

# Measuring Musical Sampling Impact Through Network Analysis

Justin Tran

Adviser: Andrea LaPaugh

## Abstract

*Musical sampling influence has only recently been studied through network structures through the basic analysis of artist-artist sampling relationships. In this paper, I integrate the use of additional properties of music sampling (such as genre, time period, and audio element sampled) to investigate patterns of influence in the musical community at large. Using the WhoSampled dataset, I investigate statistical metrics such as the most-sampled artists songs as well as the trend for musical sampling over time. I also take a more nuanced look at "influence" by providing a variety of graph centrality measurements for determining the influence of a node (representing an artist) on other nodes. This analysis resulted in a greater understanding of musical influence certain artists and genres had over other heavily-sampling artists and genres over time. The most influential genre was found to be Funk/Soul/Disco while the most influential artist of all time was James Brown. More specific influencers from different time periods were also found. We conclude with possible future research that can be applied to this network analysis of musical sampling.*

## 1. Introduction

Music sampling is the act of taking a portion of another musical piece and reusing it as an element in a new recording. Musical sampling dates as far back as the 1890's which suggest that sampling may be a product of stylistic practices rather than being a modern trend. While certain genres of music are notable for utilizing sampling heavily during certain eras such as early 90's hip-hop, the use of sampling has spread far beyond hip-hop and is being employed by a variety of music producers across a variety of genres.

Sampling will be the measure of musical influence in this paper. Specifically, influence will be partially defined by the number of times a piece is sampled by other pieces. Sampling informs listeners of the artist's level of influence on other musicians within their expected sphere of influence within a genre or outside of it. With this in mind, we can ask, "Are there certain musicians or genres that exhibit a strong influence on others in the music community?"

The primary goal of this paper is to explore relationships between artists and genres and determine which utilize sampling the most in comparison to others. In addition, we can observe intra-genre and inter-genre sampling (having a sample be used by an artist from the same genre versus another genre, respectively). By noting the intra-genre relationships, we can identify whether artists tend to sample more from other artists within their genre or if they tend to extend their musical reach to unrelated genres instead. The secondary goal is to apply our network analysis on the sampling patterns of audio element (sample vocals, bass lines, drum beats, etc.) sampling across genres to different time periods. This time-based model of analyzing sampling can uncover specific sampling practices that arose during certain eras or possibly failed to exist after a certain period, an aspect of musical sampling that is often unconsidered in past related works.

### **1.1. Sampling Influence and Copyright Law**

The importance of musical sampling influence takes on a controversial role in modern music due to the variety of lawsuits arising from unrestricted use of samples. Somoano notes that parties filing lawsuits for uncondoned sampling usage often cite the fact that their music has greatly influenced the music community to bolster the argument behind the importance of the possession of their song as billable property. [16] We can interpret these statements as saying that their sampling influence in the music community is both culturally noticeable and quantitatively measurable. By charting musical influence patterns with sampling based on a variety of factors such as time period, genre, and audio element sampled, one can better understand the artists and record companies filing lawsuits for uncondoned uses of a sample and observe whether their record does have a large influence on specific musical communities.

## 2. Related Work

### 2.1. Musical Sampling Influence Networks using WhoSampled

In past works, research has been performed on basic network analysis of sampled music that is grouped into genres at the very least. This is seen in a paper by Bryan and Wang from Stanford University's Department of Music. [7] The paper utilized the WhoSampled.com dataset to analyze musical influence and rank artists, songs, and genres based on their level of "sampling influence" throughout history then proceeded to rank the categories based on their amount of sampling analyzed via clustering and node degree. Overall, network analysis was used to indicate the relative flow of samples between genres. However, no intra-genre analysis was included which limited the potential findings that this dataset gives access to. The paper noted the complex nature of using network analysis to define "influence" in music but settled on using degree centrality as a sole measure of influence. They came to the conclusion that a unique power-law degree distribution is followed in the musical sampling world: Funk, soul, and disco music are heavily sampled by hip-hop, R&B, and electronic music when compared to the other genres that are sampled. A heavy focus was put on hip-hop, R&B, and electronic music as well while generally leaving the other sampling genres unanalyzed as the data for other genres was less rich. The paper also noticeably omits any analysis on any properties of the sampled or sampling music such as harmonics or audio elements.

Additional research by Stanford researchers Alban, Choksi, and Tsai attempted to investigate music sampling based on harmonic and timbral features such as dominant chords and qualitative emotional response. [4] It was a direct attempt to extend upon the work done by Bryan and Wang by placing a greater focus on features of the sampled works that are noticeable (but sometimes very subtle) to individuals with a deep background in music theory. The researchers identified the music by specific harmonic and timbral features rendered by the piece. Examples of "harmonic features" include "changes involving minor/major 7th chords" and "natural minor key changes". "Timbral features" include "calm, quiet, mellow". This paper's focus was clearly a far departure from the general sense of "influence" described by Bryan and Wang as this maps influence to preferential

attachment in the network graph. Each edge was scored using a product of the degrees of each node to denote influence whereas our influence metrics include multiple types of centrality and degree measures rather than choosing a single model for influence.

The focus of Bryan et. al. and Alban et. al. varies from this paper as they mainly attempted to draw associations between the presence of harmonic and timbral features and their ability to make a song more likely to be sampled to form a ranking of top features. In addition, our paper does not touch on harmonic features because music theory would require additional knowledge that is not the focus of our "influence". Timbral features are not included in our paper either due to the need for granular data tagging that would be required in our dataset which was not provided by WhoSampled. Neither of the papers touched on the quantitative popularity of sampling over time much less the specific sampling techniques used over a set number of decades.

Brandford performed a unique solo analysis of sampling done by Kanye West and applied many analysis strategies that our paper also applies. [6] One of these unique areas of analysis is time period. Brandford sorted the songs Kanye West sampled into the decades they were created in and found that a disproportionate number of samples came from 1970's tracks. It is important to note that West's samples encompassed every decade back to the 1950's. Simpler data analysis that the previous two pieces of research employed were also used by Brandford. Basic statistics on the number of samples Kanye West has used and the number of artists these samples came from were cited to ensure the reader that researching Kanye West alone could provide a rich dataset. No analysis on genres was mentioned. It was clear that the focus of analysis was more limited in this investigation as most of the analysis was placed on time period and pure numbers of occurrences rather than forming a network (which, to Brandford's defense, would be quite limited for a single artist). Our paper does borrow the idea of investigating time period from Brandford's investigation as it provides a better understanding of the types of music that sampling producers find an interest in when sampling music.

## **2.2. Alternative Measures of Influence in Music Networks**

Watson uses social network analyses from an economic and sociological perspective to pinpoint the connectivity of music networks between large metropolitan areas by analyzing their connectivity with music sampling. [21] Note that Watson's paper focuses on networks by looking at geographical areas as the nodes whereas all aforementioned works use a piece of music as a node. However, a sample usage still creates a link between nodes. Network graphs are where each node represents a metropolitan area and their number of links in the graph (amount of influence over the music industry as a whole) increases as more albums sample a song distinctly produced in their city. Nodes with greater degrees indicate more central metropolitan areas with greater influence in the music sampling network. This research applies the same principles we saw in Bryan and Wang's work but applies it to geographical areas and looks at these areas, rather than individual artists or genres, as the creators of musical influence. This is a large assumption that deviates from the popular belief that the reason a piece of music is sampled is due to the genre or artistic style provided by the track. Instead, the work believes that the unique characteristics of an area's musical community finds its way into its artists work and makes it more appealing to certain producers for sampling. Our paper deviates from this belief and analyzes influence through the traditional view of artists or genres as the creators of unique audio characteristics that make them more appealing for sampling.

A unique usage of networks was analyzed by Youngblood to analyze cultural transmission modes of music sampling in the modern era with the rise of the Internet. [22] Though the paper approaches the topic from a sociological perspective, there continues to be a focus on how sampling and how influence is spread across a network (i.e. how an artist or song becomes popular to sample). Youngblood specifically looks at the transmission of specific drum breaks through musical sampling in two different environments: Traditional cultural collaboration networks and online community networks with access to the collective knowledge of all members. The research looks at data through only three of the most (allegedly) popular sampled drum beats of all time and proceeds to analyze whether the music was sampled through a cultural collaboration network by noting time and distance. Based on the year the sample was made and the geographical location of the sampling

song's production studio, they attempted to classify which of the two environments this sampling was inspired by. Results found that sampling is less influenced by geographical location and prior collaboration between artists due to the rise of internet networks. The author claims this has led to an increase in social interactions between artists and therefore sampling in general in the modern era. Like Watson's work, this use of networks is very different from ours but it provides an alternative way to look at how influence can be derived from a song's sampling usage.

Rather than looking at influence through sampling, one can also look at influence through the number of collaborations a musician has with other musicians. Zinoviev specifically analyzes the success of musical groups as a function of the amount of collaboration between them and other musical groups as represented in a social network. [23] He hypothesizes that groups with greater public popularity benefit the most from cultural cross-pollination caused by performers moving between working on different projects and collaborating with a variety of artists. The findings suggest that average neighbors' degree affects the success of a musical group node in a graph indirectly. Zinoviev found weak, but statistically significant correlations between degrees and centrality in a collaborating network. Zinoviev also conclude that analyzing nodes with a higher centrality across multiple measures generally acted as a better predictor of success than simply making a prediction without these measures.

### **3. Approach**

In this paper, I begin by creating and analyzing a standard network graph connecting artist nodes by an edge representing an instance of a sample. This can be used to analyze the sheer volume of samples and helps us form sampling communities (k-connected subgraphs) at a basic level to help us analyze smaller groups of sampling communities. This approach allows for standard analysis of metrics such as finding the most sampled artist or song throughout history.

However, the unique method by which I am investigating musical sampling networks is through analyzing changing sampling patterns over time periods, the type of audio element sampled, as well as within genres (intra-genre) and between genres (inter-genre). This adds a unique approach giving

insight to the question of, “who samples what from which songs during which era”? The added element of analyzing the unique sampled audio elements of a song is something that has not been explored in past works.

In fact, the inclusion of audio elements is made possible by the updated WhoSampled dataset that other researchers have not had access to in the past as former pieces of research did not have this property within their datasets. Others have not tackled the subject with the focus on sampled audio elements nor sampling over time, but my approach is adequate for investigating previously unnoticed factors/properties of sampling patterns as the dataset supports the cataloging of these attributes.

## References

- [1] “Matplotlib: Python plotting.” Matplotlib. BSD Free License. [Online]. Available: <https://matplotlib.org/>
- [2] “Networkx: Software for complex networks.” NetworkX. BSD Free License. [Online]. Available: <https://networkx.github.io/>
- [3] “Whosampled.com: Sample-based music discovery database.” WhoSampled. [Online]. Available: <https://www.whosampled.com/>
- [4] M. G. Albán, V. Choksi, and S. B. Tsai, “Cs 224 w final report: Influence networks in popular music,” 2015.
- [5] V. D. Blondel *et al.*, “Fast unfolding of communities in large networks,” 2008.
- [6] M. Brandford, “Samples by the numbers a-data driven exploration of kanye west’s discography.” Available: <https://www.marcuscb.fyi/samples-by-the-numbers>
- [7] N. J. Bryan and G. Wang, “Musical influence network analysis and rank of sample-based music,” in *ISMIR*, 2011.
- [8] N. Collins, “Computational analysis of musical influence: A musicological case study using mir tools.” in *ISMIR*, 2010, pp. 177–182.
- [9] S. Collins, “M/c journal: “amen to that”,” in *M/C Journal*. Available: <http://journal.media-culture.org.au/0705/09-collins.php>
- [10] C. Dittmar *et al.*, “Audio forensics meets music information retrieval — a toolbox for inspection of music plagiarism,” in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 1249–1253.
- [11] D. F. Gleich, “Pagerank beyond the web,” *CoRR*, vol. abs/1407.5107, 2014. Available: <http://arxiv.org/abs/1407.5107>
- [12] G. Lawyer, “Understanding the influence of all nodes in a network,” *Scientific reports*, vol. 5, p. 8665, 2015.
- [13] J. C. Lena, “Meaning and membership: samples in rap music, 1979–1995,” *Poetics*, vol. 32, no. 3, pp. 297 – 310, 2004, music in Society: The Sociological Agenda. Available: <http://www.sciencedirect.com/science/article/pii/S0304422X04000294>
- [14] J. Leskovec, “Link analysis and ranking.” Stanford CS224W: Social and Information Network Analysis.
- [15] M. Peitz and P. Waelbroeck, “Why the music industry may gain from free downloading — the role of sampling,” *International Journal of Industrial Organization*, vol. 24, no. 5, pp. 907 – 913, 2006. Available: <http://www.sciencedirect.com/science/article/pii/S0167718705001682>
- [16] M. L. Somoano, “Bridgeport music, inc. v. dimension films: Has unlicensed digital sampling of copyrighted sound recordings come to an end?” *Berkeley Technology Law Journal*, vol. 21, no. 1, pp. 289–309, 2006. Available: <http://www.jstor.org/stable/24119549>
- [17] T. South, “Network analysis of the spotify artist collaboration graph,” *AMSI Vacation Research Scholarships*, 2018.
- [18] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, July 2002.
- [19] J. van Balen, M. Haro, and J. Serra, “Automatic identification of samples in hip hop music,” in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*, 2012, pp. 544–551.

- [20] T. Wang *et al.*, “Understanding graph sampling algorithms for social network analysis,” in *2011 31st International Conference on Distributed Computing Systems Workshops*, June 2011, pp. 123–128.
- [21] A. Watson, “The world according to itunes: mapping urban networks of music production,” *Global Networks*, vol. 12, no. 4, pp. 446–466, 2012.
- [22] M. Youngblood, “Cultural transmission modes of music sampling traditions remain stable despite delocalization in the digital age,” *PLOS ONE*, vol. 14, pp. 1–12, 02 2019. Available: <https://doi.org/10.1371/journal.pone.0211860>
- [23] D. Zinoviev, “Networks of music groups as success predictors,” *CoRR*, vol. abs/1709.08995, 2017. Available: <http://arxiv.org/abs/1709.08995>