Andrea Chung
Network Science Spring 2023

Analysis of Success-based Musical Genre Network Graph

Abstract

Collaboration between artists is frequent and has been more prevalent throughout the time in the music industry. Not only do musicians from the same genre frequently work together, but so do musicians from different genres. For the study, the network of musical genres are developed based on success, which is determined by the number of hit songs produced between two genres. The project's goal is to identify the elements that promote fruitful cross-genre collaboration. The research paper "Detecting collaboration profiles in success-based music genre network" by Oliviera et al. presented at the 2021 International Society for Music Information Retrieval (ISMIR) conference served as the source for the dataset and the methodology. The study shows that the collaborative nature in musical genres has been expanding constantly and detects three potential factors—Likelihood, Congruity, and Influence—behind successful genre collaboration using edge-dependent network metrics. The findings give good foundation and overview of global music market.

I. Introduction

Over the years, collaborations between artists have been increasing greatly and its importance has risen with the complexity behind the aspects of music industry. Looking at Figure 1[1], the frequency of song collaborations on the Billboard Hot 100 chart expanded substantially. Not only collaborations within the same or similar genres but also between inter-genres such as “latin” and “pop” became relevant and popular in global music market. The bridge between

[1] Ben Cerveny, “Billboard Hot 100 Collaboration Analysis,” April 2018, https://rstudio-pubs-static.s3.amazonaws.com/377689_0dc82788ae194ff0b4006f84ddae713c.html.

somewhat familiar genres with comparatively unfamiliar genres led to creating multi-dimensional revolution in the music industry[2].

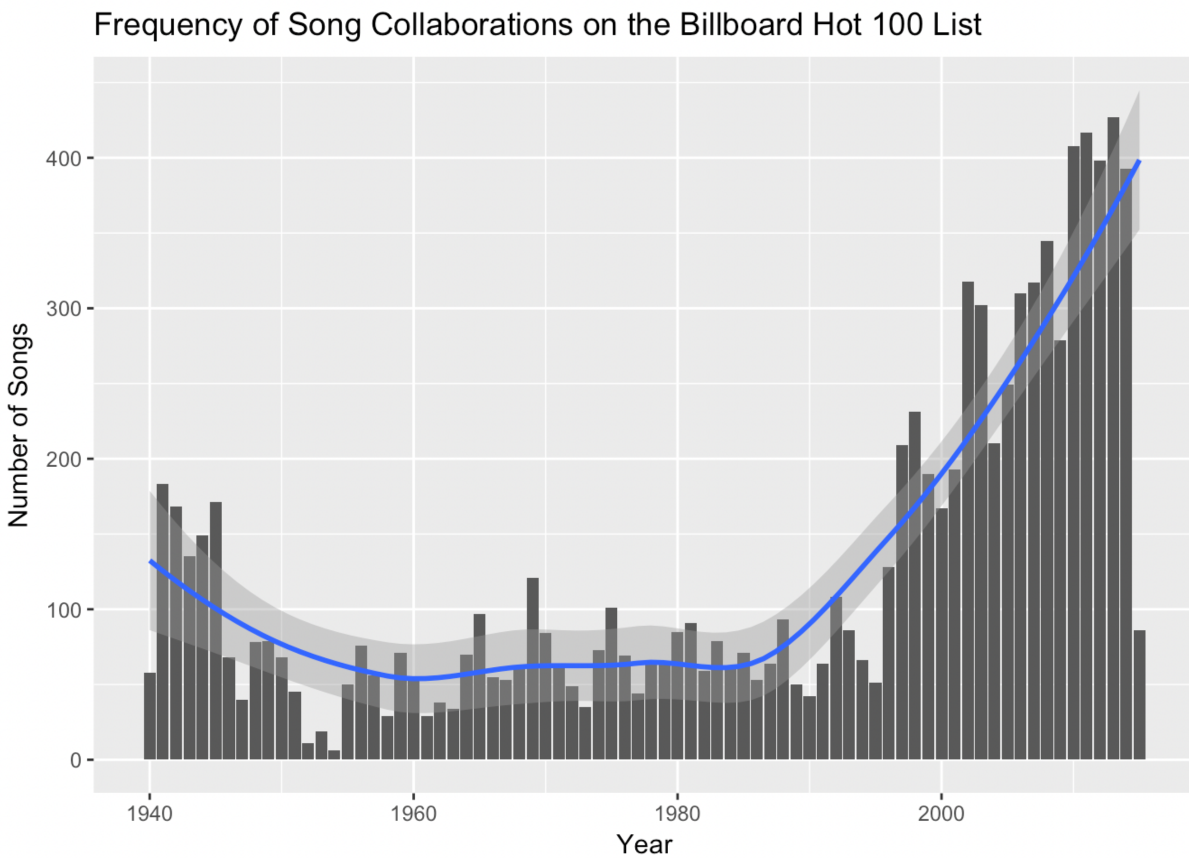


Figure 1. Frequency of Song Collaborations on Billboard Hot 100 Chart

Over the years, the recorded music market has been growing rapidly over the years. The method of listening and gaining access to music changed with the appearance of Spotify in 2008[3]. With more than 500 million users worldwide, including 205 million subscribers in more than 180 markets, Spotify is currently the most widely used audio streaming platform[4]. With the rise of such platforms, we can get easier access to data on songs, artists, its genre labels, popularity, and more

[2] Oliveira, Gabriel P., Mariana Santos, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro. "Detecting Collaboration Profiles in Success-based Music Genre Networks." In *ISMIR*, pp. 726-732. 2020.
[3] "About Spotify," Spotify, March 9, 2023, https://newsroom.spotify.com/company-info/.
[4] "About Spotify," Spotify, March 9, 2023, https://newsroom.spotify.com/company-info/.

features. In the music industry, knowing and utilizing success indicators has become essential for predicting hit songs and identifying worthwhile collaboration. To easily visualize the interconnectivity of collaboration between different artists, analysis using network graph visualization can be necessary. The study would provide its possible usage in application in recommendation or prediction model in music information retrieval (MIR) community or even in various music industry companies.

To address the problem of identifying factors behind successful genre collaboration, thorough analysis of created musical genre network was done using various network metrics and exploratory factor analysis. The following research questions were proposed to conduct the project:

Q1: How has genre collaboration progressed over the past few years?

Q2: What are potential factors in defining success behind genre collaboration?

To answer the proposed research questions, genre collaboration network data from Oliveira et al. was used. Using the data, in-depth analysis was done by applying and calculating different network metrics.

II. Prior Works

In the past, there were several collaboration network analyses. Tobin South utilized artist collaboration network to depict friendship paradox and give further analysis on eigenvector centrality[5]. The edges were created based on whether artists have worked together or not. In the paper, the success measure was popularity value provided by Spotify API. Another network graph

[5] Tobin South, "Network Analysis of the Spotify Artist Collaboration Graph," Australian Mathematical Sciences Institute (2018): 1-12.

created was network of western classical musicians by Bae et al[6]. It was created using compact disc recording data from ArkivMusic database and illustrates complex nature behind network of collaboration in western music culture and examines in multiple scales. Furthermore, Silva et al[7]. perform the study in hit song prediction model and discovers the collaboration as a significant factor in the classification model. The research states that with the added feature of collaboration, the accuracy of model prediction of hit songs increased. Also, Silva et al[8]. create success-based collaboration network at artist-song level and analyzes its effect on artists' popularity and success. However, there has not been study done on collaboration at genre perspective like the research done by Olivera et al.

III. Methods / Analysis

To address the research in global perspective, the dataset was built by using "200 Weekly Top Songs Global" charts, which show 200 songs ranked between 1 and 200 according to streaming counts in the global market provided by Spotify. Through Spotify API, various aspects relating to songs and artists can be collected. One of the elements that was mainly used was genres, which Spotify classifies and assigns to most of the artists in specific manner. For the simplicity of the analysis, the assigned genres were mapped to "super-genres[9]". For each created networks, network metrics were measured using *networkx* package to perform exploratory data analysis (EDA).

---

[6] Arram Bae et al., "The Multi-Scale Network Landscape of Collaboration," *PLOS ONE* 11, no. 3 (2016), https://doi.org/10.1371/journal.pone.0151784.
[7] Mariana O. Silva et al., "Collaboration as a Driving Factor for Hit Song Classification," *Brazilian Symposium on Multimedia and Web*, July 2022, https://doi.org/10.1145/3539637.3556993.
[8] Mariana O. Silva, Laís M. Rocha, and Mirella M. Moro, "Collaboration Profiles and Their Impact on Musical Success," *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, August 2019, https://doi.org/10.1145/3297280.3297483.
[9] Oliveira, Gabriel P., Mariana Santos, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro. "Detecting Collaboration Profiles in Success-based Music Genre Networks." In *ISMIR*, pp. 726-732. 2020.

a. Exploratory Data Analysis

	GLOBAL 2017	GLOBAL 2018	GLOBAL 2019
NODES	72	79	**89**
EDGES	564	583	**709**
AVERAGE DEGREE	15.667	14.759	**15.933**
WEIGHTED AVG. DEG	**256.917**	247.418	236.697
AVG. CLUSTERING COEF.	0.743	**0.757**	0.754
NUM. OF SELF-LOOPS	24	21	**28**

Table 1. Exploratory Data Analysis of Genre Networks (Yearly)

To examine the overall progression of genre collaboration, network characterization metrics were calculated. According to the EDA (Table 1), the number of new genres entering hit song chart increased over the three years as the number of nodes increased continuously. Also, the number of successful genre collaboration increased significantly in 2019 and it is apparent with the increase of number of edges. The average degree had subtle changes throughout the years, but average weighted degree decreased. As more genres were added into the hit song charts, more familiar genres had more collaborations with newly appearing ones. The clustering coefficient was high for each year with minute differences, which indicates that musical genres tend to cluster together, with many connections between genres within each cluster. This can be interpreted as certain genres are inclined to collaborate often and it does not change much over the years. The number of self-loops increased which represents the enlargement of collaboration within the same genre. Even though the weighted average degree decreases over time, the number of self-loops increases, which can be effect from considerable increase in number of genre collaboration in 2019.

b. Network Metrics

To study in-depth on the genre collaboration networks, selected topological-related network metrics were obtained. To measure the “success”, number of hit songs between two genres were counted. The interpretations are given by its characteristics of each metrics.

i. Weighted Degree

The weighted degree shows the number of shared hit songs between two genres. If the weight between two genres is high, it indicates that there are many in-chart or hit songs shared.

ii. Edge Betweenness (EB)

The edge betweenness centrality measures the importance of edge in a network structure based on shortest paths between all pairs of nodes that pass through the edge. In the genre collaboration network, the value can represent the importance of the connection between two genres that could act as the "bridge". If the edge between two genres has high betweenness, the collaboration is highly influential in the network and contributes to emergence of more diverse and new collaboration of genres.

iii. Neighborhood Overlap (NO)

Neighborhood overlap measures the proportion of common neighbors between two nodes relative to total number of neighbors. The percentage is measured according to the overall neighborhood size. If the edge between two genres has low NO, the relationship is relatively isolated. Comparatively, the collaboration between two genres can be insufficient in the successful background.

iv. Preferential Attachment (PA)

Preferential attachment is defined as if a node has higher degree, it has more probability of getting an edge attachment from newly introduced nodes. In the context of success-based genre collaboration network, PA can be used to see which collaboration is more probable in the future when given two genres. If the edge between two nodes has high PA, it means that two genres are highly likely to come together and create hit songs.

v. Common Neighbors (CN)

The common neighbors show the number of neighbors that two nodes share. If two genres have high number of CNs, it can be interpreted as being more likely to collaborate effectively.

vi. Resource Allocation (RA)

Resource allocation is the fraction of resource or information that a node can send to another through its common neighbors. In the context of success-based genre collaboration network, it can help to define most influential genre collaboration in the overall network. However, it can also show the imbalance of distribution among different genres. According to networkx documentation, the values of RA are to be between 0 and 1 but due to presence of self-loops in the network, there are values higher than 1.

Looking at the table below (Table 2), the genre pairs with highest number of hit songs change throughout the year. The pair of nodes with highest edge weight changes. However, interestingly, genre pairs for PA, CN, NO, and RA does not change. From this phenomenon, it can be inferred that genre "pop" is one of the most influential genres in the network and the music industry has high tendency of having more successful collaboration within the pop genre.

	GLOBAL 2017	**GLOBAL 2018**	**GLOBAL 2019**
WEIGHT	pop rap – rap	Hip hop – rap	Latin – reggaeton
EDGE BETWEENNESS	new wave – punk	Pop - sertanejo	Pop - sertanejo
PREFERENTIAL ATTACHMENT	Pop – pop	Pop – pop	Pop – pop
COMMON NEIGHBORS	Pop – pop	Pop – pop	Pop – pop
NEIGHBORHOOD OVERLAP	Pop – pop	Pop – pop	Pop – pop
RESOURCE ALLOCATION	Pop – pop	Pop – pop	Pop – pop

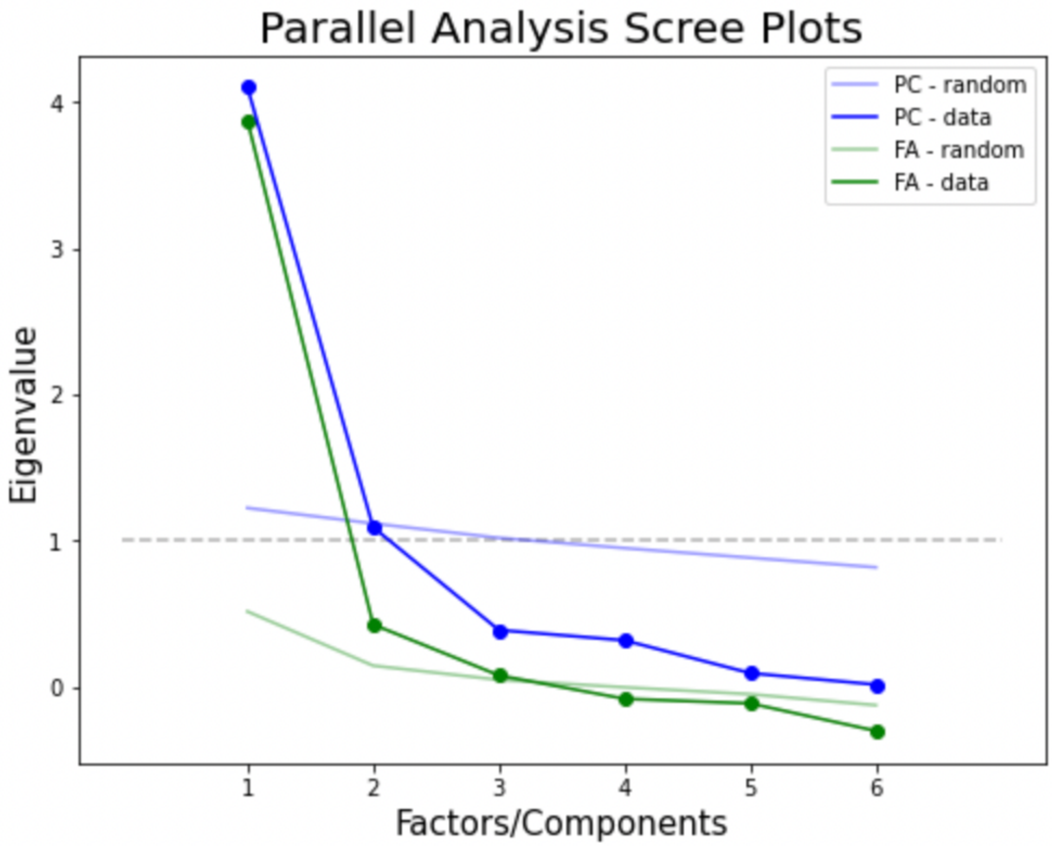
Table 2. Genre pairs with highest values for each network metrics

c. Exploratory Factor Analysis

To define the factors behind the success-based musical genres collaboration network, exploratory factor analysis (EFA) was performed. The purpose of EFA is to "reveal any latent variables that cause manifest variables to covary[10]." Through variance relationship between observable variables, hidden variables are discovered. One of the important parameters when conducting EFA is number of factors to be retained. To define the number of factors, parallel analysis was performed. "Parallel" data is generated randomly along with actual data and estimated eigenvalues are calculated[11]. The number of factors is determined when eigenvalue of simulated data is higher than that of actual data[12]. To visualize the implementation, scree plot is drawn and the number above the point where the curve flattens out shows the number of factors to be observed[13]. The plot below (plot 1) shows the example output of scree plot acquired after performing parallel analysis on edge-dependent network features data for 2018 success-based genre collaboration network. Before applying EFA, all the metric values were scaled. For additional plots, please refer to appendix.

[10] Anna B Costello and Jason Osborne, "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis," *Practical Assessment, Research, and Evaluation* 10 (2005), https://doi.org/10.7275/jyj1-4868.
[11] Ömay Çokluk and Duygu Koçak, "Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors," *Educational Sciences: Theory & Practice* 16, no. 2 (April 2016): pp. 537-551, https://doi.org/10.12738/estp.2016.2.0328.
[12] Ömay Çokluk and Duygu Koçak (2016).
[13] Anna B Costello and Jason Osborne (2005).

```
Parallel analysis suggests that the number of factors =  3
```



Plot 1. 2018 Genre Collaboration Network Graph Parallel Analysis Scree Plot

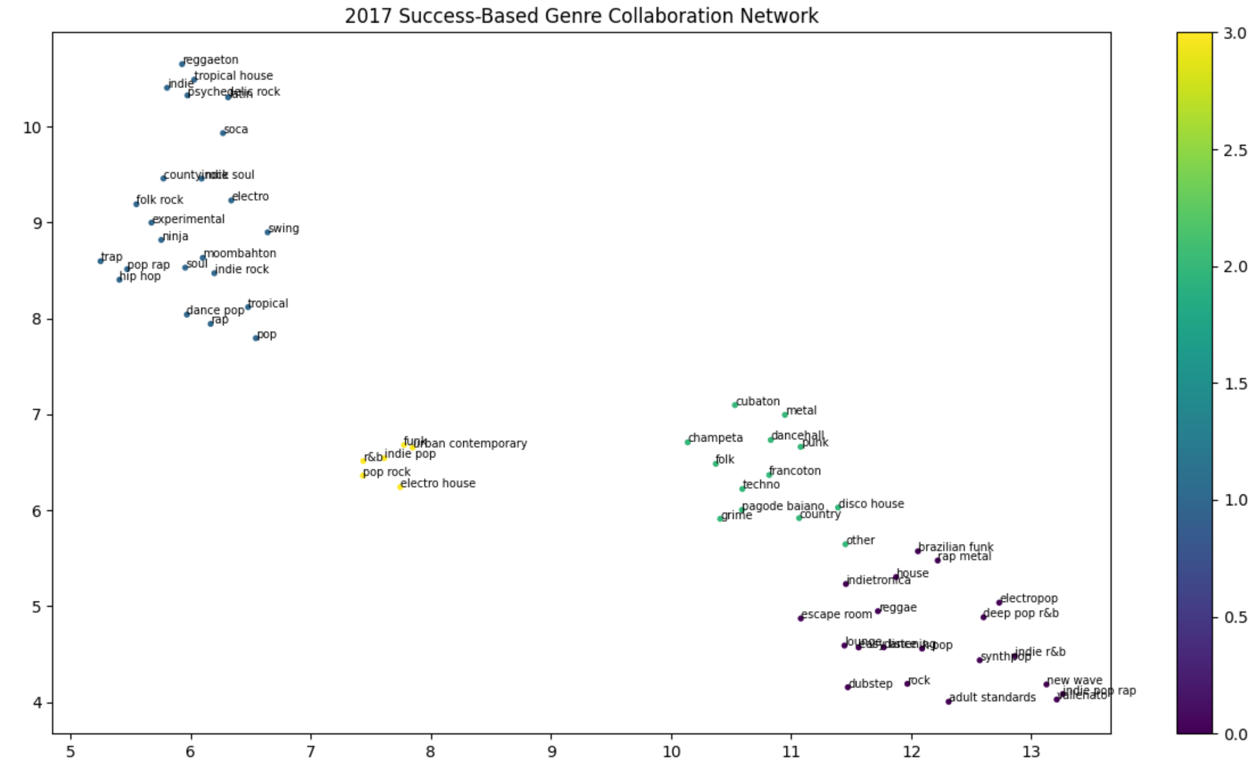
The parallel analysis for each dataset with six different network metrics from genre network for each year suggested three factor structure to draw the relationship between the features. To perform factor analysis effectively, another important parameter is rotation method. Since there is expectation that the measured features are correlated to each other, it would be more appropriate to select oblique rotation method than orthogonal rotation method[14]. Thus, "oblimin" method was used. The factor loadings are achieved and used to determine how much the observed variables explain the factors. The labeled three factors according to observed variables are as the following:

[14] Anna B Costello and Jason Osborne (2005).

1. ***Likelihood***. The factor 1 had high loadings for CN and PA with positive correlation between them. The collaboration of genres that has high CN and PA would refer to high likelihood of two genres to come together in the future. If the edge between two genres has low values or a value close to 0, it would indicate that two musical genres are not close and does not have high probability of collaborating.
2. ***Congruity***. The factor 2 had high loadings for RA and weight with positive correlation. The values of RA and weight refers to "popularity" and strong ties between the two genres. If the collaboration has high RA and weight, the frequency of the successful collaboration is high and has strong ties between them.
3. ***Influence***. The factor 3 had high loadings for NO and EB with negative correlation between them. If the edge between two genres has low NO but high EB, the collaboration is the "bridge" to emergence of new genre collaborations.

d. Cluster Analysis

To further investigate on the relationship between the musical genres in eminent background, cluster analysis was performed using KMeans clustering algorithm. To utilize the graph as the input, node embedding using Node2Vec and dimensionality reduction technique using Uniform Manifold Approximation and Projection (UMAP) were implemented and used to visualize the groupings at node level. To determine the number of clusters, silhouette score was calculated and plotted. The appropriate number of clusters to have for each genre network was four. The below plot shows the one of the outcomes of cluster analysis for 2017 success-based genre collaboration network. For additional plots, please refer to appendix.



Plot 2. Cluster Analysis Output of 2017 Success-Based Genre Collaboration Network

Looking at cluster analysis, genres from further distanced clusters tend to have less frequency of collaboration with hit songs. For example, in 2017 genre collaboration network graph, electro house and rock have weight 1, which means there is only one hit-song shared between the two. Looking at the clustering plot, electro house is placed in yellow-colored cluster when rock is part of purple-colored cluster at bottom-right. Through the cluster analysis, the overall trend of genre collaboration can be captured.

IV. Conclusion

Through this study, success-based genre collaboration network is analyzed using various edge-dependent network metrics to identify underlying factors—Likelihood, Congruity, and Influence—behind favorable outcome of association. Genre collaboration represented through network graph can help to look at overall evolution through the years. Also, deploying edge-dependent network metrics features can be advantageous to understand and pinpoint on driving factors in making genre collaboration to be successful. This finding can be beneficial to both artist and music industry companies, which they can find emerging genre collaboration in success background at global scale.

There are some limitations to this study such as considering only global market for the analysis. Looking at regional markets would help to view the interactions between musical genres at different level and give a broader view on how global market is affected by them. Another limitation is the imbalance of genre distribution between the hit songs. Since the genre “pop” is one of the traditional and prevalent genres among the groups, overpowering is inevitable. However, this limitation can be resolved if regional markets are included as further extension and apply random sampling method if needed. The outcome from the study can be the steppingstone to further analysis and possible extended research into link prediction and recommendation system. The analysis and detection of correlating success factors are valuable to music industry and provides great influence on collaboration appearances.

V. Acknowledgement

The project is inspired by research paper, “Detecting collaboration profiles in success- based music genre network” by Oliviera et al. presented in 2020 ISMIR conference. The data and overall method are from the paper with further extensions added to it.

VI. References

"About Spotify." Spotify, March 9, 2023. https://newsroom.spotify.com/company-info/.

Bae, Arram, Doheum Park, Yong-Yeol Ahn, and Juyong Park. "The Multi-Scale Network Landscape of Collaboration." *PLOS ONE* 11, no. 3 (2016). https://doi.org/10.1371/journal.pone.0151784.

Cerveny, Ben. "Billboard Hot 100 Collaboration Analysis," April 2018. https://rstudio-pubs-static.s3.amazonaws.com/377689_0dc82788ae194ff0b4006f84ddae713c.html.

Çokluk, Ömay, and Duygu Koçak. "Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors." *Educational Sciences: Theory & Practice* 16, no. 2 (April 2016): 537–51. https://doi.org/10.12738/estp.2016.2.0328.

Costello, Anna B, and Jason Osborne. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis." *Practical Assessment, Research, and Evaluation* 10 (2005). https://doi.org/10.7275/jyj1-4868.

Gabriel P. Oliveira, Mariana Santos, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro, "Detecting Collaboration Profiles in Success-based Music Genre Networks," in ISMIR, 2020, pp. 726-732.

Silva, Mariana O., Gabriel P. Oliveira, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro. "Collaboration as a Driving Factor for Hit Song Classification." *Brazilian Symposium on Multimedia and Web*, 2022. https://doi.org/10.1145/3539637.3556993.

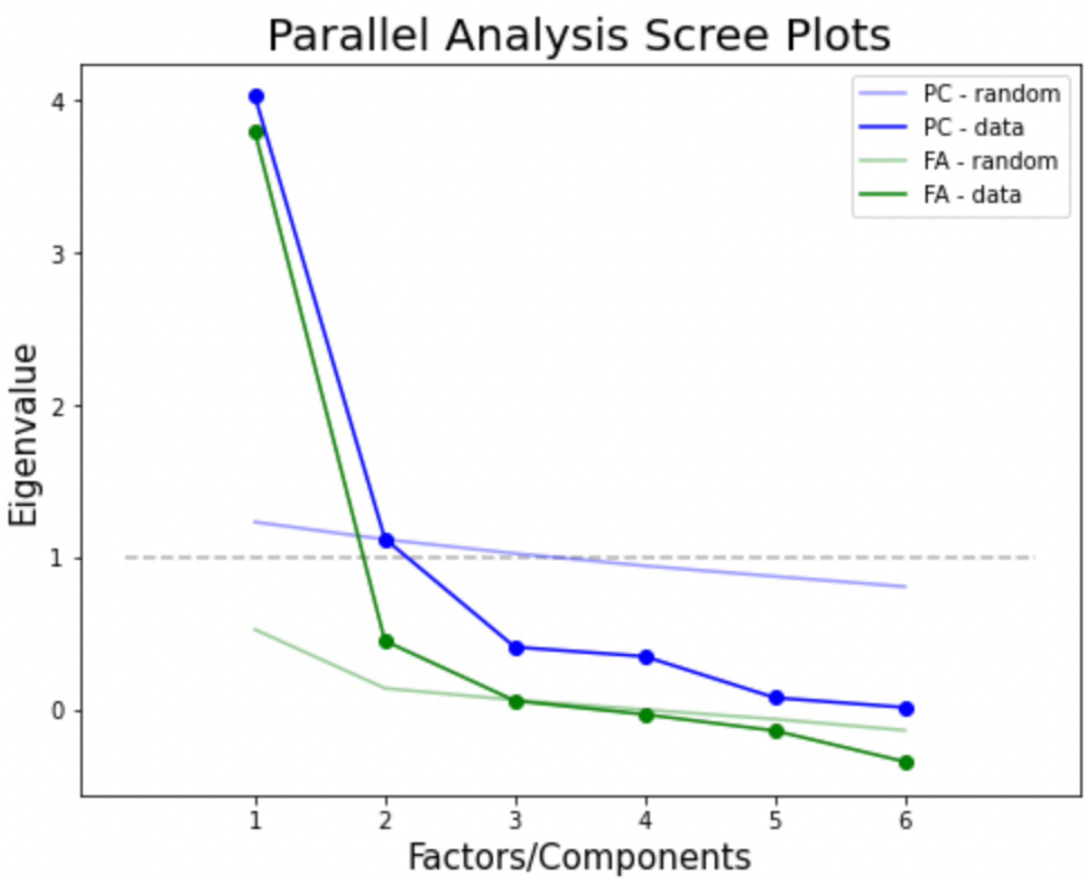
Silva, Mariana O., Laís M. Rocha, and Mirella M. Moro. "Collaboration Profiles and Their Impact on Musical Success." *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019. https://doi.org/10.1145/3297280.3297483.

South, Tobin. "Network Analysis of the Spotify Artist Collaboration Graph." Australian Mathematical Sciences Institute (2018): 1-12.

VII. Appendix

A. Plots of Scree Plots to Determine Number of Factors for EFA

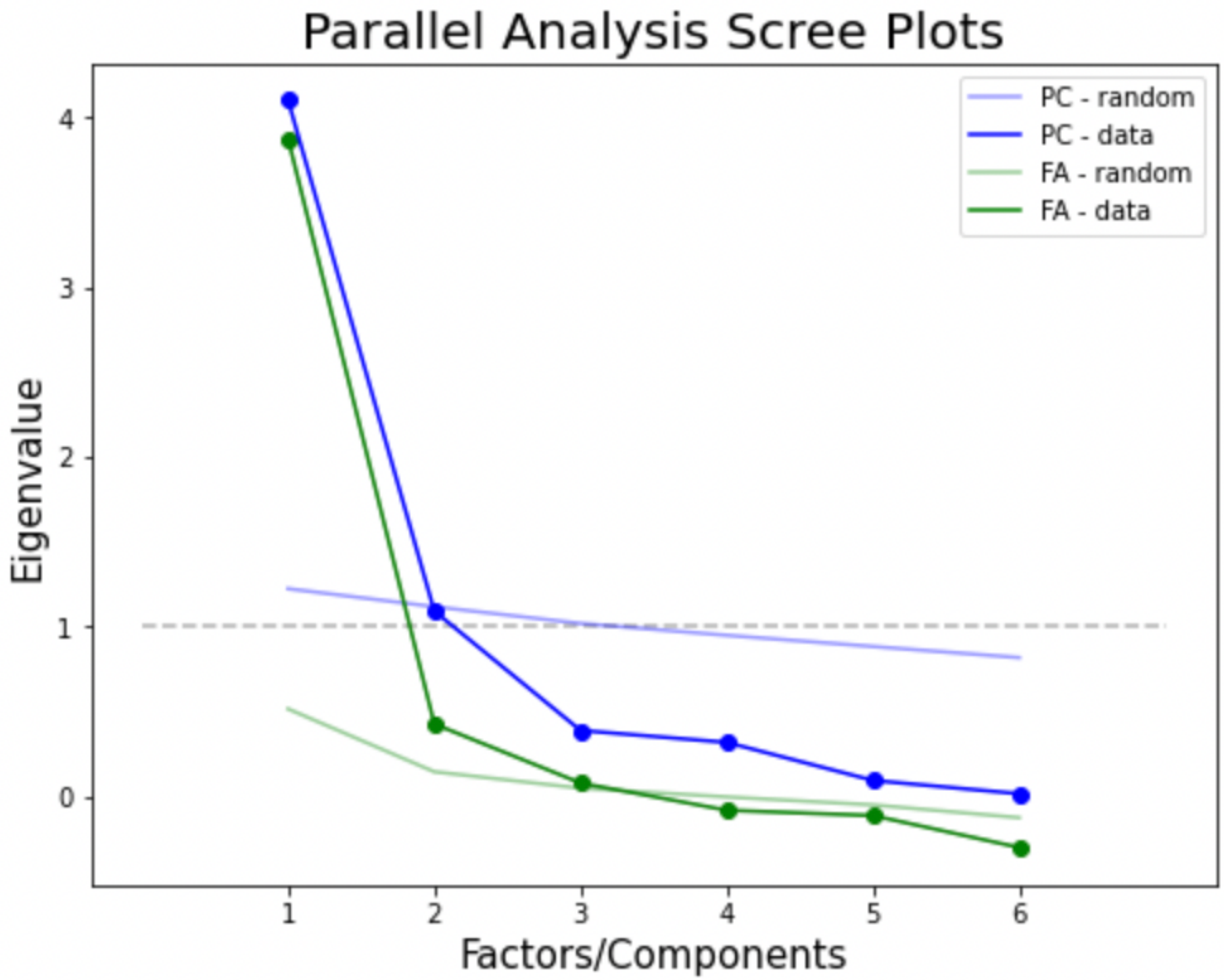
```
Parallel analysis suggests that the number of factors =  2
```



< 2017 Genre Collaboration Network Graph Parallel Analysis Scree Plot>

Andrea Chung
Network Science Spring 2023

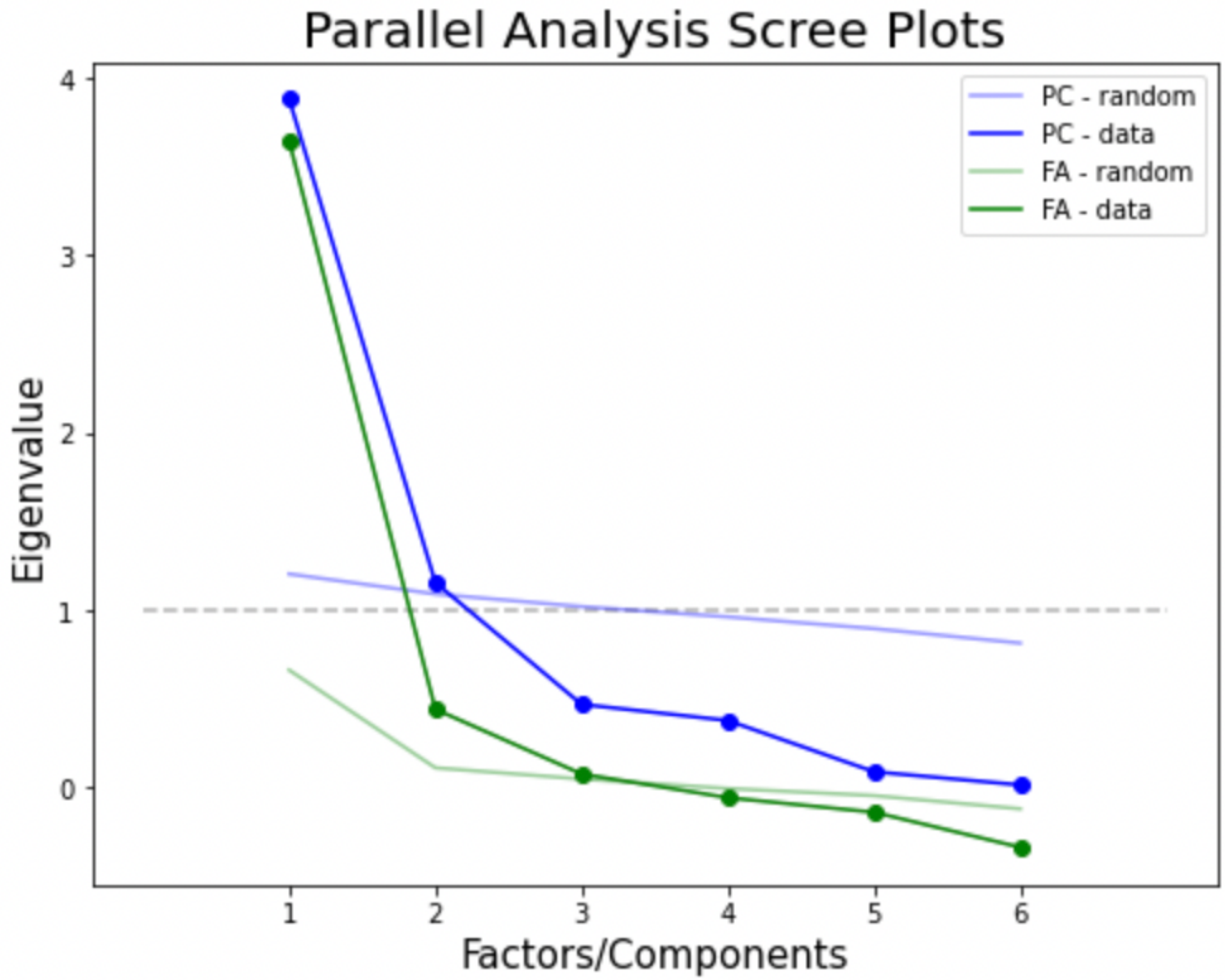
```
Parallel analysis suggests that the number of factors =  3
```



<2018 Genre Collaboration Network Graph Parallel Analysis Scree Plot>

Andrea Chung
Network Science Spring 2023

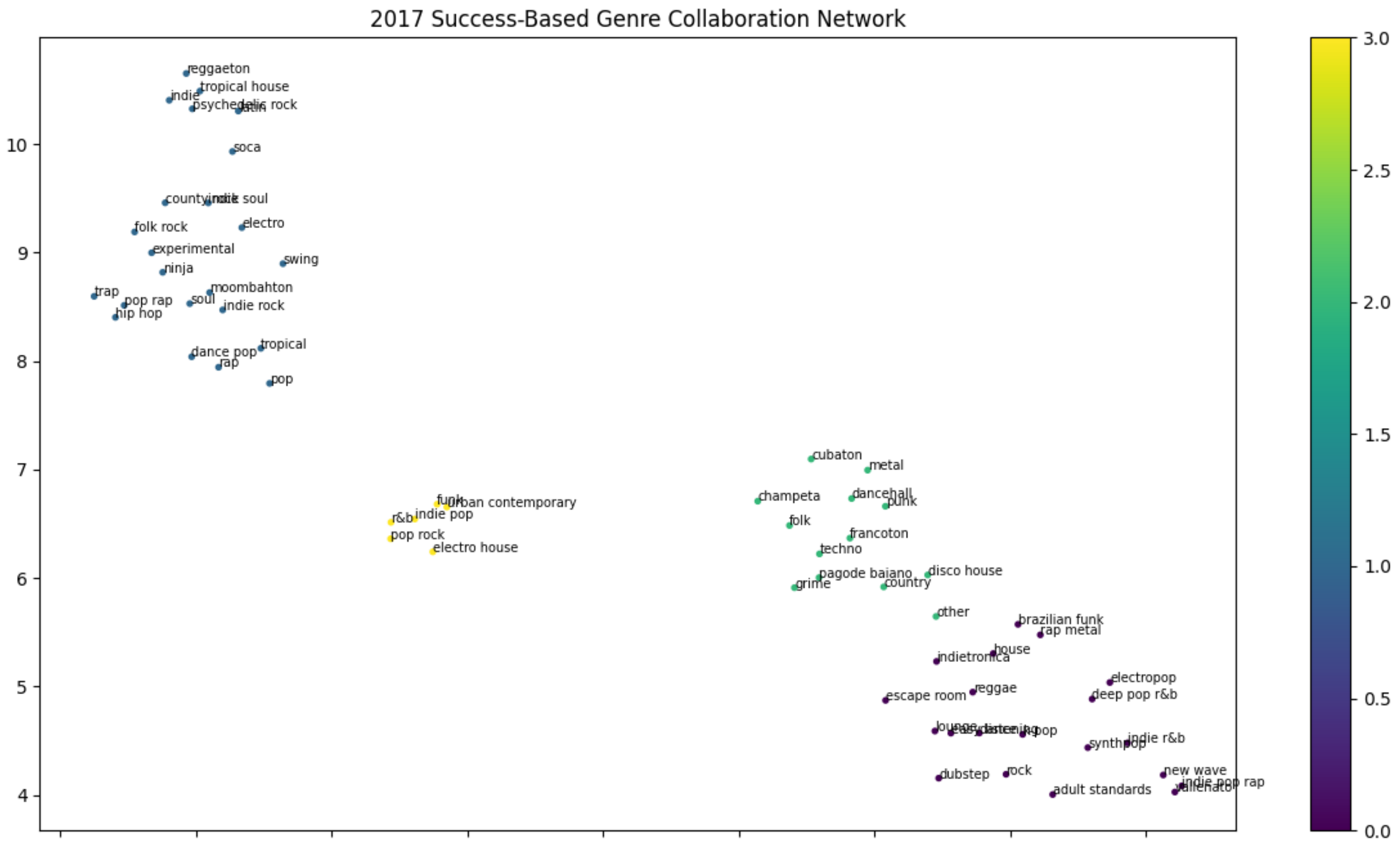
```
Parallel analysis suggests that the number of factors =  3
```



<2019 Genre Collaboration Network Graph Parallel Analysis Scree Plot>

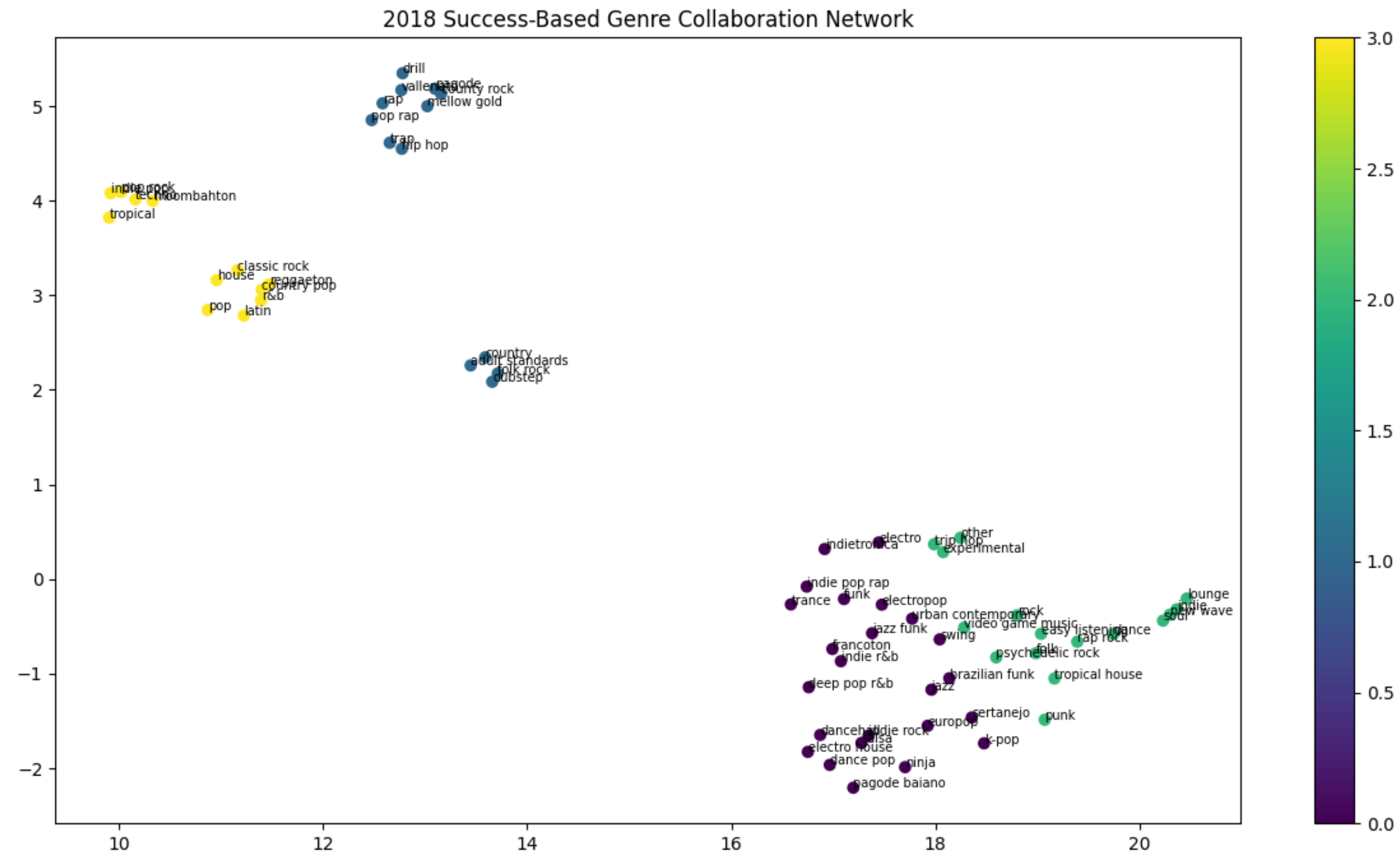
B. Plots of Cluster Analysis on Genre Collaboration Network for each Year

Andrea Chung
Network Science Spring 2023



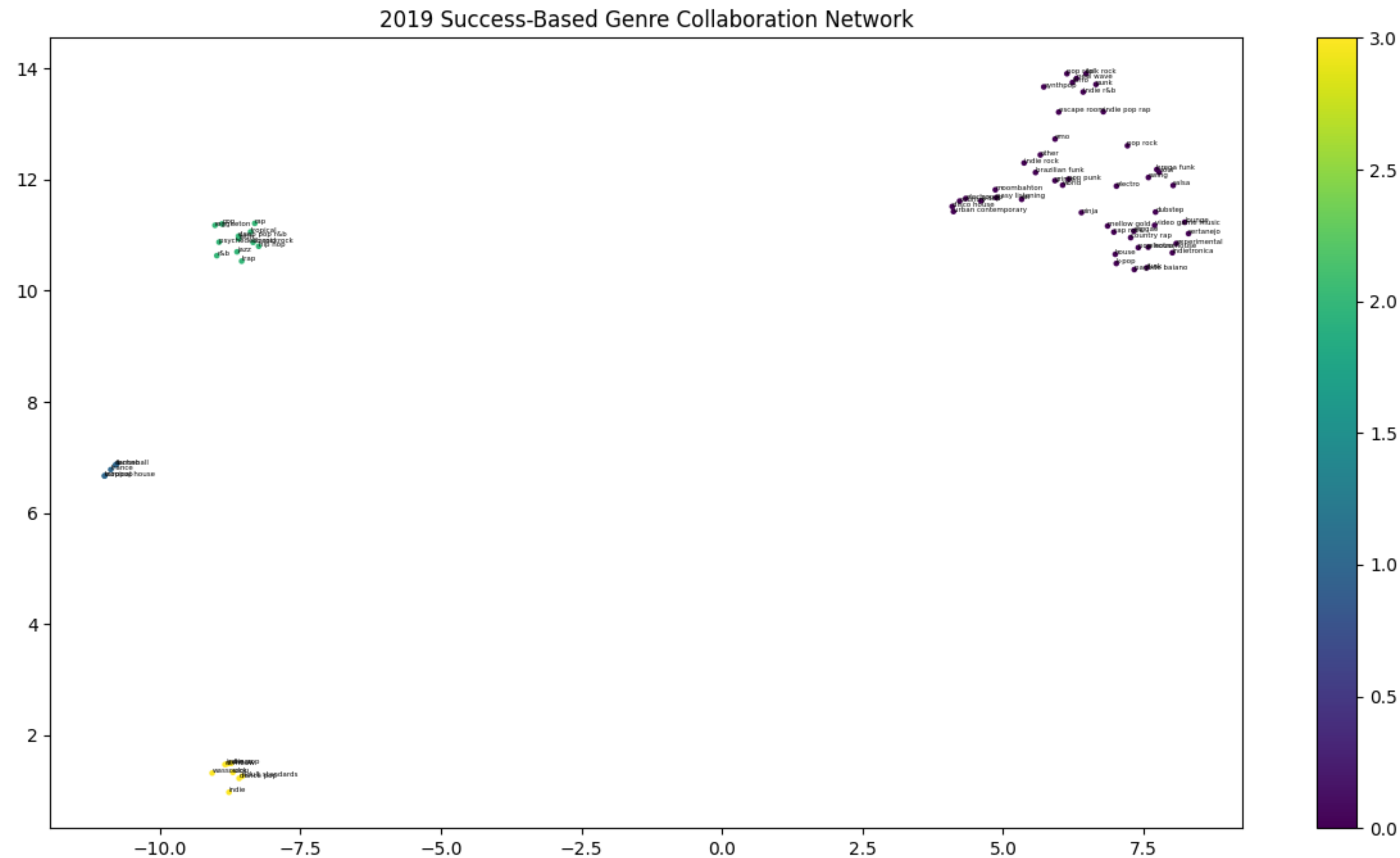
< Cluster Analysis Output of 2017 Success-Based Genre Collaboration Network>

Andrea Chung
Network Science Spring 2023



< Cluster Analysis Output of 2018 Success-Based Genre Collaboration Network>

Andrea Chung
Network Science Spring 2023



< Cluster Analysis Output of 2019 Success-Based Genre Collaboration Network>