

# QUANTIFYING DISRUPTIVE INFLUENCE IN THE ALLMUSIC GUIDE

Flavio Figueiredo

Universidade Federal de Minas Gerais  
flavioovdf@dcc.ufmg.br

Nazareno Andrade

Universidade Federal de Campina Grande  
nazareno@computacao.ufcg.edu.br

## ABSTRACT

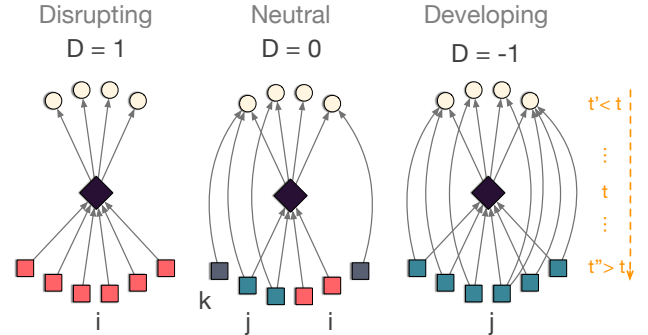
Understanding how influences shape musical creation provides rich insight into cultural trends. As such, there have been several efforts to create quantitative complex network methods that support the analysis of influence networks among artists in a music corpus. We contribute to this body of work by examining how disruption happens in a corpus about music influence from the All Music Guide. A disruptive artist is one that creates a new stream of influences; this artist builds on prior efforts but influences subsequent artists that do not build on the same prior efforts. We leverage methods devised to study disruption in Science and Technology and apply them to the context of music creation. Our results point that such methods identify innovative artists and that disruption is mostly uncorrelated with network centrality.

## 1. INTRODUCTION

*What is disruption?* To understand the concept, let us consider the careers of two famous Jazz musicians whose careers started in the 1940's: Bud Powell and Sun Ra. According to the *All Music Guide*<sup>1</sup>, both artists have been highly influential. The AllMusic biography of Bud Powell states that: "One of the giants of the jazz piano, Bud Powell changed the way that virtually all post-swing pianists play their instruments.". Similarly, the guide describes Sun Ra as a major innovator, both in his music and in his style: "[Sun Ra] surrounded his adventurous music with costumes and mythology that both looked backward toward ancient Egypt and forward into science fiction."

Disruption, as explored in this paper, is focused on differentiating artists like these even though they share similar backgrounds. While both artists may seem alike in terms of influence, the network of future artists citing them as a past influence sets these two great Jazz musicians apart by their network structure. Again according to the AllMusic guide, both artists have Thelonious Monk and Art Tatum as former influences. However, while future musicians influenced by Bud Powell also cite these two artists as influences, those following Sun Ra do not cite Thelonious

<sup>1</sup> <http://www.allmusic.org>. Quotations from June 16th 2019.



**Figure 1:** Reference networks of different focal works (diamonds), with their preceding references (circles) and posterior work (squares).  $i$  nodes (in red) reference the focal work but none of its predecessors;  $j$  nodes (in blue) reference both the focal node and its predecessors, and  $k$  nodes (in grey) link only to the focal node's references.

Monk and Art Tatum; they are mostly influenced by Sun Ra in isolation when compared to Sun Ra's past influences. In this sense, Sun Ra is primarily self-sufficient, and as a consequence, disruptive. In contrast, Bud Powell's contribution is more related to developing and consolidating an ongoing field of work. Figure 1 depicts this concept.

The figure shows an influence network and exemplifies disruption from the standpoint of the central, diamond-shaped, focal work. Links represent influence or citations. In this network, an overall influential innovator will be a node with a high in-degree, and both Bud Powell and Sun Ra fit this definition. In contrast, a disruptive node ( $D = 1$ ) is singled out and thus self-sufficient compared to its past. In our example, the focal node is cited by other nodes that tend to refer only to this single node as an influence.

To formalize the metric, let us call the focal, diamond-shaped node, as  $a$  (for artist). There are  $n_j$  nodes that reference  $a$ 's work and at least one of its predecessors, while  $n_i$  nodes reference  $a$  but none of its predecessors. There are also  $n_k$  nodes do not reference  $a$  but reference at least one of its predecessors. Funk and Owen-Smith's [5] disruption index, here called  $D$ , is measured as:

$$D = \frac{n_i - n_j}{n_i + n_j + n_k} \quad (1)$$

$D$  ranges from -1 to 1. The negative extreme captures a developing work, one that is mostly cited in conjunction with its own influences (i.e.,  $n_i = 0$  and  $n_k = 0$ ). The positive extreme captures disruption (i.e.,  $n_j = 0$  and  $n_k = 0$ ).

Motivated by recent demonstrations of the utility of  $D$  in other fields [5, 18], our work uses this index to examine disruptiveness of music artists using the AllMusic Guide. While several authors have tackled the task of understanding innovative artists, songs, and lyrics [2–4, 12, 15, 16], to the best of our knowledge, no prior effort exists that explores the disruptiveness of artists.

We opt to explore the AllMusic Guide as such a dataset has been used as a *gold standard* to other methods focused on influence [12, 16]. The guide contains a human-curated network of artists that influenced one another. This network provides us the contrasting example of Bud Powell and Sun Ra as a motivator to the importance of considering disruption. While both artists are influential, analyzing disruption unveils that one of them consolidates a stream of influences, while the other destabilizes this stream, shifting attention towards a different direction.

Our main contributions are (i) providing evidence that disruption measures provide insight when analyzing music corpora based on artist metadata, and (ii) describing disruption topologies that characterize how disruption happens in different contexts included in our corpus.

## 2. BACKGROUND AND RELATED WORK

Before presenting our dataset and results, we discuss previous work that explored different corpora to understand musical influence (Section 2.1). Next, we describe the metric that captures disruption (Section 2.2).

### 2.1 Influence Analysis in Music Corpora

Understanding musical influence is not a new endeavor [2–4, 12, 13, 15–17]. Several works have leveraged large datasets of audio and metadata to investigate historical trends in music creation quantitatively. Notably, Serrà and collaborators [15] use the Million Song Dataset to unveil historical regularities and changes in pitch transition, timbre usage, and loudness in pop music.

Several researchers have used the Billboard charts of songs most played on radios and streaming as a corpus representing western pop music. The audio and metadata about songs and artists in these charts have been used to characterize trends related to innovation, for example in lyrics [2], songs [16], and artists [16]. In particular, we point out the work of Mauch et al. [11] used timbral, tone, and harmonic information to analyze the sonic dynamics in the Billboard charts from 1960 to 2010. Their results point to three historical inflection points in the evolution of this corpus: 1964, 1983 and 1991.

A complementary approach to the analysis of aggregate trends is to examine individual artists or songs who have innovated in their context. Shalit et al. [16] use a dynamic topic model learned from audio and metadata to evaluate influence and innovation in songs from the Billboard charts. Their model identifies innovative songs and periods and suggests that overall, there is no correlation between how innovative and how influential a song is, with exceptions during the early 70s and mid-90s. For this analysis,

Shalit operationalizes innovative songs as songs hard to account for by a model trained with data from the past. On the other hand, a song is influential if its language is used by subsequent work. Authors have also studied influence for particular settings such as Electronic Dance Music [4].

### 2.2 Measuring Disruptive Influence/Innovation

Our work is inspired by the network measure proposed by Funk and Owen-Smith [5] to study technological change. Funk and Owen-Smith propose a network model and the  $D$  index, which "quantifies the extent to which an invention consolidates or destabilizes the subsequent use of the components on which it builds" [5]. The index (detailed in Eq (1)) is built on the notion that disruption should be measured by more than the number of references an invention has in subsequent work. Besides that, a measure of disruption must consider the structure of previous and subsequent work that form the context of the invention.

In their original work, the authors leverage a comprehensive database of patents to validate that their index quantifies how consolidating or destabilizing inventions are and that this information is uncorrelated with the sheer impact of innovations. More recently, Wu et al. [18] validate this same index in datasets of scientific papers and software products. Moreover, Wu et al. show that in the context of papers, software, and patents, disruptive innovation is more associated with smaller teams and work that cites prior efforts further in the past.

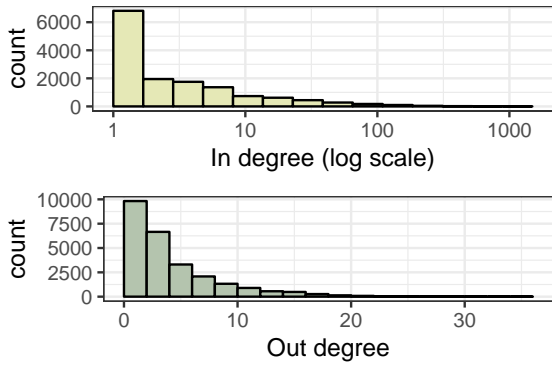
## 3. METHODOLOGY

To describe how we measure disruption in the AllMusic guide, we first detail how we identified artist pages and influence edges from the AllMusic website (Section 3.1).

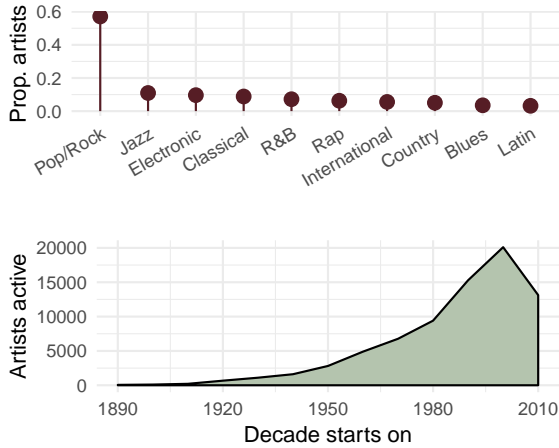
Different from other datasets where disruption has been measured [5, 18], the AllMusic guide presents a human-curated graph of influences. While from one perspective this is an advantage (e.g., provides explicit opinions of music editors), it is apparent that the information about influences is not complete, and it is likely less complete than in datasets created from patents and scientific papers. One other particularity of our network is that it widely suffers from biases towards modern occidental musicians who achieved considerable success. To tackle the disadvantages related to sparse counts, we employ a Bayesian approach to measure disruption (Section 3.2).

### 3.1 Crawling the AllMusic Guide

AllMusic is a comprehensive catalog of artists, albums and songs. The AllMusic website contains Artist Profile pages that detail an artist's biography, discography, genres, styles and links to other related artists, among other information. The list of related artists details those that are similar, have influenced, followed, or have worked with the owner of the profile page. Artists said to influence a given artist according to the site are those "that have had a direct musical influence on, or were an inspiration to, the selected artist,



**Figure 2:** Distribution of in and out node degree in the AllMusic influences dataset. For in degrees, 18,281 nodes with zero inbounding edges are omitted.



**Figure 3:** Proportion of artists with either in or out degree different from zero tagged with the 10 most popular genres, and number of artists active per decade.

as determined by our music editors<sup>2</sup>. Being a human-curated graph, there are several situations an artist influences a contemporary musician, band or singer.

We use AllMusic to create a network of influences among artists. Data were obtained by exhaustively crawling the website. The crawler started with a list of approximately 73,000 thousand AllMusic URLs present in the open MusicBrainz<sup>3</sup> database. We note that not all of these URLs were valid artist page addresses (e.g., some were wrong or deleted links). Nevertheless, we crawled the correct ones and followed their links in a snowball approach [7]. In particular, we followed links to every related artist until crawling we found no new artists. Even though we only use the influence edges in our analysis, to gather as much artists as possible, we crawled all of the similar, influenced, followed, member of egdes.

For each visited artist page, the crawler saves the artist’s name, decades of activity, genre, style, and list of influencers. The resulting set of artists from the crawler has 162,971 members, connected by 119,961 directed edges.

<sup>2</sup> <https://www.allmusic.com/faq/topic/influencedby>

<sup>3</sup> <https://musicbrainz.org/>

	Cited $a$	Did <b>not</b> cite $a$
Cited $a$ ’s influences	$n_j$	$n_k$
Did <b>not</b> cite $a$ ’s influences	$n_i$	everything else

**Table 1:** 2x2 Contingency Table used for Computing  $D$

After filtering out artists with no influencers cataloged in AllMusic, our dataset comprises of 32,568 artists connected by 119,961 directed links, where a link from artist  $a$  to  $b$  denotes that  $a$  has been influenced by  $b$ . When we consider the weakly connected components of the graph, 96% (31,279) of the nodes are in the giant component. This indicates that there is an undirected path of influence between most nodes. This is expected as major hubs, such as The Beatles, will lead to a mostly connected graph.

To characterize our graph, the in and out-degree distributions of nodes are shown in Figures 2. Complementary, in Figure 3 we show the genre distributions and active decades of the artists. Both the distribution of in and out degrees are skewed, with the distribution of in degrees being considerably more skewed and spanning a more extensive range. It is likely that influences (outgoing edges) of an artist are entered manually by editors and are kept to customary size. Bands or musicians influencing most artists are The Beatles (indegree 1,492), Bob Dylan (784), and The Rolling Stones (636), and there are 18,281 artists with no incoming edges. On the other hand, those with most extensive lists of influencers (outdegree) are Grateful Dead (36), Sonic Youth (35) and Jimi Hendrix (35). Concerning genre and epoch, our sample is biased towards Pop/Rock and artists active from the 80s to the present.

### 3.2 Measuring Disruption in Sparse Data

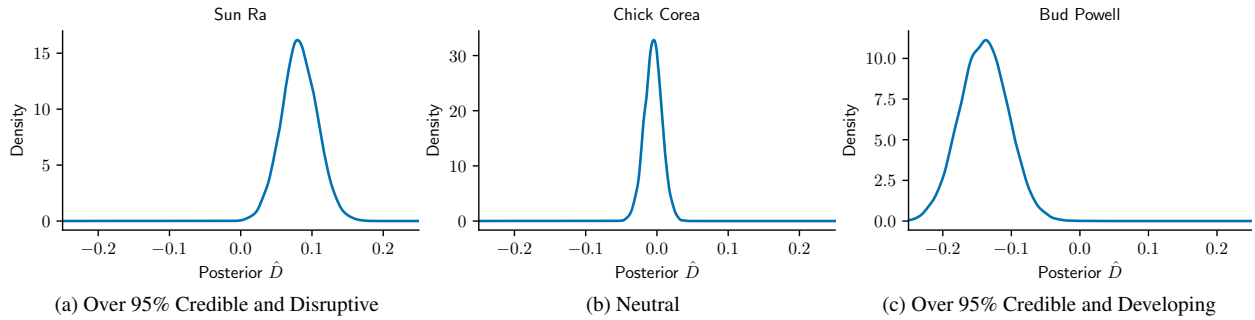
One challenge we tackle while computing disruption is how discuss results regarding  $D$  with statistical significance. Considering a focal artist  $a$ , recall that other artists may either: (1) cite  $a$  only, thus increasing  $n_i$ ; (2) cite  $a$  and  $a$ ’s past influences, thus increasing  $n_j$ ; (3) cite  $a$ ’s past influences only, increasing  $n_k$ ; or, (4) do cite  $a$  nor past influences. These choices are shown Table 1.

In some of our initial exploratory analysis, we computed disruption in the AllMusic influence graph *as is* (i.e., with no filters nor priors), and found that Eq. (1) would lead to either very high ( $D \approx 1$ ) or very low ( $D \approx -1$ ) scores when  $n_i$ ,  $n_j$  and  $n_k$  were very small. One example is the extreme case where  $n_i = 1$ ,  $n_j = n_k = 0$ . Here, the metric will unveil a biased  $D$  due to the small numbers.

To explain our Bayesian approach, initially note that Eq. (1) captures the difference between two proportions:

$$D = p_i - p_j = \frac{n_i}{n_i + n_j + n_k} - \frac{n_j}{n_i + n_j + n_k} \quad (2)$$

Here  $p_i$ ,  $p_j$  and  $p_k$  (unused) are proportions. Furthermore, the counts may be captured by a Multinomial distribution  $I, J, K \sim \text{Multinomial}(p_i, p_j, p_k, n)$ , where  $n = n_i + n_k + n_j$ . Here,  $I$ ,  $J$  and  $K$  are random variables modelling the respective counts  $n_i$ ,  $n_j$  and  $n_k$ .



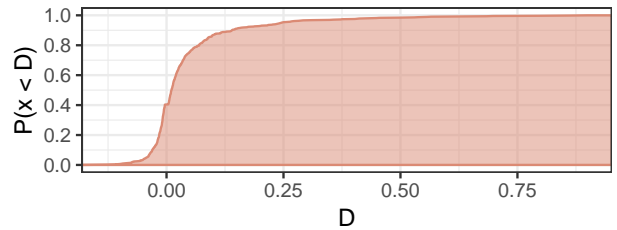
**Figure 4:** Posterior  $\hat{D}$  for three Jazz Artists. On the left we have a disruptive where 95% of posterior samples are above zero; in the middle a neutral artist; on the right a developing artist where 95% of posterior samples are below 0.

To model  $D$ , one approach would be to use the closed form for the probability mass function for  $D = I - J$ . One negative aspect of this approach is that it also requires some assumption on the joint distribution of  $I$  and  $J$  or that both are independent. Another approach would be to use statistical tests (see [1] for details). Here, classical approaches like the Binomial test for proportions have issues with small samples or assume independence in the choices that lead to the contingency table. Other options such as Fisher’s or McNemar’s test focus on comparing either rows/columns or the off-diagonal of the  $2 \times 2$  table. In our setting,  $n_j$  shares a column with  $n_i$  (see Table 1). Our Bayesian approach, discussed next, has the advantage that it does not require such closed forms or assumptions.

Given that proportions captured by a Multinomial distribution, for each node of the graph, we can apply a conjugate Dirichlet [6] prior on such a Multinomial distribution. Being a conjugate prior, the posterior will also be a Dirichlet distribution from which we can sample proportions:  $\hat{p}_i, \hat{p}_j, \hat{p}_k \sim \text{Dirichlet}(n_i + \alpha_i, n_j + \alpha_j, n_k + \alpha_k)$ . Here,  $\alpha_i, \alpha_k$ , and  $\alpha_j$  are prior hyper-parameters. These can be fine tuned by an analyst to capture prior beliefs. By sampling from this distribution, we are left with a posterior estimate of disruption that is defined as:  $\hat{D} = \hat{p}_i - \hat{p}_j$ .

Suppose we perform 10,000 of such samples. Let,  $\hat{D}$  be the vector of estimates. The average score of this vector will lead to similar results as the original one ( $\text{mean}(\hat{D}) \approx D$ ). However, using these samples, we are able to measure the *credibility* of our estimates [6]. This credibility comes from what is called the *credible-interval*, a Bayesian analogous of the confidence-interval. While a confidence-interval measures the probability that some true population statistic will fall into the range of the interval, the credible-interval is determined by the posterior samples.

To explain how we capture credibility, consider the case where  $n_i = 1$  and  $n_j = n_k = 0$ . Moreover, consider the particular choice priors (discussed later),  $\alpha_i = \alpha_k = \alpha_j = 10$ . If we measure the fraction of posterior samples greater than zero  $P(\hat{D} > 0)$ , this value is only of 0.58, even if  $p_i = 1$  and  $p_k = 0$ . Thus, our estimate is 58% credible. Credibility is thus captured by: (1)  $P(\hat{D} > 0)$  when  $D > 0$ ; and, (2)  $P(\hat{D} < 0)$  when  $D < 0$ . In other words, simply the fraction of posterior samples when  $D$  that are



**Figure 5:** CDF of the disruption  $D$  for artists where our estimation has a minimum confidence of 0.95.

either positive or negative depending on the value of  $D$ . If this fraction is above 0.95, we are over 95% credible for either the positive (disruptive) or negative (developing) case. Figure 4 shows three examples of posterior samples of  $\hat{D}$  for three artists, illustrating the cases when of disruption and development, and neutrality.

We set our hyper-parameters to the non-informative case of  $\alpha_i = \alpha_j = \alpha_k = 10$ . This choice is based on synthetic samples, where we estimated our credibility scores for different values of  $n_i - n_j$  and  $n_j - n_i$ . Via simulations using different values of  $n_i, n_j$  and  $n_k$ , we found that with these our choices credibility is over 95% only when  $|n_i - n_j| \approx 10$ . Moreover, one can notice that these priors do not bias results towards positive nor negative values of  $\hat{D}$  (i.e.,  $\alpha_i = \alpha_j$ ). We argue that this is adequate as it imposes a minimum difference between  $n_i$  and  $n_j$  to have some credibility in our estimates. Furthermore, we present disruption values  $D$  only in cases where our credibility is over 95%, using 10,000 samples per node.

Finally, being a human-curated guide, some artists will suffer from missing given the limited knowledge of the AllMusic editors. To avoid discussing such cases, we limit the analysis to nodes with at least three incoming and outgoing edges *after disruption is computed*.

#### 4. DISRUPTIVE ARTISTS

We now explore the most and least disruptive artists in our data. Figure 5 shows the distribution of disruptiveness  $D$ . This distribution is concentrated around a median value close to zero (0.01), with a longer right tail – there are more highly destabilizing artists than highly consolidating ones.

Artist	D	AM Genre	$n_i$	$n_j + n_k$
King Sunny Ade	0.90	International	9	1
Édith Piaf	0.86	Vocal	51	6
Frankie Knuckles	0.70	Electronic	30	13
The Clark Sisters	0.68	Religious	13	6
Mstislav Rostropovich	0.68	Classical	13	6
Los Tigres del Norte	0.57	Latin	14	9
John Cage	0.56	Classical	141	64
Tiësto	0.45	Electronic	10	12
Alfred Brendel	0.45	Classical	13	16
Scott Asheton	0.41	Pop/Rock	16	18
Bernard Herrmann	0.41	Stage & Screen	23	28
Converge	0.40	Pop/Rock	32	45
K.M.D.	0.38	Rap	9	15
Too Short	0.36	Rap	70	104
Darkthrone	0.35	Pop/Rock	15	28

**Table 2:** Most disruptive artists in AllMusic with at least three influences catalogued.

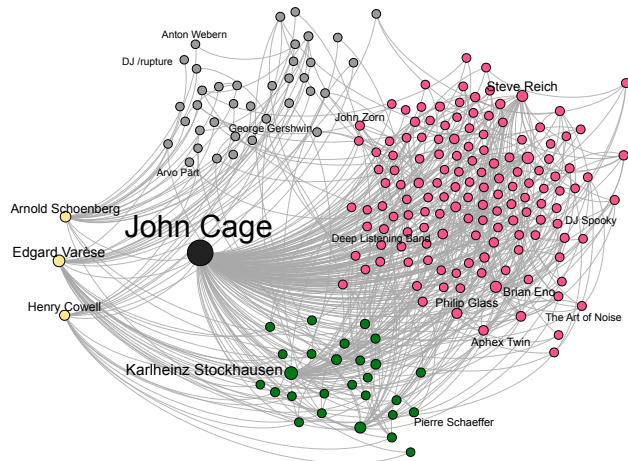
Artist	D	AM Genre	$n_i$	$n_j + n_k$
Teddy Wilson	-0.13	Jazz	7	163
John Coltrane	-0.11	Jazz	103	1307
Bud Powell	-0.11	Jazz	23	358
Philly Joe Jones	-0.10	Jazz	8	121
Geto Boys	-0.10	Rap	33	474
Sarah Vaughan	-0.09	Jazz	23	495
Pete Seeger	-0.09	Folk	22	335
Sonny Rollins	-0.08	Jazz	31	641
The Stanley Brothers	-0.08	Country	11	308
Augustus Pablo	-0.08	Reggae	1	212
Buddy Guy	-0.07	Blues	8	611
Roy Acuff	-0.07	Country	25	204
Jimmy Reed	-0.07	Blues	32	515
Oscar Peterson	-0.07	Jazz	15	411
Master P	-0.07	Rap	12	530

**Table 3:** Most consolidating artists in AllMusic with at least three influences catalogued.

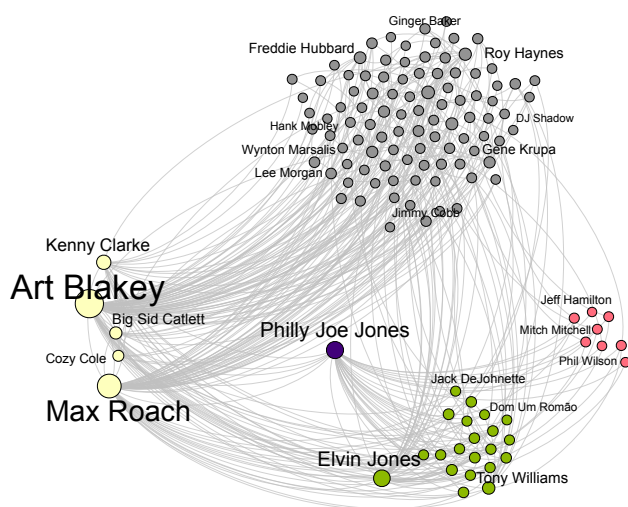
Overall, we also find disruption does not correlate with the influence of an artist. This measure is captured here by the number of other artists influenced by a certain artist (in-degree). This was measured using the linear correlation coefficient is  $\rho = -.001$ . Correlations between disruption and centrality are further detailed in the next section.

Next, in Table 2 show the 15 most disruptive in our dataset. None of the most disruptive artists are among the most influential of AllMusic, and their communities in the network are very diverse. For example, King Sunny Ade is a Nigerian musician credited for a significant contribution in the popularization of juju music worldwide. According to AllMusic, he has been influenced by other juju and highlife artists such as I. K. Dairo & His Blue Spots and Rex Lawson, while he has influenced artists from a diverse stream including Talking Heads and Trey Anastasio, both from the US. Édith Piaf is a famous French singer with a similar network structure, bridging influences from earlier French singers and an assorted group of followers spanning several decades and countries.

The remaining artists in Table 2 illustrate multiple other types of destabilizing innovation in various genres. For example, Frankie Knuckles and Too \$hort are acknowledged as pioneers of house music and gangsta rap, respectively.



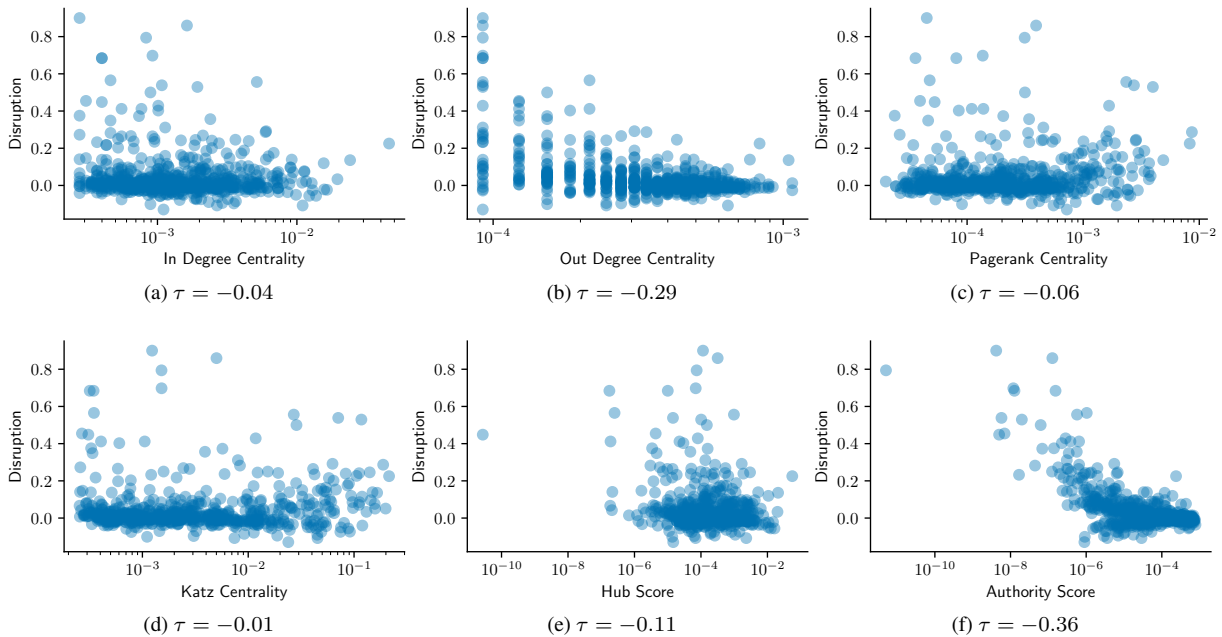
**Figure 6:** Network with John Cage as the focal node. Preceding artists are in yellow, while  $i$ ,  $j$  and  $k$  nodes are pink, green and grey, resp.



**Figure 7:** Network with Philly Joe Jones as the focal node. Preceding artists are in yellow, while  $i$ ,  $j$  and  $k$  nodes are pink, green and grey, resp.

The most influential artist in this list is John Cage, a highly inventive composer who, according to AllMusic, has influenced generations of composers, writers, dancers, and visual artists. Figure 6 shows how the three classical composers catalogued as influences of John Cage are not influences of many of his followers.

On the opposite side of the spectrum of disruption, Table 3 lists the 15 most consolidating artists, according to  $D$ . There is a predominance of jazz artists, who are 6 of the ten most consolidating artists. Jazz instrumentalists and singers whose career started after the 1930s are often characterized by our method as consolidators building on a stream of shared influences. These artists share a set of influences with many others. Figure 7 illustrates one such case for the jazz drummer Philly Joe Jones, who shares the influences of Art Blakey and Max Roach with a large number of other jazz drummers, including most subsequent work. Geto Boys and Augustus Pablo have a similar neighborhood structure in the Rap and Reggae genres.



**Figure 8:** Correlation of Disruption with six node importance measures.

Overall, the examination of artists identified as most and least destabilizing points to the expressiveness of this method. This approach highlights artists who have not necessarily influenced a large number of other artists, so this information is not readily available based on a direct quantification of impact. Moreover, our face validity analysis promptly links high valuations of disruptiveness with widely known stories of innovation. Most interestingly, these stories are diverse in their geography, genre, and epoch, even if mined from a considerably biased dataset.

## 5. DISRUPTION VS CENTRALITY

We now investigate to what degree the disruption scores provides information that is not already available through other metrics of network topology. To do so, we measure six different node importance scores using our full graph: the Indegree Centrality (normalized indegree), Outdegree Centrality (normalized outdegree), Pagerank [14], Katz [8] Centrality, Hub Scores [10], and Authority Scores [10].

We correlated each score with the disruption index of artists. Figure 8 presents the relation of each score with  $D$  together with Kendall’s rank correlation ( $\tau$ ) for each case. We resort to Kendall’s coefficient,  $\tau$ , as it addresses ties (e.g., nodes with the same in degree) and is able to uncover both linear as well as non-linear relationships [9].

The patterns in the figure and  $\tau$  scores point that most metrics do not correlate with disruptiveness. The only cases where moderate negative correlations were uncovered were: Out degree centrality ( $\tau = -0.29$ ) and Authority score ( $\tau = -0.36$ ). A small negative relation also exists for Hub scores ( $\tau = -0.11$ ). Nevertheless, these scores are moderate at best. Such a result is relevant, as it shows that disruptiveness values are not easy to recover using standard node scores from complex networks.

In summary, our results in this section combined with our discussion in the previous section, point to the relevance of measuring disruptiveness. The  $D$  index unveils insightful patterns on the AllMusic corpus that are not trivially explained by other network centrality scores.

## 6. DISCUSSION AND FUTURE WORK

In this paper, we present the first in-depth analysis of disruption in a music corpus. More importantly, we argue in favor of a Bayesian disruption index. While our analysis is limited to a dataset, the approach we here discuss is general enough to be used in other settings. In particular, we note data analysts may tune choices of hyper-parameters to their prior-beliefs for different datasets.

Our contributions bring two main implications. First, our examination of the validity of the disruption index suggests it can be applied to music corpora. Second, our analysis of disruptive artists in the AllMusic guide shows new information about artists in this guide that may be taken into account in musicological analyses.

At the same time, both of these directions call for relevant future work. In particular, further validation of the disruption index with other contexts seems very relevant, to understand its applicability to other musical traditions or to networks formed by albums or songs, for example. In complement to this direction, musicological analyses that use disruption to provide deeper insight leveraging this data is necessary to further validate the usefulness of the approach we here presented.

**Reproducibility:** We point out that our source code for data collection, analysis and for the figures and tables in this paper, as well as the dataset used, is available at: <http://github.com/flavioovdf/allmusic-disruption>.

**Acknowledgements:** This work has been partially supported by the project ATMOSPHERE (atmosphere-eubrazil.eu), funded by the Brazilian Ministry of Science, Technology and Innovation (Project 51119 - MCTI/RNP 4th Coordinated Call) and by the European Commission under the Cooperation Programme, Horizon 2020 grant agreement no 777154. Funding was also provided by the authors' individual grants from Capes and CNPq. In particular, CNPQ's Universal 2018 Grant: 421884/2018-5.

## 7. REFERENCES

- [1] Alan Agresti. *An introduction to categorical data analysis*. Wiley, 2018.
- [2] Jack Atherton and Blair Kaneshiro. I said it first: Topological analysis of lyrical influence networks. In *ISMIR*, 2016.
- [3] Nicholas J Bryan and Gen Wang. Musical influence network analysis and rank of sample-based music. In *ISMIR*, 2011.
- [4] Nick Collins. Influence in early electronic dance music: An audio content analysis investigation. In *ISMIR*, 2012.
- [5] Russell J Funk and Jason Owen-Smith. A dynamic network measure of technological change. *Management Science*, 63(3):791–817, 2016.
- [6] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [7] Leo A Goodman. Snowball sampling. *The annals of mathematical statistics*, pages 148–170, 1961.
- [8] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
- [9] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
- [10] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5), 1999.
- [11] Matthias Mauch, Robert M MacCallum, Mark Levy, and Armand M Leroi. The evolution of popular music: Usa 1960–2010. *Royal Society open science*, 2(5), 2015.
- [12] Brandon G Morton and Youngmoo E Kim. Acoustic features for recognizing musical artist influence. In *ICMLA*, 2015.
- [13] Erik Noyes, I Elaine Allen, and Salvatore Parise. Artistic influences and innovation in the popular music industry. *Frontiers of Entrepreneurship Research*, 30(15), 2010.
- [14] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
- [15] Joan Serrà, Alvaro Corral, Marián Boguñá, Martin Haro, and Josep Lluís Arcos. Measuring the evolution of contemporary western popular music. *Scientific reports*, 2, 05 2012.
- [16] Uri Shalit, Daphna Weinshall, and Gal Chechik. Modeling musical influence with topic models. In *ICML*, 2013.
- [17] Jay Summach. The structure, function, and genesis of the prechorus. *Music Theory Online*, 17(3), 2011.
- [18] Lingfei Wu, Dashun Wang, and James A Evans. Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744):378, 2019.