Bipartite network analysis of the stylistic evolution of sample-based music

Dongju Park[1] and Juyong Park[1*]

[1] Graduate School of Culture Technology, Korea Advanced Institute of Science & Technology, Daejeon, Republic of Korea
pdj333@kaist.ac.kr / juyongp@kaist.ac.kr

Abstract. In this study we present a network analysis of the communities of artists based on sampling. We construct a bipartite network between the artists who perform the sampling and the samples, then detect communities of the artists and the samples. We find that sample-based music has a clear community structure where each community features artists (nodes) with high centralities, allowing us to determine its musical style. We also define and visualize the similarities between communities representing distinct generations to observe how sample-based musical styles have evolved or been "handed off" to the posterity. This study not only enhances our understanding of sampling-based music, but also presents a novel application of network community structure to a creative enterprise such as music.

Keywords: Sample-based music; Bipartite network analysis; Community detection; Music style evolution

1 Introduction

Musical sampling is a technique used in popular music where one borrows some parts of existing recordings and incorporates them into new musical creations. Sampling can involve using any portion of a song, including the melody, drum parts, and vocals. While experimental music first began using sampling in the mid-20th century [1], it has since become extensively used in hip-hop, electronic, and pop music, particularly since the 1980s. The identity of sample-based music is profoundly related to the songs that were sampled. For instance, G-Funk, the dominant subgenre in West Coast hip-hop during the 90s, created its own rhythm by sampling George Clinton and other funk musicians [2]. Electronic music subgenres such as Jungle and Drum'n'Bass are built on the foundation of one of the most sampled songs in the world, "Amen, Brother." [3] As such, the sampling practice of an artist reflects the characteristics of the subgenre or the music community to which the artist belongs. Therefore, analyzing sampling

* corresponding author

relationships can help us comprehend the different musical styles in the in sample-based musical scene.

Music sampling data is a form of metadata about music that represents the sampling relationships between songs. It is relational data that can be analyzed using network analysis. Many studies have used network analysis to quantitatively analyze musical metadata. Notably, Bryan and Wang [4] created a musical influence network of songs based on sampling relationships and analyzed the network to identify the most influential songs in sample-based music. In another study that analyzed musical sampling using network methodology, Youngblood [5] utilized a network diffusion model to verify the hypothesis that the diffusion of drum breaks, which play a significant role as key samples in sample-based music, occurred through collaborative networks. Unlike this study that focused on drum breaks, our study considers the relationships of all samples with artists, analyzing the community structure of sample-based music that goes further beyond the influence between individual pairs of songs.

Community detection is one of the most standard methods of network analysis. It can also be applied to an influence network based on sampling relationships to discover groups within sample-based music. To do this, we take a cue from studies on citation networks. Musical sampling and academic citation are comparable in that they credit past works for the production of current works [6]. The concept of Author Bibliographic Coupling (ABC) exists in citation analysis [7], a measure of similarity between two authors who cite the same paper. When the author is replaced with an artist and the paper with a song, the similarity between two artists who sampled the same song can be defined in the same fashion.

In this study, we analyze the community structure of sample-based music by constructing an artist-sample bipartite network. The community detected in this network can be understood as reflecting a style in the sample-based musical scene. Furthermore, we define similarities between generations of communities to investigate the stylistic evolution of sample-based music, which we then visualize.

2 Materials and Methods

2.1 Data and network construction

Our data set consists of a total of 333 090 sampling cases between 1980 and 2019 procured from WhoSampled.com. Each case consists of the sampling relationship between song pairs and its metadata (artist, genre, year of release). The total number of songs included in the dataset is 296 456. Each artist's genre is set to be the most common one among the artist's songs in the data set. Since there are can be many styles within a genre, it impossible to specify an artist's musical style by the genre tag alone. To overcome this we collected the style tags shown on the on the artist' pages on Allmusic.com. Of the 42 969 sampling artists included in the sampling data, 14 124 style tags were collected. This low coverage is due to many of the artists who are relative obscure not having been tagged.

In this study, the 40-year period from 1980 to 2019 was divided into five-year intervals, yielding a total of eight generations. The songs in the data set were assigned a generation by the year they were created. Then we constructed artist-sample bipartite networks inside each generation. The bipartite networks comprise two distinct node groups (artists and songs) with edges exclusively linking nodes between the opposing groups. Since an artist may sample a song multiple times, the network is weighted.

2.2 Community detection

When dealing with bipartite networks, community detection is often performed on the one-mode projection of the bipartite network into a unipartite network [8]. Alternatively, community detection can be performed without such projection, preventing the loss of data but is not as widely used. Various modifications of 'modularity' have been proposed for bipartite networks, where modularity maximization is a popular method for community detection in unipartite networks. Modularity is an index that quantifies how many more connections are inside the community compared to random expectation. Here we utilized Barber's bimodularity [9], allowing both artist and sample nodes to be members of the same community, given as

$$Q = \frac{1}{w} \sum_i \sum_j \left(B_{ij} - \frac{d_{1,i} d_{2,j}}{w} \right) \delta_{c_{1,i}, c_{2,j}}, \tag{1}$$

where B is the biadjacency matrix of the network, and w is the sum of the weights of all edges in the network. $d_{1,i}$, $d_{2,j}$ each denotes degree of node i of type 1, and j of type 2. And $\delta_{c_{1,i}, c_{2,j}}$ is 1 when node i of type 1 and j of type 2 are in the same community, and 0 otherwise.

To maximize Barber's bimodularity, we used the Bilouvain algorithm [10], a bipartite variant of the Louvain algorithm. The Louvain algorithm is a heuristic algorithm applicable to weighted networks and is computationally efficient. In this study, the communities can contain both artists and samples in them.

2.3 Defining similarities between communities in different generations

The identity of the community can be determined from its samples; Artists resample previously sampled songs to acquire sounds similar to previous works or to demonstrate respect for senior artists. Consequently, if two communities from distinct generations share samples, it is likely that the two communities show a similar style. The similarity between two communities of distinct generations can be comuted based on this idea.

A community is a bipartite network consisting of artist nodes, sample nodes, and the edges connecting them. A network centrality can be utilized to determine the importance of the samples. The degree centrality is the most fundamental centrality, but it only takes into account the local network information and has trouble differentiating nodes with the same degree due to being an integer value. The HITS (Hypertext Induced Topic Selection) score [**?**] is a centrality that incorporates nonlocal network information, and in this study, we employ the bipartite version of the HITS algorithm. HITS is a scoring algorithm for directed unipartite networks consisting of the 'hub' score and the 'authority' score. The hub score is the sum of the authority scores of nodes that the corresponding node points to, while the authority score is the sum of the hub scores of nodes that point to the corresponding node. This can be extended to the bipartite networks [11] where the score for each node can be defined as the sum of the scores of the nodes it is connected to. This can be expressed using the formula below given as

$$p_j = \sum_{j=1}^{|U|} B_{ij} u_i ; u_i = \sum_{i=1}^{|P|} B_{ij} p_j , \tag{2}$$

where B is the biadjacency matrix of the network, and U and P are separate node sets. The final scores are normalized to 1.

To determine the similarity between two communities, we first identify the shared samples. Then we merge the two communities into a single network and calculate their HITS scores. Then we compute similarity between the two communities as the sum of the HITS scores of the shared samples:

$$HITS\_Sim(C_1, C_2) = \sum_{s \in S_1 \cap S_2} HITS_{\cup C}(s), \tag{3}$$

where $C_1 = \{A_1 \cup S_1, E_1\}$ and $C_2 = \{A_2 \cup S_2, E_2\}$. A_1 and A_2 are sets of artists and S_1 and S_2 are sets of samples. $\cup C$ is a union of communities C_1 and C_2.

3 Results

Table 1. Network information and community detection results for each generation.

Generation	# of artists	# of samples	# of edges	# of communities	bimodularity
1980-1984	1060	2362	3460	465	0.884
1985-1989	3108	5614	19975	511	0.481
1990-1994	7871	14034	53916	966	0.491
1995-1999	9964	19196	46215	1761	0.667
2000-2004	8603	20874	36254	2521	0.784
2005-2009	9691	24147	41635	2785	0.777
2010-2014	14680	34469	63569	3805	0.757
2015-2019	15185	31002	53140	4369	0.827

Table 1 shows the information on the networks belong to the different five years-long generation. While the numbers of nodes in each group exhibit an upward trend over time, the number of edges exhibits a greater degree of variation. Detection results indicate that the number of identified communities increases over time. Bimodularity, quantifying the strength of the community structure, is comparatively low during the generations of 1985–1989 and 1990–1994 but increases subsequently.

The communities we derive from this study consist of artists and samples. Thus a community is not simply a group of artists but can be considered as representing sample-based musical styles. We also identified artists and samples that played a significant role in the community (style) by calculating degree centrality, as the community evolved into another bipartite network. To better comprehend the musical styles of each community, we compiled the Allmusic style tags of each community's artists and designated the five most common tags as the community's main subgenres.

To investigate the evolution of sample-based music styles, we created a network of similarity between communities from successive generations. First, we see that the primary communities of each generation consist of more than 1% of the network's total nodes. The similarity between the primary communities of successive generations was then computed using the similarity index defined earlier. Finally, we constructed a network consisting of the the primary communities of each generation as nodes and their similarity as edge weights. We set a threshold for the edge weights to visualize only connections above a certain level of similarity. Figure 1 depicts the network visualization resulting from a threshold value of 0.2. In terms of node labeling, for instance, '2_9599' represents the 2nd community of the 1995-1999 generation. The main subgenres of the primary communities visualized in Figure 1 are presented in Table 2.

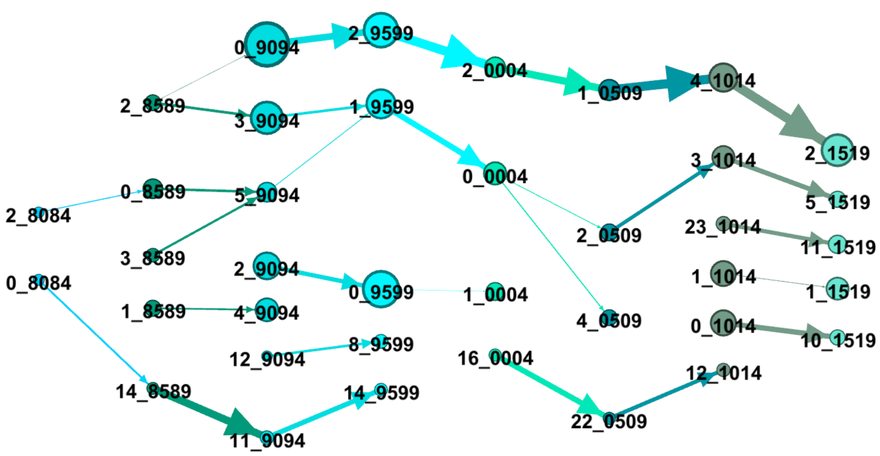

Fig. 1. Similarity network of primary communities in each generation. Edges exceeding threshold 0.2 were excluded from visualization. The node size is proportional to the number of artists belonging to each community, and the thickness of the edge is proportional to the inter-community similarity. The color of nodes signifies the generation to which they belong. In terms of node labeling, for instance, '2_9599' represents the 2nd community of the 1995-1999 generation.

In Figure 1, a significant path is observed from 0_9094 to 2_1519 (top of the figure). These communities represent Jungle and Drum'n'Bass, which are breakbeat-based subgenres of electronic music [3].('Jungle/Drum'n'Bass') Since these communities represent the largest electronic music samples of each generation, we can intuitively see that breakbeat-based music dominates sample-based electronic music. Breakbeat uses drum breaks included in funk, jazz, and R&B music, and the most famous drum break

Table 2. Main subgenres the visualized primary communities in Figure 1. The main subgenre of each community is determined by counting the style tags of artists belonging to the community.

comm.	main subgenre
0_8084	Dancehall, Roots Reggae, Ragga, Contemporary Reggae, Lovers Rock
2_8084	Golden Age, Old-School Rap, Electro, Alternative Pop/Rock, French
0_8589	Golden Age, Old-School Rap, Hardcore Rap, East Coast Rap, Club/Dance
1_8589	Club/Dance, House, Acid House, Dance-Pop, Techno
2_8589	Pop-Rap, Golden Age, East Coast Rap, Party Rap, Hardcore Rap
3_8589	Party Rap, Club/Dance, Bass Music, Quiet Storm, Modern Electric Blues
14_8589	Dancehall, Ragga, Roots Reggae, Contemporary Reggae, Lovers Rock
0_9094	Club/Dance, Jungle/Drum‘n’Bass, Techno, House, Rave
2_9094	Gangsta Rap, Hardcore Rap, West Coast Rap, G-Funk, East Coast Rap
3_9094	Hardcore Rap, Pop-Rap, Golden Age, Contemporary R&B, East Coast Rap
4_9094	Club/Dance, House, Dance-Pop, Acid House, Euro-Dance
5_9094	Club/Dance, Party Rap, Bass Music, Southern Rap, Pop-Rap
11_9094	Dancehall, Ragga, Contemporary Reggae, Reggae-Pop, Club/Dance
12_9094	Hardcore Rap, Gangsta Rap, Southern Rap, Underground Rap, Dirty South
0_9599	Gangsta Rap, Hardcore Rap, West Coast Rap, G-Funk, Pop-Rap
1_9599	Club/Dance, Turntablism, Underground Rap, Hardcore Rap, East Coast Rap
2_9599	Club/Dance, Jungle/Drum‘n’Bass, Techno, Hardcore Techno, Electronica
8_9599	Hardcore Rap, Gangsta Rap, Dirty South, Southern Rap, Adult Contemporary
14_9599	Dancehall, Contemporary Reggae, Ragga, Alternative Pop/Rock, Roots Reggae
0_0004	Alternative Rap, Hardcore Rap, Underground Rap, East Coast Rap, Turntablism
1_0004	Hardcore Rap, East Coast Rap, Gangsta Rap, Pop-Rap, West Coast Rap
2_0004	Jungle/Drum‘n’Bass, Club/Dance, Techno, Electronica, IDM
16_0004	Contemporary R&B, Club/Dance, House, French House, Pop
1_0509	Jungle/Drum‘n’Bass, Club/Dance, Garage, Breakcore, Dubstep
2_0509	Hardcore Rap, Alternative Rap, Alternative/Indie Rock, Underground Rap, French Rap
4_0509	Hardcore Rap, East Coast Rap, Alternative Rap, Trip-Hop, Club/Dance
22_0509	Pop, Dance-Pop, Adult Contemporary, Teen Pop, Contemporary R&B
0_1014	Southern Rap, Hardcore Rap, Pop-Rap, Gangsta Rap, East Coast Rap
1_1014	Pop, Alternative/Indie Rock, Club/Dance, Indie Electronic, EDM
3_1014	Hardcore Rap, East Coast Rap, Political Rap, Golden Age, Heavy Metal
4_1014	Club/Dance, Jungle/Drum‘n’Bass, Dubstep, Garage, House
12_1014	Midwest Rap, Hardcore Rap, Left-Field Rap, Alternative Rap, French Rap
23_1014	Club/Dance, House, EDM, Dubstep, Pop-Rap
1_1519	Pop, Dance-Pop, Alternative Rap, Left-Field Rap, Acappella
2_1519	Club/Dance, Jungle/Drum‘n’Bass, Dubstep, House, Garage
5_1519	Polish, Hardcore Rap, Central European Traditions, East Coast Rap, Political Rap
10_1519	West Coast Rap, Contemporary R&B, Left-Field Rap, Pop-Rap, Gangsta Rap
11_1519	Club/Dance, EDM, Pop, House, Downtempo

is “Amen, Brother” released by The Winstons. This song has been sampled the most across all communities on path. Other prominent drum breaks, such as those from Lyn Collins’ “Think (About It),” Bobby Byrd’s “Hot Pants,” and Incredible Bongo Band’s “Apache,” are commonly found on each community’s list of the top samples. Thus, the drum break, utilized primarily in breakbeat-based music, is fixed and can be seen to have been utilized throughout time.

Community 1_9599 is notable as well. The predecessors of Community 1_9599 refer to those that represent the old-school hip-hop style, including samples such as Beside’s “Change the Beat (Female Version)” and James Brown’s “Funky Drummer”. Specifically, “Change the Beat (Female Version)” is the most sampled song in the world and can be considered an iconic old-school hip-hop sample utilized in DJ scratch performances [13]. Community 1_9599 represents Turntablism and underground hip-hop styles focused on famous hip-hop DJs (‘Turntablism’, ‘Underground Rap’) and illustrates the success of Turntablism music in the late 1990s [12]. The successors of 1_9599 can be considered to be the genres that retain the essence of classic hip-hop. Therefore, it is notable that the path following Community 2_0509 is dominated by Polish hip-hop artists [14]. This suggests that in recent years, artists who inherit the old-school hip-hop style have emerged more frequently in European countries such as Poland than in the United States, the birthplace of hip-hop.

Also identified is a path connecting 0_8084 $\rightarrow$ 14_8589 $\rightarrow$ 11_9094 $\rightarrow$ 14_9599 (bottom left of the figure). These communities are synonymous with reggae music. Reggae is also a sampling-based music genre, like hip-hop and electronic music, primarily sampling prior reggae music [15]. Before 1980 when hip-hop was born, reggae was the major music type. Reggae songs such as “Funaany” by Admiral Bailey, “Full Up”, “Drum Song” by Jackie Mittoo, and “Real Rock” by Sound Dimension were frequently sampled in each community. The fact that this path was cut off in the 1995-1999 generation suggests that the sampling reggae became much less popular in the 00’s.

4 Conclusions

In this study we investigated the community structure of sample-based music post-1980 using bipartite network analysis. We constructed the sampling networks for each of the eight five year-long generations and then conducted community detection. The communities established in this manner were shown to be representing a style.

Our analysis focused on two significant ample-based music genres, electronic music and hip-hop. Jungle and Drum‘n’Bass are the subgenres of electronic music that use a style that incorporates breakbeats. Since there are only a few types of breakbeats, we showed that the Jungle / Drum‘n’Bass communities in each generation have strong ties. We also observed that the Old-school hip-hop style, which has persisted since the 1980s, is diverging into multiple branches and that non-European artists, such as those from Poland, continue to use this style into the 2000s.

In the future we intend to increase our understanding of sample-based music styles by conducting a more comprehensive analysis of the artists and genre information of the derived communities. We may also vary the time unit used to divide generations in order to conduct analyses on a different scale. Moreover, by varying the edge weight threshold when visualizing the similarity network of consecutive primary communities, we may observe the evolution of sample-based music genres in more detail. Lastly, we could investigate generation-skipping transmission of musical styles by analyzing the similarity between two communities separated by more than one generation, which could show how the phenomenon of "revival" of musical styles occurs.

Acknowledgements

This research was supported by KAIST Post-AI Research Grant and Korea Creative Content Agency funded by the Korean Government (RS-2023-00270043).

References

1. Roads, C. *Early Electronic Music Instruments: Time Line 1899-1950.* Computer Music Journal 20, no. 3 (1996).
2. Williams, Justin A. *'You Never Been on a Ride like This Befo': Los Angeles, Automotive Listening, and Dr. Dre's 'G-Funk'.* Popular Music History 4, no. 2 (2010).
3. Hockman, J., Davies, M. & Fujinaga, I. *One in the jungle: Downbeat detection in hardcore, jungle, and drum and bass.* In Proceedings of the 13th International Society for Music Information Retrieval (ISMIR), Porto (2012).
4. Bryan, N. & Wang, G. *Musical Influence Network Analysis and Rank of Sample-Based Music.* Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011. pp. 329-334 (2011).
5. Youngblood, M. *Cultural transmission modes of music sampling traditions remain stable despite delocalization in the digital age.* PloS one, 14(2), e0211860 (2019).
6. Hess, M. *Was Foucault a Plagiarist? Hip-Hop Sampling and Academic Citation.* Computers and Composition 23, no. 3. pp. 280-295 (2006).
7. Zhao, D. & Strotmann, A. *Evolution of Research Activities and Intellectual Influences in Information Science 1996-2005: Introducing Author Bibliographic-Coupling Analysis.* Journal of the American Society for Information Science and Technology 59, no. 13. pp. 2070–86 (2008).
8. Zhou, T., Ren, J., Medo, M. & Zhang, Y. C. *Bipartite Network Projection and Personal Recommendation.* Physical Review E 76, no. 4 (2007).
9. Barber, M. J. *Modularity and Community Detection in Bipartite Networks.* Physical Review E 76, no. 6 (2007).
10. Pesantez-Cabrera, P. & Kalyanaraman, A. *Detecting Communities in Biological Bipartite Networks.* Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. (2016).
11. He, X., Gao, M., Kan, M. Y., & Wang, D. Birank: Towards ranking on bipartite graphs. IEEE Transactions on Knowledge and Data Engineering, 29(1), 57-71 (2016).
12. Greasley, A. E., & Prior, H. M. Mixtapes and turntablism: DJs' perspectives on musical shape. Empirical Musicology Review, 23-43 (2013).
13. Hansen, K. F. *The Basics of Scratching.* Journal of New Music Research 31, no. 4. pp. 357–65 (2002).
14. Aniskiewicz, A.G. *Cultural Remix: Polish Hip-Hop and the Sampling of Heritage* Ph.D. dissertation, University of Michigan (2019).
15. *Dub: Soundscapes and Shattered Songs in Jamaican Reggae.* Choice Reviews Online 45, no. 03 (2007).