frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 21 January 2025
DOI 10.3389/fcomp.2024.1476996

OPEN ACCESS

EDITED BY
Rolf Bader,
University of Hamburg, Germany

REVIEWED BY
Antonio Rodà,
University of Padua, Italy
Quynh Quang Ngo,
Universität Stuttgart, Germany

*CORRESPONDENCE
Daniel Gómez-Marín
✉ dgomez@icesi.edu.co;
✉ daniel.gomez@upf.edu

RECEIVED 06 August 2024
ACCEPTED 16 December 2024
PUBLISHED 21 January 2025

CITATION
Gómez-Marín D, Jordà S and Herrera P (2025) Network representations of drum sequences for classification and generation.
Front. Comput. Sci. 6:1476996.
doi: 10.3389/fcomp.2024.1476996

Network representations of drum sequences for classification and generation

Daniel Gómez-Marín[1,2]*, Sergi Jordà[2] and Perfecto Herrera[2]

[1]Facultad de Ingeniería, Diseño y Ciencias Aplicadas, Universidad Icesi, Cali, Colombia, [2]Music Technology Group (MTG), Universitat Pompeu Fabra, Barcelona, Spain

Complex networks have emerged as a powerful framework for understanding and analyzing musical compositions, revealing underlying structures and dynamics that may not be immediately apparent. This article explores the application of complex network representations to the study of symbolic drum sequences, a topic that has received limited attention in the literature. The proposed methodology involves encoding drum rhythms as directed, weighted complex networks, where nodes represent drum events, and edges capture the temporal succession of these events. This network-based representation allows for the analysis of similarities between different drumming styles, as well as the generation of novel drum patterns. Through a series of experiments, we demonstrate the effectiveness of this approach. First, we show that the complex network representation can accurately classify drum patterns into their respective musical styles, even with a limited number of training samples. Second, we present a generative model based on Markov chains operating on the network structure, which is able to produce new drum patterns that retain the essential features of the training data. Finally, we validate the perceptual relevance of the generated patterns through listening tests, where participants are unable to distinguish the generated patterns from the original ones, suggesting that the network-based representation effectively captures the underlying characteristics of different drumming styles. The findings of this study have significant implications for music research, genre classification, and generative music applications, highlighting the potential of complex networks to provide a transparent and elegant approach to the analysis and synthesis of rhythmic structures in music.

KEYWORDS

complex networks, music, symbolic drum patterns, network similarity, genre classification, music generation, music information representation

1 Introduction

Complex networks have emerged as a powerful framework for understanding and analyzing a wide array of phenomena across diverse fields, including biology (Wild et al., 2021), social sciences (Matta et al., 2018; Óskarsdóttir et al., 2022), and technology (Kim and Sayama, 2017). Their ability to represent complex relationships and interactions through nodes and edges allows researchers to uncover underlying structures and dynamics that may not be immediately apparent. By modeling systems as networks, we can leverage advanced analytical techniques to address intricate problems, identify patterns, and facilitate the generation of new insights. This versatility makes complex networks an invaluable tool for exploring not only the intricacies of musical compositions but also the broader implications of rhythm and style across different musical genres.

There are several recent examples of how music scores can be modeled as complex networks, representing notes as nodes, and the temporal sequence as connections between these nodes. In a recent paper (Kulkarni et al., 2024), it has been shown how the study of the topology of networks created from Johann Sebastian Bach's pieces reveals the underlying organization of his music. On the one hand, the authors demonstrate that Bach's compositions exhibit a balance between complexity and simplicity that allows for efficient communication of musical information. On the other hand, by using metrics such as entropy and connectivity degree, these networks can be grouped according to their form (e.g., separating choral pieces from preludes and fugues).

This is crucial because it suggests a powerful methodology to approach the study of musical pieces. Moreover, it suggests that networks can be useful not only for comparing musical styles but also for storing information about a musical style and generating new pieces. There are examples (Liu et al., 2010; Ferretti, 2017, 2018a,b) that present the use of complex networks as mechanisms for musical analysis, encoding pieces, and generating new pieces. Our purpose is to apply these ideas and techniques, using a novel complex network representation, to the study of symbolic drum sequences, a topic that, to the best of our knowledge, has not been extensively addressed before. Our motivation is to present a simple and transparent representation of musical drum sequences and demonstrate their ability to solve problems such as musical information classification and generation.

Despite the recent use of musical representations with complex networks, the study of drum sequences through complex networks is scarce. This context interests us especially because we recognize that rhythm in general, and drum patterns in particular, have unprecedented importance in contemporary popular music, and they are present in almost all genres. We hope that the learning achieved through this research can nourish the growing knowledge of this musical phenomenon and inform future researchers and creators of musical rhythms.

This paper focuses on how to achieve style classification and rhythm generation using a complex network representation of drum sequences. With this purpose in mind, we will use two databases: one created by invited music producers for this project and another open database published by Google Magenta, known as the Groove MIDI Dataset (GMD) (Gillick et al., 2019). This paper covers four main topics: First, the proposal of a methodology for encoding drum rhythms as complex networks. Second, the study of similarities between different drumming styles through their complex networks. Third, the generation of new drum rhythms based on this representation. And finally the evaluation of our generation method with listening tests.

Our scientific research aims to advance knowledge about drum patterns through the study of network representations. We compare the advantages and disadvantages of our representation with those discussed in the literature, highlighting its computational benefits that may be relevant from various research perspectives. Additionally, we emphasize the educational value of network representations, which can be beneficial across multiple disciplines, such as cognitive science, computer science, and musicology. We appreciate the importance of both cutting-edge solutions and foundational discussions on representations and algorithms, recognizing the unique insights that network approaches can provide.

The following sections of this paper are divided as follows: In Section 2, we introduce drum sequences represented symbolically as scores (in contrast to their instantiation as sounds) and the terminology used to describe them. Then, in Section 3, we explain our methodology for processing drum sequence files and converting them into complex networks. In Section 4, we present an experiment where we convert all drum sequences in different musical styles present in the GMD dataset into complex networks, use similarity metrics to compare them (Jaccard similarity and degree), and then compare the different styles. In Section 5, we generate new patterns based on the representation of their styles as complex networks and compare the generated patterns with the studied ones, in order to validate the imitational power of this approach. We conclude the experiments in Section 6, where we present generated patterns to subjects in a listening test and ask them how similar the generated patterns are to the originals, thus validating the effectiveness of our modeling in a perceptual way. Finally, Section 7 offers a general discussion of the observed results and what our methodology and findings can offer beyond our experiments with drum sequences.

2 Symbolic drum sequences

Percussion arrangements, alongside vocal elements, have become pivotal in shaping the soundscape of global popular culture. However, their acceptance, particularly in Western contexts, has historically lagged behind that of melody and harmony (Nettl, 2005, p. 151). The evolution of global culture, driven by technological advancements and diverse musical influences, has formalized percussion sets that blend instruments from various cultural backgrounds (Dean, 2012, p. 3). This evolution culminated in the standardization of drum kits, typically comprising the essential kick drum, snare drum and hi-hat set with musicians often augmenting these kit with elements from broader percussion ensembles (Brennan, 2020).

The late 20th century saw the advent of rhythm boxes and drum machines, which leverage digital technology and the MIDI protocol to record, edit and reproduce digital scores, synthesize sounds, play pre-recorded samples, and craft intricate rhythmic patterns. The standardization of the MIDI protocol, particularly the General MIDI percussion key map (GMPKM),[1] revolutionized music composition by allowing musicians to create complex sequences and explore diverse genres with unprecedented flexibility (Loy, 1985). These advancements have not only defined musical genres through distinctive drum patterns–such as hip-hop, breakbeat or drum and bass–but have also exponentially expanded the realm of musical sub-genres. While MIDI provides a digital representation of musical scores, specifying instruments and notes without acoustic representation, it serves as a powerful tool for capturing musical ideas and facilitating the exploration of diverse musical expressions, ultimately shaping the landscape of contemporary music production and composition.

[1] https://www.midi.org/specifications-old/item/gm-level-1-sound-set

To analyze percussion arrangements, we use four distinct instrument sets based on their size and complexity:

Comprehensive 46-instrument set: this set includes all instruments in the GMPKM, summarized as 40 distinct instruments by avoiding duplicates (e.g., two kick drums, two snares, and six tom categories). Most software-based drum machines use this set.

16-instrument set: a streamlined collection of electronic and digital instruments, commonly found in classic drum machines and samplers, featuring up to 16 drum sounds (e.g., Roland TR-808, Roland TR-909, Akai MPC Live II, Elektron Machinedrum). Instruments include kick drum, snare, rimshot, clap, cowbell, closed hi-hat, open hi-hat, crash cymbal, low tom, mid tom, hi tom, low conga, mid conga, high conga, maracas and claves.

8-instrument set: this highly condensed set matches the compact size of some portable electronic drum kits and smaller drum machines (e.g., Roland TR-606, Roland TR-505, Elektron Digitakt, Arturia DrumBrute Impact, Nord Drum 3, or Teenage Engineering PO-32 Tonic). It includes kick drum, snare, rimshot, clap, open hi-hat, closed hi-hat, low tom and hi tom.

Minimal 3-Instrument Set: This represents the smallest configuration, encompassing only the kick drum, snare drum and hi-hat.

Symbolic drum sequences in the digital domain are thus expected to utilize one of these four instrumental drum sets. Given the nested relationship among these sets, we have devised an instrument mapping to convert a larger set into a smaller one, allowing for the expression of drum sequences created with a larger set using a simpler set. This mapping ensures that drum sequences created with different instrument sets can be expressed uniformly.

In MIDI, each note is accompanied by a velocity value and a time frame, known as a "tick." The velocity, coded as an integer from 0 to 127, indicates the intensity of the note, while ticks denote the position of the note with a standard resolution of at least 24 ticks per beat (TPB). Note durations in the MIDI standard are indicated by the time between a "note on" message (which includes a note-and-velocity pair associated with a tick) and a "note off" message (which indicates a note with zero velocity and a greater tick value).

For this research, drum sequences are pre-processed in terms of time and velocity to facilitate network creation. Velocity information is disregarded, focusing solely on notes with a velocity greater than 0, thus eliminating accent information. Temporal information is quantized, adjusting every note onset to a precise 16th note and disregarding note duration. Only the quantized starting times of the notes are considered, aligning with previous research in network-based music analysis (Kulkarni et al., 2024).

3 Related work

This paper focuses on the classification and generation of symbolic drum sequences using complex networks. In this section, we present related investigations that contextualize our research and establish reference values for discussing our own results.

3.1 Classification

The latest research on classification using the Groove MIDI dataset (GMD) is by Géré et al. (2024). The authors explore how to classify a subset of the GMD that includes funk, jazz, Latin, and rock music styles using two models: one based on long-short-term memory (LSTM) networks and the other based on transformers. Their main contribution is the introduction and evaluation of a symbolic music representation called the linearized rhythmic tree (LRT) and a variation known as tree-based positional encoding (TBPE), alongside common representations such as piano rolls, note tuples, and tokenization. LRTs are created by dividing MIDI data into measures, converting each measure into a rhythmic tree with notes as leaves, and then linearizing the tree. The LRT representation is a 3D matrix that is smaller than the typical piano roll representation (which is a 3D matrix of 16 steps x number of instruments x velocity) and contains more rhythmic information than the note-velocity tuples in the raw MIDI sequence. TBPEs are built on LRTs to recover the hierarchical structure of the sequence that is lost during linearization.

Both LRT and TBPE representations are valid alternatives for the models used by Géré et al. (2024). The LSTM architecture processes time-ordered sequences, while the transformer architecture requires explicit time representation. In contrast, our network representation for drum sequences maintains the temporal structure of the sequence but ignores velocity information, allowing for a more time-structured and compressed representation than LRT and TBPE.

The process described by Géré et al. (2024) involved the division of the data into three sets: 80% for training, 10% for validation, and 10% for testing. Their best F1 scores were 0.663 for the LSTM model with the piano roll representation and 0.660 for the transformer model with the TBPE representation. The authors noted that the classification accuracy, measured by the F1 scores, decreased for both models when trained with less than 80% of the data. In contrast, our classification results, detailed in Section 6, show F1 scores of 1 using 17 styles from the GMD as classes.

3.2 Generation

Research on generating symbolic drum sequences has mostly treated drums as secondary to other musical elements (e.g. Hutchings, 2017; Dahale et al., 2021; Haki et al., 2022). However, some studies focus solely on drum generation. For instance, Choi et al. (2016) used a long short-term memory (LSTM) neural network to create rock drum patterns from 60 training tracks. They reported reasonable outputs but did not provide formal evaluations. In their approach, they encoded drum sequences as token sequences, with each token representing the simultaneous drum sounds played at each 16th note, limited to 9 instruments. This representation is similar to the one we will present later, as it quantizes notes and describes simultaneous drum hits as a single event. However, our representation allows for greater flexibility in the number of instruments that can be included.

In another study, Wei et al. (2019) used a variational autoencoder (VAE) and achieved 92% cosine similarity with the training set, along with a subjective evaluation where their best model was preferred 58% of the time. They represented a one-bar drum sequence as a 46×16 binary matrix. Similarly, Makris et al. (2019) also employed LSTM for pattern generation, reporting mean similarity scores ranging from 57% to 68% in six models. Their representation is similar to that of Choi et al. (2016).

In general, the tokenization of simultaneous drum events proposed by Choi et al. (2016) and Makris et al. (2019) simplifies polyphony into a single event, normalizing the timing of simultaneous instruments to one time point and the intensity of each instrument hit to a single value. We also apply these procedures in our data preparation to create our network representation (see Section 4). Although both timing and intensity carry expressive musical information, as shown in Experiment 1, suppressing these dimensions does not hinder the accurate classification of drum sequence styles. Additionally, new drum sequences generated under these constraints can achieve high accuracy, as demonstrated in Experiment 2. In general, these generative studies establish a baseline range for the highest reported scores, between 68% and 92%. This range will help contextualize our results, which are presented below in Section 7.

4 Complex network representation of drum sequences

Here we describe our methodology to represent a drum sequence as a network. Three essential variables are taken into account: the percussion instrument set of the symbolic drum sequence (i.e. the kick, snare and hihat set in Figure 1 top left), the steps defining the minimum temporal resolution of the sequence, and the maximum number of steps within the network. The first two variables are pre-established in the symbolic sequence to streamline its representation as a network (see step indexes 0 to 15 in patterns A, B and the network representation in Figure 1). Notably, the number of steps can be customized as needed. For this research, we define the length to correspond to one bar, comprising a grouping of 16 steps.

Formally, a drum sequence can be represented as a complex directed weighted network comprising nodes and edges. In this context, a weighted network G is defined as an ordered pair $G = (V, W)$, where V represents the set of nodes (or vertices), and W is a weighting function that assigns a real nonnegative value $w(vi, vj)$ to each connected node pair ($vi\ vj$). These connected node pairs, typically referred to as edges, satisfy the conditions $vi \in V$, $vj \in V$, and $vi \neq vj$ (Umeyama, 1988).

Within the network, nodes symbolize drum events, ranging from silence (no instruments) to the occurrence of one or multiple instruments (e.g., a single low tom or a combination of a low tom and a closed hihat) to the simultaneous playing of all instruments. The total number of nodes in the network exhibits an exponential relationship with the instrument set designated for reproducing a drum sequence, calculated as $nodes = 2^{instruments}$.

Network edges represent the step-wise temporal succession of events. The weight of each edge in the network is calculated as the normalized frequency of the connection between the corresponding pair of nodes. Specifically, the weight of an edge is proportional to the number of times that connection was observed, with the weights for all edges connected to a given node summing up to 1. This normalization ensures that the edge weights represent the relative importance or likelihood of each connection, rather than just the raw counts. By scaling the weights so that they sum to 1 for each node, we can better compare the connection patterns across different nodes and networks.

Given the quantization of drum sequences in time (as discussed in the previous section), each drum event and its corresponding node in the network are associated with a specific time step, where steps indicate the events occurring every 16th note. Every step encompasses the complete set of potential nodes at a particular point in time. The array of potential nodes at a given step is referred to as a layer.

In Figure 1, a two-bar drum sequence is illustrated, featuring two distinct patterns labeled as A and B at the top, and involving three instruments: kick, snare and hihat. The depiction emphasizes the steps where instrumental onsets occur. Notably, both patterns exhibit equal instrument occurrences at steps 1, 2, 3, 5, 7, 8, 11, 12, 14, and 15, while differing at steps 0, 4, 6, 9, 10, and 13. A network representing the drum sequence, incorporating patterns A and B, is constructed by sequentially connecting drum events, as depicted in the lower part of Figure 1. Our drum network models are designed to encompass 16 steps, corresponding to one bar divided into 16 steps. Consequently, longer sequences, such as the one formed by patterns A and B, are segmented every 16 steps, with nodes connected sequentially. Nodes with a higher in-degree (representing nodes common to both patterns) are distinguished by a larger radius in Figure 1. Furthermore, nodes at steps 3, 5, 8, 12, and 15 exhibit edges with 0.5 out-degree weights, indicating that they are followed by two equally probable yet distinct nodes.

Thus, any drum sequence can be effectively represented as a directed, weighted network, with the resulting topology - comprising nodes, edges, and associated weights - reflecting the characteristics of the original sequence. Once we have set this type of representation, it is possible to study the similarity, relatedness and reproducibility of symbolic drum sequences. Subsequent sections will delve into three distinct experiments aimed at evaluating theses concepts.

5 Network similarity metrics

As we aim to introduce a novel method for representing drum sequences as complex networks, it is paramount to establish appropriate techniques for analyzing the resulting networks. Network similarity can be examined either by structural comparison or by the extraction of network descriptors followed by a comparison of these descriptors. Structural comparison generally involves assessing differences between two networks weighting these differences by their significance (Coscia, 2021). This approach is viable for our purposes, given that the networks are weighted directed networks with known node correspondence (drum sequences using the same drum set are represented by networks that potentially share identical nodes).

One powerful structural similarity metric is the Jaccard similarity index. The Jaccard index is computed for two networks created with the same instrument set: the number of coincident edges in the two networks is divided by the total number of

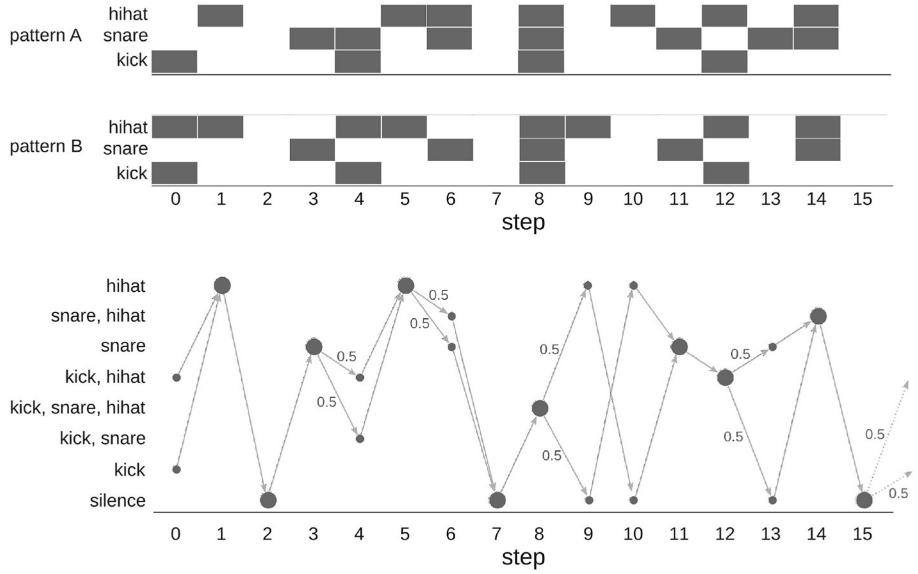

FIGURE 1
A two-bar drum sequence composed of patterns A and B, written using kick, snare and hihat instrument set **(top)** and its complex network representation **(bottom)**.

existing edges. The larger the number of common edges the larger the similarity between two networks. Maximum Jaccard similarity happens when one network is a subset of the other (A ⊂ B or B ⊂ A, meaning all connections in one network are observable in the other). Or on a middle ground, observing at least one non-common node meaning both networks do not fully overlap, forbidding inclusion in both directions A ⊄ B or B ⊄ A.

Another useful structural similarity metric is the indegree similarity. The degree of a node in a complex directed network describes the number of incoming and outgoing connections from and to other nodes. In a weighted network, indegree and outdegree can also be measured involving connection weights. Therefore, indegrees and outdegrees are measured as the both the number of connections and the sum of the weights. These metrics characterize a node's centrality and its capacity to transmit information respectively. As similar networks with overlapping nodes are expected to have similar node degrees (i.e. two different networks that have common nodes with similar degrees will be similar) a degree distance metric can be defined. The inverse of the difference between node degrees informs of the similarity between networks. In our network representation the indegree captures better the centrality given the information flow is in step wise order. Specifically, the weighted indegree of a node sums up probabilities from preceding nodes which can be larger than one (for example in Figure 1 nodes at steps 1, 5, 7, 11 and 14 all have indegree of 2 as both of the preceding nodes have an outdegree of 1) while the weighted outdegree for all nodes is always one (as outdegrees of nodes are normalized). We thus propose indegree as a very expressive indicator of centrality in our network representation.

On the other hand, descriptor-based comparison involves extracting network properties and subsequently comparing them (Berlingerio et al., 2012; Omar and Plapper, 2020). However, due to the architecture of our networks, there are specific properties that cannot be calculated. For instance, the formation of triadic closures (the creation of interconnected nodes forming a triangle) is never observed given the absence of connections between nodes within the same layer. Consequently, common descriptors such as the clustering coefficient and closeness centrality cannot be applied to our networks. Therefore, we decide to work with between-network structural comparison using the Jaccard simmilarity index and the indegree similarity (Omar and Plapper, 2020).

6 Experiment 1: network-based style classification

In this first experiment, we aim to explore the advantages of representing musical drum patterns as networks by conducting a classification experiment. We utilize the Groove Midi Dataset (GMD), which features drumming performances in various styles recorded by expert musicians in both audio and symbolic formats. The objective is to assess whether a network representations of a small batch of symbolic patterns can effectively determine the musical style they belong to. The classification method works as follows: First, we create a network representation for each musical style using 50% of the available samples for that style. Then, we compare these style networks against a network created using a small batch of test patterns, employing a simple nearest-neighbor classification procedure (Cover and Hart, 1967).

6.1 Materials

The GMD comprises drum performances in 17 musical styles, including latin, hiphop, funk, rock, jazz, reggae, gospel, pop, afrocuban, breakbeat, middleeastern, country, highlife, punk,

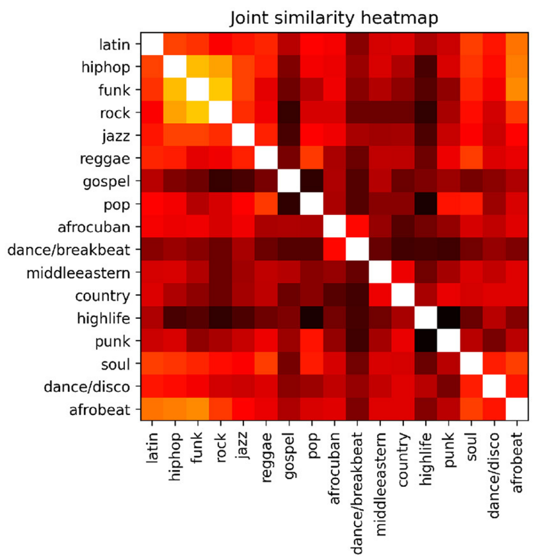

FIGURE 2
Joint similarity heatmap. All style's networks are compared among them using joint similarity.

soul, disco and afrobeat. Each style includes real-time drumming performances by professional musicians, with multiple recorded tracks showcasing different nuances of the drumming style. The dataset, however, is skewed toward popular music, with rock having the highest number of bars (4837) and gospel the least (68). We intentionally maintain this imbalance in the number of patterns to observe the performance of the network-based classification method and study the relationship between the number of patterns and classification accuracy. These drum patterns have been originally created using a virtual drum kit using 22 different sounds, and for the purpose of the experiment the 22 sounds have been remapped to three instrument kit composed of kick, snare and hihat.

6.2 Data preprocessing

To measure the relationships among style's networks, a joint similarity metric (see Section 5) using Jaccard similarity index, indegree and weighted indegree similarities is computed among all pairs producing a similarity matrix (see Figure 2). The joint similarity metric is defined as the equally-weighted sum of these three factors. To enhance qualitative comprehension of the styles, the similarity matrix is inverted to a dissimilarity matrix and then processed with a multi dimensional scaling (MDS) algorithm (Torgerson, 1952). MDS projects all the distances into a low-dimensional representation minimizing the stress between the distances of the elements in the final representation in the dissimilarity matrix. In order to enhance readability the MDS is set to project to a 2D space (see Figure 3).

Figure 2 (top left) and Figure 3 (top left quadrant) show how hiphop, funk and rock styles bear high similarity and small distances (respectively). In musical terms this means these styles have similar types of onset sequences located in similar temporal positions and with similar recurrence. This similarity can also be observed in Figure 4 as the top branch of the dendrogram contains these three same patterns. Based on the similarity metrics we use, we know that networks derived from these three drumming styles have similar nodes and edges controlling the information across different layers. Conversely, highlife, gospel and breakbeat are the style's networks with the least resemblance to the rest (0.535, 0.548 and 0.556 mean similarity respectively) located in the outer regions of the right upper and lower quadrants of Figure 3. This suggests their musical traits, embedded in their networks as nodes and connections, are the least shared with the rest of the styles. This can also be observed in Figure 2 as the darkest rows and columns belong to these styles.

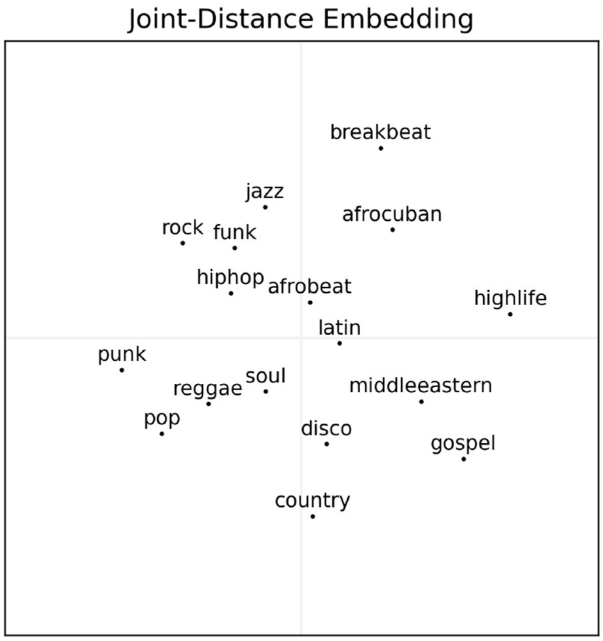

FIGURE 3
Joint distance embedding. Bidimensional representation of the styles based on network similarity.

6.3 Genre classification task

To evaluate the network's ability to represent collections of drum patterns, a classification task was conducted using a nearest-neighbor approach. The task involved computing similarities between large-sample style networks, and classifying small-sample networks based on their similarity to the large-sample ones. The efficacy of this approach was assessed by repeating multiple times the classification task and by varying batch sizes. The classification methodology involved splitting all the patterns of a style into train and test batches. Large-sample style networks are created with the train batch. A progressively larger portion of the test batch (starting from a size of 2 patterns) is used to create a small-sample network which is compared against all large-sample networks

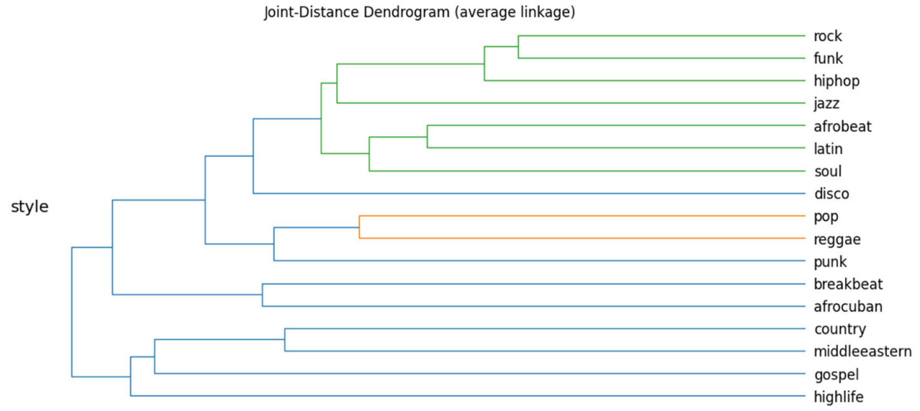

FIGURE 4
Joint-distance dendrogram (created using using the average linkage algorithm) of all style's networks.

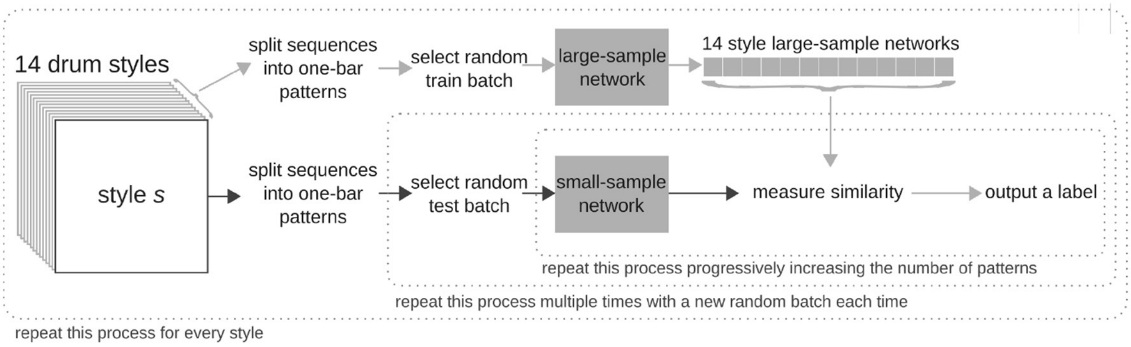
14 drum styles
style *s*
split sequences into one-bar patterns
select random train batch
large-sample network
14 style large-sample networks
split sequences into one-bar patterns
select random test batch
small-sample network
measure similarity
output a label
repeat this process progressively increasing the number of patterns
repeat this process multiple times with a new random batch each time
repeat this process for every style

FIGURE 5
Flow diagram of the classification process.

created with the test batch. To classify the test batch to a style, the maximum value in the similarity vector was utilized to identify the corresponding style. Essentially, the large-sample style network most closely resembling the small-sample network (the nearest neighbor) was designated as the class to be inferred. Although it is common practice to reserve around 2/3 of a dataset for training, here only 50% was used to increase the difficulty of the task and highlight the performance of this approach. To assess the efficacy of this classification approach, an experiment was conducted in the following stages and batch sizes, utilizing the GMD dataset (see Figure 5):

1. Selection of a style.
2. Segmentation of all patterns in the style into 50% train and 50% test.
3. Random selection of a batch of n number of patterns from the test set.
4. Creation of a small-sample test network using all patterns in the n-sized batch.
5. Calculation of the distance between the test network with the large-sample train networks of each style. This is carried out employing the joint similarity distance outlined in Section 5.
6. Identification of the small network's class based on the highest similarity value in the test distance vector.
7. Repetition of steps 3 to 6, 20 times for each batch size, with a progressive increase in batch size.
8. Repetition of steps 1 to 7, for each style in the dataset.

6.4 Results

An initial analysis was conducted to determine the optimal batch size for achieving a perfect F1 classification score. The F1 score is used as a measure of predictive performance of classification systems. It is computed using the precision and recall metrics calculated after a classification procedure. Precision is calculated as the number of correctly classified items (true positives) divided by the sum of all items classified as belonging to a label whether they are correct or not (true positives + false positives). Recall is calculated as the true positives divided by the sum of all items in the ground truth with the classification label whether they were correctly classified or not (true positives + false negatives). It was hypothesized that with an increase in batch size the F1 classification score would increase.

Figure 6 presents the progressive increment of batch size and the classification F1 score achieved. Figure 7 presents the minimum batch size to achieve a perfect F1 score. Figure 7 suggests that

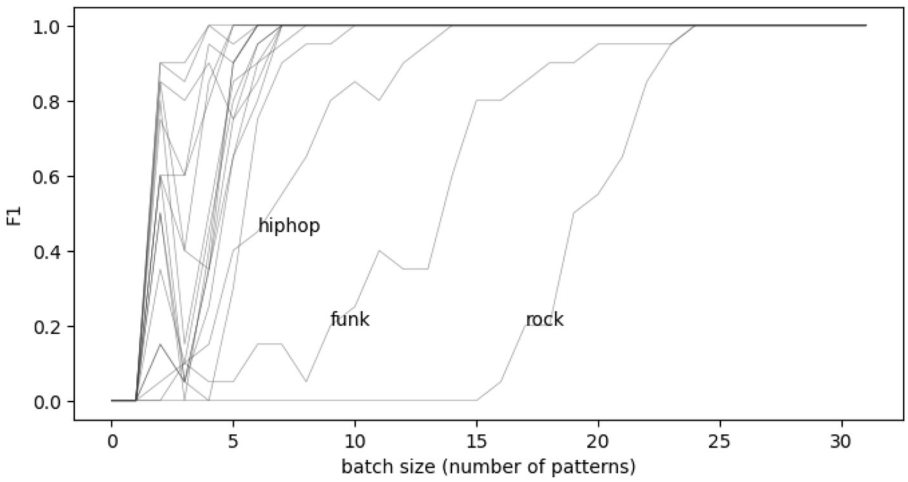

FIGURE 6
Batch size and F1 score for every style in the dataset.

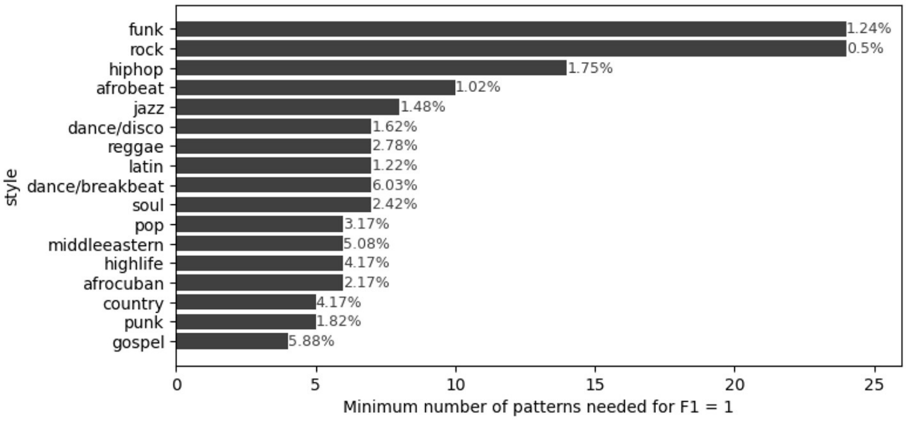

FIGURE 7
Minimum number and percentage of patterns needed to achieve perfect F1 score for every style in the dataset. The bars present the absolute value and the numbers the relative values (from the total number of patterns in the style).

rock, funk and hiphop styles need networks created with larger batch sizes in order to be correctly classified. The small-sample networks for these styles need to be constructed of at least 24, 24 and 14 random patterns respectively. However, these batch sizes represent 1.24%, 0.5%, and 1.75% of the total number of patterns respectively (see percentage in front of the bar plots in Figure 7). The larger batch sizes needed for these styles to be perfectly classified is a consequence of the similarity among their networks as described in the previous subsection (Figures 2, 3). The fact that these three styles contain patterns with similar features, and thus their networks are similar, makes the classification task require larger batch sizes in order for their distinctive features to become part of the network. When the batch sizes are large enough (equal or above the minimum number of patterns needed to 100% F1 score), the distinctive traits of each style are finally embroidered within network topology and an effective classification is obtained.

Styles not including funk, rock and hiphop (14 out of 17 styles, 82.3%) can be perfectly classified constructing networks with 10 patterns or less. In relative terms all styles need batch sizes of less than 5.88% the total number of patterns in order to construct small-sample networks that are able to classify them correctly.

On the other hand, gospel, breakbeat and middleeastern are the styles with fewer instances played in the GMD (68, 116, 118 respectively) so those are the styles that need relatively more patterns to be classified correctly (5.88%, 6.03% and 5.08% respectively).

6.5 Disambiguating funk, rock and hiphop styles studying network topology

The topology of funk, rock and hiphop style networks based on all patterns is studied to understand the relative difficulty to achieve correct classifications of their patterns. We present the three style networks created (see Figure 8 top) and the network resulting after subtracting the indegree weights of one network one from the other (rock–funk, rock–hiphop and funk–hiphop, in Figure 8 middle). Finally, in order to establish a comparison, the latin style network is subtracted from funk, rock and pop styles (Figure 8). Notice how the differences among these three styles are lower than the differences between each of them and the latin genre.

Network differences among rock, pop and hiphop are small, suggested by the small indegree differences coded as radius of the nodes (Figure 8 middle). There are specific nodes and edges of each network that make them distinctive from the other. For example in rock–funk (Figure 8 middle left) nodes representing the snare hits (MIDI instrument 38) at steps 4, 6, 10, 12 and 14 are present in rock and not in funk. A similar effect can be observed in rock–hiphop (Figure 8 middle center) at steps 6 and 9 as nodes denoting snare hits are part of rock style and not of hiphop's. The same can be said of the node representing a snare and hihat hit (MIDI instruments 38 and 42) at step 8.

On the other hand, nodes that have larger indegree in funk and hiphop and not in rock are observed (the light gray nodes symbolize negative indegrees, suggesting the subtrahend style contains a higher indegree in that node than the minuend style). These light color nodes suggest things that are not particular of rock (the minuend) but are of funk and hiphop. Again, the snare nodes (MIDI instruments 38) and in less degree the kick and snare nodes (MIDI instruments 36 and 38) can be distinctive of something that is not rock. Explicitly, snare nodes at steps 1, 5, 11 and 13 are less common in the rock network than in the counterpart networks. As well as kick and snare nodes at steps 2 and 6.

Observing the middle and bottom sections of the subtracted networks, there is not much difference among rock, funk and hiphop networks. This explains how most of the nodes containing silences and kick events (kick, kick and hihat and kick snare hihat) have the same relevance in terms of indegree among the three styles. In musical terms this is a powerful observation as the two most contrasting rhythmic features of the styles, the strongest hits of the patterns (produced by the kick and combinations) and the silences, are shared by the patterns in the three styles.

6.6 Discussion

This study demonstrates the efficacy of a multi-layered network representation (see Section 4) method for classifying drum patterns, emphasizing the importance of incorporating rhythmic information, which has often been overlooked in previous musical representations. By addressing the limitations of prior research, particularly in the context of pitch and style, we highlight the necessity of selecting network representations that prioritize the most salient musical dimensions, such as timbre and rhythm in drum sequences.

The findings reveal that the size of the drum set used significantly influences the classification task, with the minimal three-instrument set presenting the highest difficulty level. Future research should explore the potential benefits of larger instrumental sets to enhance classification accuracy and reduce the number of patterns required for effective style identification.

Our classification task, which utilized randomly selected unseen patterns, underscores the robustness of network representations in accurately categorizing drum patterns into their respective musical styles. Notably, the ability to classify certain styles with as few as four to seven patterns highlights the distinctiveness of the musical characteristics encoded within small networks. These results can be contrasted with the multiple musical genre classification approaches where perfect classification in symbolic datasets using classic machine learning methods is hardly achieved (Corrêa and Rodrigues, 2016, p. 199).

The successful encoding of musical information through basic network topology–specifically, edges and node centrality–suggests that stylistic elements are effectively captured through these connections. This insight enriches our understanding of musical networks and contributes to the broader discourse on music theory by elucidating the mechanisms that differentiate genres.

Basic network topology as edges (in Jaccard similarity metric) and node centrality (in indegrees, both weighted and non-weighted) has proven to effectively encode musical information. To the extent that by quantifying and comparing them the classification is successful. This strongly suggests that stylistic musical information is effectively captured through node connections (edges) and also by node indegree centrality (specific instruments playing at specific time positions). These findings not only enrich our comprehension of musically-inspired networks and musical styles but also contribute to the broader discourse on music theory by elucidating the underlying mechanisms that define and differentiate various genres.

In our classification task, we observe several key advantages over the findings of Géré et al. (2024). While they classify only four style classes, we successfully classify 17. This increase in the number of classes is significant, especially given the dataset's inherent imbalance, which complicates the classification of multiple styles. Géré et al. (2024) work with complete drum performance MIDI files, whereas our results demonstrate accurate classification of 17 styles using segments of no more than 25 bars. This indicates that we can achieve correct classifications for more styles with less information. Additionally, Géré et al. (2024) train their model using 80% of the data for optimal performance, while we train with only 50%, presenting a greater challenge for our models. Finally, Géré et al. (2024) report a best F1 score of approximately 63%, while our method achieves an F1 score of 100%.

In conclusion, our results indicate superior performance, as we classify four times more labels while training with less data and requiring fewer resources. Furthermore, Géré et al. (2024) emphasize the importance of representations by exploring five different methods for representing drum patterns. This highlights the relevance of our discussion on drum pattern representations. Our findings suggest that the representation we propose offers distinct advantages over those commonly used in the music information research community, contributing valuable insights to this ongoing discourse.

Furthermore, the transparent infrastructure provided by musical network representations facilitates problem-solving and enhances explainability in style disambiguation. Our analysis reveals that while certain drum events exhibit similarities across genres, the unique identities of styles such as rock, funk, and hip-hop are defined by the strategic use of specific percussive elements.

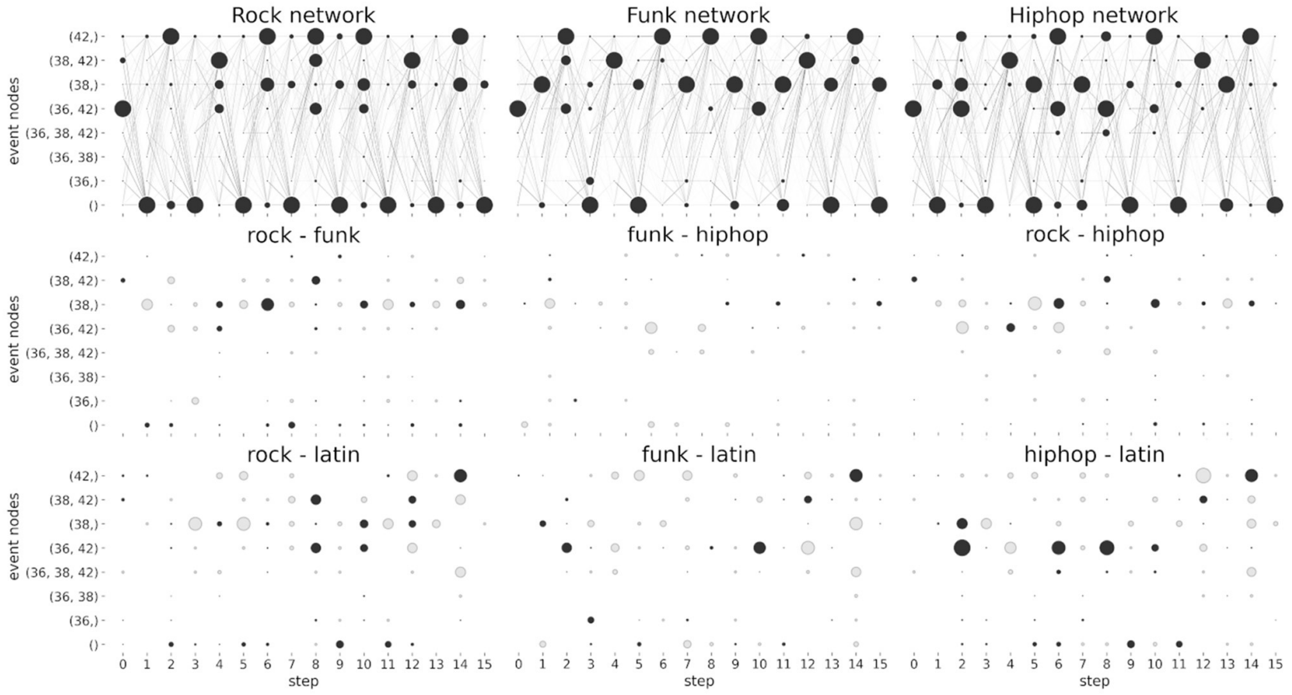

FIGURE 8
Rock, funk and hiphop networks **(top)**, rock indegree–funk indegree, funk indegree–hiphop indegree, rock indegree–hiphop indegree **(middle)**, rock indegree–latin indegree, funk indegree–latin indegree, hiphop indegree–latin indegree **(bottom)**.

Ultimately, the classification results obtained in this study have significant implications for music research and genre classification. By successfully categorizing drum patterns using network representations, we not only deepen our understanding of musical structures but also pave the way for practical applications in music analysis, recommendation systems, and music generation, all while maintaining a transparent and elegant approach to classification.

7 Experiment 2: evaluation of a network-based model for drum sequence generation

Having successfully demonstrated the efficacy of our approach in categorizing limited collections of one-bar drum sequences according to their stylistic class, we now seek to extend this methodology to the realm of generative modeling. Specifically, we propose to harness the structural properties of our style networks to produce novel one-bar drum sequences that emulate the patterns and characteristics of their archetypal counterparts. To achieve this, we will employ random walks across the network as a means of exploring the network of drumming patterns, thereby generating new sequences that are likely to conform to the stylistic features that define the original datasets (Jones, 1981; Rosvall and Bergstrom, 2008). Through a rigorous feature-based analysis (Gomez-Marin et al., 2020; Yang and Lerch, 2020), we will systematically compare the generated sequences with their original counterparts, with the hypothesis that our network-based approach will yield patterns that faithfully capture the essence of the underlying style.

7.1 Methods

7.1.1 Materials

In this experiment, we leveraged two distinct datasets, each serving a specific purpose, to create a comprehensive experimental dataset that captures the essence of studio production practices, drum performance techniques, and the concept of musical style. The first dataset utilized was the GMD (Gillick et al., 2019) used in the previous experiment. The second dataset, our custom Sano[2] and Boska[3] (S&B) dataset, was specifically devised for this task and comprises a limited selection of original patterns generated by two professional music producers using digital audio workstations (DAWs). For the S&B dataset Each producer was instructed to create a set of 10 one-bar sequences "in their style" using the instruments in the 8-instrument drum set, maintaining a constant velocity in all sequenced notes. This experimental dataset-comprising GMD and S&B- provides a wide selection of patterns designed for grasping diffeent methods to create symbolic drum seqences.

[2] https://www.discogs.com/artist/663348-Sano, https://contento.bandcamp.com/

[3] https://www.discogs.com/artist/1955792-Boska, https://xlr8r.com/news/download-one-hour-of-boska

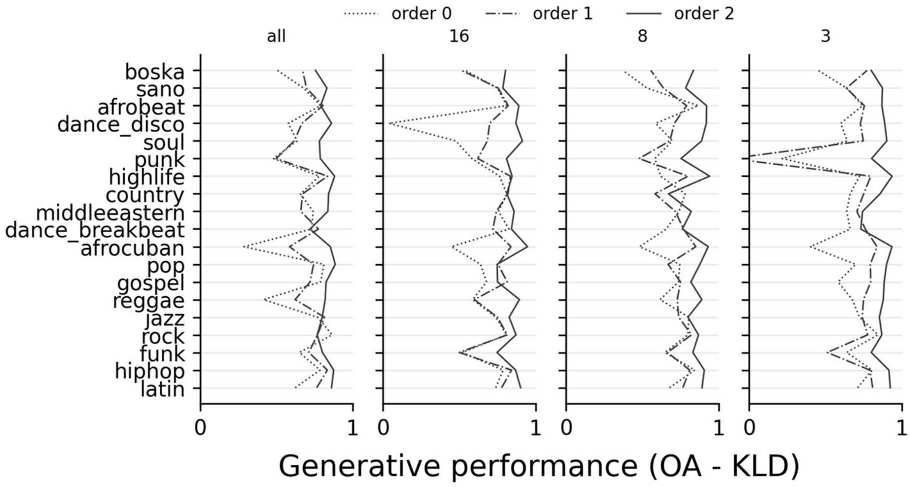

FIGURE 9
Generative performance for all styles, instrument drum sets (all, 16, 8, and 3) and Markov orders (0, 1, and 2).

7.1.2 Procedure

In order to rigorously evaluate our hypotheses, we employ zero, one, and two-order Markov processes to generate drum event nodes at each step within the network. Furthermore, we utilize diverse drum sets to progressively represent the drum patterns in the dataset, thereby diminishing the available instruments and consequently reducing the size of potential nodes in the network (refer to Section 4).

For each style in the experimental dataset, a new set of one-bar drum patterns is created. The number of generated patterns for each style matches the number of original one-bar patterns in that style, resulting in two sets of drum patterns that are the same size but distinct: one set of original patterns and one set of generated patterns.

The polyphonic descriptors for drum patterns (Gomez-Marin et al., 2020) serve as the foundation for extracting a descriptor vector from each one-bar drum sequence. These descriptors are derived from essential rhythm cognition features such as instrumental diversity, onset density, syncopation and timbre. The vectors resulting from extracting these descriptors facilitate the computation of Euclidean distances among drum patterns, thereby establishing quantifiable similarity relationships among them. This methodology extends to the comparison of all training patterns within a specific style, or any designated group of patterns.

To establish a robust evaluation metric between the patterns used for training and the patterns output by the generative process, we introduce a novel generative performance score. This score leverages on the interplay between the overlapping area (OA) and Kullback-Leibler divergence (KLD) values, as delineated by Yang and Lerch (2020). Central to this scoring mechanism is the analysis of two distributions: the Euclidean distances within the entirety of training patterns, and the Euclidean distances between the set comprising all training patterns and their generative counterparts. Through a meticulous examination of the overlapping area (OA) and the Kullback-Leibler divergence (KLD) of these distributions, we attain a comprehensive evaluation of generative fidelity. Notably, a higher OA value, constrained to a maximum of 1, signifies a closer resemblance between the generated samples and the training data, while a lower KLD value, approaching 0, underscores a superior alignment between the two sets. The proposed generative performance score, calculated as OA–KLD, encapsulates the essence of generative prowess, providing a nuanced perspective on the fidelity of the generated samples in relation to the training corpus.

TABLE 1 Generative performance summary for all drum sets and Markov orders.

	Drum set				
Order	All	16	8	3	Mean
0	0.663	0.647	0.673	0.635	0.654
1	0.71	0.731	0.711	0.703	0.714
2	0.82	0.841	0.848	0.862	0.843

7.2 Results

The generative performance score for each style, order and drum set is presented in Figure 9. It is observed that order 2 has higher generative performance score than orders 0 and 1 for the vast majority of the styles. A summary is found in Table 1. Results for orders 1 and 2 are within the baseline range (68%–92%) established by previous works reported in Section 3.2.

An analysis of variance (ANOVA) revealed a significant effect of the order on the performance score, $F_{(2,227)} = 47.1, p < 0.05$. *Post hoc* comparisons using the Tukey HSD test indicated that there was a significant difference between order 0 (M = 0.654, 95% CI [0.619, 0.689]) and order 1 (M = 0.714, 95% CI [0.683, 0.744]), $p < 0.05$. Additionally, order 0 was significantly different from order 2 (M = 0.843, 95% CI [0.829, 0.857]), $p < 0.05$. And order 1 and 2 are also significantly different $p < 0.05$. This suggests that the order has a direct effect in the quality of the generated one-bar drum sequences.

In order to evaluate the effect of the drum set used to code the network (that limits the amount of nodes in the network) an ANOVA and *post-hoc* Tukey HSD tests are used to evaluate if there is a statistical significance. The ANOVA presents no significant differences between the treatment groups [$F_{(3,\,227)} = 0.1$, $p = 0.96$]. The lack of statistical significance suggests that the drum sets used did not have a significant effect on the generative performance score. These findings indicate that encoding the drum patterns with a different drum set did not lead to statistically distinguishable outcomes in terms of the generative performance score. This highlights the similarity in the effects of the four different ways to encode the drum sequences as networks.

To evaluate the effect of the number of one-bar patterns in each style with the generative performance score, the correlations between style size and the performance score of each drum set and order is computed. Pearson's correlation is not significant (p-value > 0.05) for any drum set and Markov order. This suggests that the performance score is independent from the number of one-bar patterns used to create the networks, despite having a contrasting number of patterns within the different styles. This observation resonates with the results obtained in the previous experiment.

7.3 Discussion

The use of network representations for collections of one-bar drum sequences has enabled the generation of novel rhythmic patterns that retain the essential features of the training data. When Markov models of order 2 were employed to create the networks based on the styles of the GMD dataset, a mean generative performance score of 0.847 was achieved. Given the stringent nature of the scoring metric, which requires the generated and target distributions to be identical to attain a score of 1, this result indicates that the generated patterns successfully capture the salient characteristics of the training set without simply replicating it.

In the context of generative music, systems that can imitate the training data while introducing novel variations are highly desirable. Such approaches allow for the production of new content that aligns closely with the features of the original material, yet still leaves room for divergence. This can be likened to an apprentice learning to replicate the style of a master, but refraining from complete mimicry.

Interestingly, the choice of drum set used to encode the sequences had no discernible impact on the mean generative performance when all orders were considered. The size of the drum set directly influences the network complexity and the diversity of nodes available to describe the drum patterns. As the drum set becomes smaller, the drum patterns and their corresponding networks are simplified. Yet this simplicity does not affect the overall performance suggesting that the proposed generative model based on complex networks can be equally effective for collections of sequences expressed using a limited (i.e ten as in the S&B dataset) or extensive set of drum events (as found in GMD).

Furthermore, there was no significant correlation between the number of patterns in the training styles and the generative scores obtained. This indicates that the generative performance is independent of the size of the pattern set used to create the networks. This finding aligns with the results from the previous experiment, where the number of patterns required for perfect classification was also unaffected by the size of the pattern set used to construct the network. This speaks to the robustness of the network representation, as the objective variables (F1 score in the first experiment and generative performance score in this one) are not influenced by the scale of the pattern set employed to generate the networks. Even more, this contrasts with the prevailing notion that large datasets are needed to model the elements of a domain (Goodfellow et al., 2014; Kaplan et al., 2020).

8 Experiment 3: subjective evaluation of network-based generated drum sequences

After positively testing network representations of drum patterns in classification and generation tasks, a final experiment is devised to assess, from a human perspective, the quality of generated drum patterns. In this experiment we decide to use only the S&B dataset created by two professional music producers (see the section 7.1.1). Their styles are idiosyncratic and reflect a personal take on drum sequencing that does not fall directly into any of GMD styles, limiting possible associations of the auditory stimuli with previously known music styles. As in the two previous experiments, the small-size of the dataset (ten one-bar sequences per style) is also interesting from a generative perspective, allowing to observe if networks constructed with a small amount of one-bar drum sequences can be used to effectively imitate the style coded in them (see networks in Figure 10). The goal in this experiment is to investigate three key aspects:

- If a general difference between human-produced or generated drum patterns can be noticed by the subjects.
- If patterns produced by a human and by our system imitating her style can be differentiated by subjects.
- If subjects can differentiate patterns produced by two different humans in their own different styles.

8.1 Methods

8.1.1 Materials

As mentioned above Sano and Boska (S and B for the remaining of the paper) were instructed to craft ten one-bar MIDI drum patterns within the scope of their own artistic styles. We leveraged them to create S and B drum networks following the procedure presented above (see Section 4) and with these networks generated 10 new patterns in each style using order 2 and 8 instrument mapping (as presented in the previous experiment). This process resulted in a total of 40 MIDI drum patterns, including 10 original patterns and 10 patterns generated by our network system for each of the two producers.

All drum patterns (original and generated) were played at a tempo of 120 BPM. The audio samples used for these patterns were sourced from Roland's TR 808, 909, and 707 drum machines, which are frequently employed in electronic music production.

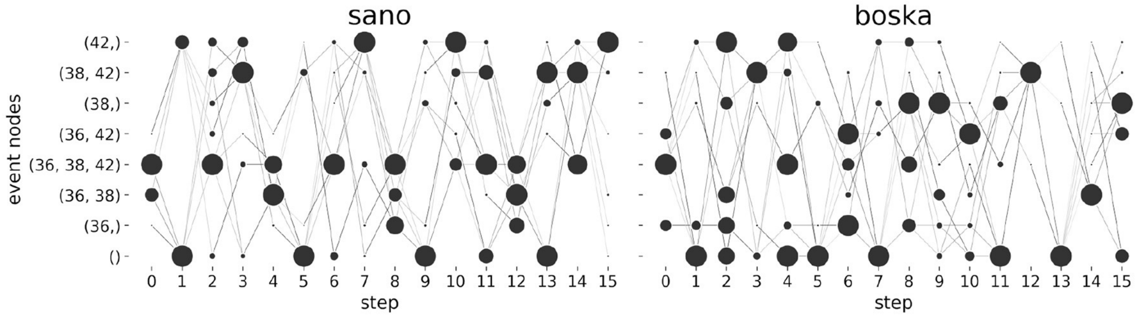

FIGURE 10
Network representations of Sano and Boska styles.

The experiment was delivered as a PureData (Puckette et al., 1996) patch and communicated to participants via email. Participants were kindly instructed to reserve a quiet environment and utilize headphones, while also being guided on how to follow the provided instructions.

8.1.2 Procedure

The experiment is divided in four sequential phases: first exposure, first assessment, second exposure, final assessment. In the first exposure, participants are invited to listen to a drum sequence context which is randomly selected from the two S and B styles. The drum context set is composed of five original randomly-selected one-bar sequences from the selected style. Simultaneously, a test set is also created containing 17 sequences:

- The remaining five original sequences of the selected style
- five randomly-selected generated sequences from the selected style
- five randomly-selected original sequences from the not-selected style
- two randomly selected sequences from the context set.

Subjects were required to listen to each drum pattern in the context set a minimum of three times before a "next" button became available, allowing them to advance to the rating phase. This requirement ensured that participants had adequately familiarized themselves with the selected drumming style.

In the first assessment phase, participants were presented with 8 test sequences randomly selected from the test set. Each drum sequence was associated with a play button and a slider ranging from 0 (“Differs Completely”) to 100 (“Resembles Perfectly”). Participants were tasked with listening to each drum sequence and rating how closely it resembled the drumming style presented during the first exposure.

The second exposure phase was identical to the first one, where subjects were asked to listen to each pattern in the context set at least three times. This additional exposure was intended to further enhance the participants’ memory of the context set.

In the final assessment phase, participants were invited to rate the remaining 9 patterns composing the test set (again they had to decide how similar the pattern seemed to those of the learned human producer). Once participants had completed this final stage, they were instructed to send an email containing a results file that was generated upon the completion of the experiment. The results file included a participant’s ID and the ratings for each test sequence.

8.1.3 Participants

In total, 47 subjects participated in the experiment, all belonging to European, Latin American and Asian backgrounds. Ten subjects were discarded because they failed to recognize control patterns as resembling the context (rated either of the two control patterns with a resemblance score below 95%). This way we ensure that the analyzed data came only from subjects that had abstracted or inferred stylistic traits of the context. The remaining 37 subjects identified themselves as female (8) and male (29), the age mean was 30.4 years with a standard deviation of 7.4.

8.2 Results

A t-test was conducted to compare the mean resemblance scores of the original pattern group and the generated pattern group. The results did not show a statistically significant difference, $t_{(184)}$ = -1.012, p = 0.313, 95% CI [0.495, 0.59]. The original pattern group (M = 0.542, SD = 0.329) scored very close to the generated pattern group (M = 0.573, SD = 0.319). This suggests the original sequences and the generated sequences, disregarding if the author was S or B, were not rated differently regarding their resemblance to the context. In general, subjects did not feel that the original sequences presented in the test were more (or less) resembling to the context than the generated sequences also presented in the test set.

In order to observe if subjects detected the drum sequences from the other producer as having less resemblance to the context set an ANOVA is carried out. A one-way ANOVA was conducted to examine the effect of the author of the drum sequences on the resemblance score given by the participants. The participants had three groups to rate: original same as style, generated same as style, and original from a different style. The results showed a statistically significant effect, $F_{(2,184)}$ = 82.174, $p < 0.05$ η^2 = 0.309 meaning that at least one of the groups is different from the others.

TABLE 2 Tukey's HSD test for the three types of patterns tested in the experiment.

Group		Mean difference	*p*
Original same as contex	Generated same as context	0.0307	>0.05
Original same as context	Original not from context	−0.3373	<0.05
Generated same as context	Original not from context	−0.368	<0.05

A *post-hoc* Tukey's HSD test was carried out to evaluate significant differences among the three authorship types (see Table 2). There is not a significant difference in resemblance scores between the original patterns in the same style as the context (M = 0.542, SD = 0.329) and the generated patterns in the same style as the context (M = 0.573, SD = 0.319). Suggesting subjects assess both original and generated patterns as resembling the patterns in the context set. Human-authored and network-based generated do not have significant differences in score. There are, however, significant differences between the resemblance scores of the original and generated sequences in the same style as the context with the scores of original sequences in a style different from the context (M = 0.205, SD = 0.265). This suggests subjects noticed patterns from a style different from the one presented in the context phase and rated them with low resemblance scores. In general, participants scored original patterns from another style as poorly resembling the context (M = 0.205) while original and generated from the same style have very similar means above 0.5 resemblance (0.542 and 0.573 respectively) suggesting subjects feel both groups of patterns positively resemble the context in a similar way in relation to the context.

8.2.1 Resemblance scores for S and B as context

To observe the results independently for each producer, and understand if any of the styles was easier to capture by the subjects one ANOVA was used to test each case.

The resemblance to S as context had three types of groups rated by the participants: original patterns from S style, generated patterns in S style and original patterns in B style. The results showed a statistically significant effect, $F_{(2,134)} = 69.876$, $p < 0.05$ $\eta^2 = 0.343$. A *post-hoc* Tukey's HSD test was carried out to evaluate significant differences among the three groups of patterns (see Table 3). There is not a significant difference in the resemblance scores between the original patterns in S style (M = 0.569, SD = 0.312) and the generated patterns in S style (M = 0.572, SD = 0.31). Participants assess both original and generated patterns as resembling the patterns in the S context set. Original patterns by S and network-based generated in S style have no significant differences in score. On the other hand, there are significant differences between the resemblance scores of the original and generated sequences in style S with the scores of original sequences in style B (M = 0.205, SD = 0.25). Subjects rated original patterns in B style with lower resemblance scores than original and generated in S style, suggesting patterns in B style were identified as being different from S style context.

TABLE 3 Tukey's HSD test for the three types of patterns tested in the experiment. Sano used as context.

Group		Mean difference	*p*
Original style S	Generated style S	0.0032	>0.05
Original style S	Original style B	−0.364	<0.05
Generated style S	Original style B	−0.367	<0.05

TABLE 4 Tukey's HSD test for the three types of patterns tested in the experiment. Boska used as context.

Group		Mean difference	*p*
Original style B	Generated style B	0.104	>0.05
Original style B	Original style S	−0.265	<0.05
Generated style B	Original style S	−0.37	<0.05

Using B as context, three groups of patterns were also defined: original patterns from B style, generated patterns in B style and original patterns in S style. The ANOVA results showed a statistically significant effect, $F_{(2,49)} = 15.874$, $p < 0.05$ $\eta^2 = 0.245$. A *post-hoc* Tukey's HSD test was carried out to evaluate significant differences among the three groups of patterns (see Table 4). There is not a significant difference in the resemblance scores between the original patterns in B style (M = 0.47, SD = 0.36) and the generated patterns in B style (M = 0.575, SD = 0.34). Participants assess both original and generated patterns as equally resembling the patterns in the B context set. Original patterns by B and network-based generated in B style have no significant differences in score. There are significant differences between the resemblance scores of the original and generated sequences in style B with the scores of original sequences in style S (M = 0.205, SD = 0.299). Subjects rated original patterns in S style with lower resemblance scores than original and generated in B style, suggesting patterns in S style were identified as being different from B style context.

Finally, while the difference in means between the original and generated sequences in the context of style B did not reach statistical significance, it nonetheless deserves attention. This pronounced behavior was not observed when style S served as the context, suggesting that style B may have been more challenging for our subjects to fully comprehend. This is indicated by the subjects' slight difficulty in relating the unheard, yet original sequences to the ones presented as the context. Previous research has explored the discrepancy between the network produced by the musical sequence information and the human perception of those sequences. They have observed that the complexity of a network encoding sequential information - as measured by degree distribution, connection strength, or entropy (Newman, 2003) - does not necessarily equate to the difficulty of humans to process those sequences. Instead, they posit that high-entropy networks with high-degree hubs and a clustered structure, as found when processing literature and music sequences, are the very structures that aid human processing (Lynn et al., 2020). Table 5 reveals that network B has more nodes with larger connection weights,

TABLE 5 Weighted degree, node density, and entropy for Sano (S) and Boska (B) networks.

Metric	Network S	Network B
Number of nodes	70	75
Weighted degree	140	150
Degree distribution mean (SD)	3.629 (1.725)	3.387 (1.632)
Weighted degree distribution mean (SD)	0.551 (0.3)	0.59 (0.3)
Entropy	34.56	33.88

a smaller mean number of connections, and less entropy than network S. This suggests that the ten sequences in the B set are indeed more diverse than those in the S set. The fact that network S exhibits fewer, better-connected nodes and higher entropy aligns with Lynn's proposition, implying that it may be more readily grasped by our subjects. Therefore, the exposure to half of the original samples in style S as context captured slightly more features of the whole style than when B was presented as context, making style S slightly easier for our subjects to grasp during the initial exposure phase.

8.3 Discussion

In the introductory phase of the experiment, participants were exposed to drum patterns created by a specific producer. This exposure appeared to enable subjects to construct a temporary conceptualization of a drumming style based on a relatively small set of patterns (five in our experiment). They formed a mental framework that encompassed the stimuli from the introduction, effectively serving as a provisional category. This mental construct was coherent enough to assess the extent to which a new stimulus aligned with or deviated from this framework. In both experiments (contexts S and B), patterns created by different authors were significantly rated as not belonging to the induced style. This suggests that style operates as a cognitive concept that allows subjects to evaluate a new pattern by the degree to which it conforms to this concept.

Out network-based generation system appears to capture essential features, structural invariances, or commonalities within the drum pattern collection. We propose that this network representation resembles the concept of style temporarily constructed in the minds of listeners, which enables them to compare patterns. This is evident when patterns generated in a given style are not perceived differently from patterns created by the human originator of the style. In other words, if a mechanism can produce stimuli validated as resembling a specific style, there might be similarities between how a style is temporarily constructed in the listener's mind and the way it is coded in the generative system (which is a network in our case). It's essential to clarify that we do not assert that our generative approach mimics the intricacies of human cognition. Still, there may be parallels between how a style is temporarily formed in a listener's mind and the network-based generative process we have introduced. In more precise and practical terms, the absence of statistical significance between groups of original patterns in style x and generated patterns in style x implies that our network-based style generation mechanism appears to be a suitable way to encode and generate polyphonic rhythm material in a specific style.

It's noteworthy that original patterns in style B, which were not part of the context exposure, were rated as less related to the context than patterns generated in the B style (means 0.47 and 0.675 respectively). However, for the experiment with S as the context, this was not the case: original patterns in the same style as the context were rated as similar to the context in a nearly identical manner as patterns generated in the S style (means 0.569 and 0.572 respectively). We propose that the phenomenon observed in the results related to B, where unheard original patterns were rated as less related to the style than patterns generated by our system, is linked to the diversity within a collection (i.e., how distinct the elements within a collection are from each other or how effectively they cover the conceptual space of a style). The lower resemblance rating of the original patterns in B style can be a consequence of the higher diversity of the sequences in style B and the resulting higher complexity of its network. We believe style B is more diverse and was not fully grasped by the subjects when listening to the 5 randomly-selected original samples presented in the exposure phases. On the other hand, the larger resemblance of generated patterns toward the context might be a consequence of the Markov process used to generate the sequences based on the information coded in the networks. Since random walks through style networks are biased toward edges with higher weights, the generated sequences retain the most common features found in the style. Therefore, such generated patterns better retain global features of the style so they are more likely to resemble it.

9 General discussion

The research presented in this paper has demonstrated the potential of using complex network representations to model and analyze music and particularly drum rhythms. By encoding drum patterns as networks, with notes as nodes and their temporal connections as edges, we were able to uncover insights into the underlying structure and organization of different drumming styles.

Through the experiments conducted in this study, several key findings emerged:

- Comparison of drumming styles: By comparing the topology of the complex networks representing drum pattern collections, using distances such as Jaccard similarity and degree similarity, we were able to effectively differentiate and group various drumming styles. This suggests that the network-based approach can serve as a powerful tool for style classification and analysis.
- Generation of new drum rhythms: Building on the insights gained from the style comparisons, we developed a method to generate new drum patterns based on the complex network representations of the original styles. The generated patterns were found to exhibit characteristics similar to the source styles, as validated through listening tests with human participants.

- Insights into musical rhythm: The network-based modeling of drum rhythms has provided a novel perspective on the underlying structure of musical rhythm. By representing the temporal sequences as network connections, we were able to uncover organizational principles that may inform our understanding of how rhythm is constructed and perceived.

These findings contribute to the growing body of research on the application of complex networks to the study of music. The successful representation and analysis of drum rhythms using this approach opens up new avenues for exploring the computational and cognitive aspects of musical rhythm, as well as potential applications in areas such as music information retrieval, composition, and performance. The transparency, simplicity and efficiency of the proposed methods suggest network representations of music are viable tools for commmon problems in music processing.

Future research directions may include expanding the dataset, exploring more advanced network-based analysis techniques, and investigating the potential of this approach for other musical elements beyond drum rhythms. Additionally, the integration of this network-based modeling with other music analysis and generation methods could lead to further advancements in the field of computational musicology. Indeed, the observation that musical sequences optimized for human perception exhibit certain network topologies (Kulkarni et al., 2024; Lynn et al., 2020) suggests that studying these network properties can lead to a better understanding of how humans relate to music and how music has evolved over time.

Overall, this study has demonstrated the value of complex network representations in the study of drum rhythms, offering a promising framework for understanding and generating musical patterns in a systematic and insightful manner.

Data availability statement

The raw data supporting the conclusions of this paper can be found in the following link: https://drive.google.com/drive/folders/1-_Ywvi5MmO1K8tGkBpN-k2eS4sxnl6TI.

Ethics statement

The studies involving humans were approved by Luisa Fernanda Prado, Universidad Icesi. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

DG-M: Conceptualization, Methodology, Data curation, Validation, Writing – original draft, Writing – review & editing. SJ: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. PH: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funded by the European Union—NextGenerationEU, Ministry of Universities and Recovery, Transformation and Resilience Plan, through a call from the Pompeu Fabra University (Barcelona). The research reported here is part of the project "IA y Música: Cátedra en Inteligencia Artificial y Música" (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program "Cátedras ENIA 2022 para la creación de cátedras universidad-empresa en IA" from the Gobierno de España.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Berlingerio, M., Koutra, D., Eliassi-Rad, T., and Faloutsos, C. (2012). Netsimile: a scalable approach to size-independent network similarity. *arXiv* [preprint] arXiv:1209.2684. doi: 10.48550/arXiv.1209.2684

Brennan, M. (2020). *Kick It: A Social History of the Drum Kit*. Oxford: Oxford University Press.

Choi, K., Fazekas, G., and Sandler, M. (2016). Text-based lstm networks for automatic music composition. *arXiv* [preprint] arXiv:1604.05358. doi: 10.48550/arXiv.1604.05358

Corrêa, D. C., and Rodrigues, F. A. (2016). A survey on symbolic data-based music genre classification. *Expert Syst. Appl.* 60, 190–210. doi: 10.1016/j.eswa.2016.04.008

Coscia, M. (2021). The atlas for the aspiring network scientist. *arXiv* [preprint] arXiv:2101.00863. doi: 10.48550/arXiv.2101.00863

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* 13, 21–27. doi: 10.1109/TIT.1967.1053964

Dahale, R., Talwadker, V., Verma, P., and Rao, P. (2021). "Neural drum accompaniment generation," in *Late-Breaking Demo Session of the 22nd International Society for Music Information Retrieval Conference.*

Dean, M. (2012). *The Drum: a History*. Lanham, MD: Scarecrow Press.

Ferretti, S. (2017). On the modeling of musical solos as complex networks. *Inform. Sci*. 375, 271–295. doi: 10.1016/j.ins.2016.10.007

Ferretti, S. (2018a). Clustering of musical pieces through complex networks: an assessment over guitar solos. *IEEE MultiMedia* 25:57–67. doi: 10.1109/MMUL.2018.2873497

Ferretti, S. (2018b). On the complex network structure of musical pieces: analysis of some use cases from different music genres. *Multimed. Tools Appl*. 77, 16003–16029. doi: 10.1007/s11042-017-5175-y

Géré, L., Rigaux, P., and Audebert, N. (2024). "Improved symbolic drum style classification with grammar-based hierarchical representations," in *Proceedings of the 25th International Society for Music Information Retrieval Conference 2024* (San Francisco, CA).

Gillick, J., Roberts, A., Engel, J., Eck, D., and Bamman, D. (2019). "Learning to groove with inverse sequence transformations," in *International Conference on Machine Learning* (New York: PMLR), 2269–2279.

Gomez-Marin, D., Jorda, S., and Herrera, P. (2020). Drum rhythm spaces: from polyphonic similarity to generative maps. *J. New Music Res*. 49, 438–456. doi: 10.1080/09298215.2020.1806887

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems* (Montreal, QC), 27.

Haki, B., Nieto, M., Pelinski, T., and Jordà Puig, S. (2022). "Real-time drum accompaniment using transformer architecture," in *Proceedings of the 3rd Conference on AI Music Creativity (AIMC 2022)* (Graz: AI Music Creativity).

Hutchings, P. (2017). Talking drums: Generating drum grooves with neural networks. *arXiv* [preprint] arXiv:1706.09558. doi: 10.48550/arXiv.1706.09558

Jones, K. (1981). Compositional applications of stochastic processes. *Comp. Music J*. 5, 45–61. doi: 10.2307/3679879

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., et al. (2020). Scaling laws for neural language models. *arXiv* [preprint] arXiv:2001.08361. doi: 10.48550/arXiv.2001.08361

Kim, M., and Sayama, H. (2017). Predicting stock market movements using network science: An information theoretic approach. *Appl. Netw. Sci*. 2, 1–14. doi: 10.1007/s41109-017-0055-y

Kulkarni, S., David, S. U., Lynn, C. W., and Bassett, D. S. (2024). Information content of note transitions in the music of js bach. *Phys. Rev. Res*. 6:013136. doi: 10.1103/PhysRevResearch.6.013136

Liu, X. F., Chi, K. T., and Small, M. (2010). Complex network structure of musical compositions: Algorithmic generation of appealing music. *Physica A* 389, 126–132. doi: 10.1016/j.physa.2009.08.035

Loy, G. (1985). Musicians make a standard: The midi phenomenon. *Comp. Music J*. 9, 8–26. doi: 10.2307/3679619

Lynn, C. W., Papadopoulos, L., Kahn, A. E., and Bassett, D. S. (2020). Human information processing in complex networks. *Nat. Phys*. 16, 965–973. doi: 10.1038/s41567-020-0924-7

Makris, D., Kaliakatsos-Papakostas, M., Karydis, I., and Kermanidis, K. L. (2019). Conditional neural sequence learners for generating drums' rhythms. *Neural Comp. Appl*. 31, 1793–1804. doi: 10.1007/s00521-018-3708-6

Matta, J., Zhao, J., Ercal, G., and Obafemi-Ajayi, T. (2018). Applications of node-based resilience graph theoretic framework to clustering autism spectrum disorders phenotypes. *Appl. Netw. Sci*. 3, 1–22. doi: 10.1007/s41109-018-0093-0

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. Champaign, IL: University of Illinois Press.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Rev*. 45, 167–256. doi: 10.1137/S003614450342480

Omar, Y. M., and Plapper, P. (2020). A survey of information entropy metrics for complex networks. *Entropy* 22:1417. doi: 10.3390/e22121417

Óskarsdóttir, M., Gísladóttir, K. E., Stefánsson, R., Aleman, D., and Sarraute, C. (2022). Social networks for enhanced player churn prediction in mobile free-to-play games. *Appl. Netw. Sci*. 7:82. doi: 10.1007/s41109-022-00524-5

Puckette, M. (1996). "Pure data: another integrated computer music environment," in *Proceedings of the Second Intercollege Computer Music Concerts* (Tachikawa), 37–41.

Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Nat. Acad. Sci*. 105, 1118–1123. doi: 10.1073/pnas.0706851105

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika* 17, 401–419. doi: 10.1007/BF02288916

Umeyama, S. (1988). An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell*. 10, 695–703. doi: 10.1109/34.6778

Wei, I.-C., Wu, C.-W., and Su, L. (2019). *Generating Structured Drum Pattern Using Variational Autoencoder and Self-Similarity Matrix*. zmir: ISMIR, 847–854.

Wild, B., Dormagen, D. M., Zachariae, A., Smith, M. L., Traynor, K. S., Brockmann, D., et al. (2021). Social networks predict the life and death of honey bees. *Nat. Commun*. 12:1110. doi: 10.1038/s41467-021-21212-5

Yang, L.-C., and Lerch, A. (2020). On the evaluation of generative models in music. *Neural Comp. Appl*. 32, 4773–4784. doi: 10.1007/s00521-018-3849-7