Universidade Federal de Santa Catarina
Departamento de Informática e Estatística

Douglas Soares Molina

CLASSIFICAÇÃO DE GÊNEROS MUSICAIS POR ANÁLISE DE GRAFOS DE VISIBILIDADE HORIZONTAL

Florianópolis
2022

Douglas Soares Molina

CLASSIFICAÇÃO DE GÊNEROS MUSICAIS POR ANÁLISE DE GRAFOS DE VISIBILIDADE HORIZONTAL

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção do Grau de Bacharel em Ciências da Computação.
Orientador: Prof. Dr. Rafael de Santiago

Universidade Federal de Santa Catarina
Departamento de Informática e Estatística

Florianópolis
2022

Douglas Soares Molina

CLASSIFICAÇÃO DE GÊNEROS MUSICAIS POR ANÁLISE DE GRAFOS DE VISIBILIDADE HORIZONTAL

Trabalho de Conclusão de Curso apresentado à Universidade Federal de Santa Catarina como requisito parcial para a obtenção do grau de Bacharel em Ciências da Computação.

Comissão Examinadora

Prof. Dr. Rafael de Santiago
Universidade Federal de Santa Catarina
Orientador

Prof. Dr. Elder Rizzon Santos
Universidade Federal de Santa Catarina

Prof. Dr. Jônata Tyska Carvalho
Universidade Federal de Santa Catarina

Florianópolis, 24 de dezembro de 2022

Dedico este trabalho a minha família e a todos aqueles que, de alguma forma, auxiliaram para a concretização desta etapa.

Agradecimentos

Primeiramente gostaria de agradecer à todos que representam e fazem parte da minha família. Principalmente aos meus pais que sempre acreditaram em mim, que me deram ensinamentos e oportunidades para ter uma educação de qualidade, e também à minha namorada Roberta, que sempre me auxiliou em momentos importantes e apoiou todas as decisões difíceis que tomei. Sem vocês nada disso seria possível. Aos demais familiares, vocês também fazem parte de tudo isso, muito obrigado! Gostaria de agradecer também a todos os profissionais da educação que fizeram parte da minha vida em algum momento, em especial o Prof. Dr. Rafael de Santiago que foi fundamental na elaboração desse trabalho. Gostaria de agradecer a Universidade Federal de Santa Catarina por tornar possível minha formação em uma instituição de rede publica e gratuita, me preparando para o exercício profissional.

Resumo

Analisando as características mais relevantes extraídas de bibliotecas de arquivos de música digital, a classificação de gênero se torna uma das formas mais comuns de categorização do conteúdo. A extração automática de atributos informativos em sinais musicais está ganhando importância devido à questão de estruturação e organização do grande número de arquivos de música disponíveis digitalmente na Web. Dentre esses atributos, o ritmo desempenha um papel muito importante na definição do estilo musical. O estudo da rítmica em sinais de áudio inclui a investigação das características de regularidade de seus eventos acústicos e transientes. Tal análise pode contribuir para o estudo da complexidade rítmica de uma música. Falta diversidade de descritores de ritmo para sinais de áudio nos estudos atuais, e o campo de processamento de sinais é restrito a técnicas baseadas em representações tempo-frequência.

Neste trabalho foi mapeado, a partir do banco de dados GTZAN, os sinais de 1000 arquivos musicais em grafos de visibilidade, onde extraiu-se algumas propriedades topológicas através dos cálculos de Modularidade, Número de Comunidades, Grau Médio e Densidade. Realizou-se uma comparação entre atributos gerados a partir de Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal para cada sinal, usando-os como entrada em um experimento de classificação baseado em redes neurais artificiais supervisionadas, onde obteve-se uma precisão semelhante à outro descritor com atributos rítmicos de intensidade de Onsets do sinal, que representam o início de todos seus eventos acústicos.

Com base nos resultados obtidos neste experimento e nos estudos mencionados, foi evidenciado que os Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal podem ser considerados como uma nova alternativa para a extração de características rítmicas para a recuperação de informações de música e podem ser usadas com sucesso em conjunto com os descritores baseado em transformadas de Fourier, em especial à aqueles de natureza de timbre.

Palavras-Chave: classificação musical, redes complexas, redes neurais.

Abstract

By analyzing the most relevant characteristics extracted from digital music archive libraries, genre classification becomes one of the most common techniques of content categorization. The automatic extraction of informative attributes in music signals is gaining importance due to the question of structure and organization of the large number of music files available digitally in the Web. Among these attributes, rhythm plays a very important role in the definition of musical style. The study of rhythm in audio signals includes the investigation of the regularity characteristics of their acoustic and transient events. Such analysis can contribute to the study of the rhythmic complexity of a music. There is a lack of diversity of rhythm descriptors for audio signals in current studies, and the field of signal processing is restricted to techniques based on tempo-frequency representations.

This work has mapped, based on the GTZAN dataset, the signal of 1000 music files into visibility graphs, where some topological properties were extracted through the calculations of Modularity, Number of Communities, Mean Degree and Density. A comparison was made between attributes generated from Natural Visibility Graphs and Horizontal Visibility Graphs for each signal, using them as input in a classification system based on supervised artificial neural networks, obtaining a similar precision rate when compared to another descriptor based on rhythmic attributes of the intensity of the signal Onsets, which represents the beginning of all its acoustic events.

Based on the results obtained in this experiment and its related studies, it was evidenced that the Natural Visibility Graphs and Horizontal Visibility Graphs can be considered as a new alternative for the extraction of rhythmic characteristics for the retrieval of music information and can be used with success in conjunction with descriptors based on Fourier transforms, especially those of timbre nature.

Keywords: music classification, complex networks, neural networks.

Sumário

Lista de figuras

1 Introdução

O processo de evolução tecnológica levou estudos à explorar técnicas de armazenamento e análise de dados de forma à ter uma representação compacta e útil para os pesquisadores. Esse fato acontece principalmente em relação ao conteúdo da área musical, onde grandes plataformas de corporações conhecidas exploram diversos algoritmos computacionais para auxiliar na busca, recuperação e recomendação de arquivos musicais, obtendo cada vez resultados mais positivos e próximos da expectativa de especialistas e usuários.

Analisando as características mais relevantes extraídas de bibliotecas de arquivos de música digital, a classificação de gênero se torna uma das formas mais comuns de categorização do conteúdo. Apesar disso, a tarefa de organização desses dados não possui um conceito muito claro e definido de forma padronizada para o senso comum. Por consequência, se torna desafiador desenvolver uma única heurística que possa ser totalmente aceita, motivando a exploração de novas técnicas que demonstram resultados interessantes dependendo de diversos atributos à serem considerados no agrupamento.

Dentro deste contexto, torna-se imprescindível o desenvolvimento de métodos computacionais automáticos de agrupamento dos dados, que busquem sumarizar informações relevantes para a classificação dos gêneros, extraindo parâmetros numéricos através de diferentes tipos de algoritmos descritores. Dentre os algoritmos mais utilizados na síntese dos atributos musicais, estão aqueles que realizam operações no domínio tempo-frequência e que estão usualmente associados com propriedades que fazem parte da percepção musical dos humanos, como análise da textura do timbre, ritmo, variação temporal e conteúdo de tons harmônicos (TZANETAKIS; COOK, 2002).

A comparação de desempenho entre abordagens é uma tarefa custosa, uma vez que diferentes taxonomias de gêneros são empregadas nas tarefas de classificação. Na última década, entretanto, bancos de dados com arquivos de áudio têm sido disponibilizados para que as comparações de desempenho sejam mais viáveis.

A proposta deste trabalho consiste na análise de dados de áudio, com base nos estudos de trabalhos relacionados e publicados até o momento, explorando um possível descritor para a classificação de gêneros musicais sob a perspectiva de grafos de visibilidade. Essa transformação permite analisar uma relação interessante entre os pontos da série temporal, onde quanto maior o grau de conexão de um determinado vértice no grafo gerado, maior é a visibilidade do seu ponto correspondente na série. Após realizar o mapeamento, a rede terá herdado as características rítmicas da série através da qualidade e quantidade de grupos encontrados, e pelo grau de conexões geradas na rede.

1.1 Objetivos

Esta seção visa apresentar os objetivos propostos pelo trabalho que será realizado.

1.1.1 Objetivo Geral

Explorar e comparar possíveis descritores para o agrupamento de gêneros musicais, com base em métodos de transformação de sinais de arquivos de áudio em grafos de visibilidade, através da análise de similaridade de propriedades topológicas das redes geradas.

1.1.2 Objetivos Específicos

- Extrair e otimizar as séries temporais de arquivos de áudio, com base na intensidade do sinal.
- Mapear as séries temporais obtidas em Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal.
- Quantificar propriedades topológicas das redes obtidas, de forma à representar numericamente características relevantes à categorização dos gêneros musicais através do cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$).
- Desenvolver uma rede neural artificial para a classificação supervisionada, que é alimentada pelos dados encontrados sobre ambos os tipos de grafos.
- Classificar os gêneros com base nos grupos encontrados, comparando resultados entre os descritores encontrados e também com descritores relacionados à esta pesquisa.
- Divulgar os resultados.

1.1.3 Metodologia

O presente trabalho se trata de uma pesquisa quantitativa e exploratória, pois realiza a análise e adaptação de pesquisas existentes, propondo uma comparação entre atributos de descritores para a classificação de gêneros musicais, de forma a observar o comportamento sobre a performance do método proposto e comparar resultados com trabalhos existentes.

Primeiramente, efetuou-se um levantamento teórico com intuito de fornecer maior credibilidade a pesquisa, com o objetivo de definir as características importantes dos sinais a serem utilizadas como descritores na rede neural.

Em seguida foi feito a análise em sinais de diversos arquivos de áudio de um banco de dados de larga escala e extração de seus atributos considerados relevantes para o estudo.

Após a conclusão da etapa de análise, foi construído um modelo de aprendizado de máquina e feita a comparação entre os diferentes descritores para o mesmo modelo de rede, analisando o resultado de acurácia e erro entre as configurações.

Por fim, foi apresentado algumas informações interessantes sobre o desempenho do método proposto e suas implicações para futuras pesquisas, concluindo toda a documentação do trabalho desenvolvido.

1.1.4 Organização do Trabalho

Este documento está estruturado em seis capítulos.

O Capítulo 1, Introdução, apresenta uma visão geral do trabalho, apontando o atual crescimento de mídias de áudio, da necessidade de classificação dos gêneros musicais e alguns métodos utilizados atualmente.

No Capítulo 2, Fundamentação Teórica, é apresentada uma revisão da literatura sobre alguns conceitos abordados nesse trabalho, como definição e extração de propriedades de Áudio, gêneros musicais, características e aplicações de Grafos, Redes Complexas e Detecção de Comunidades.

No Capítulo 3, Trabalhos Relacionados, é feita uma breve exploração de outros trabalhos e apresentação de definições para a possível análise de comparação dos resultados com o método proposto.

No Capítulo 4, Desenvolvimento, é especificado alguns pontos importantes do trabalho, apresentando algumas mudanças de heurísticas e algoritmos existentes para a possível comparação de acurácia e erro na representação dos gêneros musicais.

No Capítulo 5, Experimentos, é apresentado todos os experimentos realizados e os resultados, com comparações de importância entre os dados obtidos e análise de performance entre diferentes descritores relacionados.

No Capítulo 6, são apresentadas as considerações finais e sugestões para trabalhos futuros.

2 Fundamentação Teórica

Neste capítulo serão abordados assuntos relativos aos conceitos básicos utilizados neste trabalho, de forma à dar embasamento para a pesquisa.

2.1 Classificação de gêneros musicais

Analisando as características mais relevantes extraídas de bibliotecas de arquivos de música digital, a classificação de gênero se torna uma das formas mais comuns de categorização do conteúdo. A extração automática de informações em sinais musicais está ganhando importância (QUEIROZ; MARAR; OKIDA, 2015) devido à questão de estruturação e organização do grande número de arquivos de música disponíveis digitalmente na Web.

Apesar disso, a tarefa de organização desses dados não possui um conceito muito claro e definido de forma padronizada para o senso comum. Por consequência, se torna desafiador desenvolver uma única heurística que possa ser totalmente aceita, motivando a exploração de novas técnicas que demonstram resultados interessantes dependendo de diversos atributos à serem considerados no agrupamento.

A maioria dos trabalhos neste campo de pesquisa adota a estratégia de categorização de gêneros musicais utilizando a extração de atributos comuns de sinais musicais(ritmo, melodia e timbre) como uma de suas etapas essenciais (TZANETAKIS; COOK, 2002). O desafio na aplicação destes métodos está relacionado com a escolha de métricas adequadas de similaridade (CORRÊA, 2012), e com a obtenção de formas de representação que sejam ao mesmo tempo compactas e discriminativas.

Dentre esses atributos, o ritmo desempenha um papel muito importante na definição do estilo musical. O estudo da rítmica em sinais de áudio inclui a investigação das características de regularidade de suas transições (MELO, 2019). A análise dos transientes em sinais pode fornecer informações relevantes sobre essa regularidade e, assim, contribuir para o estudo da complexidade rítmica de uma música. A maioria dos trabalhos da área de processamento de sinais tem estudado a relação dos gêneros em música digital utilizando o histograma de batidas. Falta diversidade de descritores de ritmo para sinais de áudio nos estudos atuais, e o campo de processamento de sinais é restrito a técnicas baseadas em representações tempo-frequência.

Atributos relacionados com o timbre e variação temporal (variação do timbre ao longo do tempo), são obtidos através de técnicas de processamento de sinais como as transformadas de Fourier e análises espectrais (SALESSI, 2020). A comparação de desempenho entre abordagens é uma tarefa custosa, uma vez que diferentes taxonomias de gêneros são

empregadas nas tarefas de classificação.

Alguns estudos afirmam que outro fator essencial a ser considerado na discussão sobre classificação de gêneros musicais é que grande parte das composições encontradas na atualidade possuem elementos de mais de um gênero musical, tornando difícil definir com precisão e certeza sobre apenas um grupo (BARBEDO; AMAURI, 2007). Para lidar com esse problema pode-se utilizar técnicas de classificação multirrótulo, onde uma instância pode ser classificada em várias classes, ou até mesmo realizar uma divisão básica de gêneros em uma série de subgêneros.

De modo geral, no ramo de classificação de gêneros musicais, o sinal de áudio passa primeiramente por um pré-processamento a fim de otimizar os dados e transforma-los em uma estrutura própria aos algoritmos de extração de atributos. Feito isso, são calculadas representações numéricas da natureza tonal, rítmica, e timbrística musical que segue, na maioria dos estudos, o modelo adotado por Tzanetakis e Cook (TZANETAKIS; COOK, 2002). Ao final dessa etapa, cada sinal dos arquivos musicais estará representado por um conjunto de números que representam vetores de atributos. Finalmente, o classificador prediz o gênero ou a probabilidade de diferentes gêneros musicais a partir dos vetores de atributos, usando árvores de decisão, modelos probabilísticos e aprendizagem de máquina, onde é feito a decisão sobre um ou mais gêneros musicais, dependendo da técnica de classificação abordada. Vale ressaltar que existem técnicas de classificação que utilizam a extração automática de atributos do sinal, como no caso do ramo de aprendizagem profunda, onde o classificador é alimentado com apenas os dados de entrada, e o programa pode aprender e extrair padrões por meio de seu próprio processamento de dados.

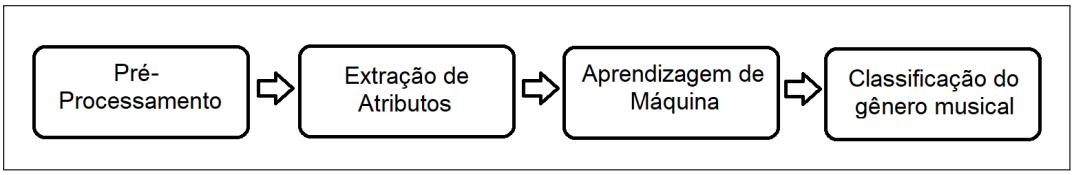


Figura 1 – Processo comum de um sistema de classificação de gêneros musicais. Fonte: Autor

2.2 Propriedades de Som e Áudio

O som é um fenômeno mecânico de propagação de vibrações em um material ou meio de transmissão(geralmente o ar), que pode possuir características perceptíveis ao ouvido humano. Essas vibrações geram ondas flutuantes de baixa ou alta pressão, com regiões de compressão e rarefação dos gases atmosféricos que se intercalam periodicamente, de acordo com a frequência em que a fonte as produz (ROSSING; FLETCHER, 2004).

Assim como todos os fenômenos de onda, o som participa de interações com o mundo através de difração, interferência, reflexão e refração. O número e a complexidade dessas

interações tornam a simulação precisa da transmissão de áudio mais difícil do que a simulação de luz, considerando intervalos de frequência de interesse para a percepção humana.

Um microfone, por exemplo, funciona como um transdutor eletroacústico, que atua como um medidor constante de variações na pressão do ar (CIPRIANI; GIRI, 2010). Ao mesmo tempo, o microfone gera um sinal elétrico correspondente ao original, no sentido de que seu fluxo de saída de tensão elétrica corresponde ao da onda sonora de entrada. Em sinais digitais, o sinal pode ser quantificado e representado por uma série de números (JR, 2014). Cada um desses valores em um sinal digital representa o valor da pressão instantânea, ou seja, o valor da pressão sonora em um dado instante. Para gerar um sinal digital, a amplitude do som é medida em intervalos regulares (taxa de amostragem) e é inteiramente analógico.

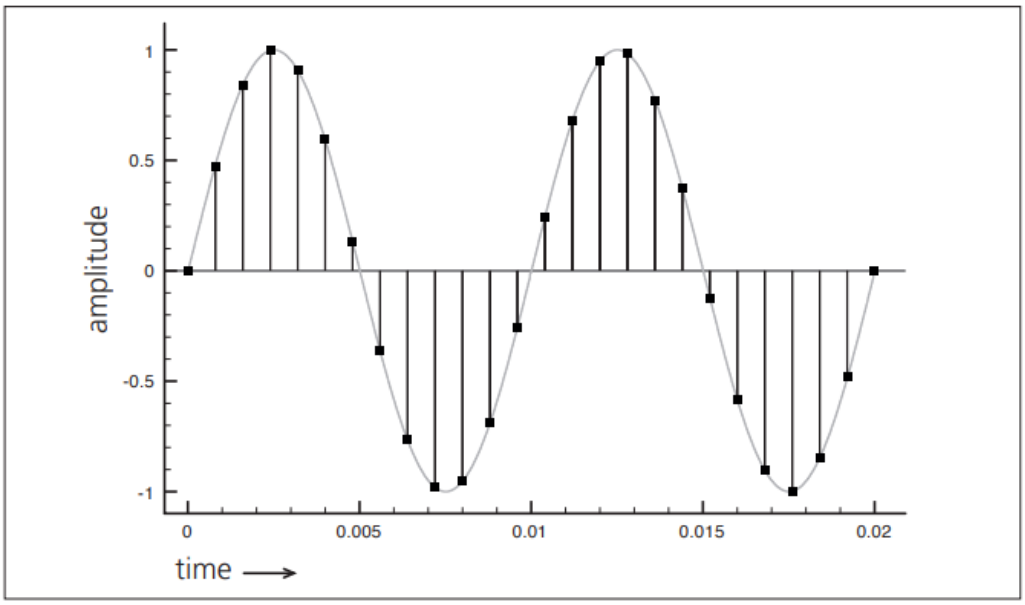


Figura 2 – Exemplo do processo de amostragem do som. Fonte: (CIPRIANI; GIRI, 2010)

Uma vez que as amostras analógicas individuais são organizadas no tempo, cada uma assume o valor da amplitude do sinal naquele instante particular, o qual é convertido em um fluxo de dados numéricos (binários). Esse processo é chamado de conversão analógico-digital. Para permitir a reprodução física do sinal convertido digitalmente, é necessário realizar o processo inverso de conversão digital para analógico, o qual pode ser enviado para um amplificador e então para os alto-falantes.

Um dos conceitos mais importantes para entender de acústica é a frequência, o qual é definida pelo número de vezes em um determinado período que um fenômeno se repete. Em sinais de áudio, a unidade de frequência mais utilizada é o Hertz(Hz). O estudo das frequências se torna particularmente útil quando é observado que qualquer sinal de banda

limitada pode ser decomposto usando a Transformada de Fourier em um número inteiro de tons puros (senoides) (SMITH, 2007).

Ao analisar ou modificar um sinal, é extremamente útil considerar o sinal como uma soma finita de partes simples linearmente separáveis em vez de um todo complexo. O sentido de audição também funciona da mesma forma: o comportamento mais importante do ouvido é separar as frequências.

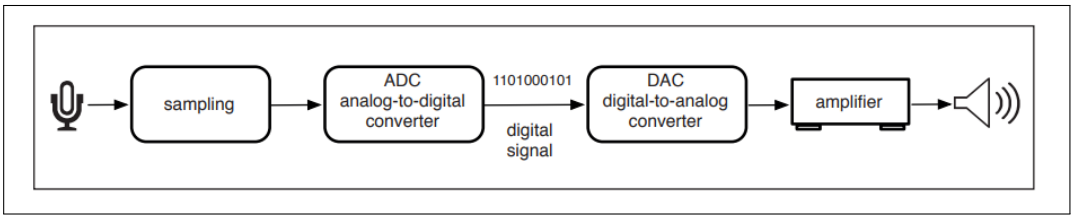


Figura 3 – Exemplo de conversão analógico-digital e digital-analógico. Fonte: (CIPRIANI; GIRI, 2010)

2.3 Extração de Atributos

A extração de atributos é uma etapa crucial dentro do desenvolvimento de sistemas de reconhecimento de padrões (COSTA et al., 2013). No caso de aplicações voltadas para a classificação de sinais de áudio musical, o processo geralmente está relacionado à extração de atributos de ritmo, melodia e timbre (TZANETAKIS; COOK, 2002).

A criação de conteúdo musical envolve na replicação e reinvenção de estruturas de composição que são influenciadas por todos os gêneros, e as músicas compartilham certas características em comum, como a presença de instrumentos similares, arranjos e distribuição das variações de frequência das vibrações de origem (SILVA, 2014).

Amostras de áudio, obtidas através da quantificação da onda sonora, não podem ser utilizadas diretamente por sistemas de analise automática. O sinal pode apresentar uma quantidade de dados muito grande que não são relevantes para a classificação. Sendo assim, o primeiro passo da maioria dos sistemas descritores é a filtragem de algumas caracterısticas dos dados do áudio para manipular informação mais significativa, a fim de reduzir a quantidade de processamento desnecessário posteriormente.

No que tange ao uso destes descritores, observa-se alguns trabalhos pioneiros que servem de base para muitos estudos. Deshpande, Singh e Nam (DESHPANDE; SINGH; NAM, 2001) realizaram o seu estudo utilizando músicas de três distintos gêneros (Rock, Clássica e Jazz), a partir de 12 coeficientes relacionados a Mel-Frequency Cepstral Coeficients (MFCC) e espectrogramas obtidos a partir do sinal do áudio. Para a tarefa de classificação, os algoritmos aplicados foram: a) K-Nearest Neighbour (KNN); b) Modelo Gaussiano; e c) Support Vector Machine (SVM). A melhor precisão de classificação entre os três gêneros foi utilizando KNNs, obtendo cerca de 75% de acurácia. O modelo

gaussiano não funcionou muito bem, evidenciando que a suposição de que a distribuição para cada categoria ser gerada por um gaussiano não está correta. SVM teve melhores resultados na identificação de música Clássica, onde foi possível distinguir entre música Clássica e Não-Clássica com uma precisão de 90%. No entanto, seu desempenho na identificação entre Rock e Clássica não foi tão bom. O classificador com KNN se saiu melhor na categorização de amostras de Clássica do que com amostras de Rock ou Jazz. Algumas músicas em particular foram classificadas erroneamente por todos os classificadores. Frequentemente, peças de Jazz que continham piano eram confundidas com Clássica pela maioria dos classificadores.

Em (TZANETAKIS; COOK, 2002), são utilizados algoritmos de identificação de padrões em sinais de voz para classificar sinais de áudio musicais. Os algoritmos com características de natureza timbrística (Cetroide Espectral, Rollof, Passagem pelo Zero, MFCCs) estão relacionadas às frequências que dão "colorido" ao som. As características tonais (FPO, UPO, IPO1, SUM, FP0) estão relacionadas com as alturas das notas musicais e remetem à sensação auditiva de afinação tonal, acordes e melodias. As características de natureza rítmica (Onsets, Histograma de Batidas, BPM) estão ligadas ao início das notas musicais, número de batidas por unidade de tempo, compasso, regularidade e pulsação da música. É através dessas características que sabemos se uma música é mais "rápida" ou mais "lenta" do que a outra.

Para o estudo deste trabalho, será utilizado e comparado descritores com atributos rítmicos construídos a partir da detecção de Onsets, que referem-se ao início de eventos acústicos do sinal. Em contraste com os estudos que se concentram na detecção de batidas e ritmo por meio da análise de periodicidades, um detector de Onset enfrenta o desafio de detectar eventos únicos, que não precisam seguir um padrão periódico, mas que tenta encontrar mudanças repentinas na dinâmica, timbre ou estrutura harmônica do sinal (GOUYON et al., 2006). Mais precisamente, será calculado o envelope de intensidade de Onset de fluxo espectral de cada sinal.

2.4 Grafos

Um grafo é uma estrutura capaz de representar informações sobre relações de conjuntos. Define-se um grafo por G = (V, E), onde V = V (G) é o conjunto finito de objetos chamados vértices e E = E(G) são pares não ordenados de vértices, denominado de arestas (WALLIS, 2007).

Por possuir uma estrutura possível de armazenar informações complexas e com conceito de fácil implementação, atualmente é encontrado para definir relações em um enorme volume de dados. É facilmente integrada com sistemas que utilizam métodos tradicionais de predição e aprendizado de máquina (JURKIEWICZ, 2009).

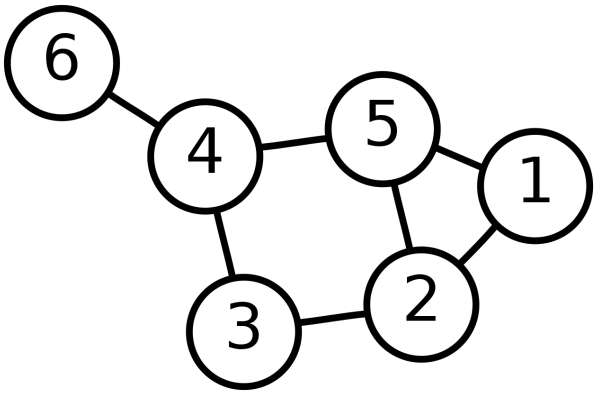


Figura 4 – Exemplo de um grafo com 6 vértices. Fonte: Autor

Os grafos que apresentam estruturas topográficas não triviais geralmente são denominados Redes Complexas, cujos padrões de conexão entre seus elementos não são completamente regulares e nem completamente randômicos (RAVASZ; BARABÁSI, 2003).

Nos últimos anos, muitos pesquisadores de áreas distintas como ciência da computação, biologia e ciências sociais, têm encontrado uma grande variedade de sistemas que podem ser representados na forma dessas redes, e seu estudo têm sido incentivado cada vez mais (NEWMAN; GIRVAN, 2004).

2.5 Métodos de Transformação de Séries Temporais em Grafos

Uma série temporal é uma sequência de pontos de dados tempo de observações, que são feitas em pontos sucessivos, em muitos casos igualmente espaçados no tempo. Desta forma, os dados de séries temporais têm uma ordenação temporal discreta natural (ZOU et al., 2019). Séries temporais cobrem uma grande variedade de variáveis potencialmente relevantes para a vida cotidiana e são muito utilizadas para análise de dados.

Em (CAMPANHARO et al., 2011), é frisado a importância de análise estatística em séries temporais e suas redes relacionadas, propondo também um mapa de uma série temporal para uma rede com uma operação inversa aproximada, possibilitando o uso de estatísticas de rede para caracterizar séries temporais e estatísticas de séries temporais para caracterizar redes.

Uma classe importante de aplicações não tradicionais da teoria de redes complexas são as redes funcionais, onde a conectividade considerada não necessariamente se refere a vértices e arestas "físicas", mas reflete inter-relações estatísticas entre a dinâmica exibida por diferentes partes do sistema sob estudar (ZOU et al., 2019).

Para tornar as séries temporais acessíveis a técnicas complexas de análise de redes e aprendizado de máquina, primeiramente é necessário encontrar uma representação de

rede adequada, ou seja, um algoritmo que defina quais são os vértices e arestas da rede. De acordo com (ZOU et al., 2019), existem pelo menos (mas não se limita a) três classes principais de abordagens comuns de redes complexas para a análise de séries temporais individuais:

- Similaridade estatística mútua ou proximidade métrica entre diferentes segmentos de uma série temporal (Redes de Proximidade)
- Probabilidades de transição entre estados discretos (Redes de Transição)
- Convexidade de observações sucessivas (Grafos de Visibilidade)

A primeira classe faz uso de semelhanças ou relações de proximidade entre diferentes partes da trajetória de um sistema dinâmico (MARWAN et al., 2009), incluindo abordagens de redes de ciclo (ZHANG et al., 2008), redes de correlações (YANG; YANG, 2008), e redes de espaço de fase baseadas em uma certa definição de vizinhos mais próximos (XU; ZHANG; SMALL, 2008).

A segunda classe representa as redes de transição, que fazem uso de ideias de dinâmicas simbólicas e processos estocásticos. Transformar uma determinada série temporal em uma rede de transição é um processo de mapeamento da informação temporal em uma cadeia de Markov, para obter uma representação compactada ou simplificada da dinâmica original.

A última classe representa os Grafos de Visibilidade, que demonstram resultados importantes em dados de séries temporais, conforme observado nos estudos de (LACASA et al., 2008), (LUQUE et al., 2009) e (DONNER; DONGES, 2012). O grafo de visibilidade e suas várias variantes têm aplicações importantes, como auxiliar em processamento de imagens (IACOVACCI; LACASA, 2019), testes estatísticos para irreversibilidade de séries temporais (LACASA et al., 2012) e também em especial na questão de classificação de gêneros musicais (MELO; FADIGAS; PEREIRA, 2020).

Uma vez que os grafos de visibilidade serão adotados na metodologia desta pesquisa, uma descrição detalhada sobre esse método será dada na próxima seção.

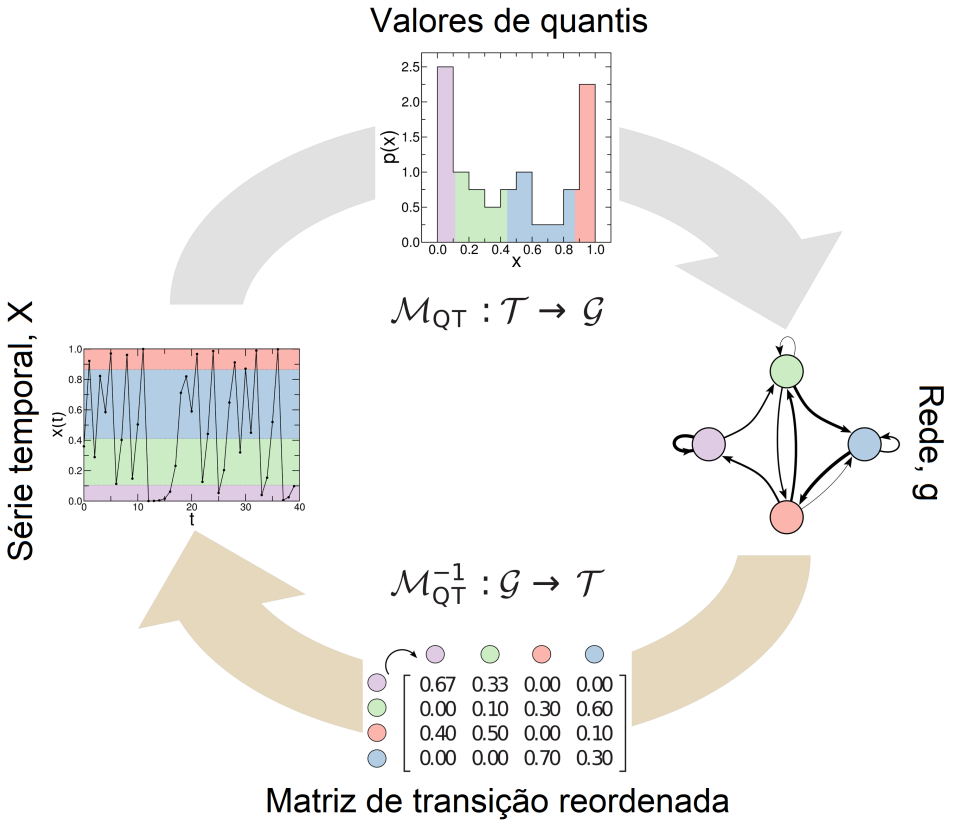


Figura 5 – Mapeamento e relação de séries temporais e redes operação inversa. Uma série temporal é dividida em quantis (sombreamento colorido) e cada quantil é atribuído a um vértice na rede correspondente. Os pares de vértices são então conectados na rede com uma aresta direcionada, onde o peso é dado pela probabilidade de um ponto em um quantil ser seguido por um ponto em outro quantil. Transições repetidas entre quantis resultam em arcos na rede com pesos maiores (representados por linhas mais grossas). Fonte: (CAMPANHARO et al., 2011)

2.6 Grafos de Visibilidade

Os grafos de visibilidade têm criado pontes entre a análise de séries temporais e a análise de redes complexas, possibilitando o uso de novas ferramentas para a compreensão de fenômenos representados por sequências temporais.

Por exemplo, o uso de grafos de visibilidade foi explorado em aplicações médicas (SUPRIYA et al., 2016), previsões de séries temporais (CHEN et al., 2014), séries financeiras (CHEN et al., 2019) e processamento de imagens (IACOVACCI; LACASA, 2019).

Podem ser definidos por redes geradas a partir de séries numéricas, onde cada ponto da série é considerado um vértice do grafo, e a ligação ou não entre dois vértices depende da "visibilidade" entre os pontos da série (LACASA et al., 2008). A visibilidade entre dois pontos é definida por um critério trigonométrico aplicado aos pontos da série. Nos

grafos de visibilidade quanto maior o grau de conexão de um determinado vértice, maior é a visibilidade do seu ponto correspondente na série, em relação à sua vizinhança.

Segundo (LACASA et al., 2008), o critério de visibilidade pode ser definido da seguinte forma: dada uma série temporal V1, V2, ..., Vn sempre haverá visibilidade entre dois pontos consecutivos da série temporal, e dois pontos arbitrários A(xa, Va) e B(xb, Vb) da série terão visibilidade mútua, se todo ponto C(xc, Vc) entre eles satisfaz a condição:

$$\frac{Vb - Vc}{xb - xc} > \frac{Vb - Va}{xb - xa} \tag{2.1}$$

Devido à natureza dos grafos de visibilidade, as divisões de vértices correspondam aos pontos mais baixos da série temporal (onde a visibilidade direta entre os pontos de dados é mais fácil) e os picos regionais da série provavelmente dividem os segmentos. Dessa forma, o resultado da segmentação da série temporal também reflete esse comportamento, o qual pode trazer relações interessantes entre as características de persistência de transientes de sinal e as características topológicas de detecção de comunidades em seus grafos associados. Isso pode sugerir, por exemplo, que gêneros musicais com picos de sinais mais definidos e repetitivos (como no caso de músicas eletrônicas) podem gerar um maior número de vértices após a transformação. Além disso, alguns vértices funcionam como "hubs"do grafo (os vértices mais conectados), representando os dados com os maiores valores na série.

Nos trabalhos de (MELO, 2019; MELO; FADIGAS; PEREIRA, 2020), foi proposto um novo descritor baseado em propriedades topológicas utilizando grafos de visibilidade, denominado de Descritor de Visibilidade em Flutuações de Variância (DVFV). O descritor apresentou ótimos resultados ao analisar cálculos de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$) para detecção de padrões de autosimilaridade nos sinais de dados musicais e classificou gêneros musicais com alta taxa de acurácia.

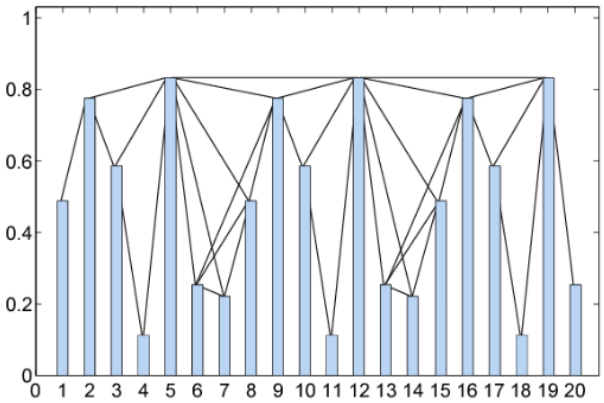


Figura 6 – Exemplo do mapeamento de uma série temporal em um grafo de visibilidade. Fonte: (LACASA et al., 2008)

2.6.1 Grafos de Visibilidade Horizontal

Os grafos de visibilidade possuem uma gama de variantes com diversas aplicações. Uma versão geometricamente mais simples foi proposta (LUQUE et al., 2009), de forma a ser computacionalmente mais eficiente do que a versão do algoritmo normal de Grafos de Visibilidade (também chamados de Grafos de Visibilidade Natural), com o foco também no mapeamento de séries temporais.

Sua visibilidade, mais restrita do que o caso geral análogo, é analisada através de retas horizontais e, por isso, é nomeado como Grafo de Visibilidade Horizontal (GVH). Ou seja, os vértices do GVH terão menos visibilidade do que no caso de um Grafo de Visibilidade Natural. Embora tal fato possa ter um leve impacto sobre os aspectos qualitativos dos grafos, a simplicidade da versão do algoritmo horizontal permite que ele seja aplicado, em resolução de um tempo computacional mais otimizado, à séries muito longas, como é o caso de alguns bancos de dados de informação musical.

Assim, estabelece-se o seguinte critério de visibilidade: dois valores arbitrários da série temporal A(xa, Va) e B(xb, Vb) terão visibilidade horizontal e, consequentemente, se tornarão dois vértices conectados por uma aresta no grafo associado, se todos os outros termos (xc, Vc) intermediários entre eles cumprirem a relação

$$xa, xb > xc \tag{2.2}$$

para todo c, tal que

$$a < c < b \tag{2.3}$$

É relevante observar também que se dois vértices tiverem visibilidade horizontal, eles também terão visibilidade natural direta (mas não o contrário) e, portanto, o Grafo de Visibilidade Horizontal de uma determinada série temporal sempre será um subgrafo de seu Grafo de Visibilidade Natural. Embora esse fato não tenha impacto nos atributos qualitativos dos grafos, o GVH normalmente terá menos estatística de forma quantitativa.

No caso de Grafos de Visibilidade Natural, o número total de verificações necessárias para obter o grafo de uma série temporal de n pontos de dados é igual a n(n-1)/2, correspondendo a uma complexidade de tempo $O(n^2)$. Já na análise de visibilidade horizontal, pode-se dar um passo adiante e assumir com segurança que nenhum ponto após um valor maior do que um valor computado será visível horizontalmente (YELA et al., 2020). Esta observação reduz efetivamente a complexidade do tempo da construção para O(n log(n)) e, no caso de sinais ruidosos (estocásticos ou caóticos), pode-se inferir que este algoritmo tem uma complexidade de tempo de caso médio O(n). No entanto, todos os pares de pontos precisam ser verificados no caso de visibilidade natural.

A simplicidade da versão horizontal do algoritmo, que é computacionalmente mais rápida que o original, deve ser usada como um teste preliminar confiável ao procurar impressões digitais determinísticas em séries temporais, como encontrar alguns elemen-

tos rítmicos mais dissipantes entre estilos musicais. Além disso, pode ser um recurso interessante na extração de atributos rítmicos em banco de dados de larga escala, com uma quantidade enorme de arquivos de áudio, que tem um requerimento de poder de processamento mais elevado.

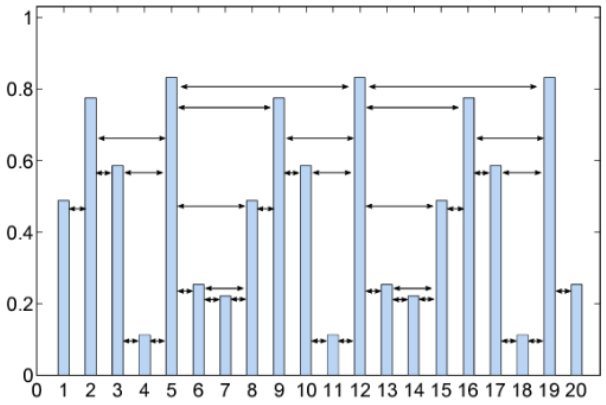


Figura 7 – Exemplo do mapeamento de uma série temporal em um grafo de visibilidade horizontal. Fonte: (LACASA et al., 2008)

2.7 Detecção de Comunidades

Um tópico de grande interesse no estudo de redes complexas e no ramo de análise de dados é a identificação de comunidades, uma vez que permite uma visão da relação funcional e estrutural de uma rede, e possui abrangentes áreas de aplicação.

As comunidades são identificadas quando há uma quantidade maior de ligações entre vértices de um mesmo subgrupo da rede, e quantidade menor de ligações entre vértices que pertencem a subgrupos diferentes (LIU; BARABÁSI, 2016).

Pode haver qualquer número de comunidades em uma determinada rede, com diversas configurações de tamanho, o que dificulta muito o procedimento de detecção desses grupos. No entanto, existem muitas técnicas sendo exploradas recentemente e há algumas propostas nesse domínio.

Os métodos de detecção de comunidade podem ser amplamente categorizados, no geral, em Métodos Aglomerativos e Métodos Divisivos. Nos métodos Aglomerativos, as arestas são adicionadas uma a uma em um grafo que contém apenas os vértices. Os métodos Divisivos seguem o oposto dos métodos aglomerativos, onde as arestas são removidas uma a uma de um grafo completo.

No intuito de mensurar o quão bem formados são os agrupamentos encontrados nas redes, foi proposto o conceito de Modularidade(Q), para estabelecer um indicador sobre as divisões das comunidades. Pode ser definida como a medida da força de divisão de uma rede em módulos (ou comunidades). Quanto maior o valor de M (modularidade), mais forte é a conexão entre os vértices dentro dos módulos.

A maximização da modularidade gananciosa (CLAUSET; NEWMAN; MOORE, 2004), utilizada no estudo deste trabalho, começa com cada vértice em sua própria comunidade e junta repetidamente o par de comunidades que causam o aumento de modularidade até obter o maior valor possível.

A partir do trabalho onde a modularidade foi proposta, iniciou-se uma corrida para otimizar e desenvolver novos algoritmos para a detecção de comunidades. Essa nova área permitiu o desenvolvimento de outras linhas de pesquisa como a detecção de padrões em comunidades, interações entre comunidades, algoritmos eficientes para identificação de comunidades, entre outros. Essas propriedades de redes podem ser utilizadas à favor da classificação de gêneros musicais, conforme é proposto e apresentado neste trabalho.

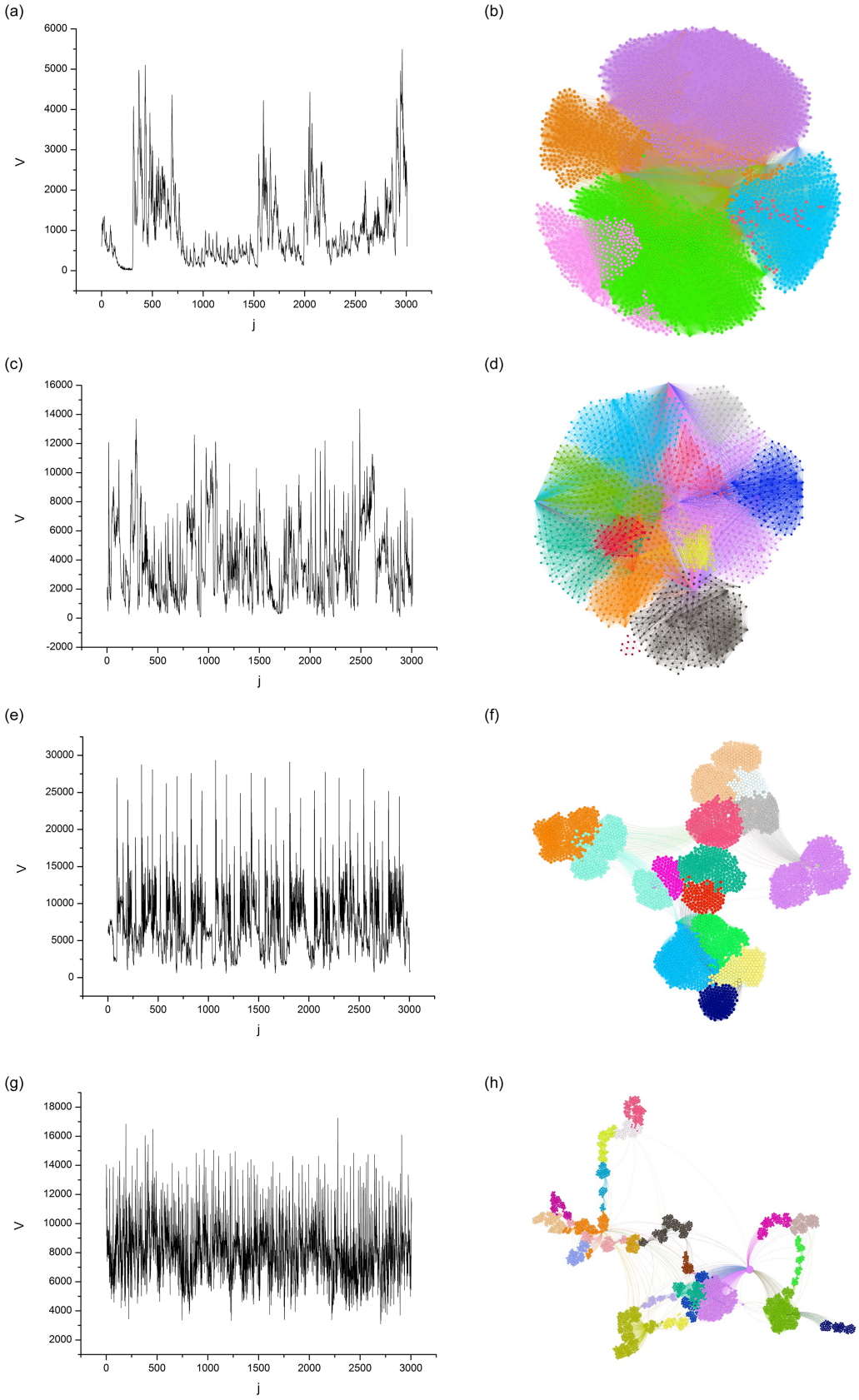


Figura 8 – Representações de sinais de áudio e seus grafos de visibilidade. As cores representam as comunidades, que são obtidas pela modularidade. (a)(b) Gênero Clássico, (c)(d) Gênero Blues, (e)(f) Gênero Pop e (g)(h) Gênero Metal. Fonte: https://doi.org/10.1371/journal.pone.0240915.g006

3 Trabalhos Relacionados

A seguinte seção apresenta alguns trabalhos relacionados que auxiliaram no desenvolvimento e conclusões deste estudo, considerando os critérios de classificação para gêneros musicais, extração de atributos de sinais musicais de diversas naturezas, análise de redes complexas e técnicas de aprendizado de máquina.

3.1 Musical genre classification of audio signals

No trabalho de (TZANETAKIS; COOK, 2002), foi proposto a classificação automática de sinais de áudio em uma hierarquia de gêneros musicais. Foram propostos três conjuntos de recursos para textura de timbre, conteúdo rítmico e melodia. O desempenho e a importância relativa dos recursos propostos foram investigados treinando classificadores estatísticos de reconhecimento de padrões usando coleções de áudio do mundo real.

O vetor de atributos consistiu com recursos de textura de timbre de 19 atributos (MFCCs, Fluxo Espectral, Taxa de Passagem Pelo Zero, Centróide Espectral, Rollof Espectral e Low-Energy), 6 atributos de conteúdo rítmico (Histograma de Batidas) e 5 atributos de conteúdo de tom (Histograma de Tom), resultando em um vetor de recursos de 30 dimensões.

O trabalho evidenciou que embora certas características individuais de atributos de classificação estejam correlacionadas, a adição de cada recurso específico melhora, de modo geral, a precisão da classificação dos grupos. Os conjuntos de recursos de conteúdo rítmico (Histograma de Batidas) e de tom (Histograma de Tom) parecem desempenhar um papel menos importante na classificação em comparação com atributos de natureza de timbre (STFT, MFCCs) em todos os casos. No entanto, os conjuntos de recursos propostos funcionam melhor do que a classificação aleatória e, portanto, fornecem algumas informações sobre gênero e sobre o conteúdo musical em geral.

Os resultados mostraram que a classificação alcançou os menores índices de acerto para os gêneros de Rock (40%) e Blues(43%), enquanto os gêneros Jazz(75%), Clássico(69%), Pop(66%) e Hiphop(64%) ficaram entre as categorias de maior acurácia. Através da análise da matriz de confusão da classificação, as taxas de resultados incorretos do sistema são semelhantes ao que um humano faria. Por exemplo, a música clássica é erroneamente classificada como música Jazz para peças com ritmo forte de compositores como Leonard Bernstein e George Gershwin. A música de Rock tem a pior precisão de classificação e é facilmente confundido com outros gêneros, o que é esperado, considerando sua natureza ampla.

De forma geral, por existir diversos conjuntos de atributos relevantes em sinais de áudio

para representar timbre, ritmo e harmonia, diferentes tipos de recuperação de similaridade são possíveis em grupos de classificadores.

3.2 Categorisation of polyphonic musical signals by using modularity community detection in audio-associated visibility network

Em 2017, (MELO; FADIGAS; PEREIRA, 2017) propõe um método para caracterizar numericamente a homogeneidade de sinais musicais polifônicos por meio da detecção de comunidades em redes de visibilidade associadas ao áudio e detectar padrões que permitam a categorização desses sinais em dois tipos de agrupamento de natureza rítmica. Observou-se que uma maior ou menor homogeneidade das magnitudes dos transientes de sinal está relacionada a uma maior ou menor modularidade de sua rede de visibilidade associada. Notou-se também que essas diferenças estão relacionadas a escolhas musicais que podem estabelecer diferenças importantes entre os estilos musicais.

No estudo foram utilizados 120 arquivos musicais das categorias Sinfônica e Percussiva, cada uma com 60 músicas. Para a música sinfônica, foram selecionadas peças de quarteto de cordas e orquestra completa. As composições incluíam concertos de Bach, sinfonias de Mozart e quartetos de Debussy, Dutileux e Ravel. A categoria Percussiva é composta por canções igualmente retiradas de seis gêneros com forte influência e execução persistente de instrumentos acústicos e de percussão eletrônica: Samba, Forró, Axé, Mangue Beat, Disco e Trance. As faixas de Samba são canções compostas para a celebração do carnaval do Rio de Janeiro de 2005 a 2014. No Mangue Beat, há uma influência da música Pop Rock Eletrônica misturada com um ritmo tradicional afro-brasileiro chamado Maracatú. As faixas da música Disco fornecem um bom panorama da cena musical dos anos 80. Axé music e Forró são ritmos e danças brasileiras tradicionalmente utilizadas em festas populares como carnaval e festas rurais na região nordeste do país. Trance representa a música eletrônica com uma batida intensa e dançante, universal e especialmente utilizada em eventos de entretenimento juvenil. As músicas Sinfônicas e Disco foram escolhidas do banco de dados GTZAN, e as faixas de Samba, Axé, Mangue Beat, Trance e Forró são do acervo pessoal do autor.

Após os experimentos, conclui-se que as redes percussivas apresentam tendência com média de alta Modularidade(0.836) e Número de Comunidades(19), além de apresentar baixo Grau Médio(15.12), por tendências que estão fortemente ligadas às escolhas musicais que influenciam o design dos transientes do sinal de cada gênero. Por outro lado, as redes sinfônicas tendem a apresentar baixa Modularidade(0.558) e Número de Comunidade(8.5), além de apresentar alto Grau Médio(46.27). Isso se deve porque as músicas Sinfônicas usam muito mais variações na dinâmica e menos persistência rítmica do que as músicas

Percussivas, resultando em sinais mais heterogêneos e grafos de visibilidade com valores de menores de Modularidade.

3.3 Music recommender system based on genre using convolutional recurrent neural networks

Em 2019, (GUNAWAN; SUHARTONO et al., 2019) foi realizado um estudo com redes neurais recorrentes convolucionais (CRNNs) para extração de recursos e relações de similaridade entre os sinais de áudio. CRNNs é uma combinação de redes neurais convolucionais (CNNs) e redes neurais recorrentes (RNNs). As CNNs são especialmente adequadas para prever recursos musicais de alto nível, como acordes e batidas, porque permitem uma estrutura hierárquica que consiste em recursos intermediários em várias escalas de tempo. Já as RNNs foram projetadas para trabalhar com dados de séries temporais, especialmente para problemas de previsão de sequência temporal. RNNs têm mais resultados ao trabalhar com sequências de palavras e parágrafos no processamento de linguagem natural.

No estudo, foi comparado o desempenho de arquiteturas de CNNs com CRNNs para classificar gêneros musicais. O modelo toma como entrada o espectrograma de quadros musicais e analisa a imagem usando CRNNs. A saída do modelo é um vetor de gêneros previstos para a música. O principal resultado do estudo é que a precisão dos CRNNs é ligeiramente superior aos métodos de CNNs que combina os domínios da frequência e do tempo e usa o mesmo número de parâmetros.

O conjunto de dados vem da coleção de músicas Free Music Archive (FMA). Para a pesquisa, foi utilizado coleções que totalizam mais de 25.000 músicas em formato mp3. Cada música tem duração de 30 segundos e foi classificada entre um conjunto de 16 gêneros musicais. Os atributos foram extraídos utilizando uma taxa de amostragem de 22050Hz, comprimento de janela de 2048 com filtro Hanning e comprimento de salto de 512.

O aplicativo de recomendação de música foi desenvolvido com a linguagem de programação Python e auxílio de bibliotecas também utilizadas no desenvolvimento deste projeto, como o Librosa, para análise do sinal e extração de atributos, e também Tensorflow e Keras para a construção dos modelos de aprendizado de máquina.

Os resultados mostraram que a classificação alcançou os menores índices de acerto para os gêneros de Eletrônico (64% com CNN e 70% com RNN) e Instrumental (79% com CNN e 75% com RNN). Por outro lado, obteve uma acurácia surpreendente próximo de 94% para ambas as redes para o gênero de Rock, onde outros trabalhos relacionados mostraram dificuldade em categorizar.

Dessa forma, evidenciou-se que um sistema de recomendação de música deve considerar as informações do gênero musical para aumentar a qualidade das classificações musicais.

Também é interessante observar que o modelo de rede utilizado e os atributos de entrada do classificador podem afetar de diferentes maneiras entre os diversos gêneros musicais.

3.4 Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain

Em 2020, o estudo de (MELO; FADIGAS; PEREIRA, 2020) demonstrou um novo método para extrair características de sinais de áudio para classificar músicas. A tese propõe um novo descritor baseado em propriedades topológicas utilizando grafos de visibilidade, denominado de Audio Signal Visibility Descriptor (ASVD). O descritor apresentou ótimos resultados ao analisar cálculos de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$) para detecção de padrões de auto-similaridade nos sinais de dados musicais e classificou gêneros musicais com alta taxa de acurácia.

A base de dados utilizada foi a Coleção de Gêneros GTZAN, o qual corresponde à mesma adotada por este trabalho. Este banco de dados consiste em dez gêneros musicais (Clássico, Jazz, Blues, Pop, Rock, Hip-hop, Metal, Disco, Reggae e Country), cada um com 100 arquivos de áudio, taxa de amostragem de 44.100 Hz e quantização de 16 bits. Esta base de dados foi proposta e tem sido utilizada em muitos estudos envolvendo a recuperação de informações musicais. A base de dados GTZAN tem se consolidado como importante referência no estudo da classificação de gêneros musicais.

No trabalho, foi realizado uma comparação mais direta de propriedades de natureza rítmica, e obteu-se uma precisão maior ou igual ao histograma de batidas em 70% dos pares de gêneros musicais, onde foi evidenciado que as quatro propriedades utilizadas da rede estavam entre as primeiras posições atribuídas pelo teste.

Esses resultados mostram que, nesse caso, a classificação dos gêneros musicais usando o ASVD para extração da atividade rítmica em vez do histograma de batidas resultou em um sistema de categorização com melhor taxa de acerto geral e individual. Essa comparação é muito importante porque é feita sobre atributos que tem sido usados como referência em muitos estudos de recuperação de informações musicais, e isso mostra como o novo método de extração de características usado no trabalho pode ser bem-sucedido em relação a um método tradicionalmente usado na literatura.

Em um sistema de classificação usando apenas o descritor de propriedades dos grafos, obteu-se uma precisão média de 39%. Foi feita a comparação das instâncias classificadas corretamente por este sistema com outro sistema utilizando apenas o histograma de batidas, e então, em uma comparação pairwise de gêneros, obteu-se uma precisão maior ou igual ao segundo sistema em 70% dos pares de gêneros musicais. Considerando um cenário com 18 atributos de processamento de sinal de áudio mais o ASVD, a precisão

média da classificação foi de 76,7%, comparável ou superior a vários estudos relacionados. Em mais um experimento de classificação usando os mesmos 18 atributos do experimento anterior, e usando o histograma de batidas em vez do ASVD, foi obtido uma precisão igual ou superior em metade dos dez grupos de gêneros musicais.

Para a simplificação e possível comparação futura de resultados, este trabalho também adotará os mesmos princípios denominados pelo descritor ASVD, relacionando as quatro propriedades de redes obtidas para ambos os tipos de Grafo de Visibilidade Natural e Grafo de Visibilidade Horizontal, demonstrando quais implicações suas definições causam na classificação de um gênero musical.

3.5 Comparativo

As diferenças mais relevantes entre os trabalhos analisados foram: os parâmetros extraídos dos sinais de áudio, a base de dados utilizada e o classificador implementado.

Percebe-se que os atributos utilizados nos trabalhos possuem naturezas distintas e a combinação variada dessas propriedades resulta em diferentes taxas de acurácia para cada gênero musical, visto que cada cada classe apresenta variadas características de timbre, tom e rítmo.

Outro ponto importante é o extenso uso de MFCCs, dos trabalhos relacionados analisados, todos concluíram que esses parâmetros são um importante avanço para a classificação de gêneros em sinais musicais. Em relação aos dados utilizados, as publicações citadas conseguiram resultados relevantes com ambos dados de base públicas e dados próprios.

A Tabela a seguir explicita alguns parâmetros importantes considerados por cada artigo citado.

Trabalho	Parâmetros	Classificadores	Base de Dados	Acurácia máxima
Tzanetakis; Cook, 2002	MFCCs, STFT, PHF, BHF	GS, GMM, K-NN	GTZAN	61%
Gunawan; Suhartono et al., 2019	MFCCs, STFT	CRNNs, CNN	FMA	71%
Melo; Fadigas; Pereira, 2020	MFCCs, BHF, ASVD	ANN	GTZAN	76.7%

Tabela 1 – Comparação entre trabalhos citados. Fonte: Autor.

4 DESENVOLVIMENTO

4.1 Banco de Dados

A pesquisa utilizou dados do banco GTZAN, criado por (TZANETAKIS; COOK, 2002), o qual oferece 100 arquivos de áudio, divididos em 10 gêneros musicais (Clássico, Jazz, Blues, Pop, Rock, Hip-Hop, Metal, Disco, Reggae, Country). Esse banco de dados tem sido usado em larga escala nas pesquisas de gêneros musicais, permitindo a análise de composição dos sinais de áudio e disponibilizando valores de propriedades de natureza timbrística, rítmica e tonal, compatíveis com entradas de sistemas de aprendizado de máquina.

Esta base de dados foi escolhida com o objetivo de favorecer a reprodutibilidade da metodologia e a comparação com outros trabalhos científicos mais recentes, tendo em vista que é uma grande base de dados pública que está em processo de estudo em pesquisas relacionadas ao reconhecimento automático de gêneros musicais.

4.2 Processos da metodologia

A proposta deste trabalho consiste na análise de dados de áudio, com base nos estudos de trabalhos relacionados apresentados anteriormente, explorando um possível descritor para a classificação de gêneros musicais sob a perspectiva de Grafos de Visibilidade Horizontal. A metodologia da tese pode ser dividida nos seguintes passos:

- Extrair e otimizar as séries temporais de arquivos de áudio, com base na intensidade do sinal.
- Mapear as séries temporais obtidas em Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal.
- Quantificar propriedades topológicas das redes obtidas, de forma à representar numericamente características relevantes à categorização dos gêneros musicais através do cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$).
- Desenvolver uma rede neural artificial para a classificação supervisionada, que é alimentada pelos dados encontrados sobre ambos os tipos de grafos.
- Classificar os gêneros com base nos grupos encontrados, comparando resultados entre os descritores encontrados e também com descritores relacionados à esta pesquisa.

- Divulgar os resultados.

De forma geral, cada arquivo de música é transformado em uma série temporal U(i) com todos os pontos da amostragem, reduzindo cada uma dessas séries em subséries V(j) com um tamanho fixo de pontos, através do cálculo de suas variâncias. Feito isso, cada nova série é transformada em dois grafos distintos através do mapeamento de visibilidade natural e horizontal. Tendo o resultado em forma de grafos, foi utilizado métodos de quantificação de propriedades de redes complexas, como a detecção de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$) que podem apresentar relações interessantes sobre a categorização dos gêneros. Após a análise dessas propriedades é construído um modelo de rede neural artificial para que possa ser realizada a classificação, utilizando aprendizado de máquina, onde é comparado os resultados entre ambos grafos de visibilidade natural e horizontal, além de comparar com o atributo de detecção de Onsets do sinal, o qual também pode evidenciar características rítmicas da música.

4.2.1 Extração da série temporal

A extração da série temporal consiste em analisar sinais de arquivos de áudio de forma otimizada, de modo à obter informações essenciais da série com o mínimo de dados possível. Estudos mostram que não existe grande diferença entre o uso de diferentes taxas de amostragem nos sinais de áudio (MELO, 2019), produzindo estatísticas muito semelhantes para fins de estudo comparativo usando a abordagem de grafos de visibilidade. A redução da taxa tem a vantagem de otimizar o tempo de processamento quando aplicado em grandes bases de dados, e também permite que diversos tipos de experimentos sejam realizados mesmo em sistemas que tenham restrições quanto ao grande número de vértices produzidos no contexto de sinais de áudio. Sendo assim, os sinais serão reduzidos à uma taxa de 11.000Hz, considerando uma duração de 30s por arquivo.

Com base nos dados obtidos, o sinal é dividido em blocos de tamanho fixo e é definido uma série de variâncias entre os pontos de cada bloco, de forma à otimizar a análise com uma representação reduzida do sinal de áudio (JENNINGS et al., 2004).

Seja Ui = U1, U2,...,Un a série temporal de amostras que representam o sinal de áudio. O número total de pontos N é a função N = A × t, onde A é a taxa de amostragem e t é a duração total do sinal. Essa série é dividida em subséries Vj = V1, V2,...,Vn de tamanho fixo de amostras com m=450 pontos cada, calculando a variância de cada subsérie, o qual é representado na equação:

$$Vj = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda}(Ui - Uj)^2}{\lambda - 1} \tag{4.1}$$

onde

$$Uj = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} Ui}{\lambda} \tag{4.2}$$

Sendo assim, cada subsérie de variância V(j) é criada com 700 pontos. Vale ressaltar que dependendo da divisão do tamanho do bloco, o processo de transformação e manipulação do sinal em redes será diretamente afetado e, por consequência, as propriedades topológicas podem demonstrar diferentes formas de agrupamento dos gêneros musicais.

4.2.2 Transformação da série V(j) em grafos de visibilidade

Cada um dos pontos da série V(j) é considerado como um vértice da rede. Dois vértices da rede são conectados por uma aresta cada vez que dois pontos da série V(j) atendem o critério de visibilidade definido por cada uma das equações de Grafo de Visibilidade Natural e Grafo de Visibilidade Horizontal.

Observe que, devido à natureza dos grafos de visibilidade, as divisões de vértices correspondam aos pontos mais baixos da série temporal (onde a visibilidade direta entre os pontos de dados é mais fácil) e os picos regionais da série provavelmente dividem os segmentos (VARELA, 2020). Portanto, o resultado da segmentação da série temporal também refletirá esse comportamento.

Esse comportamento também é observado em estudos clássicos (TZANETAKIS; COOK, 2002) dessa área, referindo essa característica pela observação da intensidade dos picos. Os autores concluem que composições de arquivos de áudio que possuem trechos musicais com batidas mais fortes e persistentes irão gerar sinais com picos mais elevados e, consequentemente, segmentar os grupos encontrados no grafo de visibilidade. Da mesma forma, quanto menor a persistência e força dos batimentos principais, maior a chance dessas amostras participarem de um mesmo grupo. (BORGES et al., 2010)

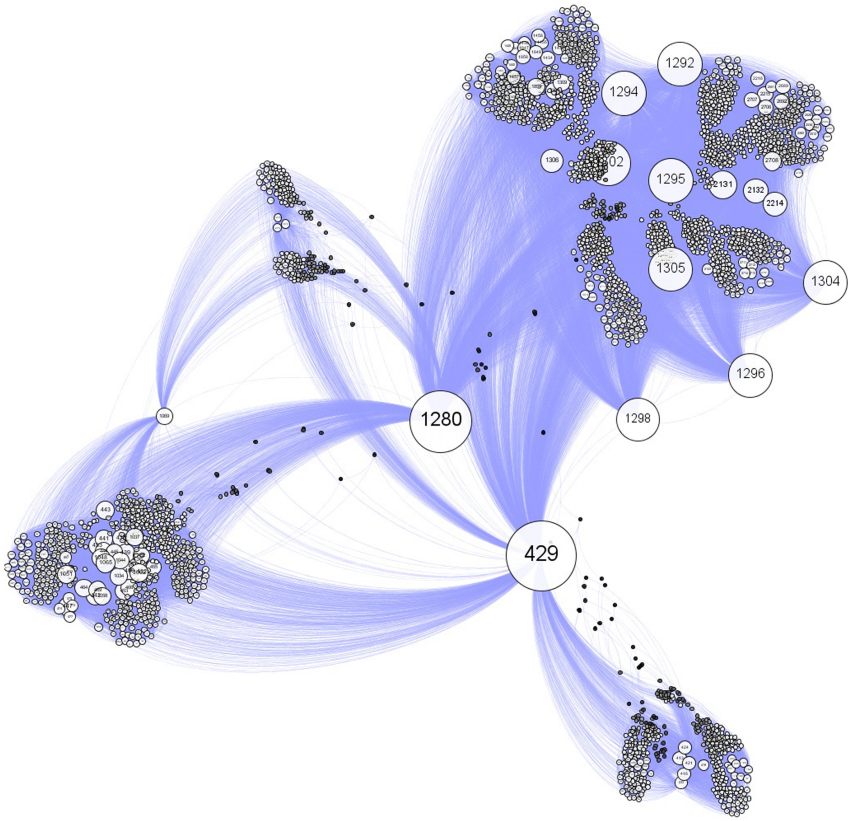


Figura 9 – Exemplo de um grafo de visibilidade gerado por sinais de áudio. Fonte: https://doi.org/10.1371/journal.pone.0240915.g003

4.2.2.1 Modularidade e Número de Comunidades

A modularidade é uma medida de estrutura de uma rede. Esta medida é projetada para estimar a força de uma divisão de uma rede em módulos (ou comunidades). Uma rede com alta modularidade possui conexões densas entre os vértices dentro dos módulos, mas conexões esparsas entre os vértices em diferentes módulos (MELO; FADIGAS; PEREIRA, 2020). Um alto valor de modularidade indica que a densidade de arestas dentro das comunidades é maior do que o esperado ao acaso, indicando uma boa partição da rede.

A modularidade é definida em (NEWMAN; GIRVAN, 2004) como

$$Q = \frac{1}{2m} \sum_{(i,j)} \left(A(i,j) - \gamma \frac{k(i)k(j)}{2m} \right) \delta(ci, cj) \tag{4.3}$$

onde m é o número de arestas, A é a matriz de adjacência de G, k(i) é o grau de i, γ é o parâmetro de resolução e δ(ci, cj) é 1 se i e j estão na mesma comunidade, caso contrário o valor é 0.

De acordo com (NEWMAN; GIRVAN, 2004), isso pode ser reduzido a

$$Q = \sum_{c=1}^{n} \left[\frac{L(c)}{m} - \gamma (\frac{k(c)}{2m})^2 \right] \tag{4.4}$$

onde a soma itera sobre todas as comunidades c, m é o número de arestas, L(c) é o número de conexões intracomunitários para a comunidade c, k(c) é a soma dos graus dos vértices na comunidade c e γ é o parâmetro de resolução.

O parâmetro de resolução define uma compensação arbitrária entre bordas intragrupo e bordas intergrupo. Padrões de agrupamento mais complexos podem ser descobertos analisando a mesma rede com vários valores de gama e combinando os resultados. A segunda fórmula é aquela realmente utilizada no cálculo da modularidade.

4.2.2.2 Grau médio

O grau de um vértice corresponde ao número total de suas arestas. Seja k(i) o grau do vértice i de uma rede. O grau médio de uma rede com N vértices é a média aritmética de k(i).

$$\langle k \rangle = \frac{1}{N} \times \sum_{i=1}^{N} k(i) \tag{4.5}$$

Este parâmetro mede a intensidade média da conectividade de cada vértice da rede. Nos grafos de visibilidade, esta medida pode ser interpretada como o nível médio de visibilidade local dos picos de sinal. Sinais em que predominam poucos picos de alta visibilidade local gerarão grafos de visibilidade com grau médio mais altos do que sinais com muitos picos de baixa visibilidade local.

4.2.2.3 Densidade

Seja N o número de vértices de um grafo. A densidade é a razão entre o número total de arestas de uma rede (m = |E|) e o maior número possível de arestas.

$$\triangle = \frac{2 \times m}{N(N-1)} \tag{4.6}$$

A densidade mede o nível geral de conectividade de rede. Nos grafos de visibilidade associados aos sinais de áudio, esta medida indica o nível de visibilidade geral destes sinais. Quanto maior o nível de persistência rítmica no sinal, menor a visibilidade geral e menor a densidade.

4.2.2.4 Extração dos atributos na prática

Para a extração dos atributos dos grafos, foi utilizado o auxílio de biblioteca NetworkX (HAGBERG; SWART; CHULT, 2008), o qual provém ferramentas para a criação, manipulação e estudo da estrutura, dinâmica e funções de redes complexas. Além disso, também foi utilizado a biblioteca Librosa (MCFEE et al., 2015), que é um pacote feito

em Python utilizado em muitos estudos para análise de áudio e música. Ele fornece os blocos de construção necessários para criar sistemas de recuperação de informações musicais, e foi essencial para à extração correta dos atributos para a classificação dos gêneros.

4.2.3 Classificação dos gêneros sob propriedades das redes obtidas

Os estudos deste trabalho apontam supostas relações interessantes de ordem topológica com a classificação dos gêneros musicais, conforme evidenciado em capítulos anteriores, através da detecção de padrões em atributos de redes complexas.

De modo geral, um sinal que possui picos intensos persistindo com pouca visibilidade irá gerar redes com menor grau médio e alta modularidade se comparado a um sinal com picos esparsos e não persistentes. Ou seja, é possível inferir que quanto menor o grau médio e maior a modularidade dos grafos obtidos, maior a auto-similaridade do sinal. (MELO; FADIGAS; PEREIRA, 2020)

As classificações realizadas nesse trabalho foram realizadas com o auxílio de redes neurais artificiais (RNA). Para isso, utilizou-se a API de aprendizagem profunda Keras, que disponibiliza uma interface produtiva e completa para soluções envolvendo aprendizado de máquina, desenvolvida em cima da plataforma de código aberto TensorFlow. Através dessa API, foi possível a definição de diferentes topologias de redes neurais para testes, assim como a customização de hiperparâmetros.

As redes neurais são modelos computacionais inspirados pelo sistema nervoso central, que são capazes de realizar o aprendizado de máquina bem como o reconhecimento de padrões em um conjunto de dados. Redes neurais artificiais geralmente são apresentadas como sistemas de "neurônios interconectados, que podem computar valores de entradas", simulando o comportamento de redes neurais biológicas.

A construção foi feita utilizando uma camada de entrada, múltiplas camadas densas escondidas e uma camada de saída. A primeira camada é a responsável por alimentar os dados de entrada na rede neural, o qual tem dimensão definida pelo número de atributos analisados. As camadas intermediárias são densas e ocultas e costumam ter dimensões maiores, visto que o volume de dados pode ser grande, para a etapa de aprendizagem. Por fim, a última camada se trata de uma camada com 10 neurônios (uma para cada gênero), a qual, a partir de uma função de ativação, fornece um resultado para a classificação de um único gênero musical.

O objetivo das redes desenvolvidas é aprender sobre as entradas fornecidas, e então classificá-las entre dez categorias de gêneros musicais. Para isso, a rede neural precisa consumir os dados de entrada iterativamente, e a cada iteração, avaliar e atualizar seus pesos internos. A avaliação de tais pesos acontece a partir de uma função de Erro(loss), calculado no modelo pelas iterações de treinamento, fornecendo um valor que descreve a

eficácia da configuração de pesos analisada. A função utilizada nas redes desse trabalho foi a Sparse Categorical Cross-Entropy (Entropia Cruzada Categórica Esparsa), que utiliza distribuições probabilísticas para o valor de Erro em classificações. A partir disso, o otimizador da rede neural fornece um método de minimizar o valor encontrado pela função de loss.

As redes neurais desenvolvidas utilizaram o otimizador Adam, o qual se trata de um método de descida de gradiente estocástico para a atualização dos pesos das redes de forma iterativa, a partir dos dados de entrada. Esse algoritmo se tornou popular na área de aprendizado de máquina (KINGMA; BA, 2014), principalmente pelo fato de atingir bons resultados rapidamente. É possível configurar um taxa de aprendizado para o otimizador, de forma a controlar quanto os pesos da rede são atualizados a cada iteração do treinamento, influenciando na velocidade com que o modelo aprende sobre os dados a serem classificados. Dessa forma, a taxa de aprendizado se faz um importante hiperparâmetro a ser configurado durante o refinamento de uma rede neural.

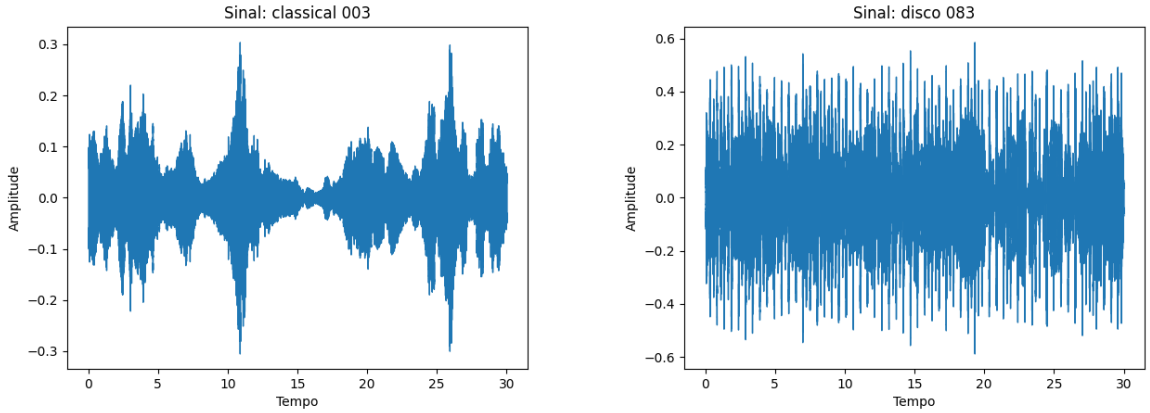
Outro fator relevante da arquitetura de uma rede neural é a função de ativação, a qual é responsável por definir a saída de um neurônio. Nesse trabalho, a função de ativação utilizada nos neurônios da primeira camada e todas as camadas intermediárias ocultas foi o retificador linear(ReLU). Por causa do efeito de retropropagação no aprendizado, cada camada do modelo calcula e envia o erro de volta às camadas entrada. Para isso, é realizado o cálculo da derivada da função de ativação utilizada, o que pode causar um problema de desaparecimento de Gradiente em alguns tipos de função, como o Sigmoid. A função de ReLU permite que o sistema tenha uma melhor convergência. Já a função de ativação utilizada nos neurônios da camada de saída foi a Softmax. Na verdade, a função Softmax é uma extensão da função Sigmoid, com ambas produzindo valores entre 0 e 1 que representam probabilidades. É comum a função Sigmoid ser utilizada em classificações binárias, enquanto a Softmax é amplamente usada em problemas multiclasse.

Durante o processo de refinamento das redes desenvolvidas, os hiperparâmetros manipulados foram: número de unidades de processamento das camadas, taxa de aprendizado, dropout, tamanho de batch e épocas de treinamento. O processo de treinamento e refinamento das redes neurais testadas foi documentado e discutido no capítulo seguinte.

5 Experimentos

Por conta do tamanho do banco de dados e número de amostras, foi necessário dividir os sinais e consequentemente aumentar o tamanho do conjunto de dados para melhorar o modelo e auxiliar a aprendizagem de máquina a extrair recursos significativos.

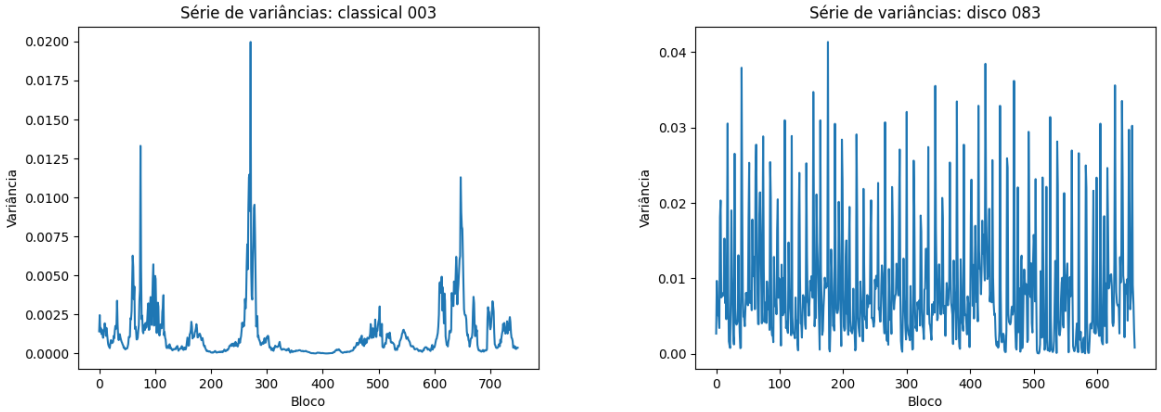
Cada uma das 1000 amostras de áudio de 30 segundos foi segmentado em 10 partes de 3 segundos, resultando em 10000 amostras de áudio no total, sendo 1000 amostras por gênero musical. Feito isso, cada uma das amostras foi transformada em uma série temporal de suas variações, e depois foi gerado um Grafo de Visibilidade Natural e um Grafo de Visibilidade Horizontal para cada série, totalizando 10000 grafos de visibilidade de cada tipo.



(a) Sinal de Clássico. Fonte: Autor.

(b) Sinal de Disco. Fonte: Autor.

Figura 10 – Comparação entre sinais de diferentes gêneros musicais. Fonte: Autor.



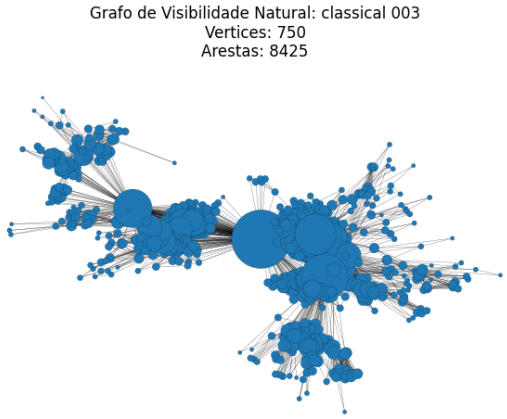
(a) Série de variância de Clássico. Fonte: Autor.

(b) Série de variância de Disco. Fonte: Autor.

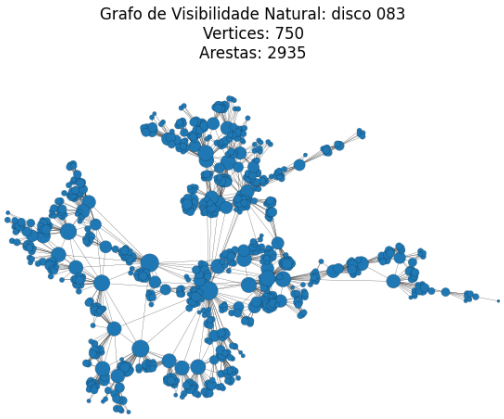
Figura 11 – Comparação entre séries de variância de diferentes gêneros musicais. Fonte: Autor.

Como pode se observar, a alta persistência rítmica encontrada no sinal de música de Disco, pelas escolhas rítmicas e instrumentais inerentes ao estilo, é mais moderado e menos persistente no estilo Clássico, onde existem variações de natureza rítmica mais evidentes.

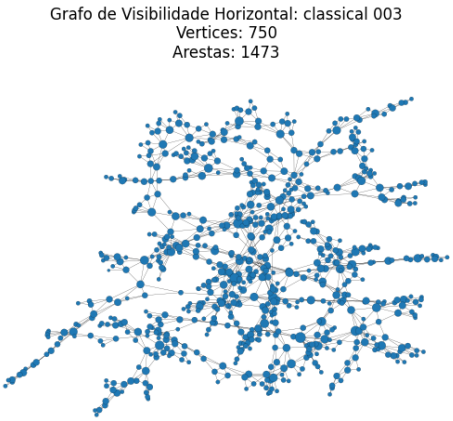
De modo geral, os sinais de arquivos de áudio que possuem trechos musicais com percussões mais fortes e persistentes geraram sinais com picos mais elevados e, consequentemente, segmentaram as possíveis comunidades encontradas nos grafos de visibilidade, como será demonstrado a seguir. Da mesma forma, quanto menor a persistência e força dos batimentos principais, maior a chance dessas amostras participarem de um mesmo grupo.



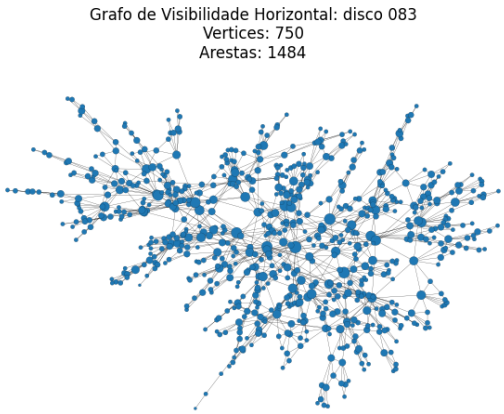
(a) Grafos de Visibilidade Natural de Clássico. Fonte: Autor.



(b) Grafos de Visibilidade Natural de Disco. Fonte: Autor.



(c) Grafos de Visibilidade Horizontal de Clássico. Fonte: Autor.



(d) Grafos de Visibilidade Horizontal de Disco. Fonte: Autor.

Figura 12 – Comparação entre tipos de grafos de diferentes gêneros musicais. Fonte: Autor.

Para todos os grafos gerados foi utilizado o modo Spring de layout, o qual simula uma representação dirigida por força da rede, tratando as arestas como molas que mantêm os vértices próximos, enquanto trata os vértices como objetos repelentes, às vezes chamados

de força antigravidade. A simulação continua até que as posições estejam próximas de um equilíbrio.

A representação do tamanho dos vértices corresponde à uma proporção linear de suas respectivas taxa de conectividade com outros vértices, calculada pelo valor de Grau. No geral, sinais que possuem picos mais definidos e que prevalecem na amplitude, como é o caso do sinal de Clássico, terão vértices com maior densidade e consequentemente o grafo será mais denso. Em sinais que os picos são mais nivelados e com pouca variação, como é o caso do sinal de Disco, terão vértices com menor densidade e consequentemente o grafo será mais esparso.

Dessa forma, os vértices que possuem maior número de arestas (representados com maior tamanho) correspondem, no geral, aos pontos mais altos de suas respectivas séries de variâncias.

As Tabelas 2 e 3 mostram a média e o desvio padrão para o cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau médio($\langle k \rangle$) e a Densidade($\triangle$) dos dois tipos de grafos de visibilidade correspondentes a 100 amostras de áudio agrupadas em 10 gêneros musicais.

Para os resultados encontrados, é evidenciado uma forte relação entre Q e Nc, onde é demonstrado de forma mais clara a diferença entre a natureza de percussão de cada estilo. Além disso, é possível observar também uma relação entre $\langle k \rangle$ e $\triangle$ de cada gênero, o que corrobora com a ideia sobre as fortes relações entre as variações locais do sinal.

Gênero	Q	σ	Nc	σ	$\langle k \rangle$	σ	$\triangle$	σ
Blues	0.808	0.037	11.71	2.046	9.448	2.593	0.013	0.004
Clássico	0.698	0.103	10.28	2.836	14.662	5.237	0.020	0.007
Country	0.809	0.035	11.31	1.791	9.583	2.391	0.013	0.003
Disco	0.839	0.017	12.54	1.388	8.255	1.041	0.011	0.001
Hip Hop	0.837	0.019	12.41	1.471	8.863	1.433	0.012	0.002
Jazz	0.797	0.035	11.35	2.002	10.501	2.948	0.014	0.004
Metal	0.826	0.021	12.23	1.752	7.347	1.035	0.010	0.001
Pop	0.829	0.028	12.44	1.788	9.132	1.534	0.012	0.002
Reggae	0.829	0.021	11.63	1.581	8.916	1.515	0.012	0.002
Rock	0.820	0.031	11.66	1.273	8.413	1.267	0.011	0.002

Tabela 2 – Media e desvio padrão das propriedades topológicas para os Grafos de Visibilidade Natural. Fonte: Autor.

Gênero	Q	σ	Nc	σ	$\langle k \rangle$	σ	$\triangle$	σ
Blues	0.879	0.009	17.08	2.235	3.951	0.016	0.005	0.000
Clássico	0.872	0.009	15.46	1.731	3.923	0.027	0.005	0.000
Country	0.880	0.008	16.93	2.263	3.953	0.017	0.005	0.000
Disco	0.892	0.008	17.97	2.290	3.958	0.011	0.005	0.000
Hip Hop	0.880	0.011	17.92	2.787	3.949	0.022	0.005	0.000
Jazz	0.878	0.007	16.43	1.846	3.941	0.021	0.005	0.000
Metal	0.883	0.009	17.96	2.153	3.958	0.010	0.005	0.000
Pop	0.881	0.009	17.38	2.441	3.954	0.012	0.005	0.000
Reggae	0.880	0.009	17.90	2.412	3.954	0.015	0.005	0.000
Rock	0.882	0.009	17.03	2.251	3.955	0.014	0.005	0.000

Tabela 3 – Media e desvio padrão das propriedades topológicas para os Grafos de Visibilidade Horizontal. Fonte: Autor.

Na observação sobre a perspectiva de Grafos de Visibilidade Natural, os valores médios para Modularidade e Número de Comunidades apresentam o gênero Clássico(menor valor) e Disco(maior valor) nos extremos da tabela. Da mesma forma, apesar da pequena diferença e leves divergências de posicionamento, gêneros como Hip-Hop e e Pop também são apresentados com altos valores na comparação, e gêneros como Jazz e Country são evidenciados com os valores mais baixos.

As médias de Gráu Médio e Densidade mantêm uma relação de certa forma oposta aos cálculos de Modularidade e Número de Comunidades, de forma que o gênero Clássico ocupa a maior posição e o gênero Disco ocupa a segunda menor posição. Os gêneros de Jazz e Country agora são posicionados nos maiores valores da tabela. Também observa-se que o gênero de Metal ocupa o menor valor para ambos os atributos, que também pode ser observado por picos bem definidos e intensos na análise do sinal. De qualquer forma, é bem evidente a diferença de grau médio entre os dois opostos da tabela, visto que suas características de natureza rítmica são de certa forma contrárias.

Para todos os quatro componentes, os gêneros de Reggae e Rock ocupam a posição intermediária. Esse tipo de organização hierárquica corrobora com a ideia de que gêneros musicais que optam por arranjos instrumentais muito “densos”, “intensos” e persistentes tendem a ocupar posições opostas a gêneros com texturas instrumentais mais ricas em dinâmicas. Em uma posição intermediária estão os estilos musicais que buscam equilibrar a essência de influências de ambos os extremos.

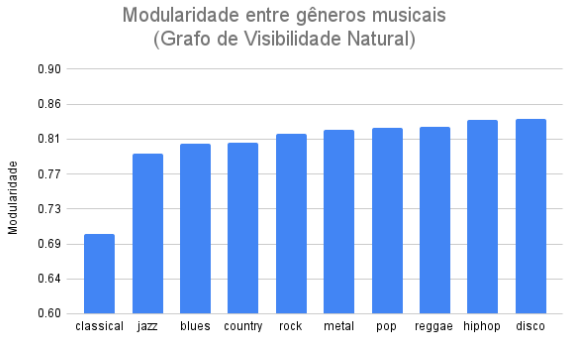
Já na observação sobre a perspectiva de Grafos de Visibilidade Horizontal, a maioria dos padrões permanecem verdadeiros, com os menores valores médios de Modularidade e Número de Comunidades ainda de gêneros como Clássico e Jazz. Porém, existe uma forte divergência do resultado para os cálculos de Grau Médio e Densidade, o qual demonstra posicionamento de gêneros de forma oposta em relação aos atributos de Grafos de Visi-

bilidade Natural, com gêneros como Clássico, Jazz e Hip-Hop ocupando as posições de valores mais baixos e com gêneros como Disco e Metal ainda ocupando os valores mais elevados.

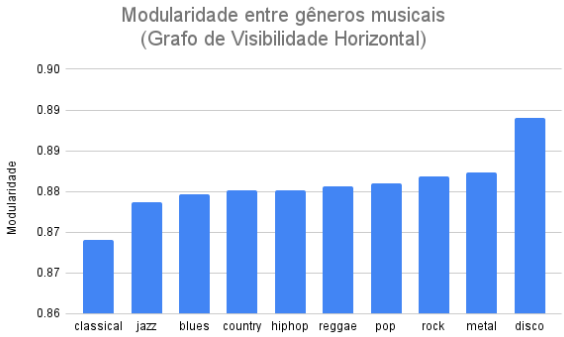
Essa divergência se deve principalmente à natureza do método de transformação do sinal em Grafos de Visibilidade Horizontal, uma vez que é uma simplificação da transformação normal de visibilidade. Pela forte divisão dos segmentos do sinal, é comum que propriedades de conectividade e densidade dos grafos seja reduzido, quando em comparação à aqueles gerados por Grafos de Visibilidade Natural. Esse comportamento também é observado pelo baixo valor de desvio padrão entre os gêneros para esses atributos na Tabela 3.

De qualquer forma, tal diferença pode evidenciar características importantes para a classificação dos gêneros, uma vez que pequenas relações de dados podem acabar sendo consideradas de menos importância pelo classificador quando utilizado Grafos de Visibilidade Natural, onde muitas outras características, além das definidas pele Grafo de Visibilidade Horizontal, são evidenciadas.

Essas relações também são demonstradas nos gráficos apresentados a seguir.

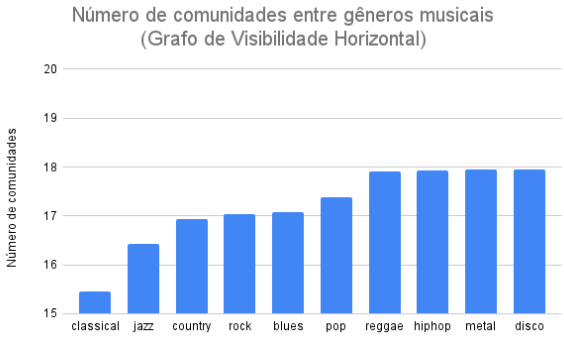


(a) Grafos de Visibilidade Natural.

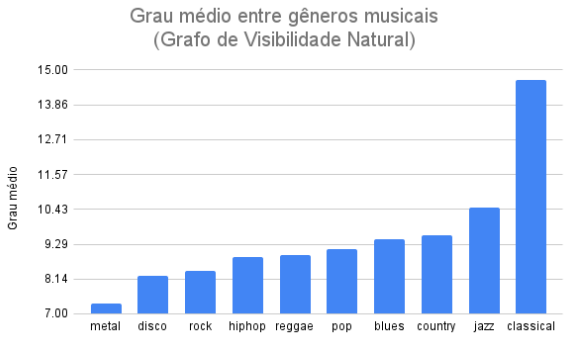


(b) Grafos de Visibilidade Horizontal.

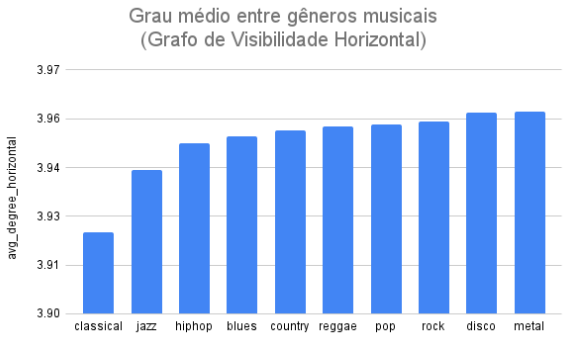
(c) Grafos de Visibilidade Natural.



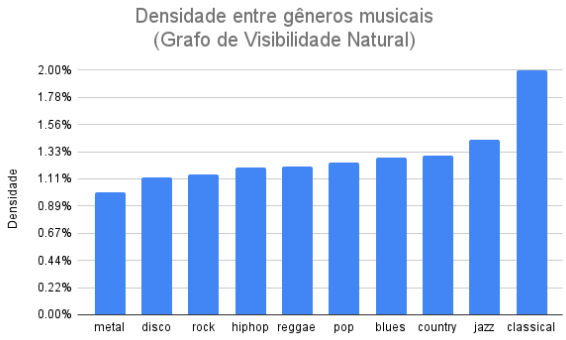
(d) Grafos de Visibilidade Horizontal.



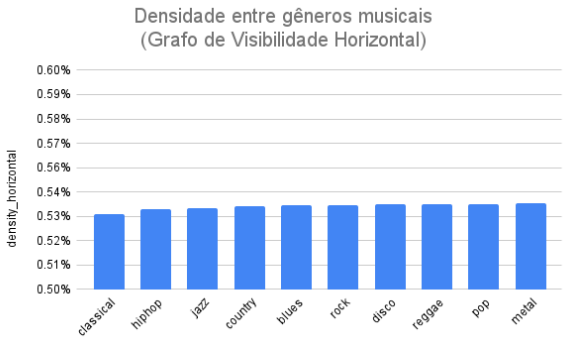
(e) Grafos de Visibilidade Natural.



(f) Grafos de Visibilidade Horizontal.



(g) Grafos de Visibilidade Natural.



(h) Grafos de Visibilidade Horizontal.

Figura 13 – Média de Modularidade, Número de Comunidades, Grau Médio e Densidade entre gêneros musicais. Fonte: Autor.

Para ambos os conjuntos de dados de Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal foi realizado um estudo de estatística descritiva para os parâmetros extraídos. Tal método pode ser utilizado para encontrar meios que são significativamente diferentes entre si. Esse tipo de informação pode evidenciar quais atributos podem ser relevantes para a classificação entre os gêneros propostos neste experimento.

Além de demonstrar noções gerais sobre o tamanho e escala do conjunto de dados, também é possível observar medidas de variabilidade entre os parâmetros. Por exemplo, observando os valores de Coeficiente de Variação na análise sob a perspectiva dos Grafos de Visibilidade Natural, é evidenciado que os atributos de Grau Médio e Densidade são fortes candidatos à ter um efeito maior no modelo que está sendo usado para prever uma determinada variável, ou seja, tais recursos podem demonstrar alta importância na classificação dos gêneros musicais. Já no caso dos dados extraídos de Grafos de Visibilidade Horizontal, o Número de Comunidades pode ser o maior candidato à ter maior importância, entre todos os parâmetros, enquanto que a Densidade não apresenta alta variação entre os grupos e provavelmente não evidenciará bons resultados para a classificação. Essa observação corrobora com os gráficos comparativos demonstrados acima.

	Modularidade	N° Comunidades	Grau Médio	Densidade
Observações(gêneros)	10	10	10	10
Soma xi	8.09	117.56	95.12	0.13
Média $\bar{x}$	0.81	11.76	9.51	0.01
Soma dos quadrados x^2i	6.56	1386.34	940.69	0.00
Variância	0.00	0.48	3.99	0.00
Desvio Padrão	0.04	0.69	2.00	0.00
Coeficiente de Variação	5.12	5.88%	21.00%	20.77%

Tabela 4 – Estatística descritiva para Grafos de Visibilidade Natural. Fonte: Autor.

	Modularidade	N° Comunidades	Grau Médio	Densidade
Observações(gêneros)	10	10	10	10
Soma xi	8.80	172.05	39.50	0.05
Média $\bar{x}$	0.88	17.21	3.95	0.01
Soma dos quadrados x^2i	7.75	2966.05	156.00	0.00
Variância	0.00	0.66	0.00	0.00
Desvio Padrão	0.00	0.81	0.01	0.00
Coeficiente de Variação	0.36%	4.72%	0.26%	0.00%

Tabela 5 – Estatística descritiva para Grafos de Visibilidade Horizontal. Fonte: Autor.

5.1 Aprendizado de máquina e classificação

O aprendizado de máquina e a classificação foram realizados com redes neurais artificiais supervisionadas levando em consideração dois cenários. No primeiro cenário, foi utilizado um vetor de atributos apenas com os atributos de grafos (Modularidade, Número de Comunidades, Grau Médio e Densidade), para ambos os tipos Natural e Horizontal. A ideia era explorar uma situação em que apenas um descritor de atividade rítmica fosse usado. Em seguida, foi realizado o aprendizado e a classificação apenas com o atributo da frequência e intensidade de Onsets, o qual refere-se ao início das notas musicais do sinal, apresentando natureza similar aos propostos pela conversão de grafos. Por questões de simplificação, os descritores que compõe os atributos gerados por Grafos de Visibilidade Natural serão chamados de DGVN (Descritor de Grafos de Visibilidade Natural) e aqueles gerados por Grafos de Visibilidade Horizontal serão chamados de DGVH (Descritor de Grafos de Visibilidade Horizontal).

No segundo cenário, que também utilizou redes neurais, foram realizados experimentos utilizando cada um dos descritores do primeiro cenário separadamente, adicionando outros parâmetros de timbre utilizados em larga escala nos descritores da atualidade. Primeiro, foi configurado um vetor de atributos juntando os atributos de grafos com 13 MFCCs que apresentam as melhores taxas de importância, conforme os estudos atuais. Em seguida, foi realizado o aprendizado de máquina combinando os atributos de Onsets também aos 13 MFCCs. Por fim, os resultados foram comparados novamente.

Durante a etapa de treinamento e testes, o banco de dados foi dividido em subconjuntos de treinamento, validação e teste. O primeiro subconjunto é utilizado no aprendizado da rede neural, com o ajuste de pesos das unidades de processamento. Já o subconjunto de validação é utilizado na análise da performance da rede durante o treinamento, possibilitando um refinamento eficiente de hiperparâmetros. Por fim, o subconjunto de teste possibilita avaliar o desempenho do classificador em dados que nunca foram vistos pela máquina. Dentre todos arquivos de áudio do banco de dados, 25% foram utilizados para testes, e dos 75% restantes, 80% formaram o conjunto de treinamento e 20% o conjunto de validação.

Para determinar a configuração de modelo e hiperparâmetros, montou-se uma entrada para uma rede neural base, e então determinou-se a melhor topologia da rede a ser utilizada, a partir da variação dos hiperparâmetros. Para a análise dos efeitos da variação de cada hiperparâmetro, o estado inicial das redes foi fixado, e com as configurações mais promissoras foram realizados os testes finais. A qualidade das redes encontradas foi definida a partir dos valores de acurácia e generalização, observados nas iterações da etapa de treinamento e validação e nos testes.

Os hiperparâmetros das redes considerados foram:

- Número de camadas redes;

- Número de unidades de processamento das células redes;
- Taxa de aprendizado;
- Dropout;
- Tamanho dos Batches;
- Número de épocas de treinamento;

Durante o período de treinamento, validação e teste, analisou-se o comportamento das redes neurais construídas, e dependendo dos resultados de acurácia e generalização, refinamentos e modificações foram feitos iterativamente. Após a exploração definiu-se utilizar uma Taxa de Aprendizado de 0.0001, Dropout de 30% nas camadas densas e Batches com tamanho de 32 amostras.

O modelo final consistiu da seguinte configuração:

- Camada de Input (número de atributos)
- 1a Camada Oculta Densa (512 neurônios, ativação="relu", Dropout 30%)
- 2a Camada Oculta Densa (256 neurônios, ativação="relu", Dropout 30%)
- 3a Camada Oculta Densa (54 neurônios, ativação="relu", Dropout 30%)
- Camada de Output (10 neurônios, ativação="softmax")

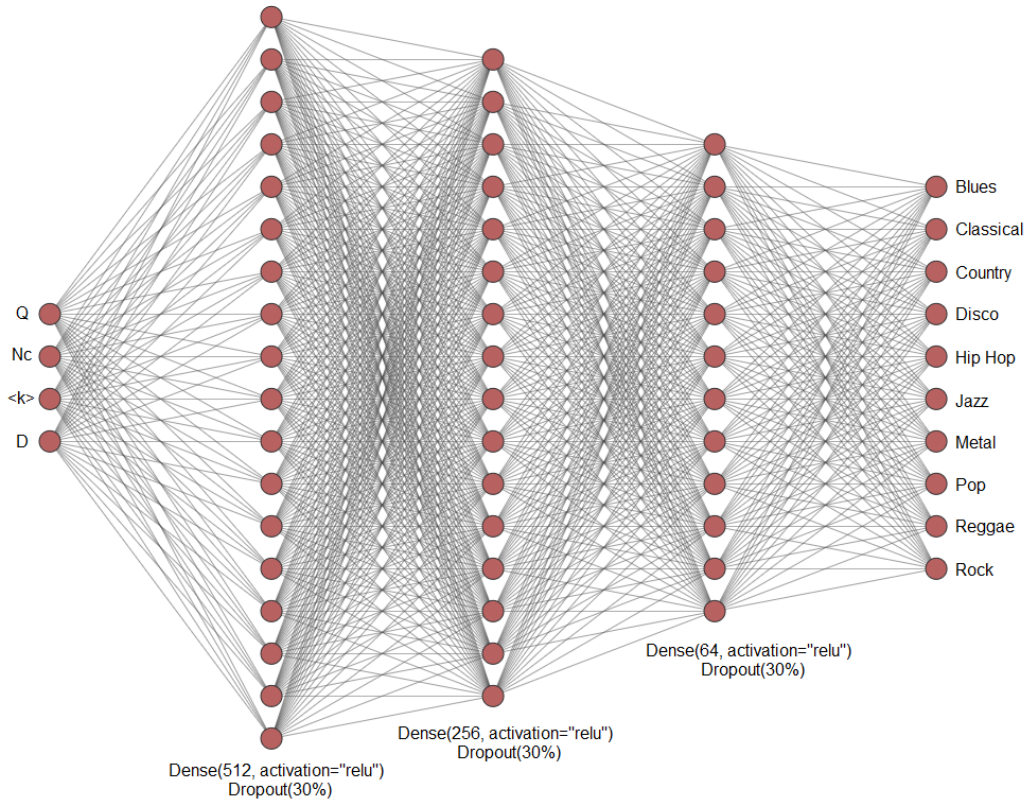


Figura 14 – Exemplo de modelo de aprendizagem apenas com parâmetros extraídos dos grafos. Fonte: Autor.

5.2 Resultados

Abaixo estão os resultados encontrados para cada combinação de atributos testada, demonstrando o erro e acurácia obtida nos testes.

Parâmetros	Épocas	Erro	Acurácia
DGVN	80	2.04	25.92%
DGVH	80	2.27	14.72%
Onset	80	2.25	34.52%
13 MFCCs	120	1.58	48.24%
DGVN + 13 MFCCs	100	1.72	55.30%
DGVH + 13 MFCCs	100	1.65	55.92%
Onset + 13 MFCCs	100	1.67	55.55%
DGVN + Onset + 13 MFCCs	100	2.30	55.92%
DGVH + Onset + 13 MFCCs	200	2.14	56.34%

Tabela 6 – Comparação de resultados dos descritores. Fonte: Autor.

Observa-se que os descritores apenas de natureza rítmica não foram suficientes para ter uma boa classificação, obtendo maiores taxas de acerto com apenas o atributo de Onsets do sinal, quando comparado com o DGVN, além de demonstrar uma diferença ainda maior se comparado ao DGVH.

Também é evidenciado que o descritor com apenas atributos de timbre (13 MFFCs) apresentou ótimo resultados por conta própria, com uma taxa de 48.24% de acurácia e erro de 1.58. Isso demonstra o quão importante é o papel desses atributos para descritores de classificação de gêneros musicais nas pesquisas atuais, conforme apresentado em trabalhos relacionados (TZANETAKIS; COOK, 2002).

Apesar disso, os descritores que utilizaram propriedades de timbre e ritmo somadas obtiveram uma taxa de acurácia maior, apesar de pouca divergência entre eles, com 55.30% de acerto para aqueles gerados por Grafos de Visibilidade Natural, 55.92% de acerto para atributos extraídos de Grafos de Visibilidade Horizontal e 55.55% de acerto para descritores de ritmo com frequência de Onsets.

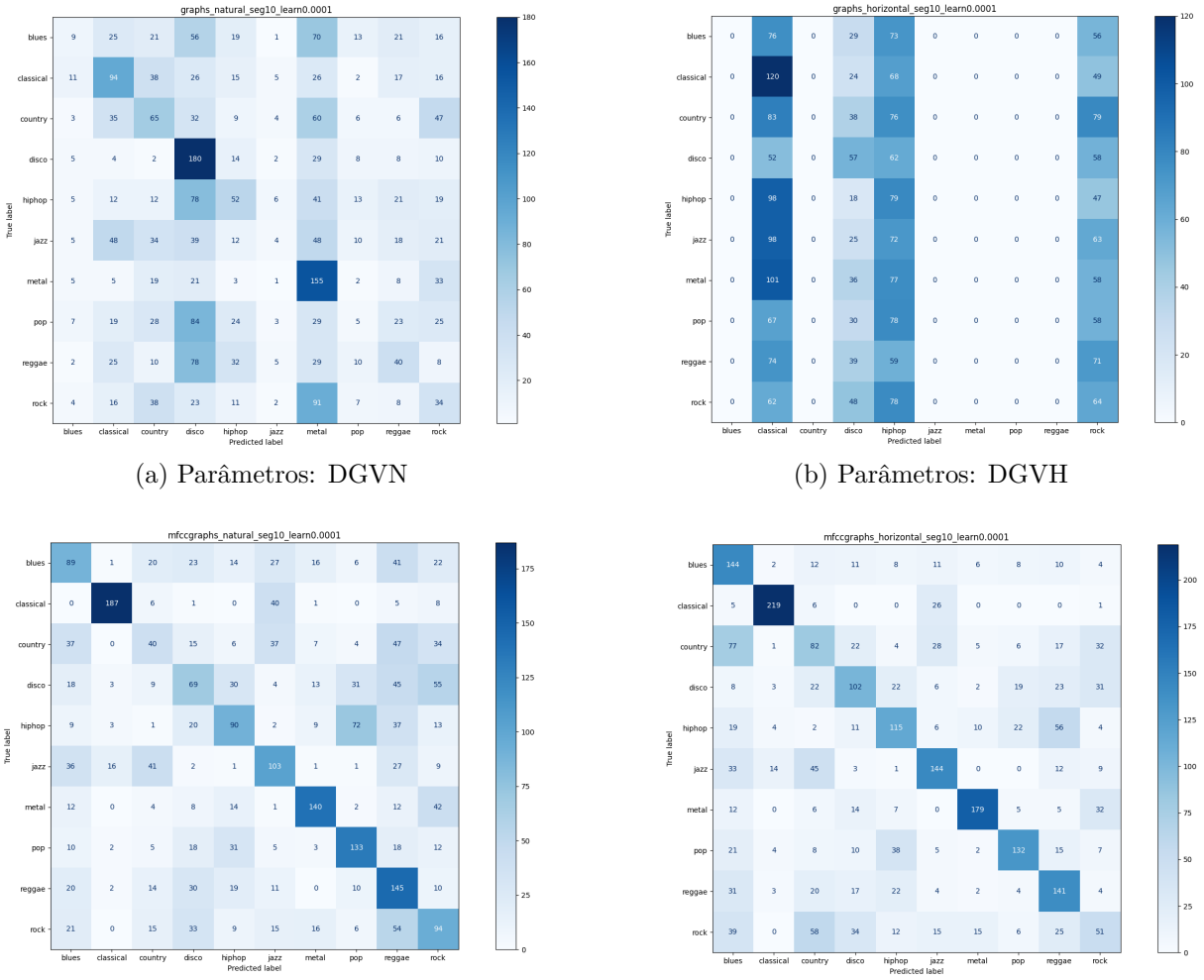
Ao combinar descritores propostos por esse trabalho junto com os atributos de Onsets, não foi possível observar uma diferença considerável na taxa de acerto, além de apresentar uma taxa maior de erro. De qualquer forma, isso pode ser consequência da quantidade de dados utilizada, e a variação de amostras entre os gêneros no banco GTZAN podem não apresentar características suficientes para que os descritores validem os dados com maior precisão.

Além disso, foi gerado a matriz de confusão e feito a comparação para as 10000 amostras de áudio (1000 por gênero). Para o aprendizado utilizando o DGVN, na diagonal principal, onde observa-se os verdadeiros positivos, é evidenciado que gêneros de Disco (18%), Metal(15.5%) e Clássico(9.4%) tiveram a maior taxa de verdadeiros positivos. Isso indica que, mesmo com uma taxa de acerto baixa nesses casos, os padrões de percussão descritos pelos atributos de grafos foram melhor interpretados pela rede neural.

As menores taxas de acerto encontradas foram Blues(0.5%), Jazz(0.4%) e Pop(0.5%), o que indica que padrões de percussão descritos pelos atributos desses grafos não são típicos o suficiente para caracterizar fortemente seu agrupamento nesse sistema particular.

Para aprendizado utilizando DGVH, na diagonal principal observa-se a maior taxa de verdadeiros positivo para gêneros de Clássico(12%), Hip-Hop(7.9%), Rock(6.4%) e Disco(5.7%). Apesar disso, é evidenciado uma grande diferença na classificação de cada gênero quando comparado com o DGVN. A maioria dos gêneros nem foram considerados pelo classificador, como é o caso de Blues, Country, Jazz, Metal, Pop e Reggae. Isso indica que, por ser um método mais simplificado, os padrões de percussão descritos pelos atributos de Grafos de Visibilidade Horizontal tendem à decidir mais fortemente em um grupo menor de informação e focam fortemente nas naturezas rítmicas do sinal.

De todas as combinações de parâmetros analisadas, foi também gerado uma matriz de confusão para o grupo que obteve a melhor acurácia (56.34%), com atributos de DGVH,



(a) Parâmetros: DGVN

(b) Parâmetros: DGVH

(c) Parâmetros: DGVN + 13 MFCCs

(d) Parâmetros: DGVH + 13 MFCCs

Figura 15 – Matriz confusão de classificação dos gêneros musicais. Fonte: Autor

Onset e 13 MFCCs.

Na diagonal principal, onde observa-se os verdadeiros positivos, é evidenciado que gêneros de Clássico(19.9%), Blues(15.6%), Metal(14.8%) e Pop(13.9%) tiveram a maior taxa de verdadeiros positivos. É interessante observar que gêneros que obtiveram as piores taxa anteriormente, em descritores de grafos, obtiveram altas taxas de acurácia em combinação com atributos de natureza rítmica, como é o caso de gêneros como Blues, Jazz e Pop. Essa relação demonstra a extrema importância em combinar atributos de diferentes características e naturezas para os descritores de classificação de gêneros musicais.

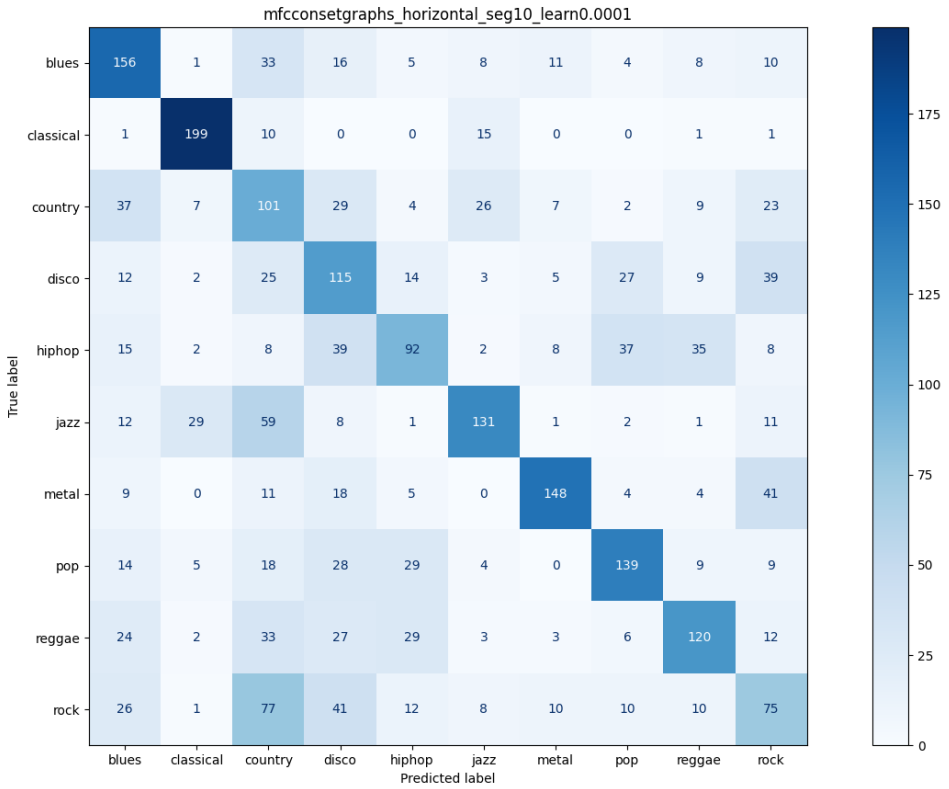


Figura 16 – Matriz confusão de classificação dos gêneros musicais com DGVH + 13 MFCCs + Onset. Fonte: Autor.

De todos os gêneros classificados, o grupo de Rock foi o que obteve a pior taxa de verdadeiros positivos, o que também é demonstrado na análise das médias de cada uma das propriedades topológicas de visibilidade horizontal nos gráficos da Figura 13, onde o gênero tomou as posições intermediárias entre todos os grupos. Essa característica refletiu na falsa classificação desses sinais principalmente no gênero de Country, o qual também assumiu posições próximas do gênero de Rock na análise das propriedades dos grafos gerados, além de ser comumente evidenciada na classificação feita por humanos.

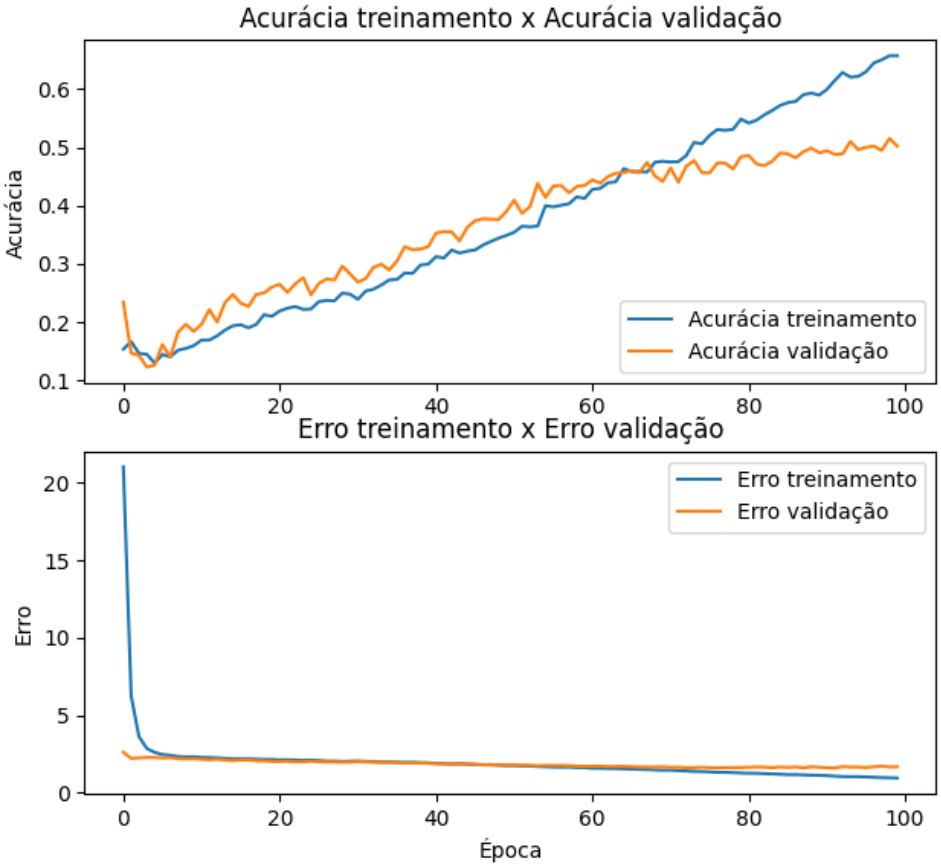


Figura 17 – Loss treinamento vs Loss validação com DGVH + 13 MFCCs + Onset. Fonte: Autor.

Para a definição da taxa de acurácia final, observou-se o comportamento da variação de precisão e erro em ambos os casos para os grupos de treinamento e validação na classificação dos dados, considerando até a época onde começa o efeito de overfit, ou seja, no momento em que há uma certa divergência entre os valores dos diferentes subconjuntos.

Observa-se que a taxa de erro no conjunto de validação, nas primeiras iterações, é relativamente menor do que no conjunto de treinamento, até atingir um ponto onde as taxas se cruzam, onde deveria ser o limite de épocas do classificador. Esse comportamento provavelmente aconteceu porque o erro de treinamento é calculado continuamente ao longo de uma época inteira, enquanto que as métricas de validação são calculadas sobre o conjunto de validação apenas quando a época de treinamento atual é concluída. Isso implica que, em média, a taxa de erro de treinamento é medida meia época antes. O tamanho da batch e o otimizador escolhido influenciam diretamente nesse cálculo e, de modo geral, quanto maior a quantidade de dados processados nas iterações da máquina, menor vai ser o erro no fim de cada iteração.

6 Considerações Finais

No presente trabalho foi realizado um levantamento teórico dos principais tópicos e métodos utilizados na classificação de gêneros musicais, além de ressaltar a importância e aplicação em diversas áreas desses problemas na atualidade. Foram apresentados alguns trabalhos relacionados que serviram de base no desenvolvimento das comparações.

A realização dos experimentos documentados permitiu extrair e analisar propriedades de grafos de visibilidade como uma nova forma de selecionar características em sinais de áudio. Além disso, a natureza e quantidade dos dados se fez um fator relevante na interpretação das saídas geradas pelas redes neurais desenvolvidas.

Após a resolução e exploração das atividades propostas, foi feita uma análise e comparação dos resultados com outros tipos de descritores e soluções já existentes, tendo em vista a compreensão sobre quais aspectos a tese pode auxiliar no problema de classificação de gêneros musicais, seja por otimização em tempo computacional ou melhorias à acurácia no resultado final do algoritmo.

Em sistemas de classificação usando apenas os dados de grafos, obteve-se a melhor acurácia de 25.92% para o DGVN e acurácia de 14.72% para DGVH. Também foi realizado a classificação adicionando elementos de natureza de timbre, os MFCCs, onde obteve-se uma precisão semelhante à outro descritor com atributos rítmicos de intensidade de Onsets do sinal, com o melhor caso atingindo 56.34% de acurácia.

Com base nos resultados obtidos neste experimento e nos estudos mencionados, foi evidenciado que os Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal podem ser considerados como uma nova alternativa para a extração de características rítmicas para a recuperação de informações de música em sinais de áudio e podem ser usadas com sucesso em conjunto com os descritores baseado em transformadas de Fourier, em especial à aqueles de natureza de timbre.

A acurácia de classificação do conjunto de recursos combinados, em alguns casos, não é significativamente aumentada em comparação com as precisões de classificação do conjunto de recursos individuais. Este fato não implica necessariamente que os recursos estejam correlacionados ou não contenham informações úteis, pois pode ser o caso de um determinado arquivo ser classificado corretamente por dois conjuntos de recursos diferentes que contêm informações de recursos diferentes e não correlacionados. Além disso, embora certas características individuais estejam correlacionadas, a adição de cada característica específica melhora, no geral, a precisão da classificação.

Para melhor entender o valor do resultado encontrado na classificação usando os descritores de grafos de visibilidade, vale ressaltar que os trabalhos citados contam com um número maior de atributos, chegando a 30 (TZANETAKIS; COOK, 2002) contra apenas

17 atributos usados neste experimento.

6.1 Trabalhos Futuros

A execução desse trabalho exigiu uma grande quantidade de dados e testes com diferentes configurações e, portanto, há diversos caminhos que podem ser explorados de forma a encontrar resultados relevantes para a classificação dos gêneros musicais.

Por exemplo, a partir da análise dos processos de treinamento, evidenciou-se que dependendo de como é dividido o sinal de entrada (apresentando séries de variância com variados números de pontos), as propriedades topológicas dos grafos gerados podem gerar comportamentos diferentes e, consequentemente, classificar os gêneros com outra perspectiva.

Além disso, o banco de dados GTZAN, apesar de ser utilizado em larga escala, ainda é considerado pequeno e com poucos dados, e certas propriedades apresentadas nesse trabalho podem ter resultados variados com diferentes níveis de amostras de sinais de áudio, além de permitir a exploração de diferentes gêneros e sub-gêneros encontrados na literatura.

A utilização de métodos mais eficientes para refinamento de redes neurais é de grande relevância, permitindo variações de modelos e cadeias de processamento com diferentes propósitos, dependendo dos atributos extraídos dos sinais. O número de atributos também pode influenciar diretamente na taxa de acurácia e erro dos sistemas, e um classificador de atributos pode ser considerado, de forma a eliminar a extração de dados desnecessários na fase de pré-processamento.

Por fim, vale ressaltar que a simplicidade da versão do algoritmo horizontal, proposto por este trabalho, permite que a extração de atributos seja realizada em resolução de um tempo computacional mais otimizado, principalmente em maiores proporções de dados. Porém, o resultado final pode perder informações relevantes, quando comparado aos classificadores gerados da mesma forma por Grafos de Visibilidade Natural, para a classificação musical.

Referências Bibliográficas

BARBEDO, J.; AMAURI, L. Automatic genre classification of musical signals. **EURASIP Journal on Advances in Signal Processing**, v. 2007, 01 2007. 16

BORGES, E. et al. Classificação do gênero musical utilizando redes neurais artificiais. In: **X Congresso Norte-Nordeste de Pesquisa e Inovação**. [S.l.: s.n.], 2010. p. 1–8. 35

CAMPANHARO, A. S. et al. Duality between time series and networks. **PloS one**, Public Library of Science San Francisco, USA, v. 6, n. 8, p. e23378, 2011. 10, 20, 22

CHEN, D.-R. et al. Predicting financial extremes based on weighted visual graph of major stock indices. **Complexity**, Hindawi, v. 2019, 2019. 22

CHEN, S. et al. The time series forecasting: from the aspect of network. **arXiv preprint arXiv:1403.1713**, 2014. 22

CIPRIANI, A.; GIRI, M. **Electronic music and sound design**. [S.l.]: Contemponet, 2010. v. 1. 10, 17, 18

CLAUSET, A.; NEWMAN, M. E.; MOORE, C. Finding community structure in very large networks. **Physical review E**, APS, v. 70, n. 6, p. 066111, 2004. 26

CORRÊA, D. C. **Inteligência artificial aplicada à análise de gêneros musicais**. Tese (Doutorado) — Universidade de São Paulo, 2012. 15

COSTA, Y. M. et al. Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores. 2013. 18

DESHPANDE, H.; SINGH, R.; NAM, U. Classification of music signals in the visual domain. In: **Proceedings of the COST-G6 conference on digital audio effects**. [S.l.: s.n.], 2001. v. 1, n. 3, p. 1–4. 18

DONNER, R. V.; DONGES, J. F. Visibility graph analysis of geophysical time series: Potentials and possible pitfalls. **Acta Geophysica**, Springer, v. 60, n. 3, p. 589–623, 2012. 21

GOUYON, F. et al. An experimental comparison of audio tempo induction algorithms. **IEEE Transactions on Audio, Speech, and Language Processing**, IEEE, v. 14, n. 5, p. 1832–1844, 2006. 19

GUNAWAN, A. A.; SUHARTONO, D. et al. Music recommender system based on genre using convolutional recurrent neural networks. **Procedia Computer Science**, Elsevier, v. 157, p. 99–109, 2019. 30

HAGBERG, A.; SWART, P.; CHULT, D. S. **Exploring network structure, dynamics, and function using NetworkX**. [S.l.], 2008. 37

IACOVACCI, J.; LACASA, L. Visibility graphs for image processing. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 42, n. 4, p. 974–987, 2019. 21, 22

JENNINGS, H. D. et al. Variance fluctuations in nonstationary time series: a comparative study of music genres. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 336, n. 3-4, p. 585–594, 2004. 34

JR, M. O. Aspectos técnicos na coleta de dados linguísticos orais. **Metodologia de Coleta e Manipulação de dados em Sociolinguística**, p. 18, 2014. 17

JURKIEWICZ, S. Grafos–uma introdução. **São Paulo: OBMEP**, 2009. 19

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014. 39

LACASA, L. et al. From time series to complex networks: The visibility graph. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 105, n. 13, p. 4972–4975, 2008. 10, 21, 22, 23, 25

LACASA, L. et al. Time series irreversibility: a visibility graph approach. **The European Physical Journal B**, Springer, v. 85, n. 6, p. 1–11, 2012. 21

LIU, Y.-Y.; BARABÁSI, A.-L. Control principles of complex systems. **Reviews of Modern Physics**, APS, v. 88, n. 3, p. 035006, 2016. 25

LUQUE, B. et al. Horizontal visibility graphs: Exact results for random time series. **Physical Review E**, APS, v. 80, n. 4, p. 046103, 2009. 21, 24

MARWAN, N. et al. Complex network approach for recurrence analysis of time series. **Physics Letters A**, Elsevier, v. 373, n. 46, p. 4246–4254, 2009. 21

MCFEE, B. et al. librosa: Audio and music signal analysis in python. In: **Proceedings of the 14th python in science conference**. [S.l.: s.n.], 2015. v. 8, p. 18–25. 37

MELO, D. d. F. P. Estudo de padrões em sinais musicais sob a perspectiva dos grafos de visibilidade. Faculdade de Educação, 2019. 15, 23, 34

MELO, D. d. F. P.; FADIGAS, I. d. S.; PEREIRA, H. B. d. B. Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain. **Plos one**, Public Library of Science San Francisco, CA USA, v. 15, n. 11, p. e0240915, 2020. 21, 23, 31, 36, 38

MELO, D. d. F. P.; FADIGAS, I. d. S.; PEREIRA, H. B. de B. Categorisation of polyphonic musical signals by using modularity community detection in audio-associated visibility network. **Applied network science**, Springer, v. 2, n. 1, p. 1–15, 2017. 29

NEWMAN, M. E.; GIRVAN, M. Finding and evaluating community structure in networks. **Physical review E**, APS, v. 69, n. 2, p. 026113, 2004. 20, 36

QUEIROZ, R. A. B. de; MARAR, J. F.; OKIDA, C. M. Investigação dos coeficientes cepstrais da frequência mel para extração de características de gêneros musicais. 2015. 15

RAVASZ, E.; BARABÁSI, A.-L. Hierarchical organization in complex networks. **Physical review E**, APS, v. 67, n. 2, p. 026112, 2003. 20

ROSSING, T. D.; FLETCHER, N. H. **Principles of vibration and sound**. [S.l.]: Acoustical Society of America, 2004. 16

SALESSI, G. d. S. T. Rotulac ao automatica de musicas usando redes neurais profundas. 2020. 15

SILVA, V. d. Classificação multirrótulo aplicada a dados musicais. Universidade Federal de Santa Maria, 2014. 18

SMITH, J. O. **Mathematics of the discrete Fourier transform (DFT): with audio applications**. [S.l.]: Julius Smith, 2007. 18

SUPRIYA, S. et al. Weighted visibility graph with complex network features in the detection of epilepsy. **IEEE access**, IEEE, v. 4, p. 6554–6566, 2016. 22

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. **IEEE Transactions on Speech and Audio Processing**, v. 10, n. 5, p. 293–302, 2002. 12, 15, 16, 18, 19, 28, 33, 35, 50, 54

VARELA, C. B. **A study of visibility graphs for time series representations**. Dissertação (B.S. thesis) — Universitat Politècnica de Catalunya, 2020. 35

WALLIS, W. D. **A beginner's guide to graph theory**. [S.l.]: Springer Science & Business Media, 2007. 19

XU, X.; ZHANG, J.; SMALL, M. Superfamily phenomena and motifs of networks induced from time series. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 105, n. 50, p. 19601–19605, 2008. 21

YANG, Y.; YANG, H. Complex network-based time series analysis. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 387, n. 5-6, p. 1381–1386, 2008. 21

YELA, D. F. et al. Online visibility graphs: Encoding visibility in a binary search tree. **Physical Review Research**, APS, v. 2, n. 2, p. 023069, 2020. 24

ZHANG, J. et al. Characterizing pseudoperiodic time series through the complex network approach. **Physica D: Nonlinear Phenomena**, Elsevier, v. 237, n. 22, p. 2856–2865, 2008. 21

ZOU, Y. et al. Complex network approaches to nonlinear time series analysis. **Physics Reports**, Elsevier, v. 787, p. 1–97, 2019. 20, 21

Apêndices

APÊNDICE A – Artigo

Classificação de gêneros musicais por análise de grafos de visibilidade horizontal

Douglas Soares Molina[1], **Rafael de Santiago**[2]

[1]Departamento de Informática e Estatística
Universidade Federal de Santa Catarina – Florianópolis – SC – Brazil

Abstract. *The automatic extraction of informative attributes in music signals is gaining importance due to the question of structure and organization of the large number of music files available digitally in the Web. Among these attributes, rhythm plays a very important role in the definition of musical style. The study of rhythm in audio signals includes the investigation of the regularity characteristics of their acoustic and transient events. This work has mapped, based on the GTZAN dataset, the signal of 1000 music files into visibility graphs, where some topological properties were extracted through the calculations of Modularity, Number of Communities, Mean Degree and Density. A comparison was made between attributes generated from Natural Visibility Graphs and Horizontal Visibility Graphs for each signal, using them as input in a classification system based on supervised artificial neural networks, obtaining a similar precision rate when compared to another descriptor based on rhythmic attributes of the intensity of the signal Onsets, which represents the beginning of all its acoustic events. Based on the results obtained in this experiment and its related studies, it was evidenced that the Natural Visibility Graphs and Horizontal Visibility Graphs can be considered as a new alternative for the extraction of rhythmic characteristics for the retrieval of music information.*

Resumo. *A extração automática de atributos informativos em sinais musicais está ganhando importância devido à questão de estruturação e organização do grande número de arquivos de música disponíveis digitalmente na Web. Dentre esses atributos, o ritmo desempenha um papel muito importante na definição do estilo musical. O estudo da rítmica em sinais de áudio inclui a investigação das características de regularidade de seus eventos acústicos e transientes. Neste trabalho foi mapeado, a partir do banco de dados GTZAN, os sinais de 1000 arquivos musicais em grafos de visibilidade, onde extraiu-se algumas propriedades topológicas através dos cálculos de Modularidade, Número de Comunidades, Grau Médio e Densidade. Realizou-se uma comparação entre atributos gerados a partir de Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal para cada sinal, usando-os como entrada em um experimento de classificação baseado em redes neurais artificiais supervisionadas, onde obteve-se uma precisão semelhante à outro descritor com atributos rítmicos de intensidade de Onsets do sinal, que representam o início de todos seus eventos acústicos. Com base nos resultados obtidos neste experimento e nos estudos mencionados, foi evidenciado que os Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal podem ser considerados como uma nova alternativa para a extração de características rítmicas para a recuperação de informações de música.*

1. Introdução

O processo de evolução tecnológica levou estudos à explorar técnicas de armazenamento e análise de dados de forma à ter uma representação compacta e útil para os pesquisadores. Analisando as características mais relevantes extraídas de bibliotecas de arquivos de música digital, a classificação de gênero se torna uma das formas mais comuns de categorização do conteúdo.

Dentro deste contexto, torna-se imprescindível o desenvolvimento de métodos computacionais automáticos de agrupamento dos dados, que busquem sumarizar informações relevantes para a classificação dos gêneros, extraindo parâmetros numéricos através de diferentes tipos de algoritmos descritores. Dentre os algoritmos mais utilizados na síntese dos atributos musicais, estão aqueles que realizam operações no domínio tempo-frequência e que estão usualmente associados com propriedades que fazem parte da percepção musical dos humanos, como análise da textura do timbre, ritmo, variação temporal e conteúdo de tons harmônicos [Tzanetakis and Cook 2002].

A proposta deste trabalho consiste na análise de dados de áudio, com base nos estudos de trabalhos relacionados e publicados até o momento, explorando um possível descritor para a classificação de gêneros musicais sob a perspectiva de grafos de visibilidade. Essa transformação permite analisar uma relação interessante entre os pontos da série temporal, onde quanto maior o grau de conexão de um determinado vértice no grafo gerado, maior é a visibilidade do seu ponto correspondente na série. Após realizar o mapeamento, a rede terá herdado as características rítmicas da série através da qualidade e quantidade de grupos encontrados, e pelo grau de conexões geradas na rede.

2. Objetivos

Explorar e comparar possíveis descritores para o agrupamento de gêneros musicais, com base em métodos de transformação de sinais de arquivos de áudio em grafos de visibilidade, através da análise de similaridade de propriedades topológicas das redes geradas.

2.1. Objetivos Específicos

- Extrair e otimizar as séries temporais de arquivos de áudio, com base na intensidade do sinal.
- Mapear as séries temporais obtidas em Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal.
- Quantificar propriedades topológicas das redes obtidas, de forma à representar numericamente características relevantes à categorização dos gêneros musicais através do cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$).
- Desenvolver uma rede neural artificial para a classificação supervisionada, que é alimentada pelos dados encontrados sobre ambos os tipos de grafos.
- Classificar os gêneros com base nos grupos encontrados, comparando resultados entre os descritores encontrados e também com descritores relacionados à esta pesquisa.
- Divulgar os resultados.

3. Método de Pesquisa

Primeiramente, efetuou-se um levantamento teórico com intuito de fornecer maior credibilidade a pesquisa, com o objetivo de definir as características importantes dos sinais a serem utilizadas como descritores na rede neural.

Em seguida foi feito a análise em sinais de diversos arquivos de áudio de um banco de dados de larga escala e extração de seus atributos considerados relevantes para o estudo.

Após a conclusão da etapa de análise, foi construído um modelo de aprendizado de máquina e feita a comparação entre os diferentes descritores para o mesmo modelo de rede, analisando o resultado de acurácia e erro entre as configurações.

Por fim, foi apresentado algumas informações interessantes sobre o desempenho do método proposto e suas implicações para futuras pesquisas, concluindo toda a documentação do trabalho desenvolvido.

4. Fundamentação Teórica

Neste capítulo serão abordados assuntos relativos aos conceitos básicos utilizados neste trabalho, de forma à dar embasamento para a pesquisa.

4.1. Classificação de gêneros musicais

A extração automática de informações em sinais musicais está ganhando importância [de Queiroz et al. 2015] devido à questão de estruturação e organização do grande número de arquivos de música disponíveis digitalmente na Web.

A maioria dos trabalhos neste campo de pesquisa adota a estratégia de categorização de gêneros musicais utilizando a extração de atributos comuns de sinais musicais(ritmo, melodia e timbre) como uma de suas etapas essenciais [Tzanetakis and Cook 2002].

De modo geral, no ramo de classificação de gêneros musicais, o sinal de áudio passa primeiramente por um pré-processamento a fim de otimizar os dados e transformalos em uma estrutura própria aos algoritmos de extração de atributos. Feito isso, são calculadas representações numéricas da natureza tonal, rítmica, e timbrística musical que segue, na maioria dos estudos, o modelo adotado por Tzanetakis e Cook [Tzanetakis and Cook 2002]. Ao final dessa etapa, cada sinal dos arquivos musicais estará representado por um conjunto de números que representam vetores de atributos. Finalmente, o classificador prediz o gênero ou a probabilidade de diferentes gêneros musicais a partir dos vetores de atributos, usando árvores de decisão, modelos probabilísticos e aprendizagem de máquina, onde é feito a decisão sobre um ou mais gêneros musicais, dependendo da técnica de classificação abordada.

4.2. Extração de Atributos

A extração de atributos é uma etapa crucial dentro do desenvolvimento de sistemas de reconhecimento de padrões [Costa et al. 2013]. No caso de aplicações voltadas para a classificação de sinais de áudio musical, o processo geralmente está relacionado à extração de atributos de ritmo, melodia e timbre [Tzanetakis and Cook 2002].

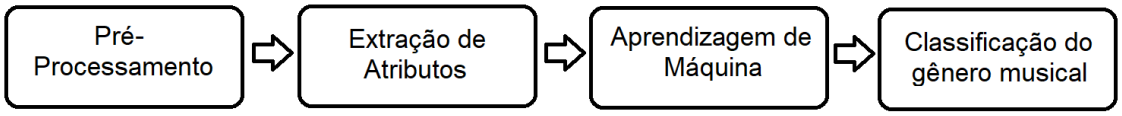


Figure 1. Processo comum de um sistema de classificação de gêneros musicais. Fonte: Autor

Amostras de áudio, obtidas através da quantificação da onda sonora, não podem ser utilizadas diretamente por sistemas de analise automática. O sinal pode apresentar uma quantidade de dados muito grande que não são relevantes para a classificação. Sendo assim, o primeiro passo da maioria dos sistemas descritores é a filtragem de algumas caracterısticas dos dados do áudio para manipular informação mais significativa, a fim de reduzir a quantidade de processamento desnecessário posteriormente.

Para o estudo deste trabalho, será utilizado e comparado descritores com atributos rítmicos construídos a partir da detecção de Onsets, que referem-se ao início de eventos acústicos do sinal. Em contraste com os estudos que se concentram na detecção de batidas e ritmo por meio da análise de periodicidades, um detector de Onset enfrenta o desafio de detectar eventos únicos, que não precisam seguir um padrão periódico, mas que tenta encontrar mudanças repentinas na dinâmica, timbre ou estrutura harmônica do sinal [Gouyon et al. 2006]. Mais precisamente, será calculado o envelope de intensidade de Onset de fluxo espectral de cada sinal.

4.3. Grafos

Um grafo é uma estrutura capaz de representar informações sobre relações de conjuntos. Define-se um grafo por G = (V, E), onde V = V (G) é o conjunto finito de objetos chamados vértices e E = E(G) são pares não ordenados de vértices, denominado de arestas [Wallis 2007].

Por possuir uma estrutura possível de armazenar informações complexas e com conceito de fácil implementação, atualmente é encontrado para definir relações em um enorme volume de dados. É facilmente integrada com sistemas que utilizam métodos tradicionais de predição e aprendizado de máquina [Jurkiewicz 2009].

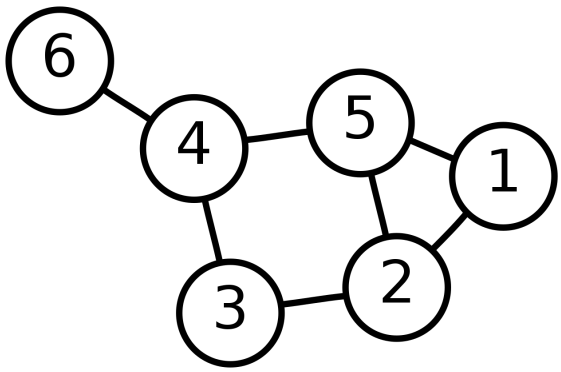


Figure 2. Exemplo de um grafo com 6 vértices. Fonte: Autor

4.4. Métodos de Transformação de Séries Temporais em Grafos

Uma série temporal é uma sequência de pontos de dados tempo de observações, que são feitas em pontos sucessivos, em muitos casos igualmente espaçados no tempo. Desta forma, os dados de séries temporais têm uma ordenação temporal discreta natural [Zou et al. 2019]. Séries temporais cobrem uma grande variedade de variáveis potencialmente relevantes para a vida cotidiana e são muito utilizadas para análise de dados.

Para tornar as séries temporais acessíveis a técnicas complexas de análise de redes e aprendizado de máquina, primeiramente é necessário encontrar uma representação de rede adequada, ou seja, um algoritmo que defina quais são os vértices e arestas da rede. De acordo com [Zou et al. 2019], existem pelo menos (mas não se limita a) três classes principais de abordagens comuns de redes complexas para a análise de séries temporais individuais:

- Similaridade estatística mútua ou proximidade métrica entre diferentes segmentos de uma série temporal (Redes de Proximidade)
- Probabilidades de transição entre estados discretos (Redes de Transição)
- Convexidade de observações sucessivas (Grafos de Visibilidade)

O grafo de visibilidade e suas várias variantes têm aplicações importantes, como auxiliar em processamento de imagens [Iacovacci and Lacasa 2019], testes estatísticos para irreversibilidade de séries temporais [Lacasa et al. 2012] e também em especial na questão de classificação de gêneros musicais [Melo et al. 2020].

Uma vez que os grafos de visibilidade serão adotados na metodologia desta pesquisa, uma descrição detalhada sobre esse método será dada na próxima seção.

4.5. Grafos de Visibilidade

Os grafos de visibilidade têm criado pontes entre a análise de séries temporais e a análise de redes complexas, possibilitando o uso de novas ferramentas para a compreensão de fenômenos representados por sequências temporais.

Podem ser definidos por redes geradas a partir de séries numéricas, onde cada ponto da série é considerado um vértice do grafo, e a ligação ou não entre dois vértices depende da "visibilidade" entre os pontos da série [Lacasa et al. 2008]. A visibilidade entre dois pontos é definida por um critério trigonométrico aplicado aos pontos da série. Nos grafos de visibilidade quanto maior o grau de conexão de um determinado vértice, maior é a visibilidade do seu ponto correspondente na série, em relação à sua vizinhança.

Segundo [Lacasa et al. 2008], o critério de visibilidade pode ser definido da seguinte forma: dada uma série temporal V1, V2, ..., Vn sempre haverá visibilidade entre dois pontos consecutivos da série temporal, e dois pontos arbitrários A(xa, Va) e B(xb, Vb) da série terão visibilidade mútua, se todo ponto C(xc, Vc) entre eles satisfaz a condição:

$$\frac{Vb - Vc}{xb - xc} > \frac{Vb - Va}{xb - xa} \tag{1}$$

Devido à natureza dos grafos de visibilidade, as divisões de vértices correspondam aos pontos mais baixos da série temporal (onde a visibilidade direta entre os pontos

de dados é mais fácil) e os picos regionais da série provavelmente dividem os segmentos. Dessa forma, o resultado da segmentação da série temporal também reflete esse comportamento, o qual pode trazer relações interessantes entre as características de persistência de transientes de sinal e as características topológicas de detecção de comunidades em seus grafos associados. Isso pode sugerir, por exemplo, que gêneros musicais com picos de sinais mais definidos e repetitivos (como no caso de músicas eletrônicas) podem gerar um maior número de vértices após a transformação. Além disso, alguns vértices funcionam como ”hubs” do grafo (os vértices mais conectados), representando os dados com os maiores valores na série.

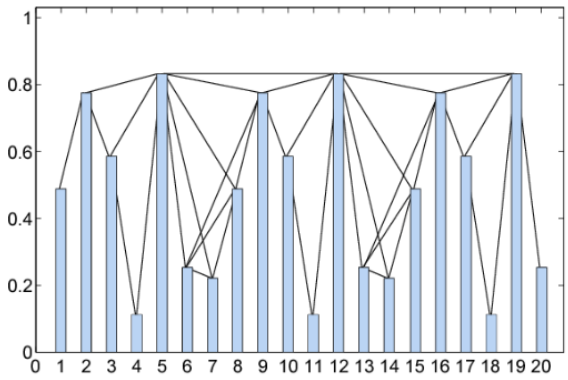


Figure 3. Exemplo do mapeamento de uma série temporal em um grafo de visibilidade. Fonte: [Lacasa et al. 2008]

4.5.1. Grafos de Visibilidade Horizontal

Os grafos de visibilidade possuem uma gama de variantes com diversas aplicações. Uma versão geometricamente mais simples foi proposta [Luque et al. 2009], de forma a ser computacionalmente mais eficiente do que a versão do algoritmo normal de Grafos de Visibilidade (também chamados de Grafos de Visibilidade Natural), com o foco também no mapeamento de séries temporais.

Sua visibilidade, mais restrita do que o caso geral análogo, é analisada através de retas horizontais e, por isso, é nomeado como Grafo de Visibilidade Horizontal (GVH). Ou seja, os vértices do GVH terão menos visibilidade do que no caso de um Grafo de Visibilidade Natural. Embora tal fato possa ter um leve impacto sobre os aspectos qualitativos dos grafos, a simplicidade da versão do algoritmo horizontal permite que ele seja aplicado, em resolução de um tempo computacional mais otimizado, à séries muito longas, como é o caso de alguns bancos de dados de informação musical.

Assim, estabelece-se o seguinte critério de visibilidade: dois valores arbitrários da série temporal A(xa, Va) e B(xb, Vb) terão visibilidade horizontal e, consequentemente, se tornarão dois vértices conectados por uma aresta no grafo associado, se todos os outros termos (xc, Vc) intermediários entre eles cumprirem a relação

$$xa, xb > xc \tag{2}$$

para todo c, tal que

$$a < c < b \tag{3}$$

No caso de Grafos de Visibilidade Natural, o número total de verificações necessárias para obter o grafo de uma série temporal de n pontos de dados é igual a n(n-1)/2, correspondendo a uma complexidade de tempo O(n²). Já na análise de visibilidade horizontal, pode-se dar um passo adiante e assumir com segurança que nenhum ponto após um valor maior do que um valor computado será visível horizontalmente [Yela et al. 2020]. Esta observação reduz efetivamente a complexidade do tempo da construção para O(n log(n)) e, no caso de sinais ruidosos (estocásticos ou caóticos), pode-se inferir que este algoritmo tem uma complexidade de tempo de caso médio O(n). No entanto, todos os pares de pontos precisam ser verificados no caso de visibilidade natural.

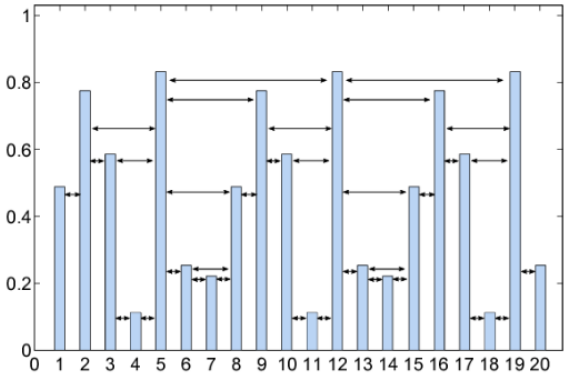


Figure 4. Exemplo do mapeamento de uma série temporal em um grafo de visibilidade horizontal. Fonte: [Lacasa et al. 2008]

4.6. Detecção de Comunidades

Um tópico de grande interesse no estudo de redes complexas e no ramo de análise de dados é a identificação de comunidades, uma vez que permite uma visão da relação funcional e estrutural de uma rede, e possui abrangentes áreas de aplicação.

As comunidades são identificadas quando há uma quantidade maior de ligações entre vértices de um mesmo subgrupo da rede, e quantidade menor de ligações entre vértices que pertencem a subgrupos diferentes [Liu and Barabási 2016].

No intuito de mensurar o quão bem formados são os agrupamentos encontrados nas redes, foi proposto o conceito de Modularidade(Q), para estabelecer um indicador sobre as divisões das comunidades. Pode ser definida como a medida da força de divisão de uma rede em módulos (ou comunidades). Quanto maior o valor de M (modularidade), mais forte é a conexão entre os vértices dentro dos módulos.

A maximização da modularidade gananciosa [Clauset et al. 2004], utilizada no estudo deste trabalho, começa com cada vértice em sua própria comunidade e junta repetidamente o par de comunidades que causam o aumento de modularidade até obter o maior valor possível.

Essas propriedades de redes podem ser utilizadas à favor da classificação de gêneros musicais, conforme é proposto e apresentado neste trabalho.

5. Trabalhos Relacionados

A seguinte seção apresenta alguns trabalhos relacionados que auxiliaram no desenvolvimento e conclusões deste estudo, considerando os critérios de classificação para gêneros musicais, extração de atributos de sinais musicais de diversas naturezas, análise de redes complexas e técnicas de aprendizado de máquina.

5.1. Musical genre classification of audio signals

No trabalho de [Tzanetakis and Cook 2002], foi proposto a classificação automática de sinais de áudio em uma hierarquia de gêneros musicais. Foram propostos três conjuntos de recursos para textura de timbre, conteúdo rítmico e melodia.

O vetor de atributos consistiu com recursos de textura de timbre de 19 atributos (MFCCs, Fluxo Espectral, Taxa de Passagem Pelo Zero, Centróide Espectral, Rollof Espectral e Low-Energy), 6 atributos de conteúdo rítmico (Histograma de Batidas) e 5 atributos de conteúdo de tom (Histograma de Tom), resultando em um vetor de recursos de 30 dimensões.

O trabalho evidenciou que embora certas características individuais de atributos de classificação estejam correlacionadas, a adição de cada recurso específico melhora, de modo geral, a precisão da classificação dos grupos. Os conjuntos de recursos de conteúdo rítmico (Histograma de Batidas) e de tom (Histograma de Tom) parecem desempenhar um papel menos importante na classificação em comparação com atributos de natureza de timbre (STFT, MFCCs) em todos os casos.

Os resultados mostraram que a classificação alcançou os menores índices de acerto para os gêneros de Rock (40%) e Blues(43%), enquanto os gêneros Jazz(75%), Clássico(69%), Pop(66%) e Hiphop(64%) ficaram entre as categorias de maior acurácia. Através da análise da matriz de confusão da classificação, as taxas de resultados incorretos do sistema são semelhantes ao que um humano faria. Por exemplo, a música clássica é erroneamente classificada como música Jazz para peças com ritmo forte de compositores como Leonard Bernstein e George Gershwin. A música de Rock tem a pior precisão de classificação e é facilmente confundido com outros gêneros, o que é esperado, considerando sua natureza ampla.

5.2. Categorisation of polyphonic musical signals by using modularity community detection in audio-associated visibility network

Em 2017, [Melo et al. 2017] propõe um método para caracterizar numericamente a homogeneidade de sinais musicais polifônicos por meio da detecção de comunidades em redes de visibilidade associadas ao áudio e detectar padrões que permitam a categorização desses sinais em dois tipos de agrupamento de natureza rítmica. Observou-se que uma maior ou menor homogeneidade das magnitudes dos transientes de sinal está relacionada a uma maior ou menor modularidade de sua rede de visibilidade associada. Notou-se também que essas diferenças estão relacionadas a escolhas musicais que podem estabelecer diferenças importantes entre os estilos musicais.

No estudo foram utilizados 120 arquivos musicais das categorias Sinfônica e Percussiva, cada uma com 60 músicas. Para a música sinfônica, foram selecionadas peças de quarteto de cordas e orquestra completa. As composições incluíam concertos de Bach, sinfonias de Mozart e quartetos de Debussy, Dutileux e Ravel.

Após os experimentos, conclui-se que as redes percussivas apresentam tendência com média de alta Modularidade(0.836) e Número de Comunidades(19), além de apresentar baixo Grau Médio(15.12), por tendências que estão fortemente ligadas às escolhas musicais que influenciam o design dos transientes do sinal de cada gênero. Por outro lado, as redes sinfônicas tendem a apresentar baixa Modularidade(0.558) e Número de Comunidade(8.5), além de apresentar alto Grau Médio(46.27). Isso se deve porque as músicas Sinfônicas usam muito mais variações na dinâmica e menos persistência rítmica do que as músicas Percussivas, resultando em sinais mais heterogêneos e grafos de visibilidade com valores de menores de Modularidade.

5.3. Music recommender system based on genre using convolutional recurrent neural networks

Em 2019, [Gunawan et al. 2019] foi realizado um estudo com redes neurais recorrentes convolucionais (CRNNs) para extração de recursos e relações de similaridade entre os sinais de áudio. CRNNs é uma combinação de redes neurais convolucionais (CNNs) e redes neurais recorrentes (RNNs). As CNNs são especialmente adequadas para prever recursos musicais de alto nível, como acordes e batidas, porque permitem uma estrutura hierárquica que consiste em recursos intermediários em várias escalas de tempo. Já as RNNs foram projetadas para trabalhar com dados de séries temporais, especialmente para problemas de previsão de sequência temporal.

No estudo, foi comparado o desempenho de arquiteturas de CNNs com CRNNs para classificar gêneros musicais. O modelo toma como entrada o espectrograma de quadros musicais e analisa a imagem usando CRNNs. A saída do modelo é um vetor de gêneros previstos para a música. O principal resultado do estudo é que a precisão dos CRNNs é ligeiramente superior aos métodos de CNNs que combina os domínios da frequência e do tempo e usa o mesmo número de parâmetros.

Os resultados mostraram que a classificação alcançou os menores índices de acerto para os gêneros de Eletrônico (64% com CNN e 70% com RNN) e Instrumental (79% com CNN e 75% com RNN). Por outro lado, obteve uma acurácia surpreendente próximo de 94% para ambas as redes para o gênero de Rock, onde outros trabalhos relacionados mostraram dificuldade em categorizar.

5.4. Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain

Em 2020, o estudo de [Melo et al. 2020] demonstrou um novo método para extrair características de sinais de áudio para classificar músicas. A tese propõe um novo descritor baseado em propriedades topológicas utilizando grafos de visibilidade, denominado de Audio Signal Visibility Descriptor (ASVD). O descritor apresentou ótimos resultados ao analisar cálculos de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$) para detecção de padrões de auto-similaridade nos sinais de dados musicais e classificou gêneros musicais com alta taxa de acurácia.

No trabalho, foi realizado uma comparação mais direta de propriedades de natureza rítmica, e obteu-se uma precisão maior ou igual ao histograma de batidas em 70% dos pares de gêneros musicais, onde foi evidenciado que as quatro propriedades utilizadas da rede estavam entre as primeiras posições atribuídas pelo teste.

Em um sistema de classificação usando apenas o descritor de propriedades dos grafos, obteu-se uma precisão média de 39%. Foi feita a comparação das instâncias classificadas corretamente por este sistema com outro sistema utilizando apenas o histograma de batidas, e então, em uma comparação pairwise de gêneros, obteu-se uma precisão maior ou igual ao segundo sistema em 70% dos pares de gêneros musicais. Considerando um cenário com 18 atributos de processamento de sinal de áudio mais o ASVD, a precisão média da classificação foi de 76,7%, comparável ou superior a vários estudos relacionados. Em mais um experimento de classificação usando os mesmos 18 atributos do experimento anterior, e usando o histograma de batidas em vez do ASVD, foi obtido uma precisão igual ou superior em metade dos dez grupos de gêneros musicais.

5.5. Comparativo

As diferenças mais relevantes entre os trabalhos analisados foram: os parâmetros extraídos dos sinais de áudio, a base de dados utilizada e o classificador implementado.

Percebe-se que os atributos utilizados nos trabalhos possuem naturezas distintas e a combinação variada dessas propriedades resulta em diferentes taxas de acurácia para cada gênero musical, visto que cada cada classe apresenta variadas características de timbre, tom e rítmo.

Outro ponto importante é o extenso uso de MFCCs, dos trabalhos relacionados analisados, todos concluíram que esses parâmetros são um importante avanço para a classificação de gêneros em sinais musicais. Em relação aos dados utilizados, as publicações citadas conseguiram resultados relevantes com ambos dados de base públicas e dados próprios.

A Tabela a seguir explicita alguns parâmetros importantes considerados por cada artigo citado.

Trabalho	Parâmetros	Classificadores	Base de Dados	Acurácia máxima
Tzanetakis; Cook, 2002	MFCCs, STFT, PHF, BHF	GS, GMM, K-NN	GTZAN	61%
Gunawan; Suhartono et al., 2019	MFCCs, STFT	CRNNs, CNN	FMA	71%
Melo; Fadigas; Pereira, 2020	MFCCs, BHF, ASVD	ANN	GTZAN	76.7%

Table 1. Comparação entre trabalhos citados. Fonte: Autor.

6. Desenvolvimento

6.1. Banco de Dados

A pesquisa utilizou dados do banco GTZAN, criado por [Tzanetakis and Cook 2002], o qual oferece 100 arquivos de áudio, divididos em 10 gêneros musicais (Clássico, Jazz, Blues, Pop, Rock, Hip-Hop, Metal, Disco, Reggae, Country). Esse banco de dados tem sido usado em larga escala nas pesquisas de gêneros musicais, permitindo a análise de composição dos sinais de áudio e disponibilizando valores de propriedades de natureza timbrística, rítmica e tonal, compatíveis com entradas de sistemas de aprendizado de máquina.

6.2. Processos da metodologia

A proposta deste trabalho consiste na análise de dados de áudio, com base nos estudos de trabalhos relacionados apresentados anteriormente, explorando um possível descritor

para a classificação de gêneros musicais sob a perspectiva de Grafos de Visibilidade Horizontal. A metodologia da tese pode ser dividida nos seguintes passos:

- Extrair e otimizar as séries temporais de arquivos de áudio, com base na intensidade do sinal.
- Mapear as séries temporais obtidas em Grafos de Visibilidade Natural e Grafos de Visibilidade Horeizontal.
- Quantificar propriedades topológicas das redes obtidas, de forma à representar numericamente características relevantes à categorização dos gêneros musicais através do cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau Médio($\langle k \rangle$) e Densidade($\triangle$).
- Desenvolver uma rede neural artificial para a classificação supervisionada, que é alimentada pelos dados encontrados sobre ambos os tipos de grafos.
- Classificar os gêneros com base nos grupos encontrados, comparando resultados entre os descritores encontrados e também com descritores relacionados à esta pesquisa.
- Divulgar os resultados.

6.2.1. Extração da série temporal

A extração da série temporal consiste em analisar sinais de arquivos de áudio de forma otimizada, de modo à obter informações essenciais da série com o mínimo de dados possível. Estudos mostram que não existe grande diferença entre o uso de diferentes taxas de amostragem nos sinais de áudio [Melo 2019], produzindo estatísticas muito semelhantes para fins de estudo comparativo usando a abordagem de grafos de visibilidade. Sendo assim, os sinais serão reduzidos à uma taxa de 11.000Hz, considerando uma duração de 30s por arquivo.

Com base nos dados obtidos, o sinal é dividido em blocos de tamanho fixo e é definido uma série de variâncias entre os pontos de cada bloco, de forma à otimizar a análise com uma representação reduzida do sinal de áudio [Jennings et al. 2004].

Seja Ui = U1, U2,...,Un a série temporal de amostras que representam o sinal de áudio. O número total de pontos N é a função N = A × t, onde A é a taxa de amostragem e t é a duração total do sinal. Essa série é dividida em subséries Vj = V1, V2,...,Vn de tamanho fixo de amostras com m=450 pontos cada, calculando a variância de cada subsérie, o qual é representado na equação:

$$Vj = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda}(Ui - Uj)^2}{\lambda - 1} \tag{4}$$

onde

$$Uj = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} Ui}{\lambda} \tag{5}$$

Sendo assim, cada subsérie de variância V(j) é criada com 700 pontos.

6.2.2. Transformação da série V(j) em grafos de visibilidade

Cada um dos pontos da série V(j) é considerado como um vértice da rede. Dois vértices da rede são conectados por uma aresta cada vez que dois pontos da série V(j) atendem o critério de visibilidade definido por cada uma das equações de Grafo de Visibilidade Natural e Grafo de Visibilidade Horizontal.

Observe que, devido à natureza dos grafos de visibilidade, as divisões de vértices correspondam aos pontos mais baixos da série temporal (onde a visibilidade direta entre os pontos de dados é mais fácil) e os picos regionais da série provavelmente dividem os segmentos [Bergillos Varela 2020]. Portanto, o resultado da segmentação da série temporal também refletirá esse comportamento.

Esse comportamento também é observado em estudos clássicos [Tzanetakis and Cook 2002] dessa área, referindo essa característica pela observação da intensidade dos picos. Os autores concluem que composições de arquivos de áudio que possuem trechos musicais com batidas mais fortes e persistentes irão gerar sinais com picos mais elevados e, consequentemente, segmentar os grupos encontrados no grafo de visibilidade. Da mesma forma, quanto menor a persistência e força dos batimentos principais, maior a chance dessas amostras participarem de um mesmo grupo. [Borges et al. 2010]

6.2.3. Modularidade e Número de Comunidades

A modularidade é uma medida de estrutura de uma rede. Esta medida é projetada para estimar a força de uma divisão de uma rede em módulos (ou comunidades). Uma rede com alta modularidade possui conexões densas entre os vértices dentro dos módulos, mas conexões esparsas entre os vértices em diferentes módulos [Melo et al. 2020]. Um alto valor de modularidade indica que a densidade de arestas dentro das comunidades é maior do que o esperado ao acaso, indicando uma boa partição da rede.

A modularidade é definida em [Newman and Girvan 2004] como

$$Q = \frac{1}{2m} \sum_{(i,j)} \left(A(i,j) - \gamma \frac{k(i)k(j)}{2m} \right) \delta(ci, cj) \tag{6}$$

onde m é o número de arestas, A é a matriz de adjacência de G, k(i) é o grau de i, γ é o parâmetro de resolução e δ(ci, cj) é 1 se i e j estão na mesma comunidade, caso contrário o valor é 0.

De acordo com [Newman and Girvan 2004], isso pode ser reduzido a

$$Q = \sum_{c=1}^{n} \left[\frac{L(c)}{m} - \gamma (\frac{k(c)}{2m})^2 \right] \tag{7}$$

onde a soma itera sobre todas as comunidades c, m é o número de arestas, L(c) é o número de conexões intracomunitários para a comunidade c, k(c) é a soma dos graus dos vértices na comunidade c e γ é o parâmetro de resolução.

O parâmetro de resolução define uma compensação arbitrária entre bordas intragrupo e bordas intergrupo. Padrões de agrupamento mais complexos podem ser descobertos analisando a mesma rede com vários valores de gama e combinando os resultados. A segunda fórmula é aquela realmente utilizada no cálculo da modularidade.

6.2.4. Grau médio

O grau de um vértice corresponde ao número total de suas arestas. Seja k(i) o grau do vértice i de uma rede. O grau médio de uma rede com N vértices é a média aritmética de k(i).

$$\langle k \rangle = \frac{1}{N} \times \sum_{i=1}^{N} k(i) \tag{8}$$

Este parâmetro mede a intensidade média da conectividade de cada vértice da rede. Nos grafos de visibilidade, esta medida pode ser interpretada como o nível médio de visibilidade local dos picos de sinal. Sinais em que predominam poucos picos de alta visibilidade local gerarão grafos de visibilidade com grau médio mais altos do que sinais com muitos picos de baixa visibilidade local.

6.2.5. Densidade

Seja N o número de vértices de um grafo. A densidade é a razão entre o número total de arestas de uma rede (m = —E—) e o maior número possível de arestas.

$$\triangle = \frac{2 \times m}{N(N-1)} \tag{9}$$

A densidade mede o nível geral de conectividade de rede. Nos grafos de visibilidade associados aos sinais de áudio, esta medida indica o nível de visibilidade geral destes sinais. Quanto maior o nível de persistência rítmica no sinal, menor a visibilidade geral e menor a densidade.

6.2.6. Extração dos atributos na prática

Para a extração dos atributos dos grafos, foi utilizado o auxílio de biblioteca NetworkX [Hagberg et al. 2008], o qual provém ferramentas para a criação, manipulação e estudo da estrutura, dinâmica e funções de redes complexas. Além disso, também foi utilizado a biblioteca Librosa [McFee et al. 2015], que é um pacote feito em Python utilizado em muitos estudos para análise de áudio e música. Ele fornece os blocos de construção necessários para criar sistemas de recuperação de informações musicais, e foi essencial para à extração correta dos atributos para a classificação dos gêneros.

6.2.7. Classificação dos gêneros sob propriedades das redes obtidas

As classificações realizadas nesse trabalho foram realizadas com o auxílio de redes neurais artificiais (RNA). Para isso, utilizou-se a API de aprendizagem profunda Keras, que disponibiliza uma interface produtiva e completa para soluções envolvendo aprendizado de máquina, desenvolvida em cima da plataforma de código aberto TensorFlow. Através dessa API, foi possível a definição de diferentes topologias de redes neurais para testes, assim como a customização de hiperparâmetros.

Durante o processo de refinamento das redes desenvolvidas, os hiperparâmetros manipulados foram: número de unidades de processamento das camadas, taxa de aprendizado, dropout, tamanho de batch e épocas de treinamento. O processo de treinamento e refinamento das redes neurais testadas foi documentado e discutido no capítulo seguinte.

7. Experimentos

Cada uma das 1000 amostras de áudio de 30 segundos foi segmentado em 10 partes de 3 segundos, resultando em 10000 amostras de áudio no total, sendo 1000 amostras por gênero musical. Feito isso, cada uma das amostras foi transformada em uma série temporal de suas variações, e depois foi gerado um Grafo de Visibilidade Natural e um Grafo de Visibilidade Horizontal para cada série, totalizando 10000 grafos de visibilidade de cada tipo.

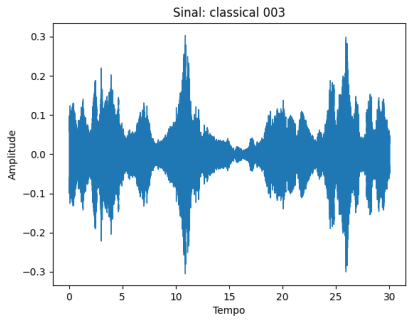


Figure 5. Sinal de Clássico. Fonte: Autor.

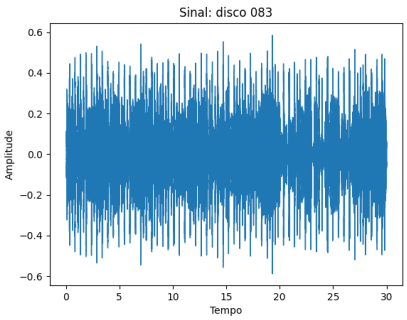


Figure 6. Sinal de Disco. Fonte: Autor.

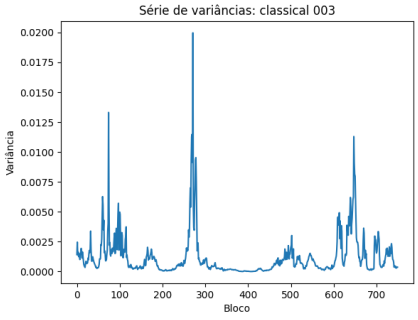


Figure 7. Série de variância de Clássico. Fonte: Autor.

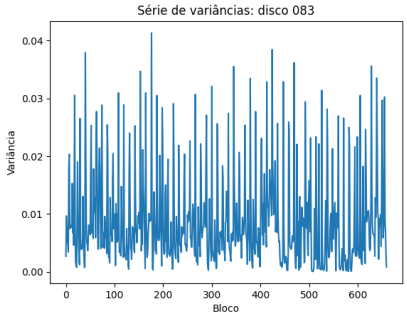


Figure 8. Série de variância de Disco. Fonte: Autor.

De modo geral, os sinais de arquivos de áudio que possuem trechos musicais com percussões mais fortes e persistentes geraram sinais com picos mais elevados e, consequentemente, segmentaram as possíveis comunidades encontradas nos grafos de visibilidade, como será demonstrado a seguir. Da mesma forma, quanto menor a persistência e força dos batimentos principais, maior a chance dessas amostras participarem de um mesmo grupo.

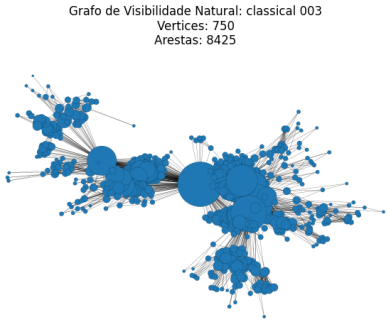


Figure 9. Grafos de Visibilidade Natural de Clássico. Fonte: Autor.

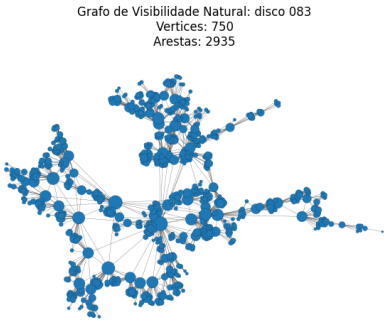


Figure 10. Grafos de Visibilidade Natural de Disco. Fonte: Autor.

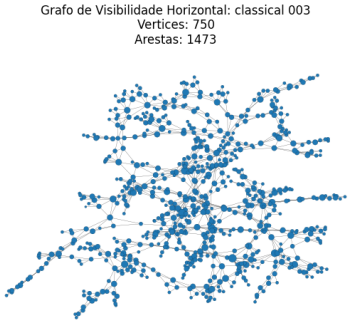


Figure 11. Grafos de Visibilidade Horizontal de Clássico. Fonte: Autor.

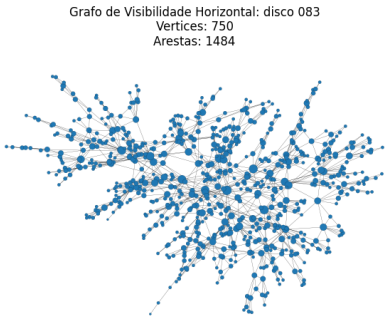


Figure 12. Grafos de Visibilidade Horizontal de Disco. Fonte: Autor.

A representação do tamanho dos vértices corresponde à uma proporção linear de suas respectivas taxa de conectividade com outros vértices, calculada pelo valor de Grau. No geral, sinais que possuem picos mais definidos e que prevalecem na amplitude, como é o caso do sinal de Clássico, terão vértices com maior densidade e consequentemente o grafo será mais denso. Em sinais que os picos são mais nivelados e com pouca variação, como é o caso do sinal de Disco, terão vértices com menor densidade e consequentemente o grafo será mais esparso.

As Tabelas 2 e 3 mostram a média e o desvio padrão para o cálculo de Modularidade(Q), Número de Comunidades(Nc), Grau médio($\langle k \rangle$) e a Densidade($\triangle$) dos dois tipos de grafos de visibilidade correspondentes a 100 amostras de áudio agrupadas em 10 gêneros musicais.

Para os resultados encontrados, é evidenciado uma forte relação entre Q e Nc, onde é demonstrado de forma mais clara a diferença entre a natureza de percussão de cada estilo. Além disso, é possível observar também uma relação entre $\langle k \rangle$ e $\triangle$ de cada

gênero, o que corrobora com a ideia sobre as fortes relações entre as variações locais do sinal.

[h]

Gênero	Q	σ	Nc	σ	$\langle k \rangle$	σ	$\triangle$	σ
Blues	0.808	0.037	11.71	2.046	9.448	2.593	0.013	0.004
Clássico	0.698	0.103	10.28	2.836	14.662	5.237	0.020	0.007
Country	0.809	0.035	11.31	1.791	9.583	2.391	0.013	0.003
Disco	0.839	0.017	12.54	1.388	8.255	1.041	0.011	0.001
Hip Hop	0.837	0.019	12.41	1.471	8.863	1.433	0.012	0.002
Jazz	0.797	0.035	11.35	2.002	10.501	2.948	0.014	0.004
Metal	0.826	0.021	12.23	1.752	7.347	1.035	0.010	0.001
Pop	0.829	0.028	12.44	1.788	9.132	1.534	0.012	0.002
Reggae	0.829	0.021	11.63	1.581	8.916	1.515	0.012	0.002
Rock	0.820	0.031	11.66	1.273	8.413	1.267	0.011	0.002

Table 2. Media e desvio padrão das propriedades topológicas para os Grafos de Visibilidade Natural. Fonte: Autor.

Gênero	Q	σ	Nc	σ	$\langle k \rangle$	σ	$\triangle$	σ
Blues	0.879	0.009	17.08	2.235	3.951	0.016	0.005	0.000
Clássico	0.872	0.009	15.46	1.731	3.923	0.027	0.005	0.000
Country	0.880	0.008	16.93	2.263	3.953	0.017	0.005	0.000
Disco	0.892	0.008	17.97	2.290	3.958	0.011	0.005	0.000
Hip Hop	0.880	0.011	17.92	2.787	3.949	0.022	0.005	0.000
Jazz	0.878	0.007	16.43	1.846	3.941	0.021	0.005	0.000
Metal	0.883	0.009	17.96	2.153	3.958	0.010	0.005	0.000
Pop	0.881	0.009	17.38	2.441	3.954	0.012	0.005	0.000
Reggae	0.880	0.009	17.90	2.412	3.954	0.015	0.005	0.000
Rock	0.882	0.009	17.03	2.251	3.955	0.014	0.005	0.000

Table 3. Media e desvio padrão das propriedades topológicas para os Grafos de Visibilidade Horizontal. Fonte: Autor.

Na observação sobre a perspectiva de Grafos de Visibilidade Natural, os valores médios para Modularidade e Número de Comunidades apresentam o gênero Clássico(menor valor) e Disco(maior valor) nos extremos da tabela. Da mesma forma, apesar da pequena diferença e leves divergências de posicionamento, gêneros como Hip-Hop e e Pop também são apresentados com altos valores na comparação, e gêneros como Jazz e Country são evidenciados com os valores mais baixos.

As médias de Gráu Médio e Densidade mantêm uma relação de certa forma oposta aos cálculos de Modularidade e Número de Comunidades, de forma que o gênero Clássico ocupa a maior posição e o gênero Disco ocupa a segunda menor posição. Os gêneros de Jazz e Country agora são posicionados nos maiores valores da tabela. Também observa-se que o gênero de Metal ocupa o menor valor para ambos os atributos, que também pode ser observado por picos bem definidos e intensos na análise do sinal. De qualquer forma, é bem evidente a diferença de grau médio entre os dois opostos da tabela, visto que suas características de natureza rítmica são de certa forma contrárias.

Para todos os quatro componentes, os gêneros de Reggae e Rock ocupam a posição intermediária. Esse tipo de organização hierárquica corrobora com a ideia de que gêneros

musicais que optam por arranjos instrumentais muito “densos”, “intensos” e persistentes tendem a ocupar posições opostas a gêneros com texturas instrumentais mais ricas em dinâmicas. Em uma posição intermediária estão os estilos musicais que buscam equilibrar a essência de influências de ambos os extremos.

Já na observação sobre a perspectiva de Grafos de Visibilidade Horizontal, a maioria dos padrões permanecem verdadeiros, com os menores valores médios de Modularidade e Número de Comunidades ainda de gêneros como Clássico e Jazz. Porém, existe uma forte divergência do resultado para os cálculos de Grau Médio e Densidade, o qual demonstra posicionamento de gêneros de forma oposta em relação aos atributos de Grafos de Visibilidade Natural, com gêneros como Clássico, Jazz e Hip-Hop ocupando as posições de valores mais baixos e com gêneros como Disco e Metal ainda ocupando os valores mais elevados.

Essa divergência se deve principalmente à natureza do método de transformação do sinal em Grafos de Visibilidade Horizontal, uma vez que é uma simplificação da transformação normal de visibilidade. Pela forte divisão dos segmentos do sinal, é comum que propriedades de conectividade e densidade dos grafos seja reduzido, quando em comparação à aqueles gerados por Grafos de Visibilidade Natural. Esse comportamento também é observado pelo baixo valor de desvio padrão entre os gêneros para esses atributos na Tabela 3.

8. Aprendizado de máquina e classificação

O aprendizado de máquina e a classificação foram realizados com redes neurais artificiais supervisionadas levando em consideração dois cenários. No primeiro cenário, foi utilizado um vetor de atributos apenas com os atributos de grafos (Modularidade, Número de Comunidades, Grau Médio e Densidade), para ambos os tipos Natural e Horizontal. A ideia era explorar uma situação em que apenas um descritor de atividade rítmica fosse usado. Em seguida, foi realizado o aprendizado e a classificação apenas com o atributo da frequência e intensidade de Onsets, o qual refere-se ao início das notas musicais do sinal, apresentando natureza similar aos propostos pela conversão de grafos. Por questões de simplificação, os descritores que compõe os atributos gerados por Grafos de Visibilidade Natural serão chamados de DGVN (Descritor de Grafos de Visibilidade Natural) e aqueles gerados por Grafos de Visibilidade Horizontal serão chamados de DGVH (Descritor de Grafos de Visibilidade Horizontal).

No segundo cenário, que também utilizou redes neurais, foram realizados experimentos utilizando cada um dos descritores do primeiro cenário separadamente, adicionando outros parâmetros de timbre utilizados em larga escala nos descritores da atualidade. Primeiro, foi configurado um vetor de atributos juntando os atributos de grafos com 13 MFCCs que apresentam as melhores taxas de importância, conforme os estudos atuais. Em seguida, foi realizado o aprendizado de máquina combinando os atributos de Onsets também aos 13 MFCCs. Por fim, os resultados foram comparados novamente.

Durante a etapa de treinamento e testes, o banco de dados foi dividido em subconjuntos de treinamento, validação e teste. O primeiro subconjunto é utilizado no aprendizado da rede neural, com o ajuste de pesos das unidades de processamento. Já o subconjunto de validação é utilizado na análise da performance da rede durante o treinamento, possibilitando um refinamento eficiente de hiperparâmetros. Por fim, o subconjunto de

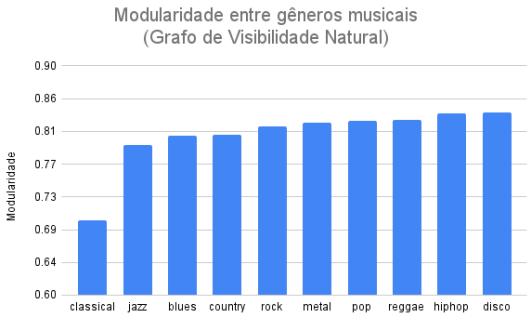


Figure 13. Grafos de Visibilidade Natural.

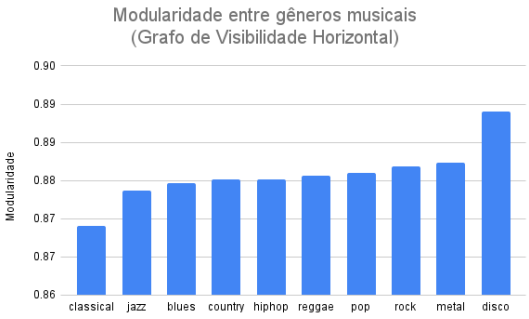


Figure 14. Grafos de Visibilidade Horizontal.

Figure 15. Grafos de Visibilidade Natural.

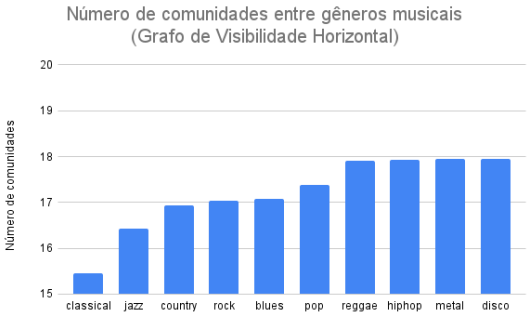


Figure 16. Grafos de Visibilidade Horizontal.

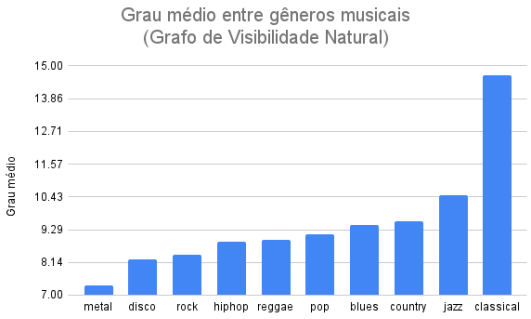


Figure 17. Grafos de Visibilidade Natural.

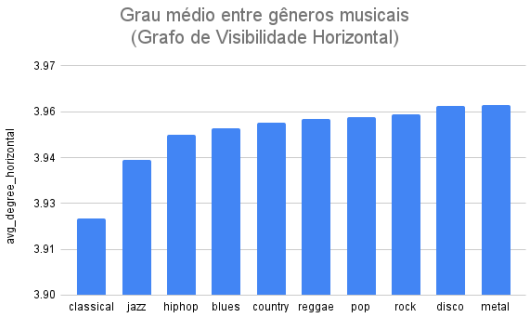


Figure 18. Grafos de Visibilidade Horizontal.

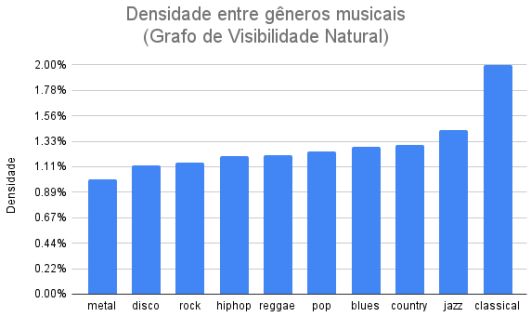


Figure 19. Grafos de Visibilidade Natural.

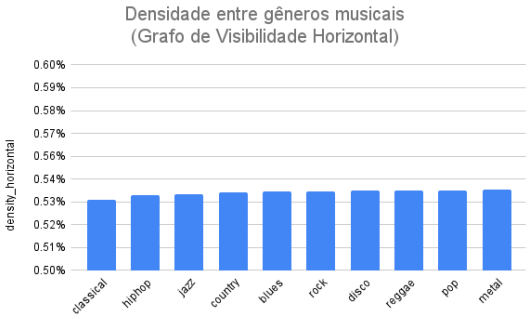


Figure 20. Grafos de Visibilidade Horizontal.

Figure 21. Média de Modularidade, Número de Comunidades, Grau Médio e Densidade entre gêneros musicais. Fonte: Autor.

teste possibilita avaliar o desempenho do classificador em dados que nunca foram vistos pela máquina. Dentre todos arquivos de áudio do banco de dados, 25% foram utilizados para testes, e dos 75% restantes, 80% formaram o conjunto de treinamento e 20% o conjunto de validação.

Os hiperparâmetros das redes considerados foram:

- Número de camadas redes;
- Número de unidades de processamento das células redes;
- Taxa de aprendizado;
- Dropout;
- Tamanho dos Batches;
- Número de épocas de treinamento;

Durante o período de treinamento, validação e teste, analisou-se o comportamento das redes neurais construídas, e dependendo dos resultados de acurácia e generalização, refinamentos e modificações foram feitos iterativamente. Após a exploração definiu-se utilizar uma Taxa de Aprendizado de 0.0001, Dropout de 30% nas camadas densas e Batches com tamanho de 32 amostras.

O modelo final consistiu da seguinte configuração:

- Camada de Input (número de atributos)
- 1a Camada Oculta Densa (512 neurônios, ativação="relu", Dropout 30%)
- 2a Camada Oculta Densa (256 neurônios, ativação="relu", Dropout 30%)
- 3a Camada Oculta Densa (54 neurônios, ativação="relu", Dropout 30%)
- Camada de Output (10 neurônios, ativação="softmax")

9. Resultados

Abaixo estão os resultados encontrados para cada combinação de atributos testada, demonstrando o erro e acurácia obtida nos testes.

Parâmetros	Épocas	Erro	Acurácia
DGVN	80	2.04	25.92%
DGVH	80	2.27	14.72%
Onset	80	2.25	34.52%
13 MFCCs	120	1.58	48.24%
DGVN + 13 MFCCs	100	1.72	55.30%
DGVH + 13 MFCCs	100	1.65	55.92%
Onset + 13 MFCCs	100	1.67	55.55%
DGVN + Onset + 13 MFCCs	100	2.30	55.92%
DGVH + Onset + 13 MFCCs	200	2.14	56.34%

Table 4. Comparação de resultados dos descritores. Fonte: Autor.

Observa-se que os descritores apenas de natureza rítmica não foram suficientes para ter uma boa classificação, obtendo maiores taxas de acerto com apenas o atributo de Onsets do sinal, quando comparado com o DGVN, além de demonstrar uma diferença ainda maior se comparado ao DGVH.

Também é evidenciado que o descritor com apenas atributos de timbre (13 MFCCs) apresentou ótimo resultados por conta própria, com uma taxa de 48.24% de acurácia

e erro de 1.58. Isso demonstra o quão importante é o papel desses atributos para descritores de classificação de gêneros musicais nas pesquisas atuais, conforme apresentado em trabalhos relacionados [Tzanetakis and Cook 2002].

Apesar disso, os descritores que utilizaram propriedades de timbre e ritmo somadas obtiveram uma taxa de acurácia maior, apesar de pouca divergência entre eles, com 55.30% de acerto para aqueles gerados por Grafos de Visibilidade Natural, 55.92% de acerto para atributos extraídos de Grafos de Visibilidade Horizontal e 55.55% de acerto para descritores de ritmo com frequência de Onsets.

Ao combinar descritores propostos por esse trabalho junto com os atributos de Onsets, não foi possível observar uma diferença considerável na taxa de acerto, além de apresentar uma taxa maior de erro. De qualquer forma, isso pode ser consequência da quantidade de dados utilizada, e a variação de amostras entre os gêneros no banco GTZAN podem não apresentar características suficientes para que os descritores validem os dados com maior precisão.

Além disso, foi gerado a matriz de confusão e feito a comparação para as 10000 amostras de áudio (1000 por gênero). Para o aprendizado utilizando o DGVN, na diagonal principal, onde observa-se os verdadeiros positivos, é evidenciado que gêneros de Disco (18%), Metal(15.5%) e Clássico(9.4%) tiveram a maior taxa de verdadeiros positivos. Isso indica que, mesmo com uma taxa de acerto baixa nesses casos, os padrões de percussão descritos pelos atributos de grafos foram melhor interpretados pela rede neural.

As menores taxas de acerto encontradas foram Blues(0.5%), Jazz(0.4%) e Pop(0.5%), o que indica que padrões de percussão descritos pelos atributos desses grafos não são típicos o suficiente para caracterizar fortemente seu agrupamento nesse sistema particular.

Para aprendizado utilizando DGVH, na diagonal principal observa-se a maior taxa de verdadeiros positivo para gêneros de Clássico(12%), Hip-Hop(7.9%), Rock(6.4%) e Disco(5.7%). Apesar disso, é evidenciado uma grande diferença na classificação de cada gênero quando comparado com o DGVN. A maioria dos gêneros nem foram considerados pelo classificador, como é o caso de Blues, Country, Jazz, Metal, Pop e Reggae. Isso indica que, por ser um método mais simplificado, os padrões de percussão descritos pelos atributos de Grafos de Visibilidade Horizontal tendem à decidir mais fortemente em um grupo menor de informação e focam fortemente nas naturezas rítmicas do sinal.

De todas as combinações de parâmetros analisadas, foi também gerado uma matriz de confusão para o grupo que obteve a melhor acurácia (56.34%), com atributos de DGVH, Onset e 13 MFCCs.

Na diagonal principal, onde observa-se os verdadeiros positivos, é evidenciado que gêneros de Clássico(19.9%), Blues(15.6%), Metal(14.8%) e Pop(13.9%) tiveram a maior taxa de verdadeiros positivos. É interessante observar que gêneros que obtiveram as piores taxa anteriormente, em descritores de grafos, obtiveram altas taxas de acurácia em combinação com atributos de natureza rítmica, como é o caso de gêneros como Blues, Jazz e Pop. Essa relação demonstra a extrema importância em combinar atributos de diferentes características e naturezas para os descritores de classificação de gêneros musicais.

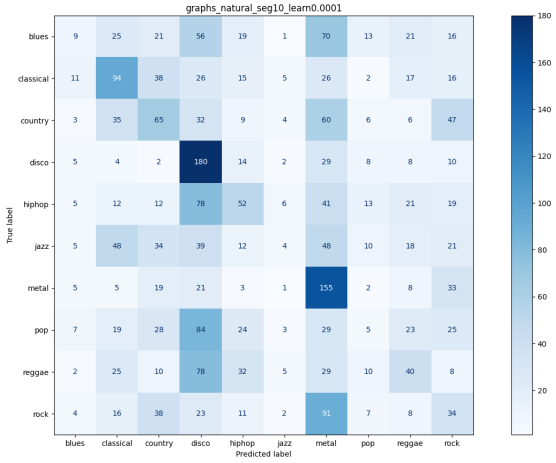


Figure 22. Parâmetros: DGVN.

Figure 23. Parâmetros: DGVH.

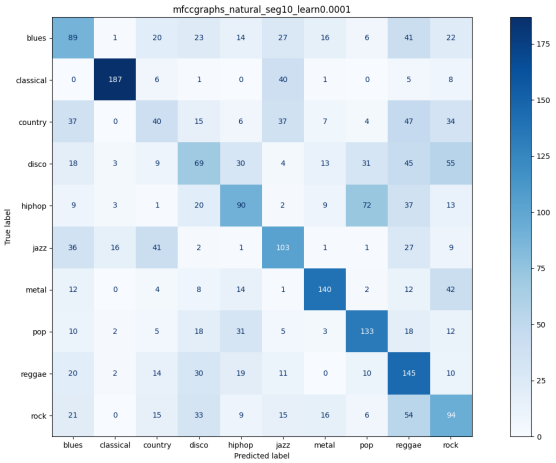


Figure 24. Parâmetros: DGVN + 13 MFCCs.

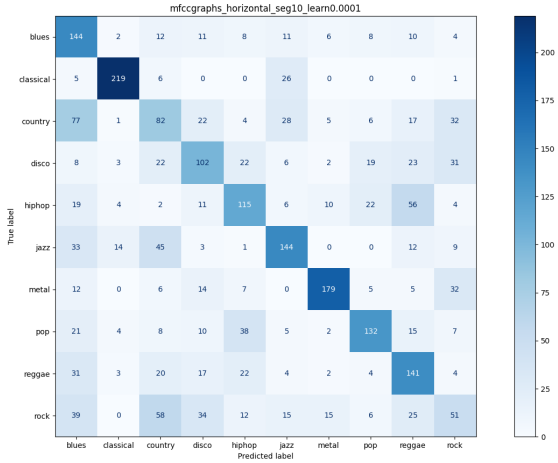


Figure 25. Parâmetros: DGVH + 13 MFCCs.

Figure 26. Matriz confusão de classificação dos gêneros musicais. Fonte: Autor

De todos os gêneros classificados, o grupo de Rock foi o que obteve a pior taxa de verdadeiros positivos, o que também é demonstrado na análise das médias de cada uma das propriedades topológicas de visibilidade horizontal nos gráficos da Figura 13, onde o gênero tomou as posições intermediárias entre todos os grupos. Essa característica refletiu na falsa classificação desses sinais principalmente no gênero de Country, o qual também assumiu posições próximas do gênero de Rock na análise das propriedades dos grafos gerados, além de ser comumente evidenciada na classificação feita por humanos.

Observa-se que a taxa de erro no conjunto de validação, nas primeiras iterações, é relativamente menor do que no conjunto de treinamento, até atingir um ponto onde as taxas se cruzam, onde deveria ser o limite de épocas do classificador. Esse comportamento provavelmente aconteceu porque o erro de treinamento é calculado continuamente ao longo de uma época inteira, enquanto que as métricas de validação são calculadas sobre o conjunto de validação apenas quando a época de treinamento atual é concluída. Isso implica que, em média, a taxa de erro de treinamento é medida meia época antes. O tamanho da batch e o otimizador escolhido influenciam diretamente nesse cálculo e, de

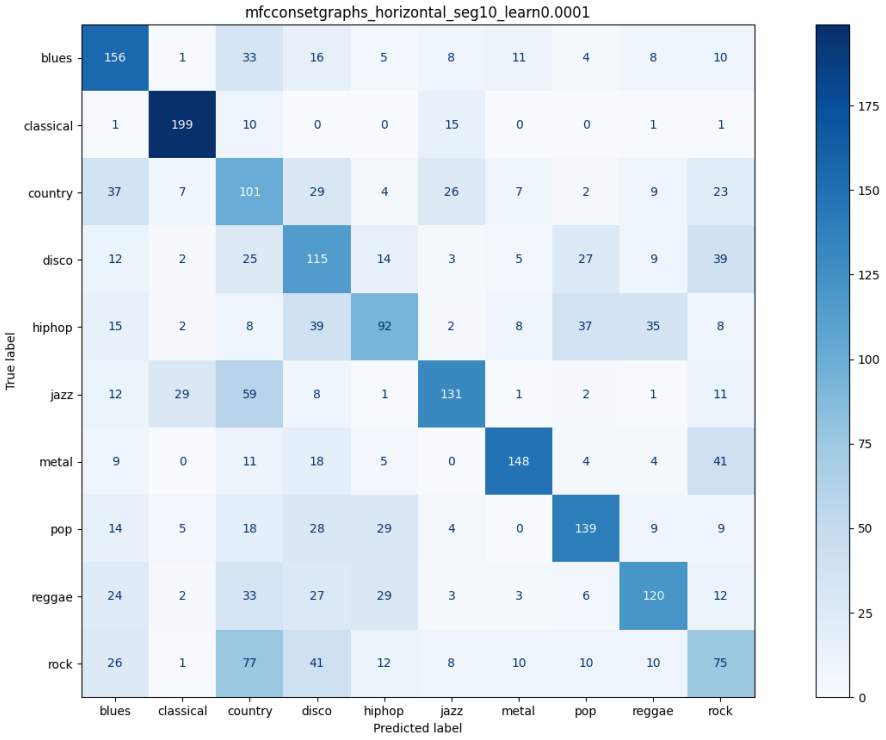


Figure 27. Matriz confusão de classificação dos gêneros musicais com DGVH + 13 MFCCs + Onset. Fonte: Autor.

modo geral, quanto maior a quantidade de dados processados nas iterações da máquina, menor vai ser o erro no fim de cada iteração.

10. Considerações Finais

A realização dos experimentos documentados permitiu extrair e analisar propriedades de grafos de visibilidade como uma nova forma de selecionar características em sinais de áudio. Além disso, a natureza e quantidade dos dados se fez um fator relevante na interpretação das saídas geradas pelas redes neurais desenvolvidas.

Em sistemas de classificação usando apenas os dados de grafos, obteve-se a melhor acurácia de 25.92% para o DGVN e acurácia de 14.72% para DGVH. Também foi realizado a classificação adicionando elementos de natureza de timbre, os MFCCs, onde obteve-se uma precisão semelhante à outro descritor com atributos rítmicos de intensidade de Onsets do sinal, com o melhor caso atingindo 56.34% de acurácia.

Com base nos resultados obtidos neste experimento e nos estudos mencionados, foi evidenciado que os Grafos de Visibilidade Natural e Grafos de Visibilidade Horizontal podem ser considerados como uma nova alternativa para a extração de características rítmicas para a recuperação de informações de música em sinais de áudio e podem ser usadas com sucesso em conjunto com os descritores baseado em transformadas de Fourier, em especial à aqueles de natureza de timbre.

A acurácia de classificação do conjunto de recursos combinados, em alguns casos, não é significativamente aumentada em comparação com as precisões de classificação do

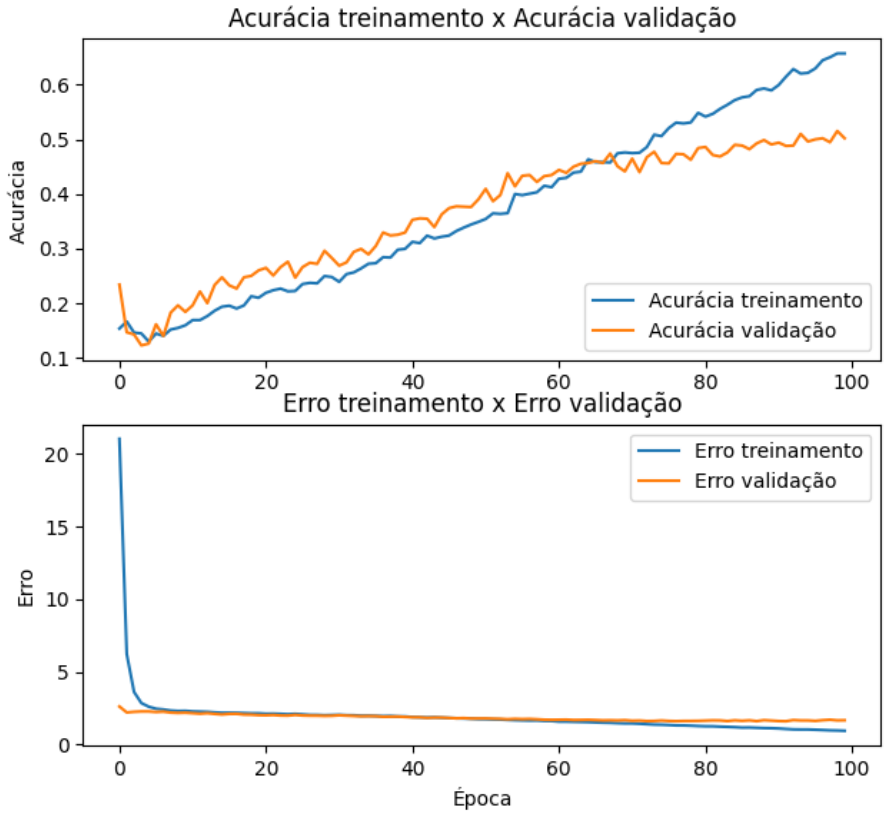


Figure 28. Loss treinamento vs Loss validação com DGVH + 13 MFCCs + Onset. Fonte: Autor.

conjunto de recursos individuais. Este fato não implica necessariamente que os recursos estejam correlacionados ou não contenham informações úteis, pois pode ser o caso de um determinado arquivo ser classificado corretamente por dois conjuntos de recursos diferentes que contêm informações de recursos diferentes e não correlacionados. Além disso, embora certas características individuais estejam correlacionadas, a adição de cada característica específica melhora, no geral, a precisão da classificação.

Para melhor entender o valor do resultado encontrado na classificação usando os descritores de grafos de visibilidade, vale ressaltar que os trabalhos citados contam com um número maior de atributos, chegando a 30 [Tzanetakis and Cook 2002] contra apenas 17 atributos usados neste experimento.

11. Trabalhos Futuros

A execução desse trabalho exigiu uma grande quantidade de dados e testes com diferentes configurações e, portanto, há diversos caminhos que podem ser explorados de forma a encontrar resultados relevantes para a classificação dos gêneros musicais.

Por exemplo, a partir da análise dos processos de treinamento, evidenciou-se que dependendo de como é dividido o sinal de entrada (apresentando séries de variância com variados números de pontos), as propriedades topológicas dos grafos gerados podem gerar comportamentos diferentes e, consequentemente, classificar os gêneros com outra perspectiva.

Além disso, o banco de dados GTZAN, apesar de ser utilizado em larga escala, ainda é considerado pequeno e com poucos dados, e certas propriedades apresentadas nesse trabalho podem ter resultados variados com diferentes níveis de amostras de sinais de áudio, além de permitir a exploração de diferentes gêneros e sub-gêneros encontrados na literatura.

A utilização de métodos mais eficientes para refinamento de redes neurais é de grande relevância, permitindo variações de modelos e cadeias de processamento com diferentes propósitos, dependendo dos atributos extraídos dos sinais. O número de atributos também pode influenciar diretamente na taxa de acurácia e erro dos sistemas, e um classificador de atributos pode ser considerado, de forma a eliminar a extração de dados desnecessários na fase de pré-processamento.

Por fim, vale ressaltar que a simplicidade da versão do algoritmo horizontal, proposto por este trabalho, permite que a extração de atributos seja realizada em resolução de um tempo computacional mais otimizado, principalmente em maiores proporções de dados. Porém, o resultado final pode perder informações relevantes, quando comparado aos classificadores gerados da mesma forma por Grafos de Visibilidade Natural, para a classificação musical.

References

Bergillos Varela, C. (2020). A study of visibility graphs for time series representations. B.S. thesis, Universitat Politècnica de Catalunya.

Borges, E., Simas Filho, E., Farias, C., Ribeiro, I., and Lopes, D. (2010). Classificação do gênero musical utilizando redes neurais artificiais. In *X Congresso Norte-Nordeste de Pesquisa e Inovação*, pages 1–8.

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.

Costa, Y. M. et al. (2013). Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores.

de Queiroz, R. A. B., Marar, J. F., and Okida, C. M. (2015). Investigação dos coeficientes cepstrais da frequência mel para extração de características de gêneros musicais.

Gouyon, F., Klapuri, A., Dixon, S., Alonso, M., Tzanetakis, G., Uhle, C., and Cano, P. (2006). An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1832–1844.

Gunawan, A. A., Suhartono, D., et al. (2019). Music recommender system based on genre using convolutional recurrent neural networks. *Procedia Computer Science*, 157:99–109.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Iacovacci, J. and Lacasa, L. (2019). Visibility graphs for image processing. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):974–987.

Jennings, H. D., Ivanov, P. C., Martins, A. d. M., da Silva, P., and Viswanathan, G. (2004). Variance fluctuations in nonstationary time series: a comparative study of music genres. *Physica A: Statistical Mechanics and its Applications*, 336(3-4):585–594.

Jurkiewicz, S. (2009). Grafos–uma introdução. *São Paulo: OBMEP*.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuno, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13):4972–4975.

Lacasa, L., Nunez, A., Roldán, É., Parrondo, J. M., and Luque, B. (2012). Time series irreversibility: a visibility graph approach. *The European Physical Journal B*, 85(6):1–11.

Liu, Y.-Y. and Barabási, A.-L. (2016). Control principles of complex systems. *Reviews of Modern Physics*, 88(3):035006.

Luque, B., Lacasa, L., Ballesteros, F., and Luque, J. (2009). Horizontal visibility graphs: Exact results for random time series. *Physical Review E*, 80(4):046103.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25.

Melo, D. d. F. P. (2019). Estudo de padrões em sinais musicais sob a perspectiva dos grafos de visibilidade.

Melo, D. d. F. P., Fadigas, I. d. S., and de Barros Pereira, H. B. (2017). Categorisation of polyphonic musical signals by using modularity community detection in audio-associated visibility network. *Applied network science*, 2(1):1–15.

Melo, D. d. F. P., Fadigas, I. d. S., and Pereira, H. B. d. B. (2020). Graph-based feature extraction: A new proposal to study the classification of music signals outside the time-frequency domain. *Plos one*, 15(11):e0240915.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.

Wallis, W. D. (2007). *A beginner's guide to graph theory*. Springer Science & Business Media.

Yela, D. F., Thalmann, F., Nicosia, V., Stowell, D., and Sandler, M. (2020). Online visibility graphs: Encoding visibility in a binary search tree. *Physical Review Research*, 2(2):023069.

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., and Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. *Physics Reports*, 787:1–97.

APÊNDICE B – Código-fonte

B.1 Transformação dás séries temporais em grafos

```
import os
import analysis_signal
from ts2vg import NaturalVG
from ts2vg import HorizontalVG
import networkx as nx

# General Setup
DATASET_PATH = "Data/genres_original"
GRAPHS_PATH = "Data/graphs_batch110/"
SAMPLE_RATE = 11000
NUM_FILES_PER_GENRE = 100

# Graph Setup
CHUNK_SIZE = 440

for i, (dirpath, dirnames, filenames) in enumerate(os.walk(DATASET_PATH)):
  if dirpath is not DATASET_PATH:
    dirpath_components = dirpath.split("\\")
    label = dirpath_components[-1]

    filenames = filenames[0:NUM_FILES_PER_GENRE]
    for f in filenames:

      file_path = os.path.join(dirpath, f)
      signal, sr = analysis_signal.load_file(file=file_path, sample_rate=
    SAMPLE_RATE)

      var_series = analysis_signal.variance_series(signal=signal,
    chunk_size=CHUNK_SIZE)
      natural_graph = analysis_signal.visibility_graph(var_series,
    NaturalVG)
      horizontal_graph = analysis_signal.visibility_graph(var_series,
    HorizontalVG)

      nx.write_pajek(natural_graph, GRAPHS_PATH + "natural/" + label + "/"
    + os.path.basename(os.path.splitext(file_path)[0]) + ".net")
      nx.write_pajek(horizontal_graph, GRAPHS_PATH + "horizontal/" + label
    + "/" + os.path.basename(os.path.splitext(file_path)[0]) + ".net")
```

B.2 Extração de atributos dos grafos

```
import os
import json
import analysis_signal
from ts2vg import NaturalVG
from ts2vg import HorizontalVG
import networkx.algorithms.community as nx_comm

# Split signal in segments
NUM_SEGMENTS = 10

# General Setup
DATASET_PATH = "Data/genres_original"
GRAPH_TYPE = "horizontal"
OUTPUT_PATH = f'Data/data_graphs_{GRAPH_TYPE}_seg{NUM_SEGMENTS}.json'
SAMPLE_RATE = 11000
DURATION = 30
NUM_FILES_PER_GENRE = 100

CHUNK_SIZE = 440

data = {
  "genre": [
    "blues",
    "classical",
    "country",
    "disco",
    "hiphop",
    "jazz",
    "metal",
    "pop",
    "reggae",
    "rock",
  ],
  f'graphs_{GRAPH_TYPE}': [],
  "labels": []
}

SAMPLES_PER_TRACK = SAMPLE_RATE * DURATION
SAMPLES_PER_SEGMENT = int(SAMPLES_PER_TRACK / NUM_SEGMENTS)

for i, (dirpath, dirnames, filenames) in enumerate(os.walk(DATASET_PATH)):
  if dirpath is not DATASET_PATH:
    dirpath_components = dirpath.split("\\")
    label = dirpath_components[-1]

```

```
    filenames = filenames[0:NUM_FILES_PER_GENRE]
    for f in filenames:

      file_path = os.path.join(dirpath, f)
      signal, sr = analysis_signal.load_file(file=file_path, sample_rate=SAMPLE_RATE)

      for s in range(NUM_SEGMENTS):
        start_sample = SAMPLES_PER_SEGMENT * s
        finish_sample = start_sample + SAMPLES_PER_SEGMENT

        var_series = analysis_signal.variance_series(signal=signal[start_sample:finish_sample], chunk_size=CHUNK_SIZE)
        if (GRAPH_TYPE == "natural"):
          graph = analysis_signal.visibility_graph(var_series, NaturalVG)
        elif (GRAPH_TYPE == "horizontal"):
          graph = analysis_signal.visibility_graph(var_series, HorizontalVG)

        num_edges = len(graph.edges)
        num_vertices = len(graph.nodes)

        communities = nx_comm.greedy_modularity_communities(graph)
        modularity = nx_comm.modularity(graph, communities)

        density = (2 * num_edges) / (num_vertices * (num_vertices - 1))
        avg_degree = sum((d for n, d in graph.degree)) / num_vertices

        data[f'graphs_{GRAPH_TYPE}'].append([
          modularity,
          len(communities),
          avg_degree,
          density,
        ])
        data["labels"].append(i - 1)
        print("{}, segment:{}".format(file_path, s+1))

with open(OUTPUT_PATH, "w") as fp:
  json.dump(data, fp, indent=4)
```

B.3 Modelo no Keras

```
import numpy as np
from sklearn.model_selection import train_test_split
from tensorflow import keras

```

```
# Training Setup
DATASET_TEST_SIZE = 0.25
DATASET_VALIDATION_SIZE = 0.2

def split_datasets(inputs, targets):
  # load data
  X = inputs
  y = targets

  # create train/test split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=
    DATASET_TEST_SIZE)

  # create train/validation split
  X_train, X_validation, y_train, y_validation = train_test_split(X_train,
    y_train, test_size=DATASET_VALIDATION_SIZE)

  return X_train, X_validation, X_test, y_train, y_validation, y_test

def build_model(input_groups):
  inputs = []

  # create inputs and concatenate
  for i, input_set in enumerate(input_groups):
    input = keras.Input(shape=(input_set.shape[1:]))
    inputs.append(input)

    if (input.shape.ndims > 2):
      model = keras.layers.Flatten()(input)
    else:
      if i == 0:
        model = input
        continue
      model = keras.layers.concatenate([model, input])

  shape = model.shape

  # 1st hidden layer
  model = keras.layers.Dense(512, activation="relu")(model)
  model = keras.layers.Dropout(0.3)(model)

  # 2nd hidden layer
  model = keras.layers.Dense(256, activation="relu")(model)
  model = keras.layers.Dropout(0.3)(model)

  # 3rd hidden layer
  model = keras.layers.Dense(64, activation="relu")(model)
```

```
  model = keras.layers.Dropout(0.3)(model)

  # output layer
  outputs = keras.layers.Dense(10, activation="softmax")(model)

  model = keras.Model(inputs=inputs, outputs=outputs)

  model.build(shape)

  return model
```

B.4 Classificação dos gêneros musicais

```
import json
import numpy as np
from tensorflow import keras
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
import matplotlib.pyplot as plot
import datetime

from model import split_datasets, build_model

# Training Setup
TRAINING_BATCH_SIZE = 32
EPOCHS = 100
LEARNING_RATE = 0.0001

NUM_DATA_SEGMENTS = 10
DATA_INPUTS = ["mfcc", "onset", "graphs_horizontal"]

data = {}
for type in DATA_INPUTS:
  with open(f'Data/data_{type}_seg{NUM_DATA_SEGMENTS}.json', "r") as f:
    data[type] = json.load(f)

genres = np.array(data[DATA_INPUTS[0]]["genre"])
inputs = {}
for type in data:
  inputs[type] = np.array(data[type][type])

targets = np.array(data[DATA_INPUTS[0]]["labels"])

def plot_history(history):
  fig, axis = plot.subplots(2)

  # create accuracy subplot
```

```
34    axis[0].plot(history.history["accuracy"], label="Acur cia treinamento")
35    axis[0].plot(history.history["val_accuracy"], label="Acur cia
        valida  o")
36    axis[0].set_ylabel("Acur cia")
37    axis[0].legend(loc="lower right")
38    axis[0].set_title("Acur cia treinamento x Acur cia valida  o")
39
40    axis[1].plot(history.history["loss"], label="Erro treinamento")
41    axis[1].plot(history.history["val_loss"], label="Erro valida  o")
42    axis[1].set_xlabel(" poca ")
43    axis[1].set_ylabel("Erro")
44    axis[1].legend(loc="upper right")
45    axis[1].set_title("Erro treinamento x Erro valida  o")
46
47    plot.show()
48
49  if __name__ == "__main__":
50    # create train, validation and test sets
51    X_train = []
52    X_validation = []
53    X_test= []
54    y_train = []
55    y_validation = []
56    y_test = []
57
58    for input in inputs:
59      xtrain, xval, xtest, ytrain, yval, ytest = split_datasets(inputs[input
        ], targets)
60      X_train.append(xtrain)
61      X_validation.append(xval)
62      X_test.append(xtest)
63      y_train.append(ytrain)
64      y_validation.append(yval)
65      y_test.append(ytest)
66
67    # build model
68    model = build_model(X_train)
69
70    # compile network
71    optimizer = keras.optimizers.Adam(learning_rate=LEARNING_RATE)
72    model.compile(optimizer=optimizer, loss="sparse_categorical_crossentropy"
        , metrics=["accuracy"])
73    model.summary()
74
75    # tensor board
76    inputs_text = ""
77    for input in DATA_INPUTS:
```

```
78      inputs_text += input
79
80  training_id = inputs_text + "_seg" + str(NUM_DATA_SEGMENTS) + "_learn" +
      str(LEARNING_RATE)
81  log_dir = "logs/fit/" + training_id
82  tensorboard_callback = keras.callbacks.TensorBoard(log_dir=log_dir,
      histogram_freq=1)
83
84  # train network
85  history = model.fit(x=X_train, y=y_train[0], validation_data=(
      X_validation, y_validation[0]), epochs=EPOCHS, batch_size=
      TRAINING_BATCH_SIZE, callbacks=[tensorboard_callback])
86
87  # evaluate the network on the test set
88  test_error, test_accuracy = model.evaluate(X_test, y_test[0], verbose=1)
89  print("Accuracy is {}. Error is {}".format(test_accuracy, test_error),)
90
91  # confusion matrix
92  prediction = model.predict(X_test)
93  cm = confusion_matrix(y_test[0], np.argmax(prediction, axis=1))
94  disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=genres)
95  disp.plot(cmap=plot.cm.Blues)
96  disp.ax_.set_title(training_id)
97  plot.show()
98
99  # plot accuracy and error over the epochs
100 plot_history(history)
```