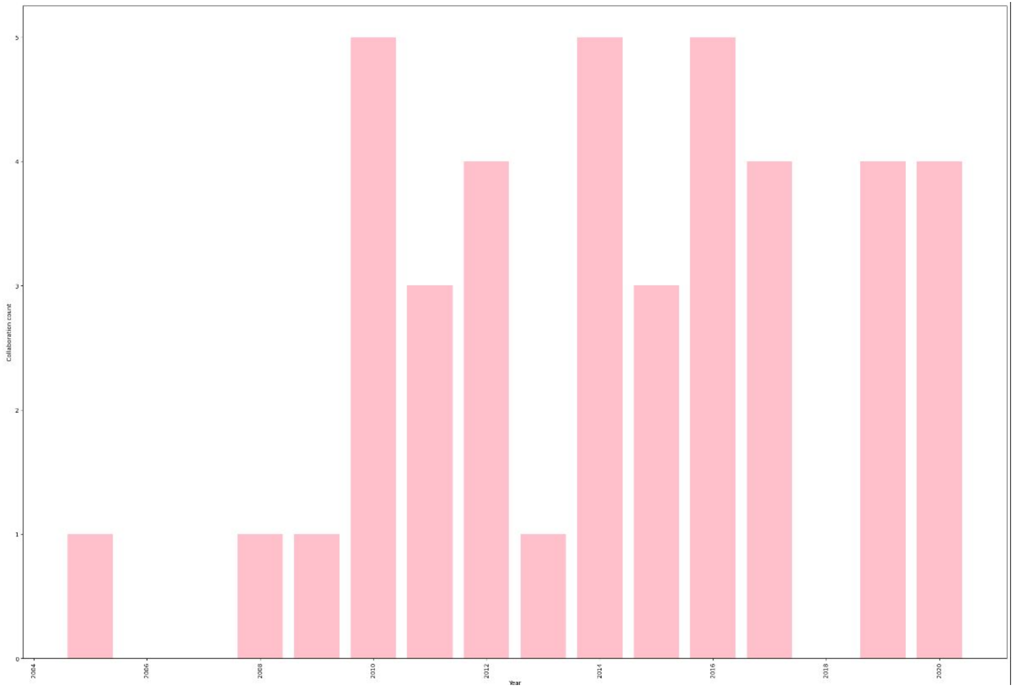

The graph is a plot of collaboration count(y-axis) vs year(x-axis).

This plot is for the Artist Rihanna

```
temporal_collab(artistyear_dict,artist_link,'Rihanna')

Worked between the years:  2006 2020
Total entries of years:  36
```

## Temporal analysis



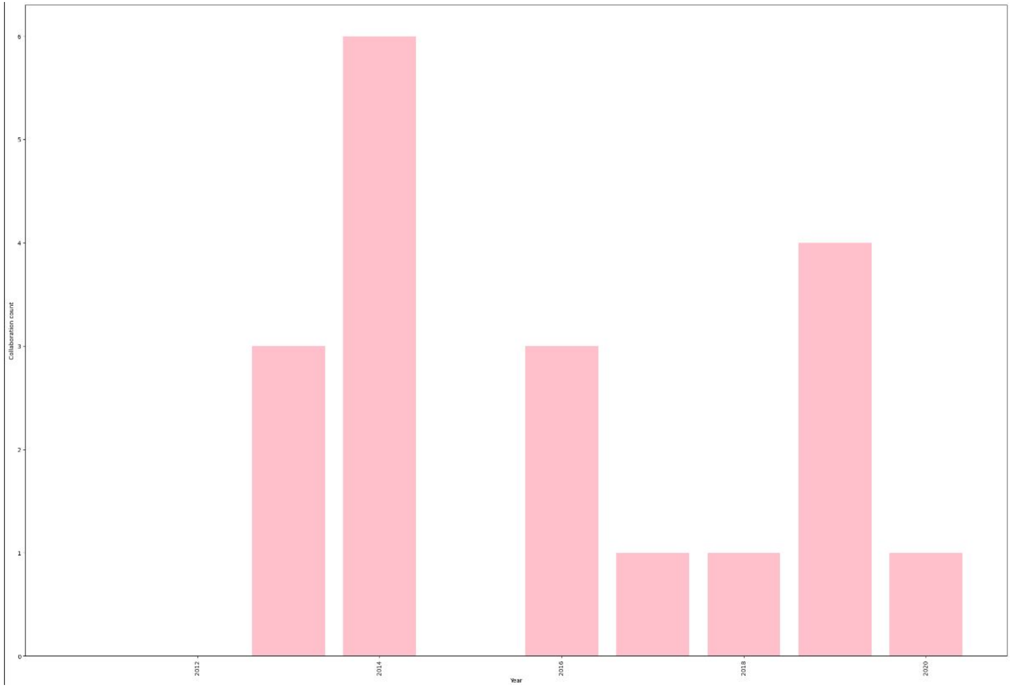

Similarly, the following plot is for the artist Chris Brown

- There are noticeable peaks in 2010 and 2015, where he collaborated on 9 and 10 songs, respectively.
- The chart also shows a gradual increase in collaborations throughout his career, with the highest number of collaborations occurring in the last few years.
- Overall, the chart suggests that Chris Brown has been highly active in collaborating with other artists throughout his career.

```
temporal_collab(artistyear_dict,artist_link,'Chris Brown')

Worked between the years:  2005 2020
Total entries of years:  101
```

# Temporal analysis




- The bar chart shows the number of collaborations that **Ariana Grande** participated in each year.
- There are noticeable peaks in 2014 and 2019, where she collaborated on 7 and 9 songs, respectively.
- The chart also shows a relatively consistent level of collaborations throughout her career, with no significant increase or decrease in collaboration count over the years.
- Overall, the chart suggests that Ariana Grande has been moderately active in collaborating with other artists throughout her career, with a relatively consistent level of collaborations over the years.

```
temporal_collab(artistyear_dict,artist_link,'Ariana Grande')

Worked between the years:  2011 2020
Total entries of years:  32
```

## Temporal analysis

- Over time, the patterns of collaboration among artists appear to change. In the early stages of their careers, artists tend to collaborate more often.
- Even well-known artists collaborate majorly during their early or mid-career stages.
- This may be due to the fact that collaboration can increase their chances of success and enable them reach out to thousands of people
- It is seen that as the artist’s popularity rises, they might reduce collaborating with other artists.

## Quantity and quality of work

| no | Code functionality | % Complete | Runs without problem (Y/N) | If there are minor issues, indicate | Individual Contribution(SRN) |
|---|---|---|---|---|---|
| 1. | Graph generation and centrality measure visualization | 100 | Y | Hard to visualise due to large data. | PES2UG20CS448 |
| 2. | Artists' gender prediction and assortativity | 100 | Y | Nearly 37% of the gender came out to be unknown | PES2UG20CS479 |
| 3. | Popularity analysis wrt genre | 100 | Y | . | PES2UG20CS453 |
| 4. | Temporal analysis | 100 | Y | Temporal analysis done only for 3 | PES2UG20CS399 |

# Top few learning

| Serial No | Top learning in this project |
| --- | --- |
| 1 | Network visualization- uses NetworkX and Matplotlib to visualize a network graph. |
| 2 | Degree centrality: calculates the degree centrality for each node in the graph, which is a measure of how many connections each node has. |
| 3 | collaboration analysis:  explores the degree of collaboration between different music genres using intra- and inter-genre link weights. |
| 4 | Correlation analysis: calculates the correlation coefficient between popularity and popularity of an artist's neighbors, which is used to examine the relationship between an artist's popularity and the popularity of their neighbors. |

# Top few learning

| Serial No | Top learning in this project |
| --- | --- |
| 5 | Gender_guesser- It is a Python library that can be used to guess the gender of a first name based on statistical data. The Detector class is used for predicting the gender of a first name based on its written form. |
| 6 | Collaboration fraction- Itis a measure that represents the fraction of collaborations between two groups of nodes in a network. It is calculated by dividing the number of edges that connect nodes from the two groups by the total number of possible edges between the two groups. |
| 7 | Assortativity is a measure of how likely nodes of similar degree are to be connected in a network. |
| 8 | Temporal analysis - Temporal statistical analysis is done to examine and model the behavior of a variable in a data set over time. |

PES UNIVERSITY

# Top unresolved challenges

| Serial No | Brief description of unresolved challenges | Type of challenge (scope/data/design/implementation / others) |
|---|---|---|
| 1 | Hard to visualise as the dataset is large in python | data |
| 2 | Many artists' gender is unknown | data |

# Reference papers



| No | Paper Title | Authors |
|---|---|---|
| 1 | Community structure in social and biological networks | M. Girvan, M. E. J. Newman, Proceedings of the National Academy of Sciences Jun 2002, 99 (12) 7821-7826; DOI: 10.1073/pnas.122653799. |
| 2 | Musical trends and predictability of success in contemporary songs in and out of the top charts | Myra Interiano, Kamyar Kazemi, Lijia Wang,Jienian Yang, Zhaoxia Yu2and Natalia L. Komarova, R.Soc.opensci.5:171274.http://dx.doi.org/10.1098/rsos.171274 |
| 3 | A network analysis of Spotify's socio- technical related artist network | Silvia Donker, International Journal of Music Business Research, April 2019, vol. 8 no. 1. |
| 4 | Network Analysis of the Spotify Artist Collaboration Graph | Tobin South, Australian Mathematical Sciences Institute, 2017-2018. URL: https://vrs.amsi.org.au/wp-content/uploads/sites/78/2018/04/tobin_south_vrs-report.pdf |