UFBA IFBA UNEB UEFS LNCC SENAI-CIMATEC FACED
PROGRAMA DO DOUTORADO MULTIINSTITUCIONAL E
MULTIDISCIPLINAR EM DIFUSÃO DO CONHECIMENTO

DIRCEU DE FREITAS PIEDADE MELO

ESTUDO DE PADRÕES EM SINAIS MUSICAIS SOB A PERSPECTIVA DOS GRAFOS DE VISIBILIDADE

TESE

SALVADOR

2017

DIRCEU DE FREITAS PIEDADE MELO

ESTUDO DE PADRÕES EM SINAIS MUSICAIS SOB A PERSPECTIVA DOS GRAFOS DE VISIBILIDADE

Tese apresentada ao Programa do Doutorado Multi-institucional e Multidisciplinar em Difusão do Conhecimento da UFBA IFBA UNEB UEFS LNCC SENAI-CIMATEC FACED como requisito parcial para obtenção do grau de "Doutor em Difusão do Conhecimento" – Área de Concentração: Modelagem da Geracão e Difusão do Conhecimento..

Orientador: Hernane Borges de Barros Pereira

Co-orientador: Inacio de Sousa Fadigas

SALVADOR

2017

Melo, Dirceu de Freitas Piedade
Estudo de padrões em sinais musicais sob a perspectiva dos grafos de visibilidade / Dirceu de Freitas Piedade Melo. -- Salvador, 2017.
87f.

Orientador: Hernane Borges de Barros Pereira.
Coorientador: Inacio de Sousa Fadigas.
Tese (Doutorado - Doutorado multidisciplinar e multiinstitucional em difusão do conhecimento) -- Universidade Federal da Bahia, IFBA, SENAI-CIMATEC, UNEB, UEFS,LNCC,FACED, 2017.

1. Redes Complexas. 2. Grafos de Visibilidade. 3. Extração de Atributos. 4. Recuperação de Informações Musicais. 5. Classificação de Gêneros Musicais. I. Pereira, Hernane Borges de Barros. II. Fadigas, Inacio de Sousa . III. Título.

ATA DE DEFESA DE TESE DO DOUTORANDO **DIRCEU DE FREITAS PIEDADE MELO** NO DOUTORADO MULTI-INSTITUCIONAL E MULTIDISCIPLINAR EM DIFUSÃO DO CONHECIMENTO

Ao vigésimo terceiro dia do mês de novembro de dois mil e dezessete, às 08:30h, reuniu-se no SENAI - CIMATEC, a Comissão Examinadora composta pelos professores doutores: Hernane Borges de Barros Pereira (Orientador), Roberto Luiz Souza Monteiro, Elias Ramos de Souza, Marcelo Albano Moret Simões Gonçalves e Eduardo Furtado de Simas Filho para julgar o trabalho intitulado **"ESTUDOS DE PADRÕES EM SINAIS MUSICAIS SOB A PERSPECTIVA DOS GRAFOS DE VISIBILIDADE",** de autoria de **Dirceu de Freitas Piedade Melo.** Após a argüição e discussão, a Banca examinou, analisou e avaliou o referido trabalho, chegando à conclusão que este foi APROVADO. Nada mais havendo a ser tratado, esta Comissão Examinadora encerrou a reunião da qual eu lavrei a presente ATA, que após lida e achada conforme, vai assinada pelos presentes e encerrada por mim, Hernane Borges de Barros Pereira.

Salvador, 23 de novembro de 2017.

Comissão Examinadora:

Prof. Dr. Hernane Borges de Barros Pereira(Orientador).....

Prof. Dr. Roberto Luiz Souza Monteiro .....

Prof. Dr. Elias Ramos de Souza.....

Prof. Dr. Marcelo Albano Moret Simões Gonçalves.....

Prof. Dr. Eduardo Furtado de Simas Filho.....

Dedico este trabalho a duas pessoas espetaculares :
meu pai Gino Frey (in memoriam) por ter me ensinado o amor pela música e a busca incessante pelo conhecimento;
minha mãe Esmeralda Melo, que tem sido para mim e outras tantas pessoas, um referencial de virtude, sabedoria e humanidade.

AGRADECIMENTOS

A Deus, pela graça de ainda poder sonhar e de ter esperança. A minha família (Esmeralda Melo, Bucka, Doka, e Pipa), pelo amor e acolhimento incondicionais. A minha esposa e filha (Rose e Ana Morena), com quem compartilho, diariamente, todas as matizes do afeto.

Ao meu orientador Prof. Hernane Borges de Barros Pereira, por acreditar em meu trabalho, pelas importantes contribuições científicas, pela amizade, e palavras de apoio. A Inácio Fadigas pela co-orientação e parceria nas publicações.

Ao Departamento de Matemática do IFBA, pelo apoio durante todas as etapas do doutorado. Ao Grupo de pesquisa em redes Fuxicos e Boatos por me ensinar que parceria científica, produtividade intelectual, e boas risadas podem conviver de forma harmoniosa.

Aos professores da *Universidad Pompeu Fabra* que me deram um extraordinário suporte acadêmico, científico e pessoal durante o meu estágio doutoral em Barcelona: Emilia Gomez Gutierrez (Chefe do Music information Research Lab), Rafael Ramirez Melendez (Chefe do Music and Machine Learning Lab), Xavier Serra (Diretor do Music Technology Group e chefe do Audio Signal Processing Lab), e Dmitry Bogdanov (Audio Signal Processing Lab).

Aos amigos-irmãos Robson e Aloísio pelos conselhos, orientações e suporte, antes e durante minha viagem a Barcelona. Aos colegas mais próximos do programa de Doutorado Multidisciplinar e Multiinstitucional em Difusão do Conhecimento, com os quais tive oportunidade compartilhar momentos especiais de aprendizagem intelectual e afetiva.

”Se o Senhor não edificar a casa,
em vão trabalham os que a edificam ...
É inútil levantar de madrugada, dormir tarde,
comer o pão conquistado com muita luta;
aos seus amados Ele o dá enquanto dormem.”
Salmo 127:1-2

RESUMO

MELO, Dirceu de Freitas Piedade. ESTUDO DE PADRÕES EM SINAIS MUSICAIS SOB A PERSPECTIVA DOS GRAFOS DE VISIBILIDADE. 88 f. Tese – Programa do Doutorado Multiinstitucional e Multidisciplinar em Difusão do Conhecimento, UFBA IFBA UNEB UEFS LNCC SENAI-CIMATEC FACED. Salvador, 2017.

O advento da tecnologia digital favoreceu um extraordinário aumento da capacidade de armazenamento e compartilhamento de arquivos de conteúdo musical, o que motivou algumas corporações a incluírem em suas plataformas, algoritmos computacionais para o gerenciamento automático de grandes bibliotecas de música digital. A classificação de gêneros musicais tem chamado a atenção como uma das formas de organização deste tipo de biblioteca, e nas últimas décadas, tem se tornado objeto de estudo de pesquisadores de um campo multidisciplinar emergente conhecido como Recuperação de Informações Musicais (MIR). A maioria dos trabalhos desse campo de pesquisa adota a estratégia de categorização de gêneros musicais usando a extração de atributos (ritmo, melodia e timbre) como uma de suas etapas essenciais. 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 de características de regularidade de seus transientes. A auto-similaridade dos sinais pode dar informações relevantes sobre essa regularidade, e desta forma, contribuir para o estudo da complexidade rítmica de uma música. A maioria dos trabalhos do campo de processamento de sinais têm estudado a auto-similaridade em música digital utilizando o histograma de batidas. Existe uma carência na diversidade de descritores rítmicos para sinais de áudio, e o campo de processamento de sinais está restrito à técnicas baseadas em representações tempo-frequência. Novos tipos de descritores poderiam colaborar com os algoritmos tradicionais, para a melhorar a extração de características rítmicas, oferecendo outro ponto de vista para essa tarefa. Esta tese propõe uma metodologia para identificar padrões de auto-similaridade em sinais de áudio, usando propriedades topológicas de redes, denominado de Descritor de Visibilidade em Flutuações de Variância (DVFV). Este descritor é constituido de: Modularidade - Q, Número de Comunidades - Nc, Grau Médio - $< k >$ e Densidade (Δ). Os resultados experimentais obtidos com o cálculo do DVFV em 1.000 grafos de visibilidade, correspondentes a 1.000 sinais, categorizados em 10 gêneros musicais, mostraram que o DVFV é capaz de detectar gráfica e numericamente, padrões de auto-similaridade em sinais classificados em gêneros musicais, de estabelecer uma relação hierárquica de categorias usando propriedades de redes, e de contribuir para que um sistema de classificação alcance precisão comparável ou superior a trabalhos correlatos.

Palavras-chave: redes complexas, grafos de visibilidade, extração de atributos, recuperação de informações musicais, classificação.

ABSTRACT

MELO, Dirceu de Freitas Piedade. STUDY OF PATTERNS IN MUSICAL SIGNALS FROM THE PERSPECTIVE OF VISIBILITY GRAPHS. 88 f. Tese – Programa do Doutorado Multiinstitucional e Multidisciplinar em Difusão do Conhecimento, UFBA IFBA UNEB UEFS LNCC SENAI-CIMATEC FACED. Salvador, 2017.

The advent of digital technology favored an extraordinary increase in the storage capacity and sharing of music content files, which motivated some corporations to include in their platforms computational algorithms for the automatic management of large digital music libraries. The classification of musical genres has attracted attention as one of the forms of organization of this type of library, and in recent decades, has become the object of study of researchers of an emerging multidisciplinary field known as Music Information Retrieval (MIR). Most of the works in this field of research adopt the strategy of categorization of musical genres using the extraction of attributes (rhythm, melody and timbre) as one of its essential stages. Among these attributes, rhythm plays a very important role in the definition of musical style. The study of rhythmic in audio signals includes the investigation of regularity characteristics of their transients. The self-similarity of the signals can give relevant information about this regularity, and thus contribute to the study of the rhythmic complexity of a song. Most of the works of the signal processing field have studied self-similarity in digital music using the beat histogram. There is a lack in the diversity of rhythm descriptors for audio signals, and the signal processing field is restricted to techniques based on time-frequency representations. New types of descriptors could collaborate with traditional algorithms to improve the extraction of rhythmic features, providing another point of view for this task. This thesis proposes a methodology to identify self-similarity patterns in audio signals, using topological properties of networks, called Variance Fluctuation Visibility Descriptor (DVFV). This descriptor consists of: Modularity - Q, Number of Communities - Nc, Average Degree - $< k >$ and Density (Δ). The experimental results obtained with the calculation of DVFV in 1.000 graphs of visibility, corresponding to 1.000 signs, categorized in 10 musical genres, showed that the DVFV is able to detect graphically and numerically, self-similarity patterns in signals classified in musical genres, establish a hierarchical relationship of categories using properties of networks, and contribute for a classification system to reach comparable or superior precision to related works.

Keywords: complex networks, visibility graphs, feature extraction, music information retrieval, classification.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

BPM	Beats Por Minuto
CD	Compact Disc
Compldyn	Complexidade da dinâmica
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DFA	Detrended Flutuation Analysis
DVFV	Descritor de Visibilidade em Flutuações de Variância
GTZAN	George Tzanetakis
IBK	Algoritmo Classificador baseado em k-NN
k-NN	k-Nearest Neighbours algorithm
MFCC	Mel Frequency Cepstral Coefficients
MIR	Music Information Retrieval
MP3	MPEG-1/2 Audio Layer 3
RMS	Root Mean Square
ROC	Receiver Operating Characteristics
STFT	Short Time Fourier transform
TxOnset	Taxa de onset
TxPassZero	Taxa de passagem pelo zero
WAV	Wave Form Audio Format
WEKA	Waikato Environment for Knowledge Analysis

LISTA DE SÍMBOLOS

Δ	Densidade de uma rede
$\overline{\Delta}$	Média de Δ
σ_{Δ}	Desvio-padrão de $\overline{\Delta}$
<k>	Grau Médio de um rede
$\overline{< k >}$	Média de $< k >$
$\sigma_{<k>}$	Desvio-padrão de $\overline{< k >}$
Nc	Número de Comunidades produzidas em **Q**
$\overline{Nc}$	Média de Nc
σ_{Nc}	Desvio-padrão de $\overline{Nc}$
Q	Modularidade de uma rede
$\overline{Q}$	Média de Q
σ_{Q}	Desvio-padrão de $\overline{Q}$

SUMÁRIO

1 INTRODUÇÃO

Segundo o dicionário Cambridge online [1] o termo música é definido como "padrão de sons feitos por instrumentos musicais, vozes, ou computadores, ou a pela combinação destes". Para aqueles que nasceram na era da informação essa definição pode soar bastante natural, pois para essa geração, diferentemente das anteriores, o uso da tecnologia digital associada à música já faz parte de sua realidade cotidiana. O surgimento da tecnologia digital possibilitou que as informações musicais pudessem ser manipuladas facilmente, além de favorecer seu registro em mídias de formato compacto. Isso trouxe para o universo da música uma gama de possibilidades nunca antes exploradas, resultando em mudanças significativas na forma de compor, produzir e compartilhar uma obra musical.

O estudo da classificação de sinais de áudio tem ganhado muita importância devido à necessidade de organização de grandes bibliotecas digitais para facilitar a busca, recuperação e recomendação de arquivos musicais na *internet* [Sturm 2013,Pampalk et al. 2002]. Por esse motivo, muitos modelos de classificação automática de gêneros musicais têm sido propostos [Ezzaidi e Rouat 2006, Jr. et al. 2006, Goulart 2012]. Em todos esses modelos pode-se observar a utilização de descritores, que são algoritmos herdados das técnicas de reconhecimento de voz, e que têm como objetivo extrair parâmetros numéricos para construir um conjunto de atributos que permitam distinguir ou associar um sinal a um agrupamento dentro do processo de classificação. Dentre os algoritmos descritores mais usados na extração de atributos, estão MFCC - *Mel Frequency Cepstral Coeficients*, Fluxo Espectral, Batimentos por Minuto (BPM), Taxa de Passagem pelo Zero. Esses algoritmos realizam suas operações a partir de representações no domínio tempo-frequência, buscando três tipos básicos de características do sinal musical: textura de timbre (qualidade e colorido musical), conteúdo rítmico (tempo, ritmo, pulsação), conteúdo tonal (notas, tons e acordes) [Tzanetakis e Cook 2002, Ahrendt e Hansen 2006, Seyerlehner et al. 2010, Berois 2008]. Particularmente, as características de natureza rítmica têm desempenhado um papel importante na definição da identidade musical [Dixon et al. 2004], podendo ser um critério decisivo para determinação do gênero musical [Jr et al. 2005].

[1] Disponível em http://dictionary.cambridge.org/dictionary/english/music.

Características como a periodicidade e estruturas repetitivas inerentes de arranjos musicais podem ser estudados através da auto-similaridade do sinal. Tzanetakis e Cook 2002 encontraram diferenças marcantes entre estilos musicais através da auto-similaridade revelada por um histograma de batidas. Jennings et al. 2004, Das e Das 2005, e Goulart 2012, estudaram diferenças entre gêneros musicais a partir da caracterização da auto-similaridade dos seus sinais usando o DFA (*Detrended Fluctuation Analysis*). Cooper e Foote 2002, Foote e Cooper 2001, Müller et al. 2011 usaram matrizes de auto-similaridade para investigar estruturas repetitivas, incluindo os aspectos ritmicos da música. Vários trabalhos constatam que o estudo de aspectos de natureza rítmica são feitos preferencialmente usando a representação tempo-frequência [Lerch 2012, Dannenberg et al. 2001, Aucouturier e Pachet 2003, Schedl et al. 2014]. Existem poucas publicações fora do campo de processamento de sinais tratando da auto-similaridade para recuperação de informações musicais. Existe uma carência na diversidade de descritores rítmicos, e o campo de processamento de sinais está restrito a técnicas baseadas em representações tempo-frequência, a exemplo dos algoritmos com base em transformadas de Fourier. Novos tipos de descritores poderiam colaborar com os algoritmos tradicionais para a melhorar a extração de características rítmicas, oferecendo outro ponto de vista para essa tarefa.

O campo de redes complexas tem mostrado um crescente desenvolvimento na representação, análise, e compreensão de vários tipos de sistemas, inclusive envolvendo informações musicais [Buldú et al. 2007, Tse et al. 2008, Jacobson et al. 2008]. Contudo, esse campo não tem apresentado trabalhos dedicados ao estudo da auto-similaridade de sinais de áudio. Dentre as publicações que adotam a modelagem de redes complexas no contexto de Recuperação de Informações Musicais, com exceção de Melo et al. 2016, e Melo et al. 2017, não foram encontrados trabalhos investigando a auto-similaridade em sinais musicais usando propriedades de redes complexas associadas à sinais de áudio, com o fim de extração de atributos para classificação. Esperando suprir essa carência, esta tese propõe um método para estimar a auto-similaridade de sinais musicais através da estrutura topológica de seus grafos de visibilidade, medida por quatro propriedades de redes. Surge então o seguinte questionamento: É possível obter extrair atributos relevantes para a classificação de sinais musicais, através da auto-similaridade estimada por propriedades topológicas de redes complexas, aplicadas a grafos de visibilidade associados a esses sinais? A fim de responder esse questionamento nós propomos nesta tese o descritor de visibilidade em flutuações de variância para sinais de áudio musicais. Este descritor estima a auto-similaridade do sinal musical através da modularidade (Q), do número de comunidades (Nc), da densidade (Δ), e do grau médio ($< k >$) do grafo mapeado a partir da visibilidade de suas flutuações de variância.

Os grafos de visibilidade (*visibility graphs*) [Lacasa et al. 2008] são 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. 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. Ao final do mapeamento, a rede terá herdado as características de auto-similaridade da série através da qualidade e quantidade de comunidades, e pelo grau de conexões geradas na rede. Depois de calculadas as quatro propriedades do descritor de visibilidade para 1.000 grafos gerados a partir de arquivos do banco GTZAN *Genre Collection* [2], serão avaliadas as suas contribuições dentro de um sistema de classificação supervisionada com dez gêneros musicais, usando primeiramente apenas o descritor de visibilidade, e depois utilizando descritores no domínio tempo-frequência (Fluxo Espectral, Loudness, BPM, *Onset Rate*, Taxa de Passagem pelo Zero, e MFCCs), um descritor usado para o estudo de correlações de longo alcance em séries temporais (Expoente DFA), e um descritor de dinâmica musical (*Dynamics Complexity*). Através da acurácia obtida na classificação, serão feitas inferências sobre escolhas musicais associadas à assinatura topológica de cada sinal musical dentro do contexto de seu respectivo gênero musical. Também serão feitas comparações entre os resultados obtidos nesta tese e trabalhos do estado da arte da recuperação de informações musicais.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

Elaborar um descritor de visibilidade em flutuações de variância de sinais musicais polofônicos para a estimativa da sua auto-similaridade usando propriedades topológicas de grafos de visibilidade associados a esses sinais, com aplicação na categorização de gêneros musicais.

1.1.2 OBJETIVOS ESPECÍFICOS

- Calcular uma representação do *loudness* do sinal , através da série de flutuações de variância;
- Transformar as séries de flutuações de variância em grafos de visibilidade, a fim de estudar suas características de auto-similaridade do ponto de vista topológico;
- Calcular a modularidade (Q), número de comunidades (Nc), grau médio ($< k >$), e

[2] Disponível em https://marsyasweb.appspot.com/download/data_sets/

densidade (Δ) desses grafos e constituir o Descritor de Visibilidade em Flutuações de Variância (DVFV);

- Modelar graficamente os grafos de visibilidade usando detecção de comunidades, a fim de explicitar características de auto-similaridade dos sinais que são transferidas para a rede durante o mapeamento;
- Estabelecer hierarquias de auto-similaridade a partir dos quatro parâmetros do DVFV;
- Classificar, em um sistema supervisionado, 1.000 sinais musicais em 10 classes de gêneros musicais usando o DVFV como atributo de complexidade rítmica;
- Comparar os resultados da classificação com experimentos correlatos, para contextualizar os resultados do DVFV dentro do estado da arte da recuperação de informações musicais (MIR);
- Estudar o ganho de informação de um vetor de atributos formado pelo DVFV, com a finalidade de avaliar o nível do DVFV em relação a descritores tempo-frequência;

1.2 ORGANIZAÇÃO DA TESE

Capítulo 2 - Revisão da Literatura

- **Seção 2.1** - Trata da classificação de gêneros musicais segundo o ponto de vista da Recuperação de Informações Musicais (MIR), campo que integra o processamento de sinais, a recuperação de informações, e a inteligência artificial, à investigação da organização automática de bibliotecas de música digital. Também serão apresentados nesta seção, os fundamentos de um sistema de classificação para música digital (pré-processamento, extração de atributos, aprendizagem de máquina, e classificação), cujos estágios serão abordados nas seções seguintes.
- **Seção 2.2** - Introduz os descritores musicais de natureza rítmica, timbrística, e tonal, segundo o modelo proposto por Tzanetakis e Cook 2002. Aqui serão enfatizados os processos de extração de atributos mais usados em pesquisa MIR, oriundos do campo de processamento de sinais de áudio na investigação de padrões de voz.
- **Seção 2.3** - Apresenta a definição de sinal polifônico para o contexto dessa tese, e diferencia o sentido do termo "polifonia" usado no texto deste trabalho, com o termo usado em teoria musical. No final da seção, será demonstrada a construção de sinais polifõnicos dentro de um sistema de gravação multi-pista.

- **Seção 2.4** - Discorre sobre os fundamentos da teoria dos grafos, começando por uma breve descrição do contexto histórico, passando pela definição matemática, e concluindo com uma discussão sobre a modelagem com redes. Na Subseção 2.4.1 serão apresentadas as redes aleatórias de forma sucinta, a fim de embasar a compreensão da definição da modularidade de Newman e Girvan 2004, que será discutida na Seção 3.5. Na Subseção 2.4.2 será feita uma abordagem geral dos métodos para transformação de séries temporais em grafos, e uma abordagem específica, mais detalhada, para o método adotado nesta tese: grafos de visibilidade (*visibility graphs*) (Subseção 2.4.3). Na última parte (Subseção 2.4.4) serão mostrados trabalhos científicos nos quais as redes complexas são utilizadas para resolver problemas dentro do campo de Recuperação de Informações Musicais. Nesta seção será identificada a carência que esta tese se propõe a preencher: a falta de trabalhos que usam redes complexas com o fim de extrair características de sinais de áudio a partir de dados não-simbólicos.

- **Seção 2.5** - Define que o termo auto-similaridade usado por Tzanetakis e Cook 2002 para estudar padrões em sinais de áudio, será o mesmo adotada nesta tese, diferentemente do sentido usado em fractais geométricos, fractais estatísticos, e correlações longo alcançe em séries temporais.

Capítulo 3 - Metodologia

- **Seção 3.1** - Explica as particularidades técnicas da base de dados GTZAN, e apresenta o motivo da escolha desta base de dados.

- **Seção 3.2** - Descreve, de uma forma geral, os processos usados para a realização da parte experimantal da tese: Cálculo das flutuações de variância do sinal de áudio, Mapeamento dessas flutuações em Grafos de Visibilidade, Cálculo das características do Descritor de Visibilidade, Cálculo dos descritores tempo-frequência, Classificação de gêneros musicais (incluindo aprendizagem e treinamento de máquina), e seleção de atributos. Nesta seção é apresentado um fluxograma que ilustra todas as etapas do percurso metodológico.

- **Seção 3.3** - Descreve cálculo da série $V(j)$ - conjunto de pontos usados para representar o sinal de áudio de uma determinada música. A partir desta seção $V(j)$ será considerada como a representação das flutuações de variância do sinal, que em Jennings et al. 2004 é aproximada pelo desvio-padrão em caixas fixas, e interpretada como flutuações de *loudness*. Para este estudo, cada ponto da série $V(j)$ será visto como o resumo do comportamento das flutuações de intensidade dentro de um intervalo de 10 ms.

- **Seção 3.4** - Apresenta a transformação de uma série $V(j)$ em um grafo de visibilidade $G(V(m), V(n))$, de forma mais específica e resumida, uma vez que o detalhamento foi dado na Seção 2.4.3. Aqui serão apresentadas algumas características do mapeamento que irão ser evidenciadas e melhor discutidas na Seção 4.1.

- **Seção 3.5** - Apresenta o cálculo do Descritor de Visibilidade em Flutuações de Variância (DVFV) como a proposta desta tese para descrever a auto-similaridade de sinais de áudio através de quatro propriedades de redes complexas: $< k >$ grau médio, Δ densidade, Nc número de comunidades, e Q modularidade.

- **Seção 3.6** - São apresentados aqui, os métodos de cálculo dos descritores tempo-frequência: MFCCs (*Mel Frequency Cepstral Coefficients*), Fluxo Espectral, *Loudness*, *Onset Rate*, Taxa de Passagem pelo Zero, Complexidade da Dinâmica, Histograma de Batidas, Expoente DFA, Batimentos pos minuto (BPM).

- **Seção 3.7** - Define e discute fundamentos dos algoritmos utilizados para aprender e testar estratégias de classificação a partir das informações dadas pelos descritores DVFV e Tempo-Frequência. Foram quatro os algoritmos utilizados: J48, Naive Bayes, Multiclasss classifier, IBK, e k-Star. Nesta seção são descritas as características do software WEKA - plataforma utilizada para realizar a modelagem computacional.

- **Seção 3.8** - Explana o método utilizado para estabelecer o ranking do ganho de informação obtido pelos atributos DVFV e tempo-frequência, através da taxa de ganho.

Capítulo 4 - Resultados e Discussão

- **Seção 4.1** - Mostra resultados da modelagem gráfica, utilizando o algoritmo de detecção de comunidades, em grafos criados a partir de alguns sinais musicais. Apresenta exemplos onde a modelagem gráfica dos grafos de visibilidade, associada ao DVFV, é utilizada para diferenciar sinais de gêneros musicais distintos.

- **Seção 4.2** - Apresenta os resultados do cálculo do DVFV para 1.000 grafos de visibilidade, do teste de hipótese sobre a diferenças entre as médias de $< k >$, Q, Nc, Δ, e propõe hierarquias de auto-similaridade usando as quatro componentes do DVFV.

- **Seção 4.3** - Mostra os resultados da classificação usando apenas o descritor de visibilidade (**Seção 4.3.1**), usando o DVFV junto com descritores tempo-frequência, e usando um descritor por vez (Seção 4.3.2).

- **Seção 4.4** - Realiza a seleção de atributos e mostra o ranking do ganho de informação para o DVFV e descritores tempo-frequência.

- **Seção 4.5** Realiza a comparação com trabalhos correlatos. Na Subseção 4.5.1 compara a hierarquia de gêneros musicais usando o DVFV com a hierarquia usando DFA. Na subseção **Subseção 4.5.2** compara os resultados da classificação desta tese com trabalhos que utilizaram a mesma base de dados.

Capítulo 5 - Considerações Finais Nesta seção são feitas as considerações finais, são relatadas as principais contribuições e os trabalhos futuros.

2 REVISÃO DA LITERATURA

2.1 CLASSIFICAÇÃO DE GÊNEROS MUSICAIS

Tzanetakis e Cook 2002 definem gêneros musicais como rótulos criados e utilizados por seres humanos para categorizar e descrever o vasto universo da música. Para esses autores, os gêneros musicais não têm definições e limites rigorosos, principalmente por se tratar de um tema que exigiria uma complexa interação entre especialistas, público, *marketing* e fatores histórico-culturais. Este fato tem levado alguns pesquisadores a sugerirem a definição de um novo esquema de classificação de gêneros musicais puramente para fins de pesquisa MIR (*Music Information Retrievial*) [Pachet e Cazaly 2000]. Barbedo e Lopes 2007 abordam em sua pesquisa que, apesar de não haver unanimidade nesse tipo de classificação, os membros de um determinado gênero compartilham características em comum como estrutura rítmica, instrumentação, conteúdo tonal e outros. O grau de arbitrariedade e incoerência na classificação da música em gêneros foi discutido por Pachet e Cazaly 2000, onde foram comparadas três diferentes taxonomias de gênero da internet. Barbedo e Lopes 2007 afirmam que outro fator importante a ser considerado na discussão sobre classificação de gêneros musicais é que grande parte das canções produzidas na atualidade possuem elementos de mais de um gênero musical, e que para lidar com esse problema pode-se utilizar uma divisão básica de gêneros em uma série de subgêneros capazes de abranger classes intermediárias. Ahrendt e Hansen 2006 fizeram uma experiência comparando a classificação humana com a classificação automática para 11 gêneros musicais, chegando a um resultado de 57% de acerto para classificação humana e 48% para a classificação computacional. Já o experimento de Gjerdingen e Perrott 2008 revela um acerto humano de 70% para um experimento com dez gêneros musicais. Um aprofundamento sobre essa questão pode ser encontrado em Seyerlehner et al. 2010. Conclui-se daí que realizar a classificação em gêneros musicais não se constitui tarefa trivial, nem para o sofisticado sistema de identificação do cérebro humano.

A Figura 2.1 apresenta os passos fundamentais para a realização da classificação de gêneros musicais. Essas etapas têm sido largamente utilizadas em trabalhos ligados à área

de computação e recuperação de informações musiciais (MIR), tais como Tzanetakis e Cook 2002, Panagakis et al. 2008, e Lykartsis e Lerch 2015, e servem como estrutura para muitos dos trabalhos reportados em Schedl et al. 2014,e Lerch 2012. O sinal de áudio musical passa primeiramente por um pré-processamento a fim de entregar à etapa de extração de atributos uma representação adequada aos seus algoritmos. Na segunda etapa, são calculadas representações numéricas da natureza tonal, rítmica, e timbrística da música, seguindo, na maioria das vezes, o modelo adotado em Tzanetakis e Cook 2002. Ao final dessa etapa, cada sinal musical estará representado por um conjunto de números denominado de vetor de atributos. Nas etapas finais, 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, e o pós-processamento é usado para alcançar decisão por um único gênero musical. No final, o resultado da aprendizagem pode ser representado em uma matriz de confusão que consiste em uma representação matricial onde se pode identificar facilmente os verdadeiros e falsos positivos da classificação. A Extração de Atributos será apresentada com maiores detalhes na Seção 2.2, enquanto a Aprendizagem e Classificação serão melhor explanadas em uma parte da metodologia descrita na Seção 3.7.

Figura 2.1: Esquema básico de um sistema de classificação de gêneros musicais. Fonte: Autor.

2.2 EXTRAÇÃO DE ATRIBUTOS

Desde sua publicação, o trabalho de Tzanetakis e Cook 2002 tem sido usado como referência para a classificação de gêneros musicais [Aucouturier e Pachet 2003, Bergstra et al. 2005, Müller 2007, Andén e Mallat 2011]. Nesse trabalho são utilizados algoritmos de identificação de padrões em sinais de voz para classificar sinais de áudio musicais. Esses algoritmos foram divididos quanto às caracterísiticas timbrísticas, tonais e rítmicas. As características de natureza timbrística estão relacionadas às frequências que dão “colorido” ao som. É esse tipo de característica que diferencia por exemplo o som “àspero” de uma guitarra com efeito de

distorção bem característica do estilo *rock*, do timbre "doce" de uma flauta transversal. As características tonais estão relacionadas com as alturas das notas musicais e remetem à sensação auditiva de afinação tonal, acordes e melodias. Estilos musicais como o Jazz e a música clássica possuem muito mais variações no espectro tonal que estilos como Pop, Reggae e Metal. As características de natureza rítmica estão ligadas ao número de batidas por unidade de tempo, compasso, regularidade e pulsação da música. É através dessa característica que sabemos se uma música é mais "rápida" ou mais "lenta" do que a outra. Intuitivamente nós marcamos esse aspecto musical batendo o pé ou as mãos durante a execução de uma música. Um ouvinte com algum treino em percepção musical não teria muitas dificuldades para identificar a maioria desses aspectos. Em Tzanetakis e Cook 2002, os algoritmos utilizados para identificar os atributos timbrísticos, rítmicos, e tonais, utlizam os algoritmos conforme descritos a seguir:

- Atributos de Textura Timbral

  Centróide Espectral - centro de gravidade do espectro de magnitude do STFT (Transformada curta de Fourier). O centróide é uma medida do formato do espectro. Valores altos de centróide correspondem a texturas "mais brilhantes" decorrentes de frequências mais altas.

  ***Rollof* Espectral** - frequência abaixo da qual 85% da distribuição de magnitude está concentrada. O *roloff* é outra forma de medir a forma espectral.

  Fluxo Espectral - diferença quadrática entre as magnitudes normalizadas de distribuições espectrais sucessivas. O fluxo espectral é uma medida da quantidade de mudança espectral local.

  Passagem pelo Zero - quantifica o número de cruzamentos pelo zero do sinal no domínio do tempo, e está associada à distância entre dois picos ou vales. É um atributo que mostra em que medida amostras sucessivas possuem sinais diferentes. Essa taxa pode ser usada para estimar a complexidade do sinal. Altos valores indicam a presença de muitos picos e vales. A taxa de passagem pelo zero também mostra uma sensibilidade especial para sons vocais e percussivos, e valores altos dessa taxa também podem indicar maior presença desses sons nos sinais.

  MFCCs (*Mel Frequency Cepstral coefficients*) ou Coeficientes Cepstrais da Frequência Mel - são a descrição compacta da forma do envelope espectral de um sinal de áudio baseado na escala Mel. A escala Mel é uma escala de frequência não-linear que procura se aproximar à escala de percepção humana. A escala Mel foi introduzida em 1937 por Stevens et al. 1937 com uma abordagem diferente da escala dB (decibéis), que ao invés de medir a intensidade e a potência do sinal, busca uma associação entre a frequência da

nota percebida pelo ouvido humano e a sua frequência fundamental (o pitch). Deste modo a escala Mel busca uma representação qualitativa para a resposta humana ao estímulo sonoro, permitindo uma estimativa da percepção de quão grave ou agudo é um determinado tipo de som. Desde a sua introdução, os MFCCs têm sido amplamente utilizados no campo de processamento de sinais de fala e também têm encontrado aplicação no processamento de sinais de música. No contexto da classificação de sinais de áudio, verificou-se que um pequeno subconjunto de MFCCs resultantes já contém a informação principal do sinal. Na maioria dos casos, o número de MFCCs usados varia de 4 a 20 [Lerch 2012]. Em aplicações de segmentação de música e fala normalmente são usados 13 MFCCs.

- Atributos de Textura Rítmica

  Histograma de Batidas - Os sistemas de detecção automática de batidas em sinais de áudio, fornecem uma estimativa da execução e da força do seu ritmo principal. Além desse recursos para caracterizar gêneros musicais, podem ser utlizadas nos vetores de atributos, informações como a regularidade do ritmo, a relação da batida principal com as sub-batidas e a força relativa dos sub-batimentos com o ritmo principal. Uma das estruturas comuns de detecção automática de batidas, consiste em uma decomposição de banco de filtros, seguida por um passo de extração de envelope e, finalmente, um algoritmo de detecção de periodicidade que é usado para detectar o atraso em que o envelope do sinal é auto-similar. O cálculo de recursos para representar a estrutura rítmica da música baseia-se na transformada wavelet (TW), que é uma técnica para analisar sinais que foram desenvolvidos como uma alternativa ao STFT para superar seus problemas de resolução [Tzanetakis e Cook 2002].

  A Figura 2.2 mostra o diagrama de fluxo do algoritmo de análise de batidas. O sinal é primeiro decomposto em uma série de bandas de freqüência de oitava usando a transformada discreta Wavelet. Após esta decomposição, o envelope de amplitude do domínio do tempo de cada banda é extraído separadamente. Isto é conseguido através da aplicação de retificação de onda completa, filtragem de passagem baixa e downsampling para cada banda de frequência de oitava. Após a remoção da média, os envelopes de cada banda são então somados e é calculada a autocorrelação do envelope resultante. Os picos dominantes da função de auto-correlação correspondem às várias periodicidades do envelope do sinal. Esses picos são acumulados em todo o arquivo de som em um histograma de batidas onde cada bin corresponde ao atraso de pico, isto é, o período de batida em batimentos por minuto (BPM). Em vez de adicionar um, a amplitude de cada pico é adicionada ao histograma da batida. Dessa forma, quando o sinal possui muita auto-similaridade (batida forte), os picos de histograma serão maiores [Tzanetakis e Cook 2002].

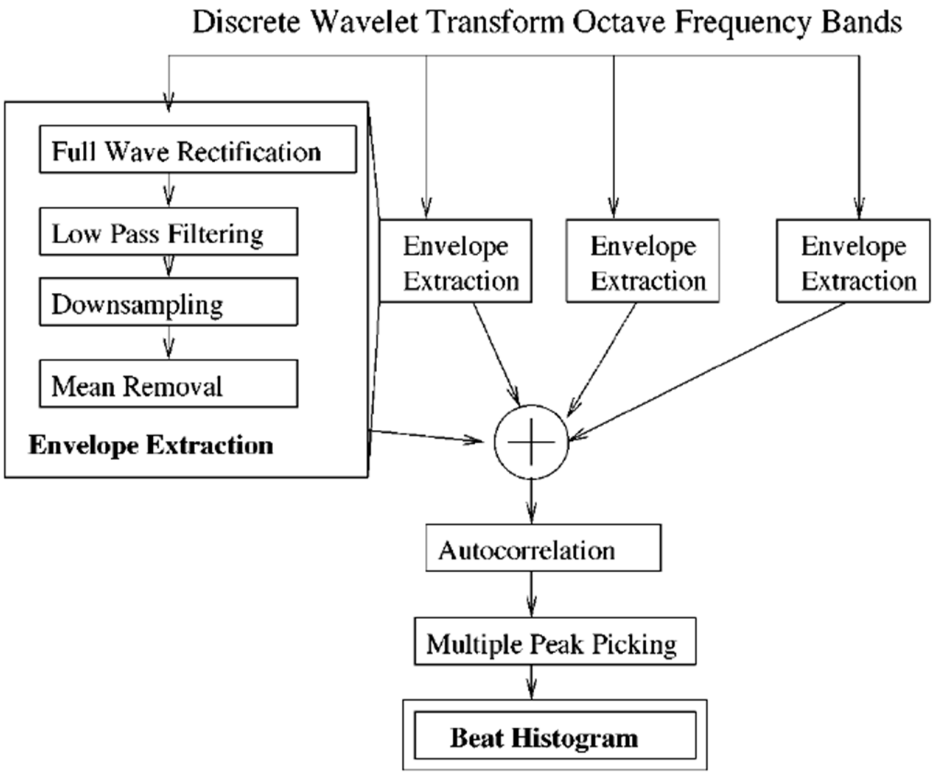


Figura 2.2: Diagrama para o cálculo do Histograma de Batidas. Fonte:Tzanetakis e Cook 2002.

A0, A1 - amplitude relativa dividida pelas somas das amplitudes do primeiro e do segundo picos do histograma. A amplitude relativa " é uma medida de quão distintivas são as batidas comparadas com o resto do sinal" [Jr et al. 2005];

RA - taxa da amplitude do segundo pico dividido pela amplitude do primeiro pico. Essa taxa "expressa a relação entre a batida principal e a primeira batida auxiliar" [Jr et al. 2005];

P1, P2 - períodos do primeiro e segundo picos em BPM, " indicam quão rápida é a música" [Jr et al. 2005];

SUM - soma total do histograma (para capturar a força da batida da música). "A soma das bandas do histograma é uma medida de força da auto-similaridade entre as batidas, a qual é um fator de quão rítmica uma musica parece ser." [Jr et al. 2005].

- Atributos de Textura Tonal

FPO - Período do pico máximo do histograma dobrado, correspondente aos acordes da tônica ou da dominante. Este pico será maior para músicas que não têm muitos mudanças harmônicas.

UPO - Período do pico máximo do histograma desdobrado, que corresponde ao âmbito

de uma oitava da tonalidade dominante da música.

IPO1 - Intervalo de tom entre os dois picos mais proeminentes do histograma dobrado. Isso corresponde à relação principal do intervalo tonal. Para peças musicais com estrutura harmônica simples, esse recurso terá valor 1 ou -1 correspondente ao quinto ou quarto intervalo (tônico-dominante).

SUM - Soma geral do histograma. Esta característica é uma medida da força da detecção tonal.

FP0 - Período do pico máximo do histograma dobrado. Isso corresponde à classe do tom principal da música.

2.3 SINAIS POLIFÔNICOS

Do ponto de vista da teoria musical o termo polifonia (do grego *polyphonía*) está associado ao movimento de duas ou mais linhas melódicas (ou vozes) independentes, porém inter-relacionadas em uma composição musical, através da técnica do contraponto. A monofonia, por sua vez, utiliza apenas uma linha melódica. Nesta tese, adotamos a mesma concepção de Lerch 2012, que considera o termo *sinal polifônico* para designar a representação de múltiplas vozes em um único sinal elétrico, onde cada uma delas é caracterizada por uma frequência fundamental, sem compromisso com a definição restrita da teoria musical no que diz respeito à inter-relação das vozes segundo a arte do contraponto. Assim, por exemplo, o sinal que representa a gravação de um conjunto de choro com flauta, pandeiro, e violão, ou ainda um duo de violinos, serão considerados sinais polifônicos, sem levar em consideração o discurso estético entre as linhas melódicas do arranjo musical. Consequentemente, a gravação de um tenor solo, ou um violão solo, serão considerados sinais monofônicos.

Dependendo do treinamento musical do ouvinte, ele poderá detectar mais ou menos melodias independentes no âmbito da criação polifônica. O cérebro humano pode ser capaz de identificar as vozes simultâneas de um sinal polifônico de forma bastante precisa, sem perder a noção do todo da obra musical. No universo da análise de sinais de áudio, a identificação de frequências fundamentais correspondentes às vozes simultâneas de um sinal polifônico tem sido um desafio para os pesquisadores que se dedicam à detecção de múltiplos tons em sinais polifônicos. Algoritmos computacionais para a separação de instrumentos, identificação de melodias, estimação do campo harmônico, e identificação de acordes têm sido experimentados em vários trabalhos desta área. Maiores detalhes podem ser encontrados em Lerch 2012.

Certamente, para a maioria das pessoas, a experiência de ouvir música se dá a partir

de sinais polifônicos. Esse tipo de sinal pode ser obtidos através da gravação simultânea de todos os sinais em em única pista, ou por meio da gravação multi-pista, que representa uma grande parte dos registros da atualidade. Nesse sistema, cada fonte sonora é gravada em um pista (ou faixa), separadamente, e depois mixadas (misturadas) em uma única pista. A Figura 2.3 mostra o resultado da gravação de um arranjo musical em um sistema multi-pista, com o sinal polifônico resultante da mixagem de seis pistas (bateria, contrabaixo, violoncello, violinos 1 e 2, e oboé).

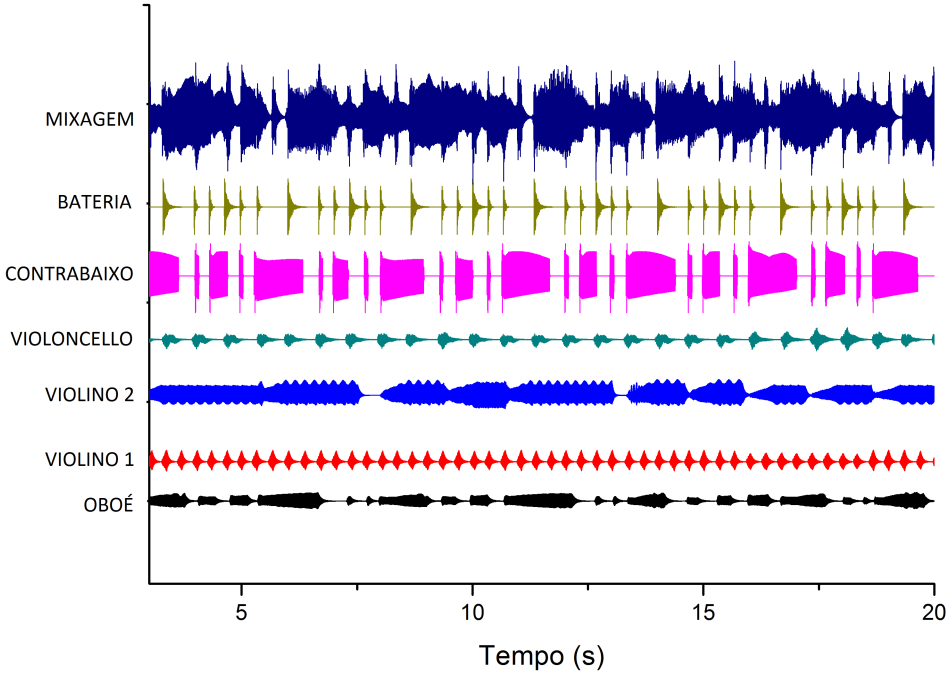
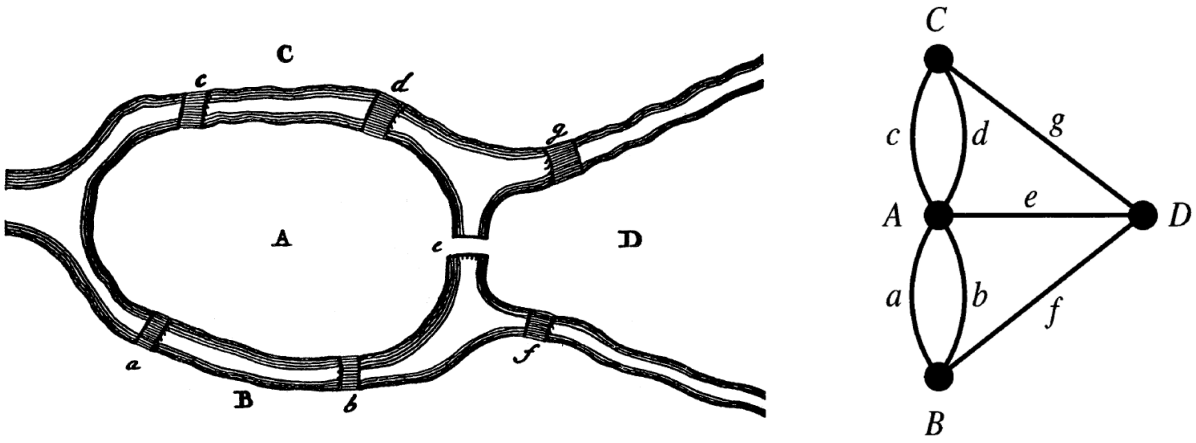


Figura 2.3: Sinais de áudio gravados em um sistema multipista. Fonte: Autor.

2.4 GRAFOS E REDES

Muitos autores consideram que o estudo da ciência de redes teve seu início no século XVIII com o surgimento da teoria dos grafos [Newman et al. 2011]. O início do uso dos grafos é atribuído ao matemático Leonhard Euler (1707-1783), que utilizou um sistema constituído de nós e conexões através de um diagrama que representava o mapa da cidade (Figura 2.4), para resolver um tradicional enigma que envolvia o tráfego em sete pontes na cidade de Könisberg. A primeira demonstração matemática desta solução foi apresentada por Euler para os membros da Academia Petersburgo em 26 de Agosto, 1735, e publicado em 1736 no artigo entitulado *Solutio Problematis ad Geometriam Situs Pertinentis* (A solução a um problema relacionado com a geometria da posição) [Alexanderson 2006].



(a) (b)

Figura 2.4: (a)As sete pontes da cidade prussiana de Königsberg segundo a representação de Euler em seu artigo de 1736; (b) A topologia das quatro massas de terra de Königsberg e as sete pontes representadas por um grafo com quatro vértices e sete arestas. Fonte: Hopkins e Wilson 2004.

Matematicamente, um grafo G é um par ordenado (V, E) formado por um conjunto de vértices V, e um conjunto de arestas E, juntamente com uma função de incidência que associa a cada aresta de G, um par não-ordenado de vértices de G (não necessariamente distintos). Se e é uma aresta e u e v são vértices de tal forma que $\psi(e) = (u, v)$, diz-se que e está unindo u e v, e que estes são as extremidades de e. Ainda segundo Bondy e Murty 1976:

> os grafos são assim chamados porque podem ser representados graficamente, e essa representação gráfica é que nos ajuda a compreender muitas de suas propriedades. Cada vértice é indicado por um ponto, e cada extremidade por uma linha que une os pontos representando suas extremidades. Se uma aresta possui indicação de sentido (uma seta) ela é chamada de arco. Um grafo que possui arcos é chamado de digrafo ou grafo dirigido.

A Figura 2.5 traz ilustrações de grafos dirigidos e não dirigidos.

Depois de dois séculos, na década de 1960, dois matemáticos, Paul Erdös e Alfred Rényi, introduziram um novo conceito que permite o estudo dessas redes, a teoria dos grafos aleatórios. A ideia dos dois matemáticos foi combinar os conceitos de teoria dos grafos com ferramentas da teoria da probabilidade.

Uma outra importante propriedade presente nas redes sociais foi descoberta em 1967 quando Stanley Milgram, interessado na estrutura da sociedade americana, descobriu, por meio de um experimento com correspondências, que a distância média entre duas pessoas quaisquer é próxima de seis. Este fenômeno passou a ser conhecido como efeito mundo pequeno. Existem, em seguida, dois trabalhos que proporcionaram uma grande revolução no estudo das redes complexas por meio de grafos. Foram os artigos publicados por Watts e Strogatz 1998 e por

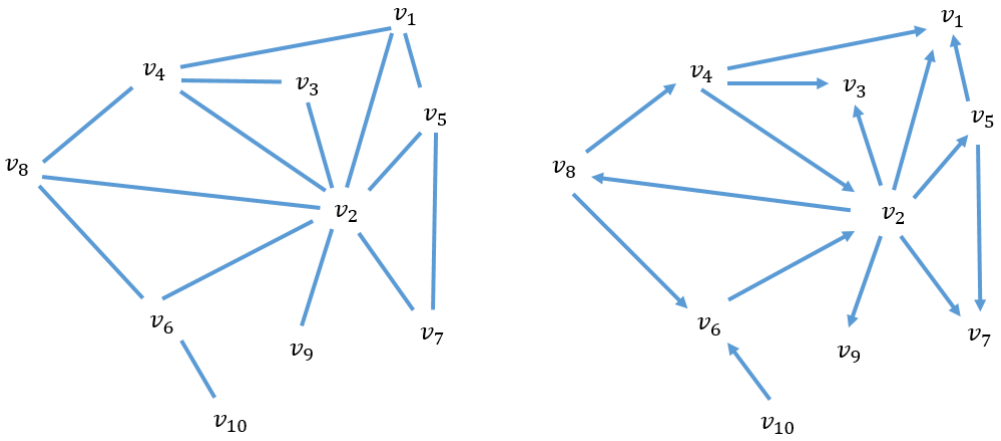


Figura 2.5: (a) grafo não-dirigido; (b) grafo dirigido. Fonte: Autor.

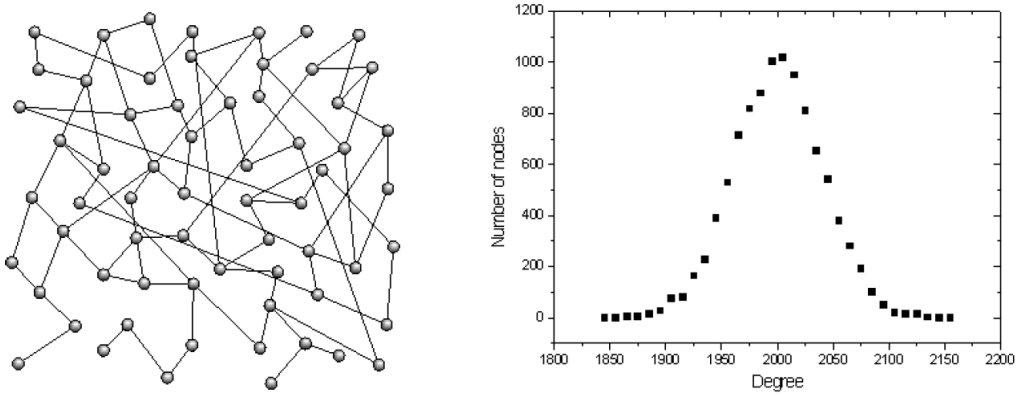
Barabási e Albert 1999. A partir daí, os grafos passaram a ser a base matemática da nova teoria das redes complexas [Newman et al. 2011].

A teoria dos grafos se afirmou como uma teoria elegante, profunda e poderosa, e em certa medida se distanciou das questões empíricas. Por outro lado, a nova ciência de redes tem direcionando seu foco para: (1) modelagem de redes reais, onde são conciliadas questões teóricas com empíricas; (2) uma perspectiva que vê as redes como objetos não estáticos, mas envolvidos no tempo; (3) redes como sistemas dinâmicos, usando uma abordagem que busca entender as redes não apenas como objetos topológicos, mas também como uma estrutura sobre a qual sistemas dinâmicos distribuídos são construídos. 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 como redes, e têm constatado que existe um grande manancial de conhecimento a ser descoberto através do estudo dessas redes [Newman et al. 2011].

2.4.1 REDES ALEATÓRIAS

As redes aleatórias, criadas por Solomonoff e Rapoport 1951 e Erdos e Rényi 1960, são consideradas como um dos modelos clássicos de redes, juntamente com as redes Mundo Pequeno [Watts e Strogatz 1998] e as redes Livres-de-escala [Barabási e Albert 1999]. Para formar uma rede aleatória segundo o modelo que ficou conhecido como Erdös-Renyi, são fixados n vértices onde cada uma das $\frac{1}{2}n(n-1)$ arestas são apresentadas com a mesma probabilidade de p de conexão, de forma aleatória e independente. A distribuição de graus desta rede é uma distribuição Binomial ou de Poisson, no caso de n tendendo ao infinito (Figura 2.6). As redes

aleatórias de um modo geral apresentam pequenos valores de coeficiente de aglomeração e caminho mínimo médio [Newman et al. 2006]. Pesquisadores da atualidade têm direcionado o foco para o estudo de redes reais, buscando modelos distintos das redes aleatórias para melhor compreender as suas propriedades. Apesar disso, Costa et al. 2007 ressaltam a importância teórica das redes aleatórias para a realização de estudos comparativos.



(a) (b)

Figura 2.6: (a) rede aleatória de Erdös-Rényi; (b) distribuição de graus da 10 redes aleatórias Erdös-Rényi 10.000 vértices e $p = 0,2$. Fonte: Costa et al. 2007.

2.4.2 MÉTODOS PARA MAPEAMENTO DE SÉRIES TEMPORAIS EM REDES

A análise de séries temporais tem como alvo principal a elaboração de modelos estatísticos para descrever a dinâmica de um dado fenômeno representado a partir de uma sequência de observações no tempo. Dentre os modelos que foram desenvolvidos podemos citar: modelos auto-regressivos integrados de média móvel (ARIMA) [Wei et al. 2006], modelos que usam redes neurais [Refenes et al. 1994], análise multifractal [Kantelhardt et al. 2002], e modelos não-lineares [Kantz e Schreiber]. Uma nova abordagem, que vem emergindo para se aliar às técnicas tradicionais, utiliza o mapeamento de séries temporais em redes. Com esse novo foco o estudo da dinâmica de fenômenos temporais, pode beneficiar-se de todo conhecimento desenvolvido para a análise topológica de redes complexas. Dentre os vários métodos desenvolvidos para realizar esse tipo de mapeamento, destacamos:

- *Correlation networks*: Dada uma série arbitrária, são definidos vetores x_i em um espaço de fase n-dimensional de variáveis embutidas. Os vértices da rede não dirigida associada à essa série temporal são os vetores x_i. Se o coeficiente de correlação de Pearson r_{ij} entre x_i e x_j for maior que um dado valor r, esses pares de vetores estarão conectados, e, portanto, será definida uma aresta entre eles [Yang e Yang 2008], [Gao e Jin 2009].

- *k-Nearest Neighbor Networks*: onde os vértices v_i são também observações vetoriais (estados) dentro de um espaço de fase, e o critério de conexão utilizado considera que um vértice v_i está ligado a outros k vértices v_j, cujas distâncias d_{ij} em relação aos seus k vizinhos mais próximos, sejam mínimas [Donner et al. 2011].

- *Cicle Networks*: método que usa propriedades topológicas de séries temporais pseudo-periódicas, onde os vértices são os vários ciclos identificados na série [Zhang e Small 2006].

- Mapeamento Direto e Inverso: onde Campanharo 2011 propõe um modelo que utiliza o mapeamento direto e inverso (Figura 2.7), usando uma função de auto-correlação, e sugerindo o estudo de séries temporais a partir de propriedades de redes complexas.

- Grafos de Visibilidade (*Visibility Graphs*): constroem as redes utilizando a convexidade de observações sucessivas segundo o critério de visibilidade [Lacasa et al. 2008]. Este método se destaca dos demais que, ao invés de uma abordagem estatística, utiliza uma abordagem que é adaptada de um princípio simples da geometria computacional. Uma vez que os grafos de visibilidade serão adotados na metodologia desta nesta tese, uma descrição detalhada sobre esse método será dada na Seção 2.4.3.

Outras técnicas de mapeamento de séries temporais em redes podem ser encontrados em Small et al. 2009, Donner et al. 2011, e Yang e Yang 2008.

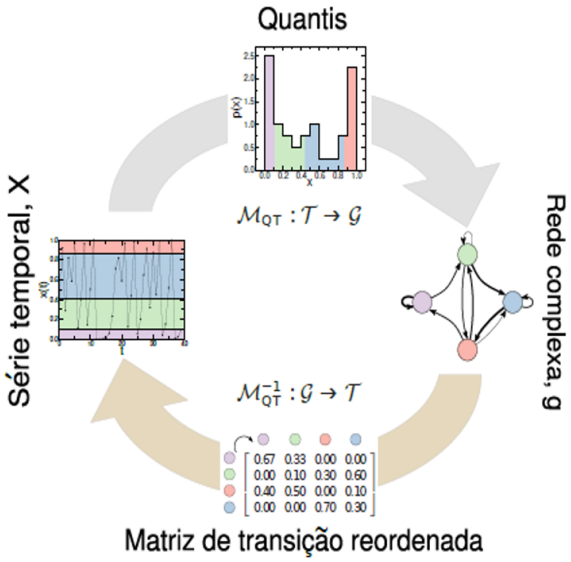


Figura 2.7: Mapeamento direto e inverso. Fonte: Campanharo 2011.

2.4.3 GRAFOS DE VISIBILIDADE

Os grafos de visibilidade têm criado pontes de ligação 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. Através dessas pontes, tem sido possível estudar temas de vários campos de pesquisa como: estruturas auto-similares [Lacasa et al. 2009] e fractais em séries temporais [Nunez et al. 2012], índices de bolsas de valores [Stephen et al. 2015], e tráfego de pacotes de informações [Andjelković et al. 2015]. Esses estudos, em particular, têm mostrado o potencial dos grafos de visibilidade para detecção de tendências locais em séries temporais.

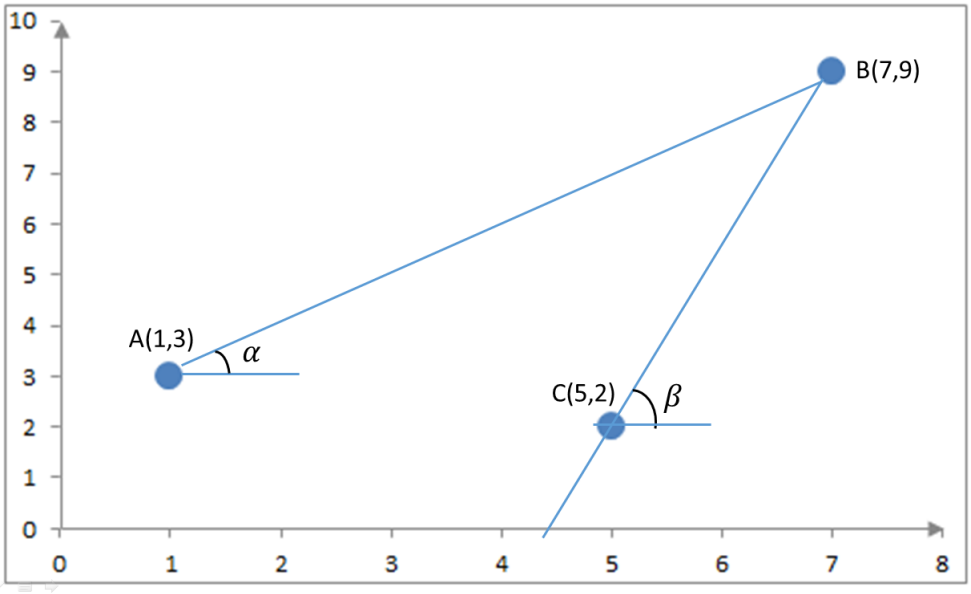
Segundo Lacasa et al. 2008 o critério de visibilidade é definido da seguinte maneira: dada uma série temporal $\{V_1, V_2, ..., V_n\}$ sempre existe visibilidade entre dois pontos consecutivos da série temporal, e dois pontos arbitrários $A(x_a, V_a)$ e $B(x_b, V_b)$ da série terão visibilidade mútua, se todo ponto $C(x_c, V_c)$ entre eles satisfaz a condição:

$$\frac{V_b - V_c}{x_b - x_c} > \frac{V_b - V_a}{x_b - x_a}. \tag{11}$$

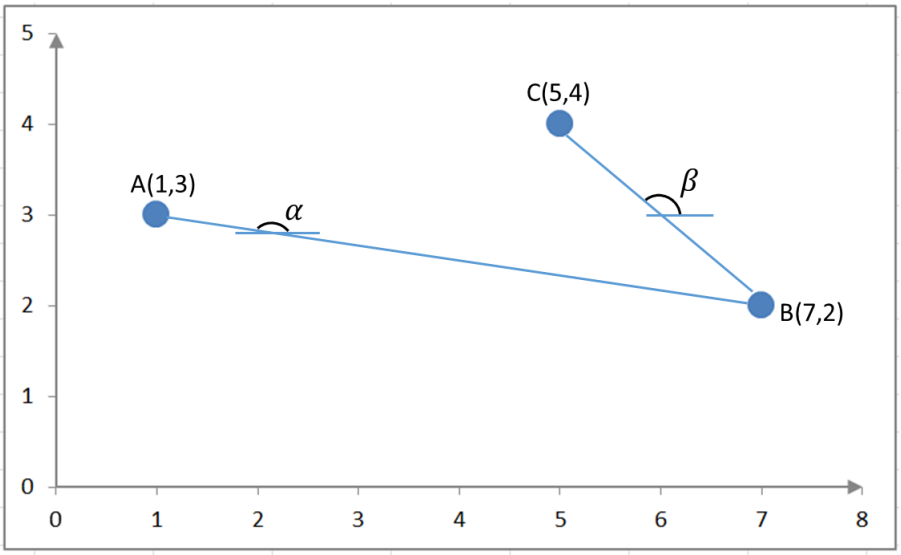
A Equação 11(a) estabelece uma comparação entre a inclinação da reta BC ($tg(\beta)$) e a inclinação da reta BA ($tg(\alpha)$) (Figura 2.8). Toda vez que $tg(\beta) > tg(\alpha)$ existe visibilidade entre os pontos A e B da série, e no grafo será criado uma aresta $E(V_a, V_b)$, caso contrário não haverá visibilidade entre A e B na série, tampouco será criada uma aresta conectando A e B no grafo.

Lacasa et al. 2008 sugere que imaginemos os pontos A e B como picos de duas montanhas. O ponto A só "enxerga" o ponto B se não existir uma terceira montanha C entre A e B interrompendo a linha reta de visualização AB. Na Figura 2.8(a) podemos ver um exemplo onde há visibilidade entre os pontos A e B. Observamos neste caso que a "linha de visão" AB não é interrompida pela "montanha" C. Calculando as tangentes dos ângulos das retas AB e BC em relação ao eixo ox temos $tg(\beta) = 3,5 > tg(\alpha) = 1$, o que, segundo a Equação 11, satisfaz o critério para a existência de visibilidade entre A e B. A Figura 2.8(b) mostra que a "linha de visão" entre A e B é interrompida pela altura de C. Nesse caso $tg(\beta) = -1 < tg(\alpha) = -0,17$, o que corresponde a dizer que não há visibilidade entre os pontos A e B.

A Figura 2.9 (a) mostra uma série de oito pontos no plano cartesiano, e em seguida a Figura 2.9 (b) mostra a criação da rede de visibilidade a partir dessa série, passo a passo, descrevendo a ligação entre os vértices. Cada ponto P_i da série se transforma em um vértice V_i



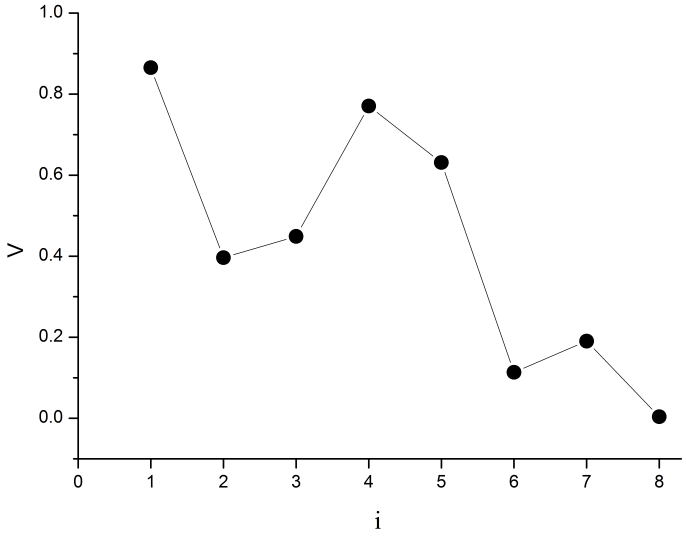
(a) A e B satisfazem o critério de visibilidade, basta notar que $tg(\beta) > tg(\alpha)$.



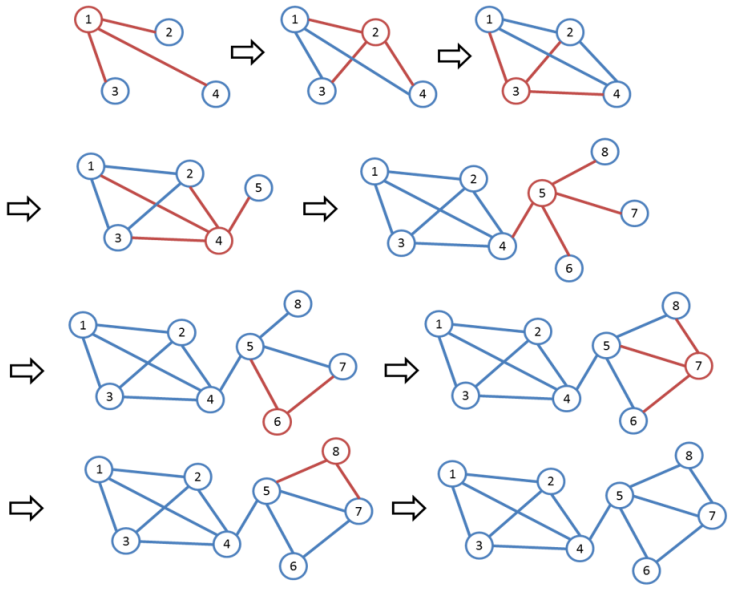
(b) A e B não têm visibilidade mútua uma vez que $tg(\beta) < tg(\alpha)$.

Figura 2.8: Ilustração da visibilidade entre pontos de uma série em dois casos. Fonte: Autor.

da rede, e estes vão se ligando a outros vértices segundo o critério de visibilidade aplicado à P_k, em cronológica até o último ponto. O primeiro grafo apresenta a visibilidade de P_1, gerando as arestas $V_1(1,2)$, $V_2(1,3)$ e $V_3(1,4)$ no grafo. O vértice P_2 traz as ligações correspondentes à visibilidade de P_2, e assim por diante. O último grafo de visibilidade (Figura 2.9), correspondente ao mapeamento de toda série.



(a) Representação cartesiana de uma série de oito pontos



(b) Grafo de visibilidade da série representada em (a)

Figura 2.9: Ilustração da construção de uma rede de visibilidade a partir de uma série de oito pontos. Fonte: Autor.

2.4.4 REDES COMPLEXAS NO ESTUDO DE INFORMAÇÕES MUSICAIS

Muitos pesquisadores da área de recuperação de informações musicais (MIR) têm utilizado redes complexas dentro dos mais variados contextos. Buldú et al. 2007 utilizam os características fundamentais de redes para analisar o gosto musical de usuários a partir de suas *playlists*, onde os vértices são os títulos das músicas e as arestas ocorrem entre o título da música A e o título da música B, se elas aparecem em mais de uma *playlist*. Tse et al. 2008 e Liu et

al. 2010 constroem redes com base na análise de padrões em notas musicais de composições, e utilizam propriedades universais encontradas nestas redes para estabelecer regras de composição algorítmica. Itzkovitz et al. 2006 estudam padrões universais de harmonia através de *motifs* de redes complexas. Park et al. 2015 estuda a topologia e evolução da rede de compositores da música clássica ocidental, a partir de meta-dados de arquivos de áudio, associando informações sobre o autor, período e estilo. Correa et al. 2010 tratam de classificação de gêneros musicais utilizando ritmos extraídos de uma base de dados MIDI e transformados em redes complexas, onde cada célula rítmica representa um nó, enquanto as sequências de notas definem as ligações entre os nós segundo um modelo de Markov. Jacobson et al. 2008 combinam análise de áudio e estrutura de redes para identificar comunidades de artistas do myspace. Em nenhum dos trabalhos descritos anteriormente, foi encontrada a extração de atributos de sinais de áudio musicais utilizando parâmetros de redes complexas para classificação de gêneros musicais.

2.5 AUTO-SIMILARIDADE DE SINAIS MUSICAIS

Nesta seção iremos exclarecer o termo auto-similaridade usado nesta tese, que se identifica com o sentido adotado por Tzanetakis e Cook 2002, e difere daquele utilizado para se referir à uma das propriedades dos fractais. Para evidenciar a diferença entre os usos do termo, vamos fazer uma breve revisão e contextualização.

Em geometria fractal objetos com propriedade de auto-similaridade estrita ou determinística, são aqueles constituídos de porções que podem ser identificadas como a réplica do todo, ou entes geométricos para os quais existe um ponto onde todos os seus vizinhos contém uma cópia da figura por inteiro [Mandelbrot e Pignoni 1983]. A Figura 2.10 mostra fractais com auto-similaridade estrita.

O conceito de auto-sumilaridade pode ser usado também para se referir à auto-similaridade estatística, que é uma extensão do conceito de auto-similaridade quando as partes não são estritamente iguais ao todo, mas guardam fortes semelhanças estatísticas em várias escalas. Séries temporais que exibem trechos invariantes em várias escalas são consideradas estruturas auto-similares. A Figura 2.11 mostra o exemplo de uma sére temporal de batimentos cardíacos como exemplo de fractal estatístico.

Devido a estruturação repetitiva inerente à obras musicais, é muito comum encontrar a recorrência de trechos em sinais de áudio de música. Jennings et al. 2004, Streich e Herrera 2005, Goulart 2012 estudam a auto-similaridade através das correlações em várias escalas usando o método DFA. Cooper e Foote 2002, Foote e Cooper 2001, Müller et al. 2011 usam

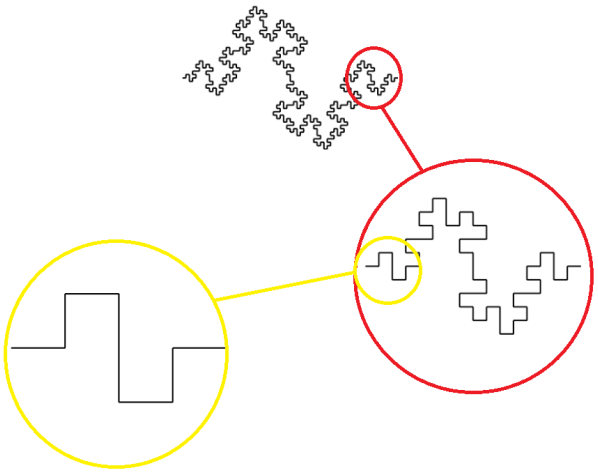


Figura 2.10: Objeto fractal e a ampliação de duas porções que evidenciam a auto-similaridade determinística. Fonte: Autor.

o termo auto-similaridade para designar padrões musicais repetitivos que visualizam a forma musical e a estrutura rítmica através de matrizes de auto-similaridade. A Figura 2.12 apresenta a identificação da forma musical de uma obra gravada em formato de áudio. Nesta matriz, a auto-similaridade do sinal é visualizada em uma representação bidimensional do tempo, obtida pelo cálculo da similaridade dois a dois de atributos como os MFCCs. O arquivo de áudio é representado como um quadrado, onde cada lado é proporcional ao comprimento da peça, e o tempo corre da esquerda para a direita, bem como de baixo para cima. Deste modo, o canto inferior esquerdo do quadrado corresponde ao início da peça, e o canto superior direito ao final. No quadrado, o brilho de um ponto é proporcional à semelhança de áudio nos tempos i e j (i: linha; j:coluna da matriz). Regiões semelhantes são brilhantes enquanto regiões diferentes são escuras.

Tzanetakis e Cook 2002 usam o termo auto-similaridade para se referir à homogeneidade da intensidade de picos. Os autores consideram que gravações de de áudio que possuem trechos musicais com batidas muito fortes e persistentes irão gerar sinais com muita auto-similaridade, e quanto menor a persistência e força dos batimentos principais, menor será a auto-similaridade. Para os autores, essa percepção ocorre durante o cálculo da função autocorrelação, durante o processo de construção do histograma de batidas.

Tzanetakis e Cook 2002 consideram que o sinal será mais auto-similar à medida que os picos do histograma de batida são maiores. Na Figura 2.13 podemos observar que:

> o canto superior esquerdo, rotulado como Clássico, é o histograma de batidas de um trecho de *La Mer* de Claude Debussy. Devido à complexidade dos múltiplos instrumentos da orquestra, não existe uma forte auto-semelhança e não há um pico dominante claro no histograma. Picos mais fortes podem ser vistos no canto inferior

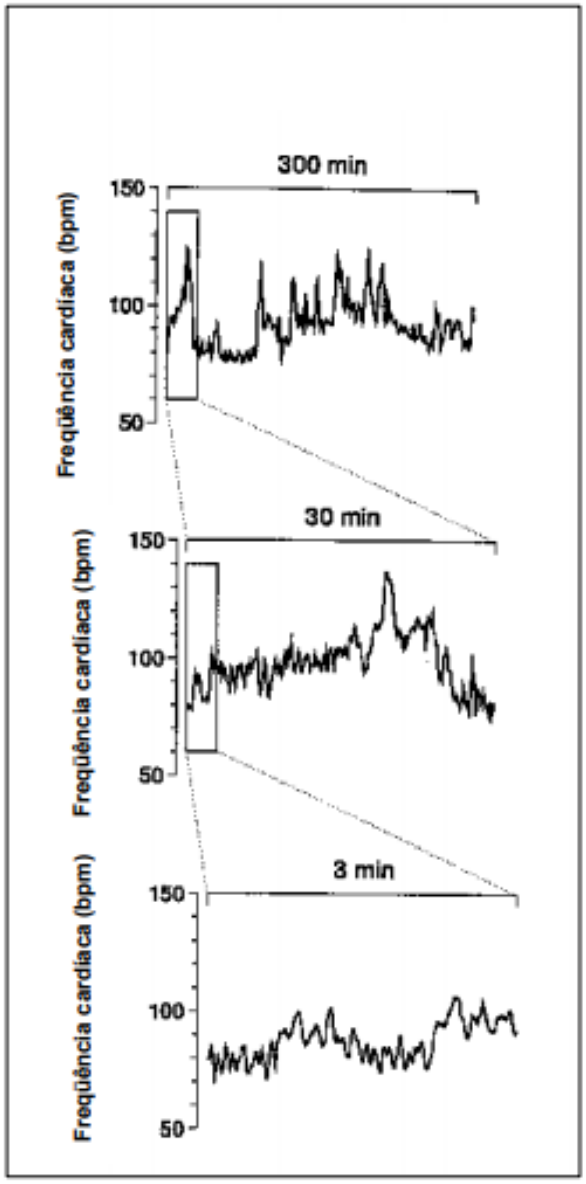


Figura 2.11: Frequência de batimentos cardíacos com ampliação de trechos que ilustram a auto-similaridade estatística. Fonte: Goldberger et al. 2002.

esquerdo, rotulado de jazz, que é um trecho de uma performance ao vivo de Dee Dee Bridgewater. Os dois picos correspondem ao ritmo da música (70 e 140 bpm). O histograma de batidas do gênero Rock é mostrado no canto superior direito onde os picos são mais pronunciados por causa da batida mais forte desse estilo musical. Os picos mais altos do canto inferior direito indicam a forte estrutura rítmica de uma música HipHop de Neneh Cherry [Tzanetakis e Cook 2002].

Os picos representados no histograma de batidas são calculados a partir da funçao de auto-correlação aprimorada (detalhes em Atributos de textura rítmica, na Seção 2.2). Esses picos:

correspondem aos intervalos de tempo em que o sinal é auto-similar (mais parecido

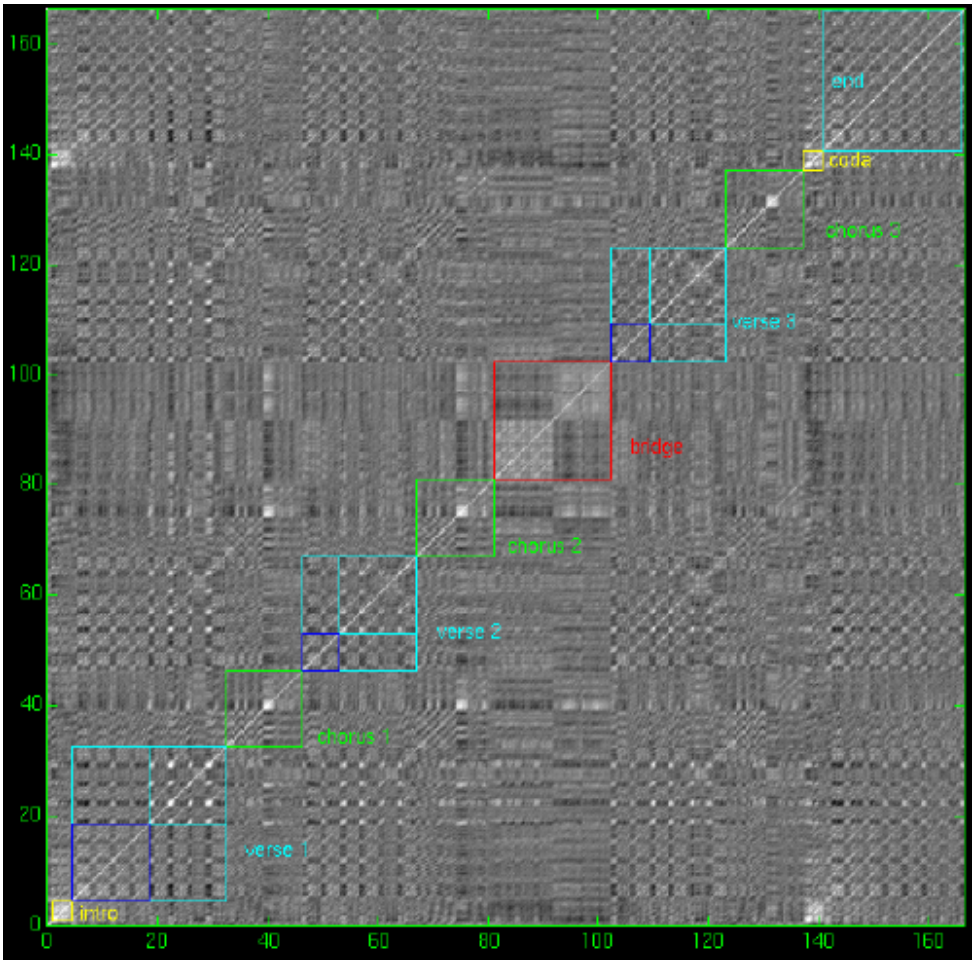


Figura 2.12: Matriz de auto-similaridade de um sinal de áudio musical onde podem ser vistas as partes que compõem a estrutura da forma musical: verse, chorus, bridge, e end. Fonte: Foote e Cooper 2001.

> com ele próprio). Os atrasos de tempo dos picos no intervalo direito do tempo correspondem, para a análise do ritmo, às periodicidades de batida [Tzanetakis e Cook 2002].

Nesta tese, nós adotamos um conceito de auto-similaridade na mesma linha de Tzanetakis e Cook 2002, uma vez que a estimamos com base na visibilidade dos picos locais predominantes, considerando que a auto-similaridade dos picos é maior na proporção que os sinais musicais representam batidas intensas e persistentes. A diferença é que, enquanto Tzanetakis e Cook 2002 usam o histograma de batidas, nós estudamos a auto-similaridade dos sinais através de propriedades topológicas de grafos de visibilidade. A Figura 2.14 mostra a representação das flutuações de variância de dois sinais musicais conforme a metodologia aplicada nesta tese, cujos detalhes serão aprofundados no Capítulo 3. Na Figura 2.14 (a) temos um sinal bastante homogêneo, baixa visibilidade local, e forte auto-similaridade, enquanto na Figura 2.14 (b) está representado um sinal heterogêneo, com alta visibilidade local e com baixa auto-similaridade.

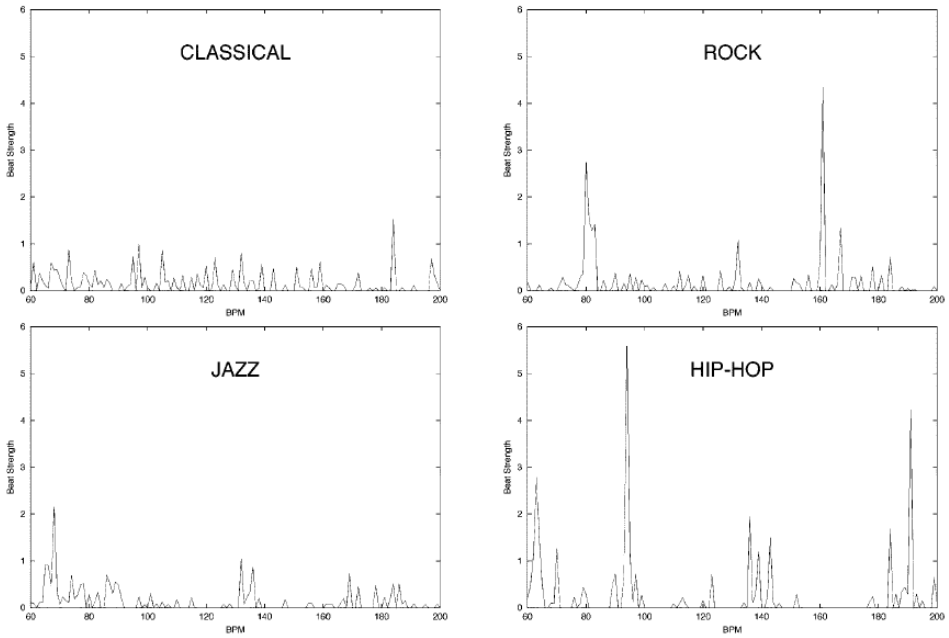
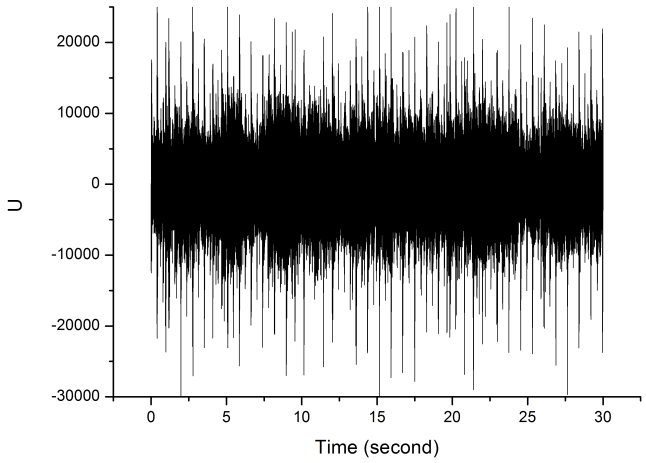
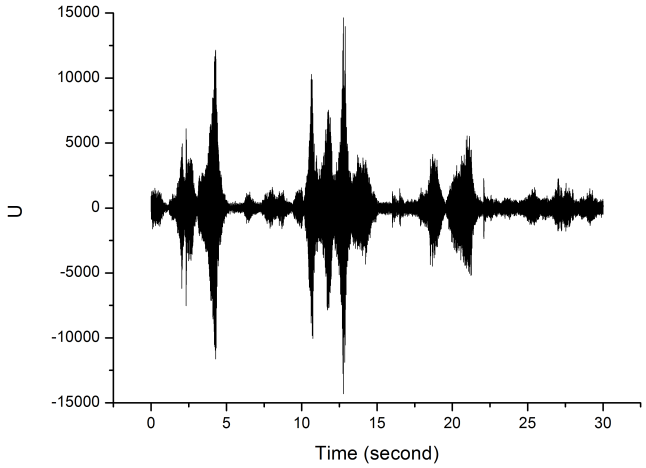


Figura 2.13: Histograma de batidas de quatro sinais musicais. Fonte: Tzanetakis e Cook 2002.



(a)



(b)

Figura 2.14: (a) sinal musical com forte auto-similaridade (estilo heavy-metal); (b) sinal musical com fraca auto-similaridade (estilo clássico). Fonte: Autor

3 METODOLOGIA

3.1 BASE DE DADOS

Os dados utilizados na pesquisa são 1000 arquivos de áudio digital da base GTZAN Genre Collection [1],constituída de dez gêneros musicais. Cada música tem duração de 30 segundos, quantização de 16-bit, e taxa de amostragem de 22.050 Hz. Desde sua publicação em Tzanetakis e Cook 2000, o banco GTZAN *Genre Collection* tem sido usado em muitos trabalhos envolvendo recuperação de informações musicais [Tzanetakis e Cook 2002, Tsunoo et al. 2009, Panagakis et al. 2009]. Esta base da dados foi escolhida com o objetivo de favorecer a reprodutibilidade da metodologia e a comparação com outros trabalhos científicos, uma vez que esta é uma das bases de dados públicas mais usadas em pesquisas relacionadas ao reconhecimento de gêneros musicais, como afirma Sturm 2013. Apesar do mesmo autor, após um estudo detalhado, apontar várias limitações neste banco de dados [Sturm 2012, Sturm 2013], os resultados preliminares mostram pouca influência destas limitações na aplicação da metodologia proposta.

3.2 DESCRIÇÃO DO MÉTODO

A metodologia a ser utilizada consiste em uma combinação de técnicas que são aplicadas nas seguintes etapas:

- Cálculo da série de flutuações de variância do sinal de áudio;
- Mapeamento da série de flutuações de variância em Grafos de Visibilidade;
- Cálculo das propriedades de Grafos para construir o Descritor de Visibilidade;
- Classificação de gêneros musicais.

[1] Disponível para download em https://marsyasweb.appspot.com/download/data_sets/

A Figura 3.1 apresenta um fluxograma com as etapas básicas da metodologia. Primeiro, cada música em formato de áudio de 30 s, é transformada em uma série temporal $U(i)$ com 330.000 pontos e guardada em um arquivo no formato txt. Depois, de cada uma dessas séries transformada em uma subsérie $V(j)$ com 3.000 pontos, através do cálculo do desvio-padrão em caixas de tamanho fixo. Na terceira fase cada série $V(j)$ se transforma em um grafo (V(m), V(n)) através do mapeamento de visibilidade [Lacasa et al. 2009]. Na fase seguinte são calculadas quatro propriedades topológicas para cada grafo usando a plataforma Gephi 0.9.0 [2]. Essas propriedades são as componentes do Descritor de Visibilidade em Flutuações de Variância através de grafos áudio-associados. Após a análise dessas propriedades é aplicado um sistema de classificação baseado em árvore de decisão utilizando a ferramenta WEKA [3] consiste em uma plataforma livre constituída de um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados, de onde será escolhido o melhor classificador para o conjunto de dados.

3.3 CÁLCULO DA SÉRIE $V(J)$

O cálculo da série de flutuações de variância $V(j)$ segue o procedimento adotado em Jennings et al. 2004, onde cada arquivo de áudio é tomado com uma taxa de amostragem $SR = 11.025Hz$. Essa taxa de amostragem tem apresentado excelentes resultados em estudos de correlações de longo alcançe em séries temporais de sinais musicais como em Streich e Herrera 2005, Melo 2013, Berois 2008. No trabalho de Melo et al. 2016 é feito um experimento comprovando que as taxas de amostragem $44.110Hz, 22.050Hz$, e $11.025Hz$ produzem estatísticas muito semelhantes para fins de estudo comparativo de sinais de áudio usando a abordagem de redes complexas. A vantagem de adotar essa taxa é, por um lado, o ganho de tempo de processamento para grandes bases de dados, e por outro lado, a garantia da reprodutibilidade do experimento mesmo em plataformas de código aberto que tenham restrições quanto ao grande número de vértices produzidos no contexto de sinais de áudio. Tomando $SR = 11.025Hz$ em sinais de 30 s temos uma série numérica com 330.750 pontos. Nesta tese nós transformamos cada arquivo de áudio em séries $\{U(i)\} = \{U(1); \cdots U(n)\}$ com $n = 330.000$ pontos. Deste modo evitamos problemas com números não inteiros na definição das caixas.

A série $U(i)$ é segmentada em m intervalos não-sobrepostos de $10ms$, correspondentes a caixas de tamanho $\lambda = 110$. O desvio-padrão $V(j)$ é calculado para cada caixa

[2] Gephi é um software gratuito e open-source que realiza a visualização e a exploração de todos os tipos de grafos e redes, disponível em https://gephi.org/

[3] WEKA (*Waikato Environment for Knowledge Analysis*), disponível em https://www.cs.waikato.ac.nz/ml/weka/.

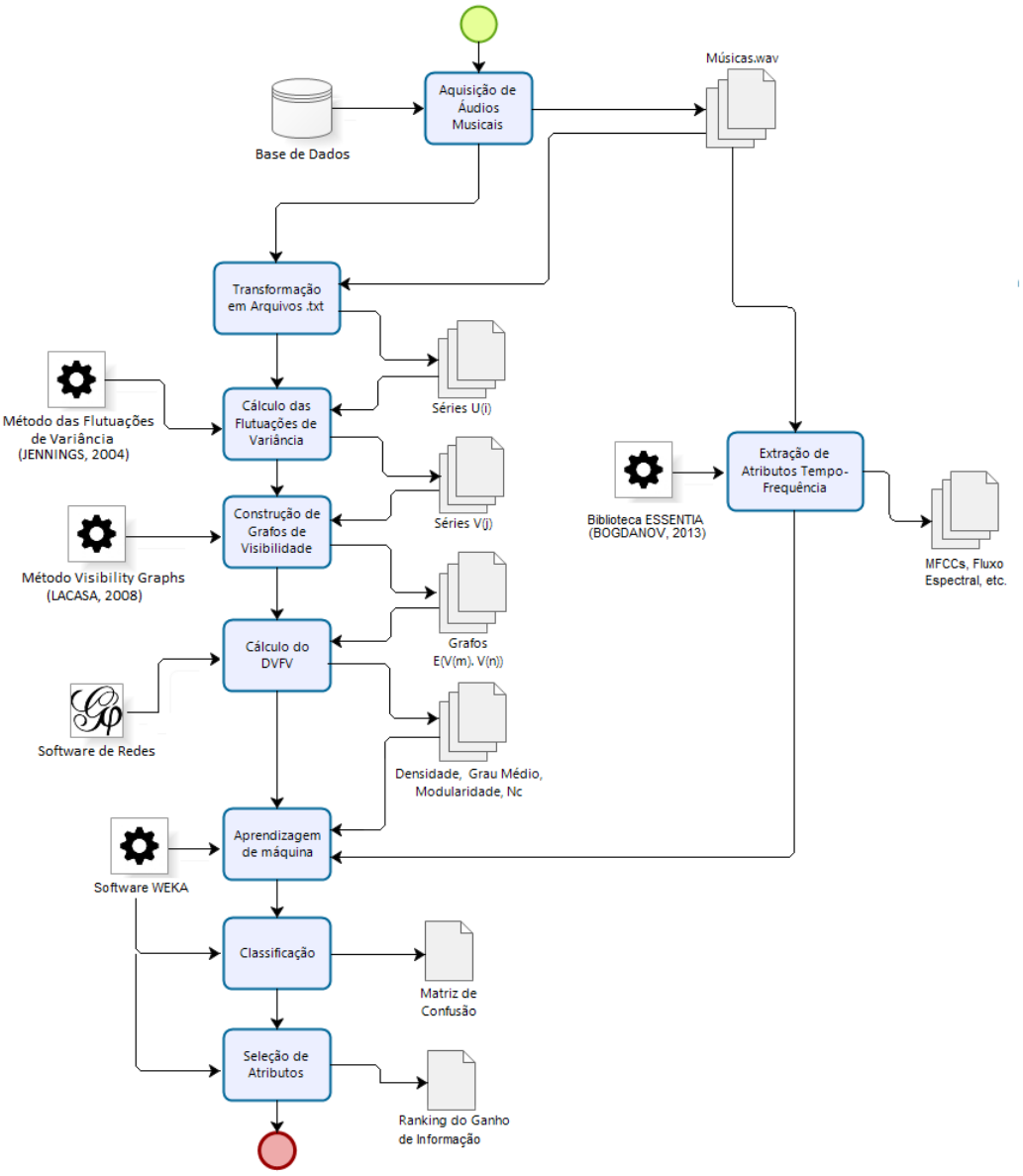


Figura 3.1: Fluxograma da metodologia utilizada na Tese. Fonte: Autor.

$j = 1, \cdots, 3.000$. Na j-ésima caixa temos:

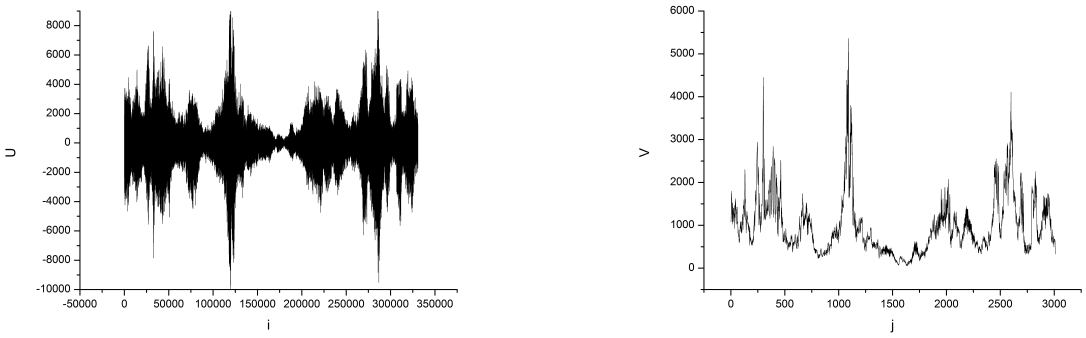
$$V(j) = \sqrt{\frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} (X(i) - \overline{X}_j)^2}{\lambda - 1}}, \tag{14}$$

onde a media é dada por:

$$\overline{X}_j = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} (X(i))}{\lambda}. \tag{16}$$

Com isso cria-se uma sub-série $V(j)$ =$\{V(1), V(2), \cdots, V(m)\}$, com $m = 3.000$ amostras. Nesta tese iremos nos referir à série de desvios-padrão $V(j)$ como série de flutuações de variância.

A Figura 3.2 ilustra a série correspondente a um trecho de 30s de um sinal musical polifônico transformado em uma série $U(i)$ e a sua respectiva série de flutuações de variância V(j).

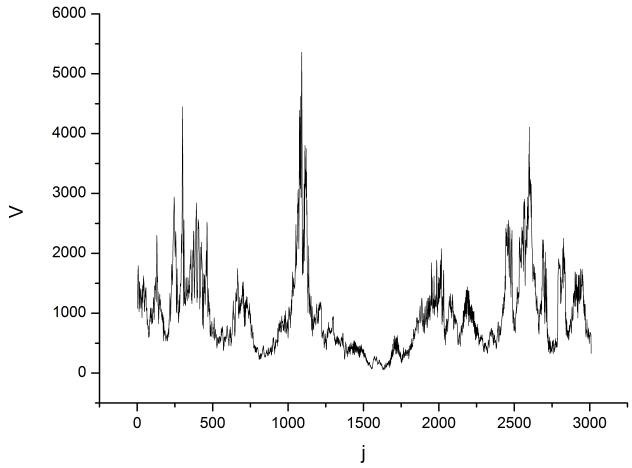


(a) (b)

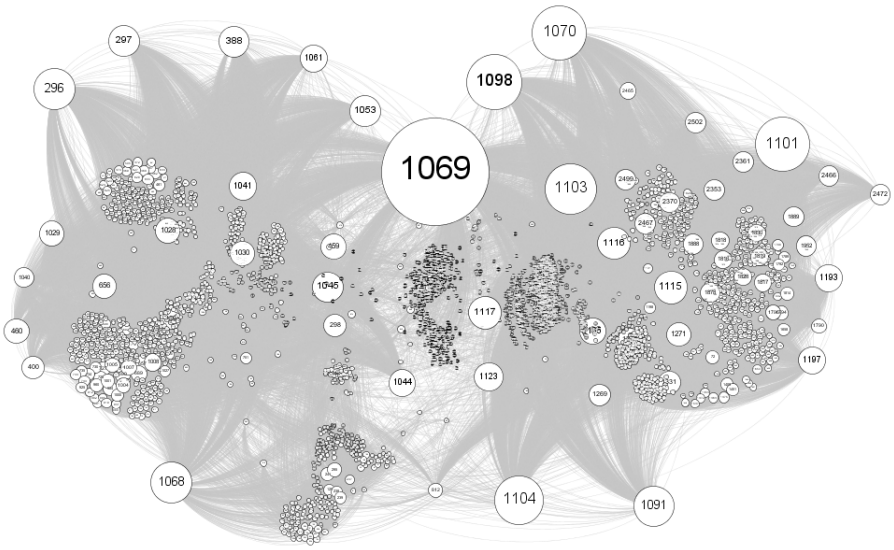
Figura 3.2: (a) Série correspondente à amostragem de 30s da cantata *Ich steh mit einem Fuss im Grabe* BWV 156 de J.S. Bach, (b) Série de flutuações de variância do sinal representado em (a). Fonte: Autor.

3.4 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 toda vez que dois pontos da série $\{V(j)\}$ atendem o critério de visibilidade definido pela Equação 11. Depois que este critério é aplicado para cada um dos pontos da série, obtemos um grafo com padrão hamiltoniano, conectado, e com arestas $E(V(m), V(n))$ onde $m = 1 \cdots 3000$, $n = 1 \cdots 3000$ e $n \neq m$. A Figura 3.3 mostra a série $V(j)$ correspondente à cantata BWV 156 de J.S. Bach e o seu respectivo grafo de visibilidade. Os números que aparecem nos rótulos dos vértices do grafo correspondem à posição j de cada ponto $V(j)$ da série de variâncias, e os tamanhos dos vértices são proporcionais ao respectivo número de arestas incidentes. Podemos notar na Figura 3.3 que o ponto de maior visibilidade da série ($j = 1069$) aparece como o maior vértice do grafo, correspondendo a 2869 arestas.



(a)



(b)

Figura 3.3: (a) Série de flutuações de variância da cantata BWV 156 de J.S. Bach com 3000 pontos, (b) e seu respectivo grafo de visibilidade com 3000 vértices e 54.188 arestas. Fonte: Autor.

3.5 DESCRITOR DE VISIBILIDADE

Nesta seção, apresentaremos quatro propriedades de redes que serão utilizadas sobre o grafo de visibilidade para mensurar a auto-similaridade dos sinais de áudio. A este conjunto

de propriedades, nós denominamos de descritor de visibilidade de flutuações de variância. Este descritor irá capturar a auto-similaridade dos sinais através da visibilidade em torno de seus picos locais, embutida na força das ligações e na qualidade das comunidades determinadas por seus respectivos vértices no grafo de visibilidade. Pensando em termos extremos, um sinal que possui muitos picos persistindo com pouca visibilidade irá gerar grafos com menor grau médio e maior modularidade se comparado a um sinal com picos esparsos e não persistentes. Nestes termos, quanto menor o grau médio e maior a modularidade, maior a auto-similaridade do sinal.

O descritor de visibilidade é constituído das seguintes propriedades: grau médio ($< k >$), densidade (Δ), modularidade (Q), e número de comunidades (Nc). A qualidade das comunidades é calculada pela modularidade e o número de comunidades, e a força das ligações dos picos locais será estimada pelo grau médio e pela densidade.

- **Grau** (k): é a quantidade de arestas de cada vértice da rede.
- **Grau Médio** $(< k >)$: é a média dos graus de todos os vértices que fazem parte da rede. Seja k_i o i-ésimo grau do vértice de uma rede. O grau médio de uma rede com N vértices é a média aritmética dos k_i

$$< k >= \frac{1}{N} \times \sum_{i=1}^{N} k_i \tag{18}$$

- **Densidade** (Δ) : Seja n o número de vértices de uma rede. A densidade é a razão entre o número total de arestas e o número máximo de arestas de uma rede.

$$\Delta = \frac{2 \times E}{n(n-1)} \tag{20}$$

- **Modularidade** (Q) A modularidade é uma medida de estrutura das redes. Esta medida foi projetada para estimar a força da divisão de uma rede em módulos (ou comunidades). Redes com alta modularidade têm conexões densas entre os vértices dentro dos módulos, mas ligações esparsas entre vértices em diferentes módulos [Newman e Girvan 2004]. Um valor alto de modularidade indica que a densidade dos *links* dentro das comumidades é maior que o esperado ao acaso, indicando uma boa partição da rede. Segundo Newman e Girvan 2004, a modularidade é definida por:

$$Q = \frac{1}{2m} \sum_{(i,j)} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{22}$$

onde i e j são os vertices da rede; A_{ij} representa o número de arestas entre i e j; k_i e k_j são a soma das arestas ligadas a i and j; m é a soma de todas as arestas da rede. $\delta(c_i, c_j)$ é a função delta de Kronecker (Equação 24); onde c_i e c_j são as comunidades dos vértices.

$$\delta(c_i, c_j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{24}$$

O cálculo da modularidade pode ser interpretado como a comparação entre a densidade das conexões de um dado conjunto de vértices e a densidade das conexões sobre o mesmo conjunto tomados aletoriamente. Os desvios sistemáticos calculados pela Equação 22 nos permitem definir a qualidade das partições da rede. Partindo do pressuposto que as redes aleatórias não exibem estruturas de comunidades, quanto menor o desvio entre a densidade intrínseca da rede e a densidade tomada aleatóriamente, menor a modularidade da rede.

A maximização da modularidade é feita através do método Louvain [Blondel et al. 2008]. Esse método usa dois estágios que se repetem iterativamente da seguinte maneira: (1) inicialmente cada vértice é considerado como uma comunidade. O ganho de modularidade é calculado para cada vértice i removendo este vértice de sua comunidade nativa C para as comunidades de sua vizinhança. Ao final da visita à sua vizinhança, o vértice i irá de integrar àquela comunidade em que adquirir maior ganho de modularidade. Caso não haja ganho ele permanece na comunidade inicial; (2) cada nova comunidade será considerada um vértice, e uma nova rede é griada com esses "super-vértices". Essas etapas são repetidas até que a máxima modularidade seja alcançada, e uma hierarquia de comunidade seja produzida. O ganho de modularidade é calculado pela Equação 26 e a Figura 3.4 ilustra as etapas do processo maximização.

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \tag{26}$$

onde $\sum_{in}$ é a soma das arestas dentro da comunidade C; $\sum_{tot}$ é a soma das arestas incidentes nos vértices em C; $k_{i,in}$ é a soma das arestas de que ligam i aos vértices em C; k_i é a soma das arestas que incidem no vértice i; m é a soma das arestas que ligam o vértice i aos vértices em C, e m é a soma de todas as arestas da rede.

Existem vários algoritmos de detecção de comunidades, a exemplo de infomap, que

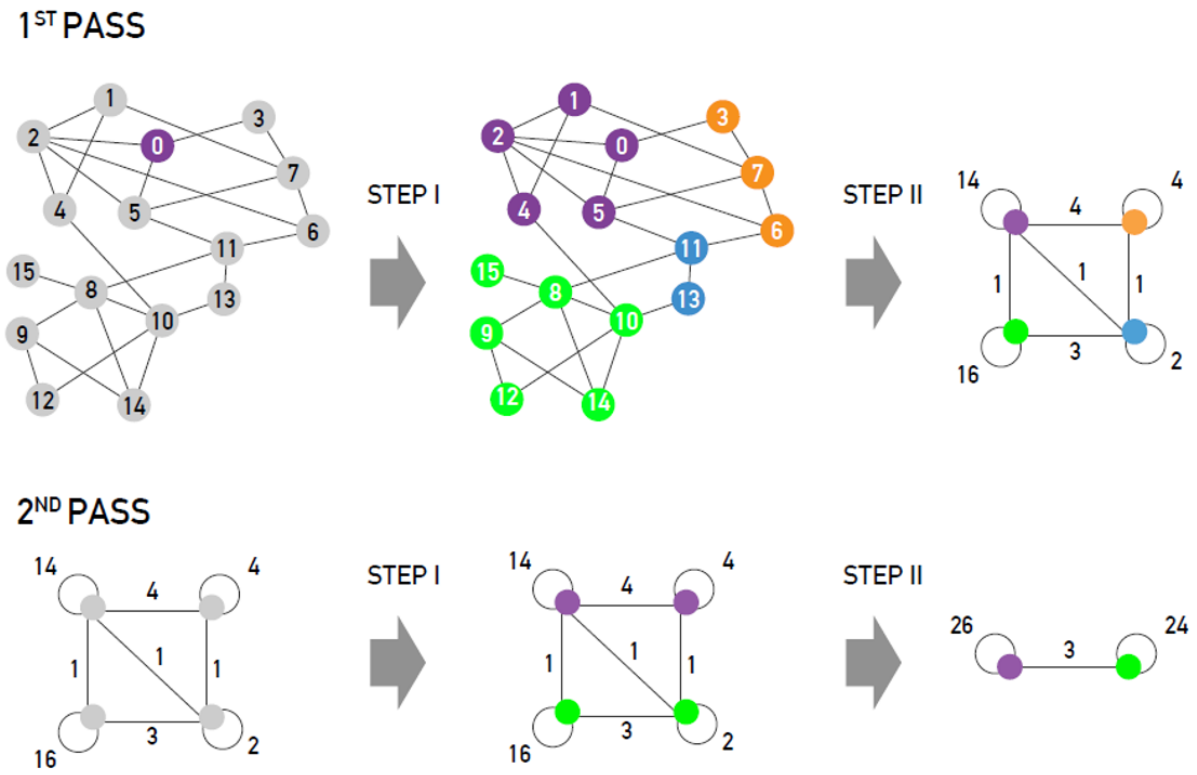


Figura 3.4: Estágios para a maximização da modularidade. Fonte: Blondel et al. 2008.

usa a minimização de uma grandeza de fluxo e tem complexidade computacional de O(LlogL); o algoritmo guloso (*greedy algorithm*), com complexidade computacional O(N^2); e o algoritmo Blondel (Método Louvain), com complexidade O(L). Os dois últimos realizam a maximização da modularidade [Barabási 2016]. A escolha do método Louvain deveu-se a dois fatores: a eficiência, em termos de tempo, e ao fato de estar implementado no Gephi, permitindo uma fácil utilização por parte de pesquisadores que queiram replicar o experimento.

3.6 DESCRITORES TEMPO-FREQUÊNCIA

- **Taxa de Passagem pelo Zero** quantifica o número de cruzamentos pelo zero do sinal no domínio do tempo, e está associada à distância entre dois picos ou vales. É um atributo que mostra em que medida amostras sucessivas possuem sinais diferentes.

$$Z_t = \frac{1}{n-1}\sum_{i=2}^{n} |sign(x_i) - sign(x_{i-1})| \tag{28}$$

onde,

$$sign(x_i) = \begin{cases} 1, & \text{se } x_i \geq 0, \\ 0, & \text{se } x_i < 0. \end{cases} \tag{29}$$

A Equação 29 é a função sinal, cujo valor é 1 para argumentos positivos e 0 para negativos, e x_i é o domínio temporal do sinal por quadro t.

Essa taxa pode ser usada para estimar a complexidade do sinal. Altos valores indicam a presença de muitos picos e vales, e portanto uma alta complexidade. Valores mais baixos indicam uma baixa complexidade e um "sinal mais comportado" [Silva 2014]. A taxa de passagem pelo zero também mostra uma sensibilidade especial para sons vocais e percussivos, e valores altos dessa taxa também podem indicar maior presença desses sons nos sinais.

- **Expoente DFA (α_{DFA})** é um coeficiente que mede o nível de correlações de longo prazo de uma série temporal usando a Análise de Flutuações Destendenciadas (*Detrended Fluctuation Analysis* - DFA), proposta por PENG 1994. Uma adaptação do DFA para o estudo de sinais de áudio musicais foi proposta por Jennings et al. 2004, onde um coeficiente é apresentado para calcular os desvios da lei de potência em caixas de vários tamanhos. Com este coeficiente, gêneros como o *Techno Dance* estão associados a valores baixos do Expoente DFA, como conseqüência de uma baixa correlação de longo alcance na série temporal de seu sinal de áudio. Por outro lado, gêneros musicais como *Hindustani* [4] e *New Age* [5], apresentam um alto valor do coeficiente DFA, como reflexo de uma alta correlação de longo alcance em suas séries. O Expoente DFA também é conhecido como *danceability* [Streich e Herrera 2005]. A seguir, descrevemos o procedimento usado por Jennings et al. 2004 para implementar a função α_{DFA}, seguindo os seguintes passos:

  (i) Determinação da série $V(j)$, a partir da série $U(i)$, onde V é o desvio-padrão. O *loudness* sonoro está relacionado com a variância $(V(j)^2)$ do sinal musical Jennings et al. 2004;

  (ii) Integração de $V(j)$, gerando a série $Y(j)$;

  (iii) Regressão Linear da série $Y(j)$;

  (iv) Cálculo da Função DFA;

  (v) Cálculo do Expoente DFA.

[4] Estilo de música clássica indiana tradicionalmente encontrada no norte da Índia https://prabhatsamgiita.org/estudo-musical/a-musica-hindustani-e-seus-estilos.

[5] Gênero musical nascido de uma estética que visa induzir uma sensação de calma interior, muito utilizada nos campos meditativo e holístico https://www.allmusic.com/genre/new-age-ma0000002745.

Conforme descrito em Jennings et al. 2004: A série $U(i)$ representa o sinal de áudio, com $i = 1 \cdots N$. O número total de pontos N é uma função

$$N(t) = S_r \cdot t, \tag{30}$$

onde S_r é o *sample rate*, cujo valor é aproximadamente 11 Khz e o tempo t é dado em segundos.

Em primeiro lugar o conjunto $\{U(i)\} = \{U(1), ..., U(N)\}$ é segmentado em m blocos ou caixas não-sobrepostas de tamanho λ. Nesta pesquisa, foi adotado um tamanho de caixa λ igual 110 *samples*, que está associado a um tempo de $10ms$. Deste modo cada caixa j é formada por um subconjunto de $U(i)$, tomando-se $\{U(110(j-1)+1), ..., U(110j)\}$, com j variando de 1 a m e i variando de $110(j-1)+1$ a $110j$. Escrevendo de forma mais detalhada temos,

$$\text{caixa } j = 1\text{: } \{U(1), U(2), \cdots, U(110)\},$$
$$\text{caixa } j = 2\text{: } \{U(111), U(112), \cdots, U(220)\},$$
$$\cdots,$$
$$\text{caixa } j = m\text{: } \{U(110(m-1)+1), \cdots, U(110m)\}, \text{ onde } N = 110m.$$

Para cada caixa $j = 1 \cdots m$ é calculado o desvio-padrão

$$V(1) = \sqrt{\frac{\sum_{1}^{\lambda}(U(i)-\overline{U}_1)^2}{\lambda-1}}$$
$$V(2) = \sqrt{\frac{\sum_{\lambda+1}^{2\lambda}(U(i)-\overline{U}_2)^2}{\lambda-1}}$$
$$\cdots$$
$$V(m) = \sqrt{\frac{\sum_{(m-1)\lambda+1}^{m\lambda}(U(i)-\overline{U}_m)^2}{\lambda-1}}$$

Então para a j-ésima caixa temos:

$$V(j) = \sqrt{\frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda}(U(i)-\overline{U}_j)^2}{\lambda-1}}, \tag{32}$$

onde a média é dada por

$$\overline{U}_j = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda}(U(i))}{\lambda} \tag{34}$$

Com isso, criamos uma nova série $V(j) = \{V(1), V(2), \cdots, V(m)\}$, com N/λ amostras. Esse conjunto se assemelha a séries temporais não-estacionárias limitadas (*bounded time series*) que têm relação com a intensidade média do som em cada bloco. Uma vez que esse tipo de série não revela tendências através do DFA, efetuamos então a integração de $V(j)$, com o objetivo de criar uma série temporal ilimitada (*unbounded time series*)

$$Y(m) = \sum_{j=1}^{m} V(j), \tag{36}$$

O processo de integração é fundamental na computação do processo DFA. Em séries limitadas, o Expoente DFA seria sempre igual a zero quando temos escalas de tempo de grande porte.

A série gerada por $Y(m)$ é também dividida em subséries sobrepostas de tamanho τ. Cada subsérie é deslocada de uma única amostra em relação à subsequência anterior.

Para cada bloco de comprimento τ é removida a tendência linear $\hat{y}_k$, através da regressão

$$\hat{y}_k = b_1(x) + b_0, \tag{38}$$

onde,

$$b_0 = \overline{y} - b_1\overline{x} = \frac{1}{\tau} \cdot (\sum(y_i) - b_1(\sum(x_i))) \tag{40}$$

e

$$b_1 = \frac{\sum(x_i - \overline{x})(y_1 - \overline{y})}{\sum(x_i - \overline{x})^2}. \tag{42}$$

Em seguida, é computada a média do quadrado residual para cada bloco k,

$$D(k, \tau) = \frac{1}{\tau} \cdot \sum_{m=0}^{\tau-1} (y(k+m) - \hat{y}_k(m))^2 \tag{44}$$

Finalmente a flutuação DFA é dada pela raiz quadrada da média dos $D(k, \tau)$ em todos os K blocos:

$$F(\tau) = \sqrt{\frac{1}{K} \sum_{i=1}^{k} (D(k, \tau))} \tag{46}$$

A flutuação DFA é uma função de τ, ou seja, é função da variável de tempo em foco, denominada de janela ou intervalo de interesse. O objetivo do DFA é revelar propriedades de correlação em diferentes escalas de tempo. O processo então é repetido sobre diferentes valores de τ inseridos na janela de interesse. Essas escalas de tempo estão relacionadas com o sinal musical, abrangendo desde pulsações de alto nível até padrões rítmicos mais simples [Streich e Herrera 2005]. Deste modo a janela de interesse irá computar variações de *loudness* dentro de janelas temporais que têm relação com a atividade rítmica do sinal estudado. Nesta tese foram utlizadas janelas de interesse com tamanhos de τ de 31 a 909, que correspondem a intervalos de tempo 310 ms a 9,09 s, segundo a configuração adotada em Melo 2013.

O Expoente DFA é definido como a inclinação em um gráfico log x log de F sobre τ. Para calcular o Expoente DFA, primeiramente foi adotada a Equação 48, que representa a taxa de variação de F em função de t para cada valor de i. No cálculo de α para valores pequenos de τ é necessário um ajuste no denominador. A proporção que τ cresce, a influencia da correção torna-se desprezível [Streich e Herrera 2005].

$$\alpha(i) = \frac{\log_{10} F(\tau_{i+1}) - \log_{10} F(\tau_i)}{\log_{10}(\tau_{i+1} + 3) - \log_{10}(\tau_i + 3)} \tag{48}$$

Finalmente, o Expoente DFA é computado pela Equação 50.

$$\alpha_{DFA} = \frac{\sum_{i=1}^{32} \alpha(i)}{32} \tag{50}$$

- **Complexidade da Dinâmica** é calculada pelo desvio médio absoluto da estimativa geral do nível de *loudness* na escala de decibéis (dB). Esta medida está relacionada ao alcance dinâmico e à quantidade de flutuação na intensidade presente em uma gravação (Figura 3.5). Streich et al. 2006 toma como base Vickers 2001 para definir a complexidade da dinâmica da seguinte forma:

$$C_{dyn} = \frac{1}{M} \sum_{i=0}^{M-1} |V_{dB}(i) - L|, \tag{52}$$

onde o *loudness* global L é uma média ponderada de todos os M níveis instantâneamente estimados por

$$L = \sum_{i=0}^{M-1} w(i).V_{dB}(i), \tag{54}$$

e o *loudness* instantâneo convertido em dB é

$$V_{dB}(i) = 20.log_{10}(V_{rms}(i)), \tag{56}$$

$$w(i) = \frac{u(i)}{\sum_{i=0}^{M-1} u(j)}, \tag{58}$$

e

$$u(j) = 0.9^{-V_{dB}(j)}. \tag{60}$$

Este algoritmo traz uma estimativa da complexidade simplificada, ou seja, a rapidez das mudanças e a variação periódica de volume não são explicitamente levadas em consideração, antes, o algoritmo usa como indicador apenas a distância média do volume global. Deste modo, uma alta complexidade corresponde a grandes distâncias, e vice-versa. A Figura 3.5 mostra como a complexidade da dinâmica varia em diferentes estilos musicais.

- **Onset Rate** calcula o número de *onsets* por segundo em um trecho de áudio. A taxa de *onsets* é baseada em um método conhecido como função de Conteúdo de Alta Freqüência (HFC), cuja equação é mostrada abaixo:

$$D_H[n] = \sum_{k=0}^{N} k|X_k[n]|, \tag{62}$$

onde $k|X_k[n]$ é o valor do k^{th} bin de $X[n]$ (STFT do sinal x no tempo n). De acordo com Brossier et al. 2004, o HFC é mais bem sucedido na detecção de *onsets* de percussivos do que não-percussivos, como os emitidos por violinos e flautas.

- **Fluxo Espectral** calcula a quantidade de mudanças no espectro no decorrer do tempo através da diferença entre duas amplitudes normalizadas de sucessivas distribuições es-

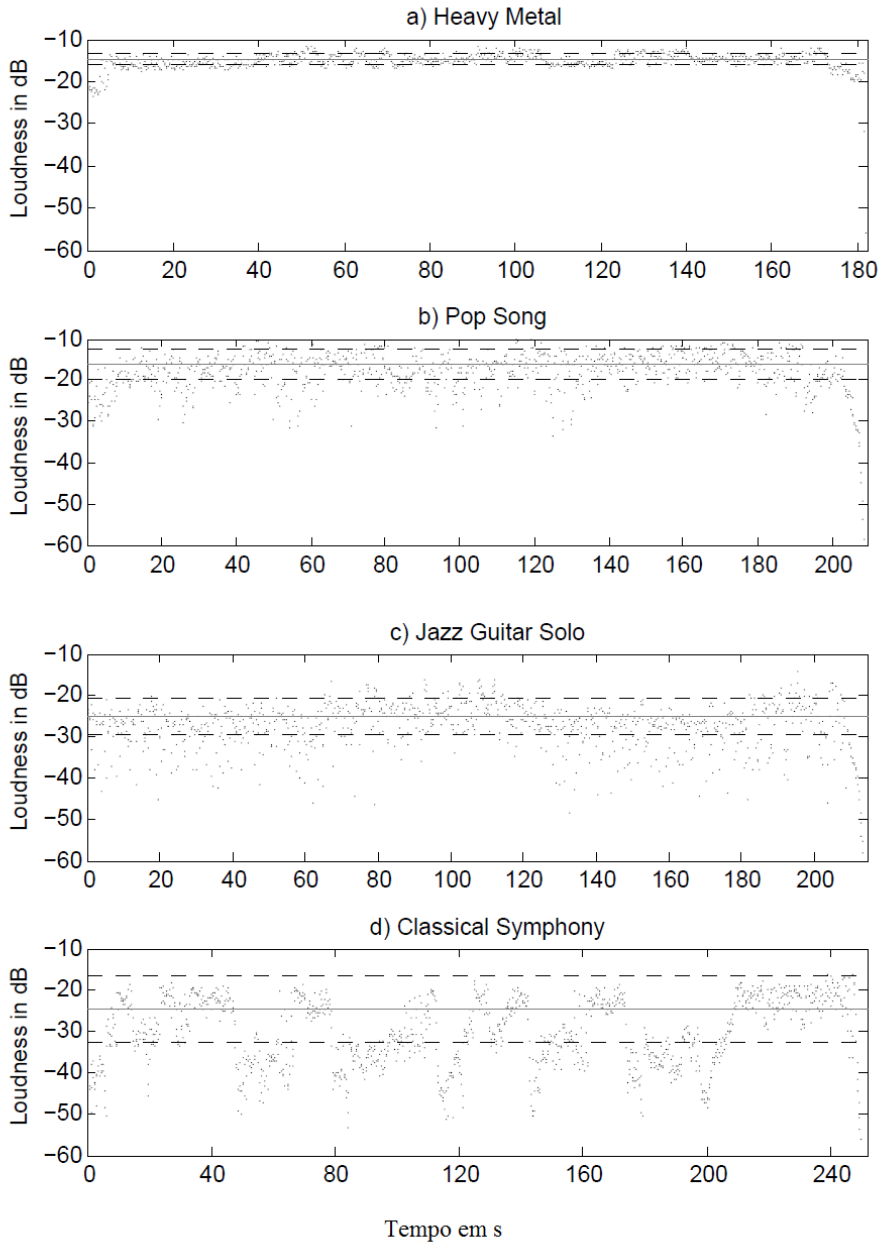


Figura 3.5: *Loudness* instantâneo - Equação 56 - (pontos), *loudness* global - Equação 54 - (linha sólida cinza), e margem de distância média do nível de *loudness* global - Equação 52 (linhas trace-jadas) para quatro sinais de áudio.

pectrais. O fluxo espectral dá uma estimativa da rapidez com que a magnitude das componentes de frequência varia. Ela é definida pela Equação 64.

$$SF(t) = \sum_{n=1}^{N}(|Y_t[n] - Y_{t-1}[n]|)^2 \tag{64}$$

onde $Y_t[n]$ é o valor normalizado da transformada de Fourier na janela *t*.

É um atributo que é sensível às variações na música, podendo detectar se uma música apresenta mudanças bruscas. Essas mudanças bruscas podem ser influenciadas pela atividade percussiva presente no arranjo musical. O fluxo espectral tem sido aplicado em muitos trabalhos como função de detecção de *onsets* em sinais musicais [Dixon 2006, Bello et al. 2005, Jensen e Andersen 2003].

- **Loudness** é definido como uma entidade relacionada à percepção do som, enquanto a intensidade é uma entidade relacionada à característica física do som, cuja magnitude pode ser medida numericamente. Portanto, em seu sentido essencial, o loudness tem uma natureza subjetiva. Stevens 1957 propôs uma maneira de quantificar o loudness ao estabelecer uma relação entre a sensação de percepção do som e a intensidade do som através da Equação 66)

$$\psi(I) = kI^{0.67}, \tag{66}$$

onde $\psi(I)$ representa a magnitude da sensação subjetiva dada pelo estímulo do som, I é a magnitude do estímulo físico, α é o expoente do estímulo dado por uma pressão sonora de 3.000 Hz, e k é uma constante de proporcionalidade que depende das unidades utilizadas.

- **Batimentos Por Minuto (BPM)** é a média dos valores BPM mais salientes que representam periodicidades no sinal (o BPM médio). O conjunto de características para representar a estrutura do ritmo baseia-se na detecção das periodicidades mais salientes do sinal. O sinal é primeiro decomposto em uma série de frequência de bandas de oitava usando a transformada discreta wavelet (DWT). Após esta decomposição, o envelope de amplitude do domínio do tempo de cada banda é extraído separadamente. Isto é conseguido aplicando a retificação de onda completa, filtragem de passa-baixa e decimação (*downsampling*) para cada banda de frequência de oitava. Após a remoção média, as envoltórias (envelopes) de cada banda são então somados e a autocorrelação da envoltória

soma resultante é calculada. Os picos dominantes da função de autocorrelação correspondem às várias periodicidades da envoltória do sinal. Estes picos são acumulados em todo o arquivo de som em um histograma de batidas, onde cada compartimento corresponde ao intervalo de pico, ou seja, o período de batida em batimentos por minuto (bpm) [Tzanetakis e Cook 2002].

- **Mel Frequency Cepstral Coefficients (MFCCs)** Os coeficientes cepstral de frequência de Mel (MFCC) são atributos motivados perceptivamente que também são baseados na transformada de tempo curto de Fourier (STFT). Depois de tomar a amplitude logarítmica do espectro de magnitude, as caixas da transformada rápida de Fourier (FFT) são agrupadas e alisadas de acordo com a escala de frequência Mel com percepção motivada. Finalmente, para descorrelacionar os vetores de características resultantes é realizada uma transformada discreta de cosseno. Nesta tese, são utilizados 13 coeficientes MFCC [Tzanetakis e Cook 2002].

3.7 ALGORITMOS DE APRENDIZAGEM E CLASSIFICAÇÃO

Nesta seção, abordaremos os algoritmos da plataforma WEKA versão 3.16.13, utilizados na aprendizagem de máquina e classificação de gêneros musicais desta tese. O WEKA é uma plataforma livre desenvolvida pela universidade de Waikato, e dispõe de uma coleção de algoritmos de aprendizagem de máquina escritos em JAVA para tarefas de mineração de dados. O WEKA conta também com ferramentas para pré-processamento, classificação, regressão, agrupamento, regras de associação, seleção de atributos e visualização de dados [Hall et al. 2009]. No WEKA, é possível escolher qual o melhor resultado entre vários algoritmos tradicionais na literatura, dentre eles foram selecionados para esta tese: o J48, o *Naive Bayes*, o *Multiclasss classifier*, o IBK, e o *k-Star*.

- **O Algoritmo J48** - realiza a classificação a partir de árvores de decisão. Podemos definir uma árvore de decisão como uma estrutura simples constituída de vértices e folhas. Os vértices são bifurcações onde são tomadas decisões binárias - é neste ponto que os atributos são testados. As folhas são os locais onde aparecem os resultados das decisões. A Figura 3.6 mostra uma árvore de decisão usando o J48 para a classificação de 60 músicas igualmente distribuídas segundo as classes Classical, Jazz e Hiphop, usando a modularidade (Q), e a densidade (Δ) como atributos de classificação. As árvores de decisão se consagraram como uma técnica muito eficiente e muito utilizada em problemas de classificação. Um dos motivos para que esta técnica seja bastante utilizada é a facilidade de

entender sua estrutura de regras. Árvores de decisão utilizam a abordagem dividir-para-conquistar, ou seja decompõem um problema complexo em subproblemas mais simples, aplicam repetidamente a mesma estratégia a cada subproblema, dividem o espaço definido pelos atributos em subespaços, e finalmente associam a eles uma classe. A maioria dos algoritmos computacionais baseados em árvores de decisão foram desenvolvidos com base na árvore ID3 [Quinlan 1986] e no seu sucessor a árvore C4.5 [Quinlan 1993].

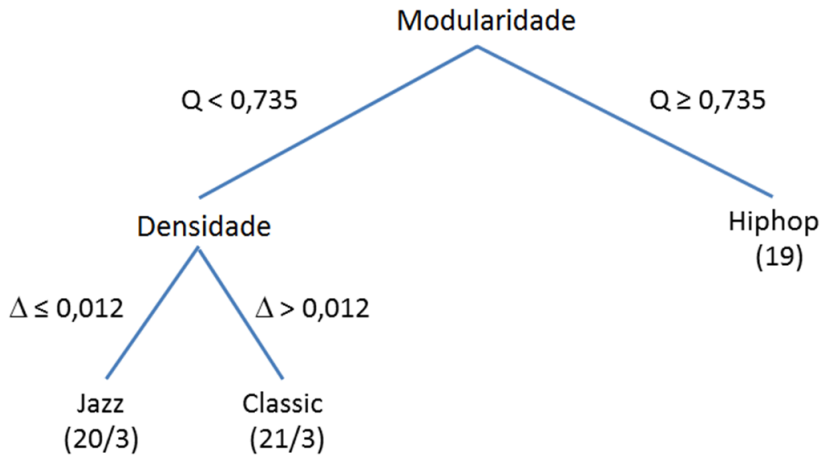


Figura 3.6: Árvore de decisão J4.8 usando o modo 10-*fold crosss validation*. Fonte: Autor.

- ***Naive Bayes*** - é um algoritmo baseado no teorema de Bayes. O termo Naïve (ingênuo) foi adotado no nome porque o algoritmo considera que os atributos são independentes,ou seja a ocorrência de um evento não influencia a ocorrência do outro. O termo Bayes vem de uma técnica estatística (probabilidade condicional) baseada no teorema de Thomas Bayes. Segundo o teorema de Bayes, é possível encontrar a probabilidade de um certo evento ocorrer, dada a probabilidade de um outro evento que já ocorreu (Equação 68). Em outras palavras, um classificador de Bayes ingênuo assume que a presença ou ausência de uma característica particular, não está relacionada com a presença ou ausência de qualquer outro elemento, tendo em conta a classe variável [Duda et al. 2012].

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \tag{68}$$

onde $P(B|A)$ é a probabilidade de B dado A, $P(A)$ é a probabilidade de A, e $P(A \cap B)$ é a probabilidade da ocorrência simultânea de A e B.

- ***Multiclasss classifier***

Sistemas de classificação envolvendo mais de duas categorias ou classes são denominados

de sistemas multiclasse. Uma das formas de resolver problemas de classificação multiclasse é transformá-las em múltiplos sistemas binários ou de duas classes. O *Multiclasss classifier* é um algoritmo de meta-aprendizagem do WEKA que transforma um sistema multiclasse em vários sistemas binários, usando um dos seguintes métodos:

1. Um - contra - todos (*One-versus-rest*) [Baeza-Yates e Ribeiro-Neto 2013]:

2. Classificação em pares usando votação para prever (*Pairwise classification*) [Park e Fürnkranz 2007].

3. Códigos de correção de erros exaustivos (*Exhaustive error-correcting codes*) [Witten et al. 2016]

4. Códigos de correção de erros selecionados aleatoriamente (*Randomly selected error-correcting codes*) [Dietterich e Bakiri 1995]

Witten et al. 2016 descreve a transformação de um sistema multiclasse em um conjunto sistemas binários da seguinte maneira:

> Para cada classe, um conjunto de dados é gerado contendo um cópia de cada instância nos dados originais, mas com um valor de classe modificado. Se o instância tem a classe associada ao conjunto de dados correspondente é marcada "sim"; caso contrário, "não". Em seguida, são criados classificadores para cada um desses conjuntos de dados binários, classificadores esses que produzem um intervalo de confiança em suas previsões - por exemplo, a probabilidade estimada de que a classe é "sim". Durante a classificação, uma instância de teste é alimentada em cada classificador binário, e a classe final é aquela associada ao classificador que prevê o "sim" com mais confiança. Claro, esse método é sensível à precisão dos intervalos de confiança produzidos pelos classificadores: se alguns os classificadores têm uma opinião exagerada de suas próprias previsões, o resultado sofrerá.

O seguinte exemplo também é dado por Witten et al. 2016, ilustra a tranformação multiclasse para binária:

> Suponha que ele pertence a classe a, e que as previsões dos classificadores individuais são 1 0 1 1 1 1 1 (respectivamente). Obviamente, comparando esta palavra de código com as da Figura 3.7 (b), O segundo classificador cometeu um erro: ele previu 0 em vez de 1 (não no lugar de sim). No entanto, comparando os bits previstos com a palavra de código associada com cada classe, a instância está claramente mais próxima de a do que em qualquer outra classe. Isso pode ser quantificado pelo número de bits que devem ser alterados para converter a palavra-código predita em uma das palavras da Figura 3.7 (b): a distância de Hamming ou a discrepância entre os strings de bits, é 1, 3, 3 e 5 para as classes a, b, c e d, respectivamente. Podemos concluir com segurança que o segundo classificador cometeu um erro e identificar corretamente a como a classe verdadeira da instância.

Transforming a multiclass problem into a two-class one: (a) standard method and (b) error-correcting code.

Class	Class vector	Class	Class vector
a	1 0 0 0	a	1 1 1 1 1 1 1
b	0 1 0 0	b	0 0 0 0 1 1 1
c	0 0 1 0	c	0 0 1 1 0 0 1
d	0 0 0 1	d	0 1 0 1 0 1 0
(a)		(b)	

Figura 3.7: Transformação de um sistema multiclasse em um sistema de duas classes. Fonte: [Witten et al. 2016].

- **IBk** - é a implementação do WEKA para o classificador de vizinhos mais próximos conhecida como $k - NN$. O Nearest Neighbor (NN) é um algoritmo de classificação largamente utilizado no estudo de reconhecimento de padrões. O seu princípio de funcionamento consiste em determinar o vizinho mais próximo de uma determinada instância [Witten et al. 1999]. O $k - NN$ (k-Nearest Neighbor) é um algoritmo que realiza a classificação de um dado elemento em concordância com as respectivas classes dos k ($k > 1$) vizinhos mais próximos. O algoritmo calcula a distância de determinado elemento i para cada elemento j da base e então estabelece uma ordem aos elementos desta base, partindo do mais próximo para o mais distante. Dentre os elementos ordenados, apenas os k primeiros são escolhidos [Martins et al.]. Esses k elementos irão servir de parâmetro para a regra de classificação. Por exemplo, no algoritmo $k - NN$ onde $k = 7$, serão utilizados os sete elementos mais próximos da instância e com base nas classes destes sete elementos, será definida a classe do elemento de teste [Martins et al.]. O IBk utiliza, por padrão, $k = 1$, significando que apenas o elemento do treinamento mais próximo da instância será selecionado para classificação.

- ***k-Star***

  Semelhantemente ao IBk, o *k-Star* também é um algoritmo baseado em intâncias que utiliza a técnica dos vizinhos mais próximos ($k - NN$). A diferença entre eles é que o *k-Star* utiliza o conceito de entropia para definir sua métrica de distância. A entropia é calculada através da complexidade da transformação de uma instância em outra. O cálculo da complexidade é feita em duas etapas: primeiro um conjunto finito de transformações que mapeiam instâncias para instâncias é definido, depois é calculada a distância de *Kolmogorov*. Essa distância consiste no comprimento da cadeia mais curta que liga duas instâncias.

> Esta abordagem concentra-se em uma única transformação (a mais curta), de muitas possíveis transformações. O resultado é uma medida de distância que é muito sensível a pequenas mudanças no espaço da instância e que não resolve bem o problema de suavidade. A distância k* definida abaixo tenta lidar com esse problema, somando todas as possíveis transformações entre duas instâncias. [Cleary e Trigg 1995]

3.8 SELEÇÃO DE ATRIBUTOS E GANHO DE INFORMAÇÃO

Uma das formas de mensurar a importância relativa dos atributos dentro de um sistema de classificação é através do cálculo do ganho de informação de cada um desses atributos em uma estrutura baseada em árvore de decisão. Nesta seção nós mostramos um ranking onde os atributos usados na Seção 4.3.2 são ordenados de acordo com o ganho estimado em função da entropia da informação em cada nó de uma árvore C4.5.

Neste trabalho, nós usamos o algoritmo $Ranker{+}GainRatioAttributeEval$ do WEKA 3.6.9, que identifica os atributos mais significativos em uma árvore J4.8 (versão WEKA da renomada C4.5). Antes de apresentarmos os resultados, iremos mostrar brevemente os fundamentos do cálculo de seleção de atributos usando o algoritmo $GainRatio$ segundo a abordagem de Karegowda et al. 2010.

Podemos definir uma árvore de decisão como uma estrutura simples constituída de nós e folhas. Os vértices são bifurcações onde são tomadas decisões binárias - é neste ponto que os atributos são testados. As folhas são os locais onde aparecem os resultados das decisões. O ganho de informação é usado para medir e selecionar o atributo de teste em cada nó da árvore. A informação esperada necessária para classificar uma determinada amostra s de um conjunto S, é dada por:

$$I\left(S\right) = -\sum_{i=1}^{m} p_i \times log_2\left(p_i\right), \tag{70}$$

onde p_i é a probabilidade de uma amostra arbitrária pertencer à classe C_i e é estimado por s_i/s.

A entropia, ou a informação esperada do atributo A com v valores distintos, é dada por

$$E\left(A\right) = -\sum_{i=1}^{m} I\left(S\right) \frac{s_{1i} + s_{2i} + \cdots + s_{mi}}{s}, \tag{72}$$

onde s_{ij} é o número de amostras da classe C_i em um subconjunto $S_j.S_j$.

O ganho de informação do atributo A por nó é calculado por

$$Gain\left(A\right) = I\left(S\right) - E\left(A\right). \tag{74}$$

A árvore J4.8 usa uma taxa de ganho normalizada que é dada pela Equação 76.

$$SplitInfo_A\left(S\right) = -\sum_{i=1}^{v}\left(\left|S_i\right| / \left|S\right|\right) log_2\left(\left|S_i\right| / \left|S\right|\right) \tag{76}$$

O valor da Equação 76 representa a informação gerada através da divisão do conjunto de dados de treinamento S em v partições correspondentes a v resultados de um teste no atributo A. A taxa de ganho de informação é definida pela Equação 78.

$$GainRatio\left(A\right) = Gain\left(A\right) / SplitInfo_A\left(S\right) \tag{78}$$

No final o atributo com a maior relação de ganho é selecionado como o atributo de divisão. Os vértices não terminais da árvore gerada são considerados como atributos relevantes [Karegowda et al. 2010].

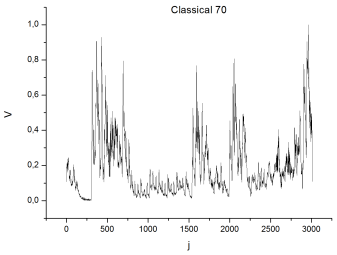
4 RESULTADOS E DISCUSSÃO

Neste capítulo, mostramos como o descritor de visibilidade em flutuações de variância (DVFV) foi utilizado sobre a base de dados para identificar padrões de auto-similaridade em sinais, e a partir dessa informação, hierarquizá-los, classificá-los em categorias musicais, e comparar o desempenho do DVFV em relação a descritores largamente usados na literatura. Na Seção 4.1, apresentamos sinais de quatro gêneros musicais e seus respectivos grafos de visibilidade, e mostramos, através da modelagem gráfica e dos índices do DVFV, que existem diferenças de auto-similaridade entre gêneros musicais associadas à diferenças topológicas indicadas graficamente pela detecção de comunidades. Na Seção 4.2, apresentamos os resultados do cálculo do DVFV para 1.000 redes, e apresentamos o potencial do DVFV para detecção de padrões, e para a organização hierárquica de gêneros musicais em função da auto-similaridade. Em seguida, na Seção 4.3, nós realizamos os processos de aprendizagem de máquina e classificação, e apresentamos os resultados para duas configurações experimentais: (a) utilizando apenas o DVFV como vetor de atributos (Seção 4.3.1); (b) usando o DVFV junto a algoritmos de descrição tempo-frequência usados em processamento de sinais (Seção 4.3.2). Na Seção 4.4, nós calculamos o ganho de informação em um sistema baseado em árvore de decisão usando o DVFV e descritores tempo-frequência, e a partir de um ranking construído com a taxa de ganho, avaliamos a relevância do descritor do DVFV dentro desse sistema. Por fim, na Seção 4.5 comparamos os resultados da nossa proposta com trabalhos correlatos. Na Subseção 4.5.1 comparamos a hierarquia de auto-similaridade usando DVFV com uma hierarquia utilizando DFA. Na Subseção 4.5.2 nós comparamos os resultados da classificação desta tese com vários resultados de classificações usando descritores tempo-frequência.

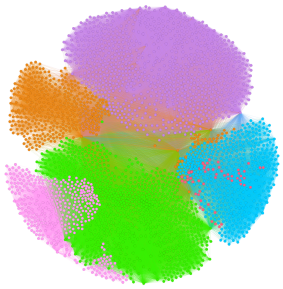
4.1 GRAFOS DE VISIBILIDADE ÁUDIO ASSOCIADOS

Para cada um dos sinais de áudio do banco GTZAN nós calculamos uma série de flutuações de variância e depois nós geramos um grafo para cada série, totalizando 1.000 grafos de visibilidade. A Figura 4.1 apresenta quatro séries V(j) representando músicas de gêneros musicais distintos, e seus respectivos grafos de visibilidade. As diferentes cores nos grafos são

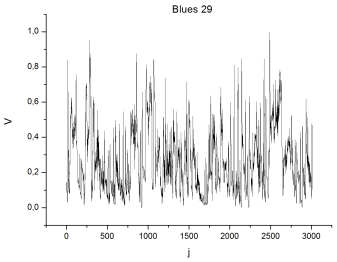
clusters identificadas através do algoritmo de detecção de comunidades - modularidade [Blondel et al. 2008] - implementado no Gephi. Para todos os grafos utilizamos o modo aleatório, que é a opção dada pelo software para a produção de uma melhor decomposição, e a resolução padrão de 1.0 (maiores detalhes em Lambiotte et al. 2008). Também foram utilizadas as mesmos opções de visualização e distribuição de vértices. O critério de parada do algoritmo foi a melhor visualização dos *clusters*. No resultado final podemos observar que os grafos com uma maior quantidade de comunidades (ou classes de modularidade) possuem grafos com nódulos menores e mais espalhados para facilitar a visualização. Através da Figura 4.1 notamos uma correspondência entre características de persistência dos transientes dos sinais, e características topológicas da detecção de comunidades em seus grafos associados. Observamos que a medida que a auto-similaridade aumenta, a modularidade e o número de comunidades também aumentam, enquanto o grau médio e a densidade diminuem. Essa característica sugere que os parâmetros do descritor de visibilidade podem ser usados para hierarquizar um conjunto de sinais em função da auto-similaridade de seus transientes. Para testar a plausibilidade dessa conjectura nós precisamos verificar se esse comportamento se sustenta para um conjunto maior de sinais e gênero musicais. É isso que faremos na Seção 4.2.



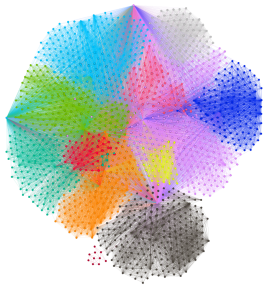
(a) Série Clássico



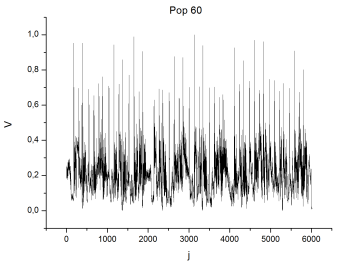
(b) Grafo de Visibilidade Clássico (Q=0,613, Nc=6, $< k >$=64,10, $\Delta(\%) = 2,1$)



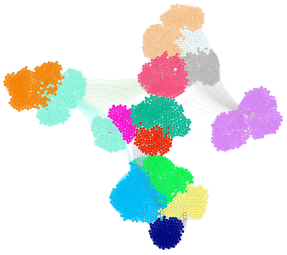
(c) Série Blues



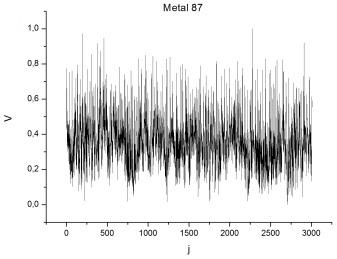
(d) Grafo de visibilidade Blues (Q=0,762, Nc=12, $< k >$=27,71, $\Delta(\%) = 0,9$)



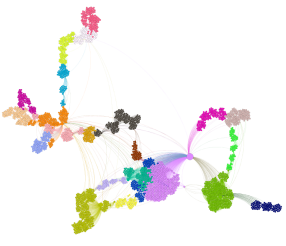
(e) Série Pop



(f) Grafo de visibilidade Pop (Q=0,836, Nc= 14, $< k >$=20,09, $\Delta(\%) = 0,7$)



(g) Série Metal



(h) Grafo de Visibilidade Metal (Q=0,881, Nc=22, $< k >$=10,19, $\Delta(\%) = 0,3$)

Figura 4.1: Séries $V(j)$ (à esquerda) e seus respectivos grafos de visibilidade (à direita). As cores nos grafos são as comunidades, obtidas a partir da modularidade. Fonte: Autor.

4.2 HIERARQUIA SEGUNDO A AUTO-SIMILARIDADE

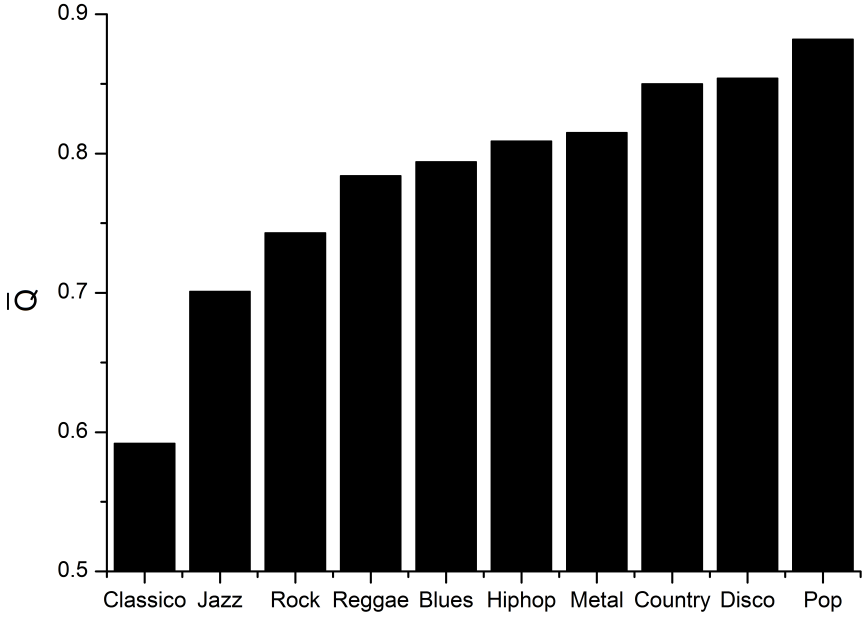
A Tabela 4.1 apresenta a média e o desvio-padrão da modularidade, do número de comunidades, do grau médio, e da densidade, dos grafos de visibilidade correspondentes a 100 amostras de áudio agrupadas em 10 gêneros musicais.

Tabela 4.1: Média e desvio-padrão de propriedades topológicas de grafos de visibilidade. Q (modularidade), Nc (número de comunidades), $< k >$ (grau médio), Δ (densidade). Fonte: Autor.

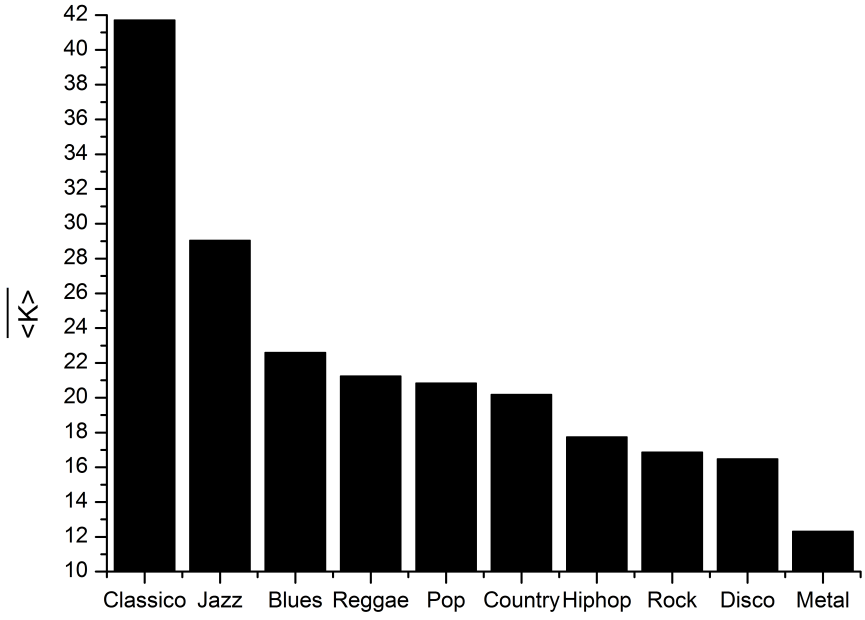
	$\overline{Q}$	σ_Q	$\overline{Nc}$	σ_{Nc}	$\overline{< k >}$	$\sigma_{<k>}$	$\overline{\Delta}(\%)$	σ_Δ
Classical	0,592	0,120	9,57	3,24	41,71	12,89	1,41	0,48
Jazz	0,701	0,083	12,25	3,40	29,04	9,12	0,97	0,37
Blues	0,794	0,068	21,23	3,36	22,60	9,93	0,75	0,34
Reggae	0,784	0,096	13,95	2,75	21,24	5,21	0,70	0,21
Pop	0,882	0,041	15,71	4,75	20,83	5,54	0,69	0,21
Country	0,850	0,056	18,75	3,57	20,18	6,94	0,68	0,26
Hiphop	0,809	0,077	14,61	2,85	17,75	3,93	0,59	0,16
Rock	0,743	0,085	12,49	3,13	16,86	4,33	0,56	0,17
Disco	0,854	0,052	19,29	2,85	16,48	4,5	0,55	0,17
Metal	0,815	0,073	16,31	3,39	12,3	2,7	0,44	0,12

Se colocarmos as médias do grau médio ($\overline{< k >}$) e da densidade ($\overline{\Delta}$) em ordem decrescente, obteremos a mesma sequência de gêneros musicais. Entre as ordens crescentes de $\overline{Q}$ e $\overline{Nc}$ houve diferença no posicionamento apenas para três agrupamentos (Blues, Hiphop e Pop). Para as quatro componentes do DVFV, os valores médios de Clássico, Jazz, Rock, Reggae, e Pop preservaram-se na mesma posição quando colocadas em ordem decrescente para $\overline{< k >}$ e $\overline{\Delta}$, e crescente para $\overline{Q}$ e $\overline{Nc}$. Se usarmos $\overline{Q}$ e $\overline{< k >}$, segundo essa ordem, para pensar uma hierarquia, podemos considerar as Figuras 4.2 (a) e (b) como representações de gêneros musicais em ordem crescente de auto-similaridade. Deste modo, o gênero musical que possui sinais menos auto-similares é o Clássico, basta notar a grande diferença dos valores de $\overline{Q}$ e $\overline{< k >}$ do gênero clássico em relação a todos os outros. Entre os gêneros cujos sinais tem maior auto-similaridade estão o metal, disco, e hiphop. O Jazz é o gênero mais próximo ao Clássico, porém com uma diferença considerável entre eles. Para todos os quatro índices, o Reggae ocupa a posição intermediária. Este tipo de organização hierárquica corrobora com a ideia de que gêneros musicais que optam por arranjos instrumentais muito "densos", "intensos", e "persistentes", possuem sinais com maior auto-similaridade, e tendem a ocupar posições opostas a gêneros musicais com texturas instrumentais mais ricas em dinâmica, e portanto, com menor auto-similaridade em seus sinais. Em posição intermediária estão estilos musicais que buscam o equilíbrio das influências estéticas dos dois extremos. Em muitos casos podem não haver diferenças importantes entre os agrupamentos estabelecidos por Q, $< k >$, Δ, e Nc. Para estudar

esse aspecto, realizamos um teste de hipóteses para comparação de duas médias.



(a) Média da Modularidade



(b) Média do Grau Médio

Figura 4.2: Q (a) e $< k >$ (b) médios calculados a partir de 100 redes de visibilidade rotuladas em 10 gêneros musicais. Fonte: Autor.

A Tabela 4.2 mostra o percentual de pares de gêneros musicais que possuem diferenças significaticas segundo o teste Tuckey, adotando um intervalo de confiança de $0,95$. Notamos que a maioria dos pares de gêneros musicais possuem diferenças significativas segundo o teste de diferenças entre médias populacionais. A Figura 4.3 apresenta os resultados do teste Tuckey para $\overline{Nc}$. Neste figura, 39 dos 45 pares de gêneros submetidos à hipotese de diferença significativa, acusaram resposta positiva. Dentre os pares deste teste que tiveram a hipótese rejeitada estão Metal e Hipop, Rock e Blues, e Country e Jazz. Esse teste coloca a representação dos agrupamentos num perspectiva realista em relação à distinção de gêneros musicais. Até este ponto não foram utilizadas as técnicas necessárias para fazer a distinção de agrupamentos. Até agora foram detectados indícios de tendências que podem ser úteis na classificação de gêneros musicais. A Seção 4.3 irá mostrar resultados que oferecerão uma discussão com base em algoritmos que irão utilizar as informações do DVFV para realizar a classificação propriamente dita.

Tabela 4.2: Pares de gêneros musicais com diferenças significativas para agrupamentos formados com as componentes do DVFV, segundo o teste Tuckey. Fonte: Autor

Descritor de Visibilidade	**Pares com diferença Significativa** (%)
Nc	86,7
Q	77,8
$< k >$	75,6
Δ	71,1

4.3 CLASSIFICAÇÃO

Nesta seção, nós estudamos a classificação dos gêneros musicais do banco GTZAN em duas configurações. Na primeira, nós montamos o vetor de atributos utilizando apenas o DVFV (grau médio $< k >$, a densidade Δ, a modularidade Q, e o número de comunidades Nc). Na segunda, além do DVFV, utilizamos algoritmos tempo-frequência discutidos na Seção 3.6, totalizando um vetor com 22 atributos (Tabela 4.7). Para ambas as configurações nós usamos um sistema de classificação supervisionado, ou seja, aquele cujos dados são considerados com uma classe pré-estabalecida, neste caso, o gênero musical. Todas as etapas de pré-processamento, classificação e seleção de atributos foram realizadas com o suporte dos algoritmos da plataforma de mineração de dados WEKA (discutidos na Seção 3.7). Dentre os algoritmos classificadores disponíveis no WEKA nós utilizamos: k-Star e IBk, Multiclass classifier, Naive Bayes, e a árvore de decisão J48. Na fase de treinamento e teste nós utlizamos, para todos os classificadores, o modelo de validação cruzada com 10 dobras (10-*fold croos validation*). Isso significa que

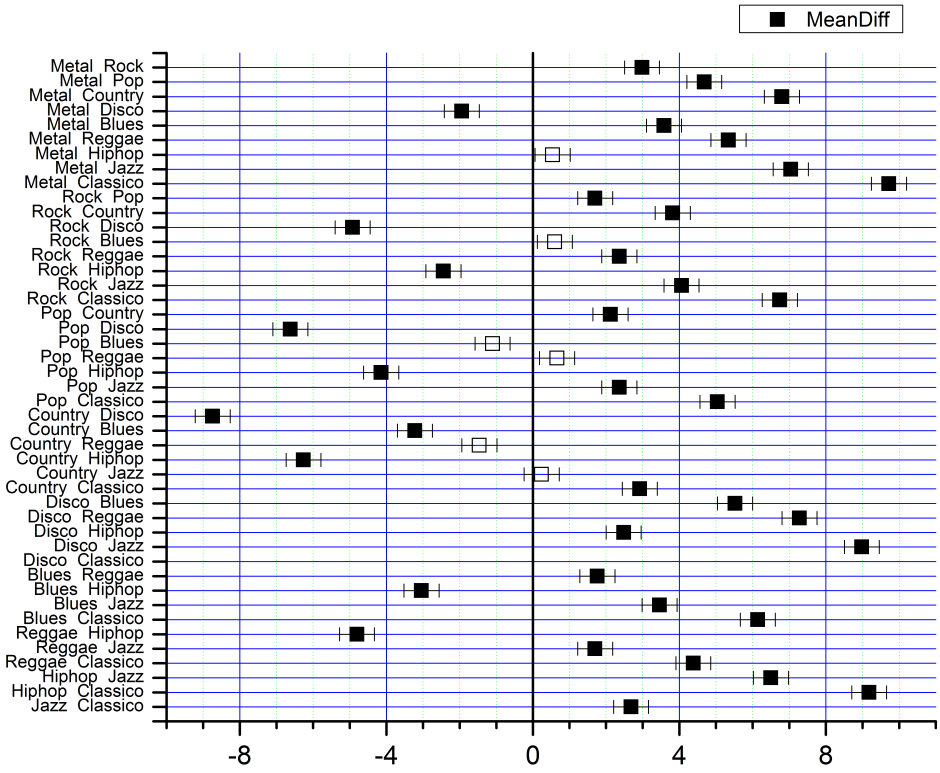


Figura 4.3: Diferença do NC médio entre pares de gêneros musicais segundo o teste Tuckey. Os boxes de cor preta representam as diferenças estatisticamente significativas, e os boxes brancos diferenças não-significativas. Fonte: Autor

o conjunto de dados foi dividido em 10 partes, das quais nove foram usadas para treinamento (ou aprendizagem), e uma para teste. Esse processo foi repetido 10 vezes, e, em cada vez, foi escolhida uma parte diferente para ser testada. A implementação computacional usada para extrair os atributos tempo-frequência foi realizada com recursos disponíveis na biblioteca Essentia [1] [Bogdanov et al. 2013].

Nos resultados a seguir serão informadas a acurácia e a área ROC da classificação. A acurácia mostra as instâncias corretamente classificadoas para cada gênero musical. A área ROC (*Receiver Operating Characteristic*), é um indicador de sensibilidade versus especifidade que estima o quanto as distribuições dos verdadeiros positivos e falsos positivos estão separadas. Uma área ROC igual a 1 indica que houve uma discriminação perfeita entre as classes, e uma área 0,5 significa que o teste é inválido, ou seja, que não houve distinção. Quanto mais próxima de 1 for a área ROC, melhor a qualidade da classificação.

[1]Essentia é uma biblioteca C++ de código aberto para análise e recuperação de informações músicais baseada em áudio disponíveis em https://essentia.upf.edu

4.3.1 USANDO APENAS O DESCRITOR DE VISIBILIDADE - DVFV

Nesta configuração nós usamos um vetor de atributos composto por $< k >$, Δ, Q, e Nc. A Tabela 4.3 mostra a acurácia e a área ROC da classificação de 1.000 arquivos de áudio em 10 gêneros musicais. O algoritmo classificador com melhor performance foi o *k-Star* com uma acurácia de $39,2\%$ e uma área ROC de $0,806$.

Tabela 4.3: Resultado da classificação utilizando apenas o DVFV. Fonte: Autor.

Classificadores	**Acurácia Média(%)**	**Área ROC Média**
k-Star	39,2	0,806
MultiClassClassifier	38	0,787
Naive Bayes	36,9	0,773
Árvore J48	34,8	0,690

A matriz de confusão do classificador k-Star é apresentada na Tabela 4.4. Nesta matriz, as colunas representam a classe predita e as linhas a classe atual. Na quarta linha e coluna, por exemplo, temos em negrito 47 músicas corretamente classificadas para o gênero Reggae. Esse número é chamado de verdadeiro positivo. Na mesma linha, sem o negrito, estão os falsos negativos, ou seja, músicas que são de fato do gênero Reggae mas foram classificadas como outros estilos, como por exemplo Classico (4), Pop (21), e Disco (1). Os verdadeiros negativos para o gênero Reggae são os elementos da diagonal nxn com $n \neq 4$. Os falsos positivos para o gênero Reggae são os elementos a_{ij} da coluna $j = 4$, que possuem a linha $i \neq 4$, como por exemplo o elemento a_{54}, que apresenta 13 falsos positivos de Reggae para o gênero Blues. Podemos notar nessa matriz que Metal e Clássico são os gêneros que, além de alcançarem o maior índice de acerto (68% ambos), não tiveram falsos negativos, nem falsos positivos entre si. Isso indica que em ambos os casos existem tendências de auto-similaridade de sinais muito bem definidas e que foram capturadas pelo descritor de visibilidade. Olhando o índice de acerto das músicas classificadas como Rock (19%), podemos notar que o algoritmo de classificação decidiu que 21% das músicas desse gênero deveriam ser classsificadas como Metal. Isso mostra que as afinidades entre Rock e Metal, que muitas vezes confunde a classificação feita por humanos, está retratada nesses resultados.

4.3.2 USANDO O DVFV E DESCRITORES TEMPO-FREQUÊNCIA

Nesta seção nós usamos um sistema de classificação utilizando um vetor de atributos constituído de 22 componentes que inclui o DVFV e vários descritores tempo-frequência (primeira coluna da Tabela 4.7). A Tabela 4.5 apresenta os resultados da classificação para os 10 gêneros musicais da base de dados GTZAN. O melhor classificador foi o Multi-class Classifier

Tabela 4.4: Matriz de confusão da classificação usando os descritores de visibilidade. Fonte: Autor.

	Classificado como									
	Clas	**Jazz**	**Hip**	**Regg**	**Blu**	**Disco**	**Coun**	**Pop**	**Rock**	**Metal**
Clas	**68**	18	0	2	3	0	5	3	1	0
Jazz	21	**39**	2	9	7	0	13	2	5	2
Hip	1	5	**17**	6	2	23	3	16	12	15
Regg	4	6	4	**47**	8	1	6	21	3	0
Blu	11	20	5	13	**7**	9	5	4	10	16
Disco	1	0	7	0	2	**56**	1	7	10	16
Coun	5	18	3	9	6	1	**32**	6	11	9
Pop	5	3	8	21	0	2	11	**39**	5	6
Rock	2	7	8	11	6	8	13	5	**19**	21
Metal	0	1	0	1	4	13	4	4	5	**68**

com uma acurácia média de $75,2\%$ e área ROC média de $0,948$.

Tabela 4.5: Resultado da classificação utilizando o DVFV + Descritores Tempo-Frequência. Fonte: Autor.

Classificadores	**Acurácia Média (%)**	**Área ROC Média**
Multiclass Classifier	75,20	0,948
IBk	69,30	0,829
k-Star	69,10	0,960
Naive Bayes	66,40	0,939
Árvore J48	61,60	0,820

A Matriz de confusão da Tabela 4.6 mostra a acurácia da classificação para cada gênero musical. Os gêneros Clássico e Metal apresentaram os dois maiores índices de acerto, com Metal alcançando 90% de verdadeiros positivos. Clássico, Country e Reggae foram as classes que tiveram o menor número de falsos positivos. O gênero Blues alcançou a pior percentual de acerto (47%). Neste caso houve um peso maior de falsos negativos distribuídos entre apenas dois gêneros: o Jazz (11%), e o Rock (16%). Isto significa que o algoritmo de classificação reconheceu como Jazz ou Blues 27% das instâncias rejeitadas como Rock. Situação semelhante ocorre entre Rock e Metal, nesse caso com 15% de falsos negativos atribuídos a Metal (linha 9 coluna 10). O maior número de falsos positivos ficaram com o gênero Rock, e também a segunda pior acurácia. Uma especulação para justificar esse resultado já foi apresentada, para o mesmo estilo, na Seção 4.3.1. Comparando os resultados da classificação usando apenas o DVFV (Tabela 4.3) com os resultados do sistema completo (Tabela 4.5), podemos observar que entre maiores percentuais ficaram com Clássico e Metal, e entre os menores, os estilos Blues e Rock. Os valores intermediários também encontraram equivalência. Isso indica que o sistema de classificação usando apenas o atributo de complexidade rítmica (DVFV) conseguiu detectar tendências que se confirmaram no sitema completo.

De um modo geral, houve uma excelente relação "perdas-e-ganhos" entre verdadeiros positivos e falsos positivos da classificação, uma vez que a área ROC média ficou em torno de 95%. Na seção 4.5.2 serão apresentadas comparações que permitirão uma melhor avaliação do percentual de acerto da classificação.

Tabela 4.6: Matriz de confusão do Classificador Multiclass.Fonte: Autor

	Classificado como									
	Clas	**Jazz**	**Hip**	**Regg**	**Blu**	**Disco**	**Coun**	**Pop**	**Rock**	**Metal**
Clas	**88**	4	0	0	3	1	0	1	3	0
Jazz	8	**70**	2	0	10	1	1	0	6	2
Hip	1	2	**84**	0	4	5	0	0	1	3
Regg	0	2	0	**79**	0	0	7	12	0	0
Blu	4	11	5	2	**47**	9	0	0	16	6
Disco	1	0	9	0	5	**73**	1	0	9	2
Coun	0	1	0	8	0	0	**85**	5	0	1
Pop	0	0	0	5	0	0	8	**87**	0	0
Rock	2	8	2	0	11	13	0	0	**49**	15
Metal	0	0	0	0	2	3	0	0	5	**90**

A Figura 4.4 mostra os melhores resultados da classificação dos 1.000 arquivos de áudio usando apenas um atributo por vez. Os melhores classificadores foram $k - Star$ para Δ, $< k >$, Q, Nc, Fluxo Espectral (FluxoSpectr), *loudness*, e $MFCC2$; IBK para DFA, BPM, e *Onset Rate* (TxOnset); *Multiclass Classifier* para Complexidade da Dinâmica (Compldyn); e Naive Bayes para *Zero Crossing Rate*. (TxPassZero).

4.4 SELEÇÃO DE ATRIBUTOS

Neste experimento nós usamos além do DVFV: 13 MFCCs, BPM, Loudness, Expoente DFA, Complexidade da Dinâmica, Fluxo Espectral, *Onset Rate*, e Taxa de Passagem pelo Zero. A Figura 4.5 mostra o ranking das melhores taxas de ganho por tipo, dentre os 24 atributos dos usados no experimento. Dentre os 13 MFCCs destacamos aquele que alcancou o melhor índice foi o MFCC 8. O resultado mostra que as duas melhores taxas de ganho ficaram com dois descritores de timbre (MFCC8 e Taxa de passagem pelo zero - ZeroCr), sendo que a taxa de passagem pelo zero também pode ser considerada como um descritor de similaridade que se aproxima dos descritores rítmicos. Em terceiro lugar aparece a modularidade como representante do descritor de visibilidade, superando tradicionais descritores do campo de processamento de sinais de áudio como Fluxo Espectral, Taxa de Onsets e Loudness. Os outros descritores de visibilidade ($< k >$, Nc, e Δ), ocupam respectivamente a sexta, a sétima, e a oitava posições, alcançando taxas semelhantes entre si, e estando à frente do descritor de autosimilaridade (Expoente DFA)

e mais outros três descritores tempo-frequência. No geral, os descritores de visibilidade ocuparam uma boa posição em relação aos descritores de timbre, e uma excelente posição em relação aos descritores de rítmo e de auto-similaridade. Tendo em vista que nesse experimento foram usados um número relativamente limitado de descritores de natureza rítmica, como por exemplo o histograma de batidas, podemos inferir que os descritores propostos nesta pesquisa possuem um grande potencial para dar informações relevantes a um sistema classificatório, e podendo até superar a contribuição dada por descritores tradicionalmente usados da literatura.

4.5 COMPARAÇÃO COM TRABALHOS CORRELATOS

4.5.1 COMPARAÇÃO COM HIERARQUIA DE GÊNEROS MUSICAIS USANDO DFA

A Figura 4.6 representa a dispersão entre os Expoentes DFA e os graus médios e modularidades de grafos de visibilidade associados aos 1.000 arquivos do banco GTZAN. O coeficiente de Pearson da dispersão apresentada na Figura 4.6 (a) $(\rho = 0,53)$ indica que existe uma correlação positiva e moderada entre α_{DFA} e $< k >$. Entre o Expoente DFA e a modularidade (Figura 4.6a) também ocorre uma correlação de Pearson moderada, porém negativa $(\rho = -0,59)$ (Figura 4.6 (b)). Os resultados das correlações entre α_{DFA} e Δ (correlação positiva), e α_{DFA} e Nc (correlação negativa), são bastante semelhantes àqueles encontrados para

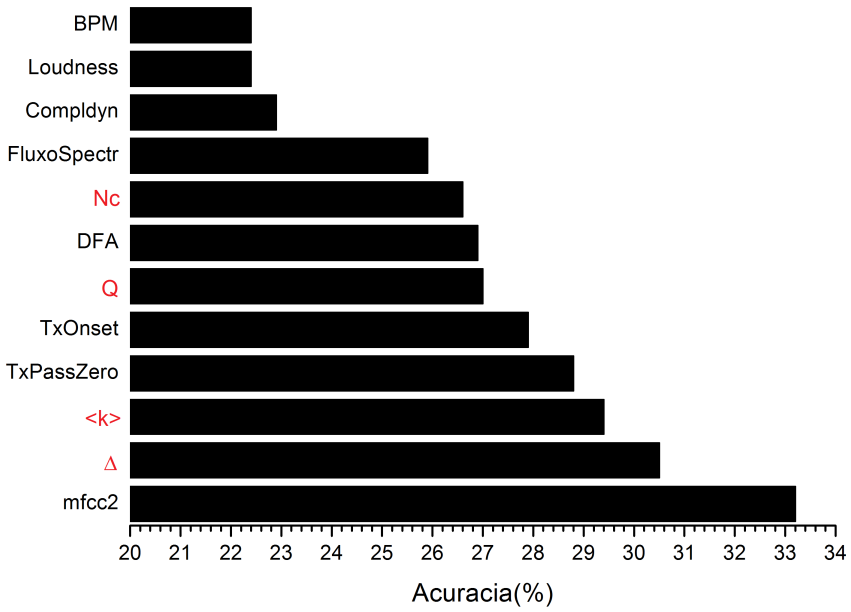


Figura 4.4: Acurácia média da classificação usando cada um dos atributos separadamente. Os propostos por esta tese aparecem na cor vermelha. Fonte: Autor.

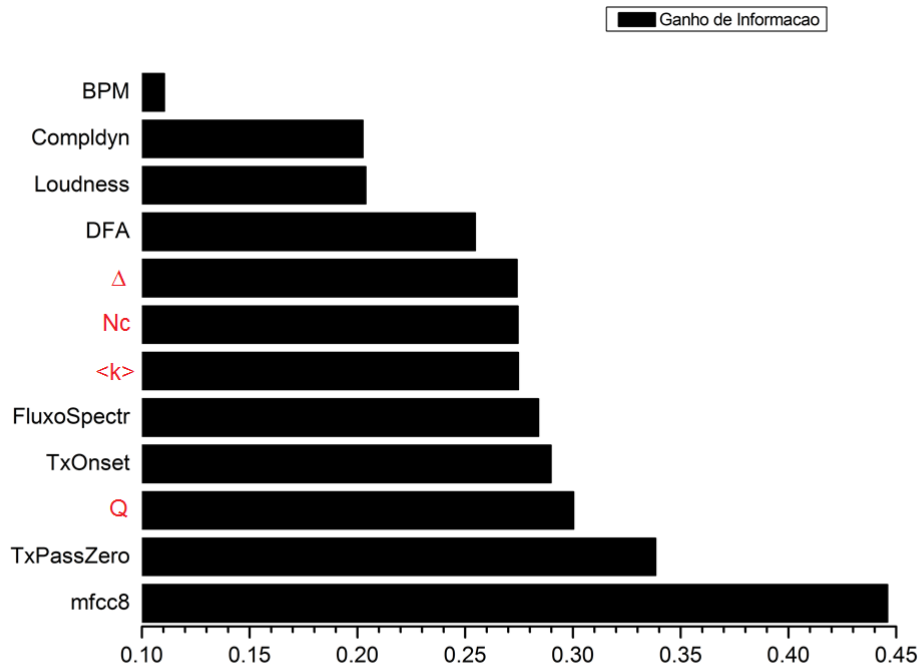


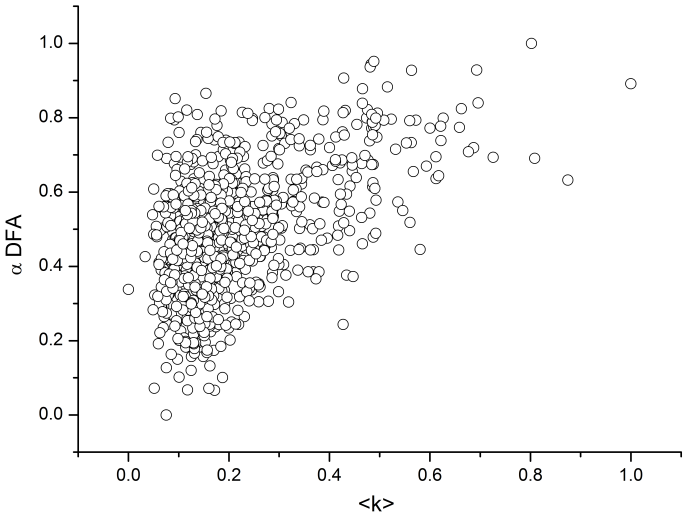
Figura 4.5: Ganho de informação de descritores de áudio. Os descritores de visibilidade (DVFV)aparecem na cor vermelha. Os demais (na cor preta)são descritores tempo-frequência. Fonte: Autor.

$\alpha_{DFA} \times < k >$, e $\alpha_{DFA} \times Q$, respectivamente. Esses resultados indicam uma correlação não desprezível entre o descritor de visibilidade e o Expoente DFA.

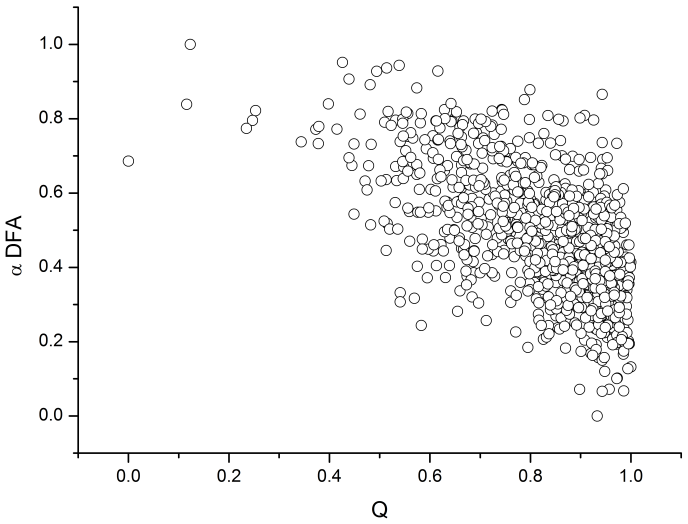
Comparando os resultados de α_{DFA} encontrados por Melo 2012 para a mesma base de dados desta tese (Figura 4.7), e os valores de $< k >$ para dez gêneros musicais apresentados na Figura 4.2 (b)), podemos perceber que em ambos os casos temos o estabelecimento de hierarquias com uma estrutura bastante semelhante, e às vezes com posições idênticas. Em ambas as hierarquias nota-se uma ordem decrescente que parte de gêneros mais eruditos (clássico, jazz) termina em estilos com mais dançantes e com grande persistência percussiva (metal, disco e hiphop). Nos dois casos o gênero clássico ocupa a primeira posição, destacando-se das demais em magnitude.

4.5.2 COMPARAÇÃO COM OUTROS SISTEMAS DE CLASSIFICAÇÃO USANDO A BASE GTZAN

Uma vez que a extração de características rítmicas no sistema de classificação desta tese foi efetuada através do conjunto formado por DVFV, Expoente DFA, e BPM, nós escolhemos fazer uma comparação com o trabalho de Lykartsis e Lerch 2015. Essa escolha foi



(a)



(b)

Figura 4.6: Dispersão entre o Expoente DFA (α_{DFA}) e: (a) o grau médio $< k >$; (b) a modularidade Q dos grafos de visibilidade. Fonte: Autor.

motivada pelo fato de ser um trabalho onde a classificação é feita usando apenas o histograma de batidas. A acurácia média obtida em Lykartsis e Lerch 2015 foi de $59,3\%$, usando 231 atributos. Apesar do percentual de acerto maior que os $44,4\%$ de nosso experimento, podemos considerar que nossa pesquisa alcançou um resultado competitivo tendo em vista que nossos resultados foram obtidos com o uso de apenas 4 atributos rítmicos (ver Tabela 4.7).

Para a configuração DVFV+Descritores Tempo-Frequência, que alcançou uma acu-

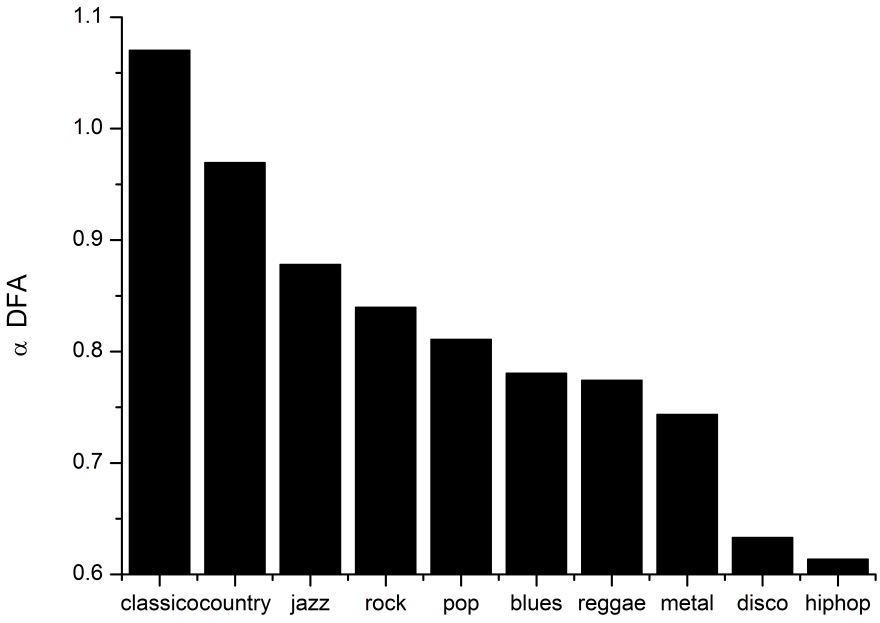


Figura 4.7: Expoente DFA de 1.000 sinais musicais do banco GTZAN. Fonte: Melo 2012.

rácia média de $75,4\%$, notamos que este resultado superou àqueles encontrados em alguns trabalhos com a mesma base de dados, a exemplo de Tzanetakis e Cook 2002 (61%), Holzapfel e Stylianou 2008 (74%), Lidy e Rauber 2005 ($74,5\%$), Jr et al. 2007 ($58,07\%$), e é comparável aos trabalhos de Li et al. 2003 ($78,5\%$), Lidy et al. 2007 ($76,8\%$), e Panagakis et al. 2008 ($78,2\%$), que também usaram a mesma base. Para melhor entender o valor do resultado encontrado na classificação usando os descritores de visibilidade em flutuações de variância vale ressaltar que os trabalhos citados contam com um número maior de atributos, chegando a 80 [Benetos e Kotropoulos 2008], contra apenas 22 atributos usados em nosso experimento. A Tabela 4.7 mostra o vetor de características desta tese (primeira coluna), e o vetor utilizado em Tzanetakis e Cook 2002 (segunda coluna). Em comparação podemos notar que nosso trabalho não utiliza descritores do aspecto tonal, e utiliza para estudar a complexidade rítmica apenas o DVFV. Em relação ao timbre nesta tese utilizamos as médias do Fluxo Espectral e da Taxa de Passagem pelo Zero, ao invés de usar a média e a variância, a fim de seguir o mesmo procedimento utilizado nos trabalhos que usamos referência.

A Figura 4.8 traz um gráfico de barras que nos permitiu comparar os resultados da taxa de acerto da classificação desta tese (barras pretas) (Tabela 4.6), com os resultados do trabalho de Tzanetakis e Cook 2002 (barras brancas) para cada gênero musical. Nesta figura podemos ver que, em ambos os experimentos, o Blues e o Rock alcançaram os menores índices

Tabela 4.7: Composição do Vetor de Atributos usados em dois Trabalhos

	Atributos (Quantidade)	
Tipologia	**Nossa Proposta**	Tzanetakis e Cook 2002
Timbre	MFCCc (13) Fluxo Spectral (1) Taxa de Passagem pelo Zero (1)	MFCCs (10) Fluxo Espectral (2) Taxa de Passagem pelo Zero (2) Centróide Espectral (2) Rollof Espectral (2) Low-Energy (1)
Complexidade Rítmica	DVFV (4)	Histograma de Batidas (6)
Tom (Pitch)		Histograma de Picos (5)
Detecção de Onsets	*Onset Rate (1)*	
Dinâmica	Loudness (1) Complexidade da Dinâmica (1)	
Total de Atributos	22	30

de acerto, enquanto os gêneros Clásssico, Hiphop e Pop ficaram entre as categorias de maior acurácia. Em termos relativos o nosso trabalho obteve uma taxa de acerto superior para todos as categorias, com exceção do gênero Jazz. Para sete classes houve uma diferença igual ou superior a 19%. Para Country e Metal, em particular, houve uma diferença notável na quantidade de verdadeiros positivos: 32 e 37%, respectivamente. Este resultado mostrou que a classificação de gêneros musicais usando o DVFV para extração da complexidade rítmica, em substituição do histograma de batidas, resultou em um sistema de categorização com melhor taxa de acerto global e individual.

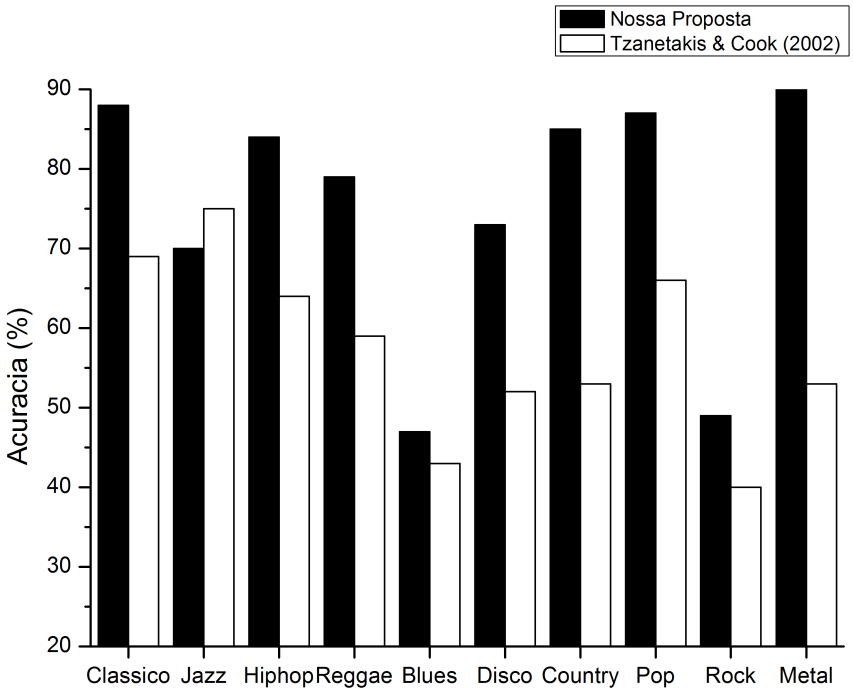


Figura 4.8: Resultado da classificação de gêneros musicais do banco GTZAN para a nossa proposta, e para o experimento de Tzanetakis e Cook 2002. Fonte: Autor.

5 CONSIDERAÇÕES FINAIS

Nesta tese nós apresentamos um novo método para representar numericamente (usando propriedades de redes), e representar (usando modelagem gráfica de redes) a complexidade rítmica em sinais musicais polifônicos, estimando a auto-similaridade de seus transientes com base no descritor de visibilidade em flutuações de variância.

Antes de explicarmos detalhadamente a metodologia e mostrarmos os resultados empíricos, nós realizamos uma revisão sistemática da literatura abordando os problemas e métodos usados na classificação de gêneros muscais, e na extração de atributos, dentro do contexto da pesquisa de recuparação de informações musicais. Definimos o conceito de sinais polifônicos usado nesta tese em comparação com o termo usado na teoria musical. Revisamos a origem, os fundamentos, e a tipologia de redes complexas sob um recorte suficientemente dimensionado para a compreensão da metodologia proposta. Apresentamos alguns métodos usados para transformar séries temporais em redes, dedicando uma maior profundidade para o mapeamento por grafos de visibilidade (*visibility graphs*). Expusemos o modo como a modelagem de redes complexas tem sido usada no campo de informações musicais, e mostramos que não foram encontrados trabalhos onde são utilizadas propriedades de redes para a detecção de aspectos ritmicos em sinais musicais usando elementos não simbólicos na formação da rede. Recuperamos diferentes conceitos de auto-similaridade encontrados na literatura e definimos o termo conforme usado no contexto desta tese.

Nós iniciamos a apresentação dos resultados experimentais, mostrando que as diferenças de padrões de sinais podem ser notados através dos diferentes níveis de auto-similaridade. Esses diferentes níveis foram mostrados graficamente através da representação das comunidades detectadas em seus grafos de visibilidade, e numericamente através dos parâmetros do descritor de visibilidade. Ficou bastante claro nos quatro exemplos iniciais, que séries com baixa auto-similaridade (séries menos "densas", com "picos" seguidos de "vales"), como a *série classica*, geraram grafos com menor número de comunidades, e modularidade, e maior grau médio e densidade, do que séries com maior auto-similaridade (séries mais "densas"), como a *série metal*. Mostramos que a representação gráfica dos sinais classico, blues, pop, e metal, e

seus respectivos grafos de visibilidade, sugerem uma espécie de hierarquia crescente de auto-similaridade. Em seguida, apresentamos resultados mostrando que as tendências apresentadas inicialmente para os quatro sinais, se confirmaram para a base de dados completa. O teste de hipóteses ANOVA para diferenças entre as médias populacionais de Q, Nc, Δ, e $< k >$, resultou em diferenças significativas para mais de 71% dos pares gêneros musicais.

A partir das médias de Q, Nc, Δ, e $< k >$, de cada classe musical, nós propusamos uma hierarquia de auto-similaridade crescente para Q e Nc, e uma decrescente para Δ e $< k >$, que podem ser comparadas a possíveis hierarquias feitas a partir da percepção humana. Depois disso, fizemos experimentos de classificação: usando apenas o DVFV, e alcançamos $39,2\%$ de acurácia; combinando o DVFV com 18 atributos tempo-frequência, e chegamos a uma precisão de $75,2\%$ na classificação; usando um atributo por vez, totalizando 24 processos de aprendizagem de máquina e classificação, e conseguimos um resultado onde a densidade e o grau médio ficaram entre as três melhores acurácias, com $30,5\%$ e $29,4\%$, respectivamente. No ranking do ganho de informação o descritos de visibilidade alcançou a terceira melhor taxa com a modularidade, que ficou abaixo apenas da taxa de passagem pelo zero e do mfcc8. Todos as componentes do DVFV ficaram à frente dos descritores de complexidade rítmica do experimento (BPM e Expoente DFA). Ao compararmos os resultados encontrados nesta tese com trabalhos correlatos, encontramos resultados comparáveis no caso da classificação usando apenas o DVFV ($44,4\%$), e no experimento usando o DVFV como único atributo de complexidade ritmica, junto a 18 atributos tempo-frequência, encontramos alguns resultados com uma taxa de classificação inferiores ao nosso (faixa de 58 a $74,5\%$), e outros com valores comparáveis (faixa de 76 a $78,5\%$).

Ao detalharmos a comparação para um trabalho bastante referenciado no campo de recuperação de informações musicais (Tzanetakis e Cook 2002), nós encontramos um valor superior para precisão média ($75,2$ comparado a 61%), e acurácias relativas superiores para nove dos 10 gêneros musicais. Nesta última comparação, tivemos a oportunidade de testar nosso sistema de classificação, que usou o DVFV como descritor da complexidade rítmica, frente ao sistema de classificação usado por Tzanetakis e Cook 2002, que tem o histograma de batidas fazendo o mesmo papel. Concluímos que neste caso o DVFV conseguiu contribuir positivamente para que o sistema alcançasse um resultado superior ao de Tzanetakis e Cook 2002. Isso é bastante relevante pois o histograma de batidas tem sido adotado por muitos trabalhos como descritor de complexidade ritmica, e o resultado apresentado nesta tese pode contribuir para a inserção de um novo recurso para detecção de padrões de natureza rítmica.

Concluimos que o descritor de visibilidade foi capaz de: detectar padrões de auto-

similaridade gráfica e numericamente, estabelecer hierarquias de gêneros musicais através de seus parâmetros, contribuir de forma relevante para um sistema de classificação supervisionado de gêneros musicais, alcançando acurácias comparáveis ou superiores a trabalhos do estado da arte, e obter uma posição respeitável, entre algoritmos consagrados da literatura, no ranking de taxa de ganho de informação.

5.0.1 CONTRIBUIÇÕES

Nesta tese foi apresentado o primeiro uso de propriedades topológicas de redes complexas, especificamente a modularidade, o grau médio, o número de comunidades, e a densidade, como atributos em um sistema de classificação, junto a descritores baseados em tranfromadas de Fourier, para a classificação de gêneros musicais. Também foi realizado nesta tese, o segundo experimento utilizando o mapeamento de sinais de áudio polifônicos em grafos de visibilidade, para o estudo de recuperação de informações musicais, com o uso de dados não-simbólicos na construção das redes. Foi introduzida a modelagem gráfica de redes utilizando a detecção de comunidades para identificar diferenças de auto-similaridade em sinais de mais de duas classes de gêneros musicais. Realizou-se a utilização inédita da modularidade, do grau médio, do número de comunidades, e da densidade de redes complexas, na hierarquização de mais de dois gêneros musicais.

5.0.2 TRABALHOS FUTUROS

Nesta tese foi utilizado o algoritmo de detecção de comunidades de Newman 2006 e a otimização de Blondel et al. 2008. Nos próximos trabalhos gostaríamos de testar outros algoritmos de detecção de comunidades como Muff et al. 2005, Duch e Arenas 2005, e Massen e Doye 2005. O experimentos desta tese foram feitos utilizando a base de dados GTZAN. Gostaríamos de ver a performance do DVFV em bases como Latin Music Database [1], que abranje ritmos brasileiros e latinos, Balroom [2] que é uma base de dados dedicada a ritmos dançantes, MIREX 2009 [3] que é base para muitas pesquisas em *audio mood classification.*

[1] http://www.ppgia.pucpr.br/silla/lmd/index.html
[2] http://mtg.upf.edu/ismir2004/contest/rhythmContest/
[3] //www.musicir.org/mirex/2009/index.php/Audio_Music_Mood_Classificatione

REFERÊNCIAS

AHRENDT, P.; HANSEN, L. K. **Music genre classification systems-a computational approach**. Tese (Doutorado) — Technical University of DenmarkDanmarks Tekniske Universitet, Department of Informatics and Mathematical Modeling Institut for Informatik og Matematisk Modellering, Cognitive SystemsKognitive systemer, 2006.

ALEXANDERSON, G. About the cover: Euler and konigsbergs bridges: A historical view. **Bulletin of the american mathematical society**, v. 43, n. 4, p. 567–573, 2006.

ANDÉN, J.; MALLAT, S. Multiscale scattering for audio classification. In: **ISMIR**. [S.l.: s.n.], 2011. p. 657–662.

ANDJELKOVIĆ, M.; GUPTE, N.; TADIĆ, B. Hidden geometry of traffic jamming. **Physical Review E**, APS, v. 91, n. 5, p. 052817, 2015.

AUCOUTURIER, J.-J.; PACHET, F. Representing musical genre: A state of the art. **Journal of New Music Research**, Taylor & Francis, v. 32, n. 1, p. 83–93, 2003.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação-: Conceitos e Tecnologia das Máquinas de Busca**. [S.l.]: Bookman Editora, 2013.

BARABÁSI, A.-L. **Network science**. [S.l.]: Cambridge university press, 2016.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **science**, American Association for the Advancement of Science, v. 286, n. 5439, p. 509–512, 1999.

BARBEDO, J. G. A.; LOPES, A. Automatic genre classification of musical signals. **EURASIP Journal on Applied Signal Processing**, Hindawi Publishing Corp., v. 2007, n. 1, p. 157–157, 2007.

BELLO, J. P. et al. A tutorial on onset detection in music signals. **IEEE Transactions on speech and audio processing**, IEEE, v. 13, n. 5, p. 1035–1047, 2005.

BENETOS, E.; KOTROPOULOS, C. A tensor-based approach for automatic music genre classification. In: IEEE. **Signal Processing Conference, 2008 16th European**. [S.l.], 2008. p. 1–4.

BERGSTRA, J.; CASAGRANDE, N.; ECK, D. Two algorithms for timbre and rhythm-based multiresolution audio classification. In: **Proceedings of ISMIR**. [S.l.: s.n.], 2005.

BEROIS, M. H. Detecting and describing percussive events in polyphonic music. **Master thesis. Universitat Pompeo Fabra, Spain**, 2008.

BLONDEL, V. D. et al. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2008, n. 10, p. P10008, 2008.

BOGDANOV, D. et al. Essentia: An audio analysis library for music information retrieval. In: **ISMIR**. [S.l.: s.n.], 2013. p. 493–498.

BONDY, J. A.; MURTY, U. S. R. **Graph theory with applications**. [S.l.]: Macmillan London, 1976.

BROSSIER, P.; BELLO, J. P.; PLUMBLEY, M. D. Real-time temporal segmentation of note objects in music signals. In: **Proceedings of ICMC 2004, the 30th Annual International Computer Music Conference**. [S.l.: s.n.], 2004.

BULDÚ, J. M. et al. The complex network of musical tastes. **New Journal of Physics**, IOP Publishing, v. 9, n. 6, p. 172, 2007.

CAMPANHARO, A. **Dualidade entre análise de séries temporais e de redes complexas**. Tese (Doutorado) — PhD thesis, Instituto Nacional de Pesquisas Espaciais, University of Bergen, Sao José dos Campos, 2011.

CLEARY, J. G.; TRIGG, L. E. K*: An instance-based learner using an entropic distance measure. In: **12th International Conference on Machine Learning**. [S.l.: s.n.], 1995. p. 108–114.

COOPER, M. L.; FOOTE, J. Automatic music summarization via similarity analysis. In: **ISMIR**. [S.l.: s.n.], 2002.

CORREA, D. C.; SAITO, J. H.; COSTA, L. da F. Musical genres: beating to the rhythms of different drums. **New Journal of Physics**, IOP Publishing, v. 12, n. 5, p. 053030, 2010.

COSTA, L. d. F. et al. Characterization of complex networks: A survey of measurements. **Advances in physics**, Taylor & Francis, v. 56, n. 1, p. 167–242, 2007.

DANNENBERG, R. B. et al. Panel: New directions in music information retrieval. In: **ICMC**. [S.l.: s.n.], 2001.

DAS, A.; DAS, P. Classification of different indian songs based on fractal analysis. **Complex Systems**, [Champaign, IL, USA: Complex Systems Publications, Inc., c1987-, v. 15, n. 3, p. 253, 2005.

DIETTERICH, T. G.; BAKIRI, G. Solving multiclass learning problems via error-correcting output codes. **Journal of artificial intelligence research**, v. 2, p. 263–286, 1995.

DIXON, S. Onset detection revisited. In: CITESEER. **Proceedings of the 9th International Conference on Digital Audio Effects**. [S.l.], 2006. v. 120, p. 133–137.

DIXON, S. et al. Towards characterisation of music via rhythmic patterns. In: **ISMIR**. [S.l.: s.n.], 2004.

DONNER, R. V. et al. Recurrence-based time series analysis by means of complex network methods. **International Journal of Bifurcation and Chaos**, World Scientific, v. 21, n. 04, p. 1019–1046, 2011.

DUCH, J.; ARENAS, A. Community detection in complex networks using extremal optimization. **Physical review E**, APS, v. 72, n. 2, p. 027104, 2005.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. [S.l.]: John Wiley & Sons, 2012.

ERDOS, P.; RÉNYI, A. On the evolution of random graphs. **Publ. Math. Inst. Hung. Acad. Sci**, v. 5, n. 1, p. 17–60, 1960.

EZZAIDI, H.; ROUAT, J. Automatic musical genre classification using divergence and average information measures. **World Academy of Science, Engineering and Technology**, v. 15, 2006.

FOOTE, J.; COOPER, M. L. Visualizing musical structure and rhythm via self-similarity. In: **ICMC**. [S.l.: s.n.], 2001. v. 1, p. 423–430.

GAO, Z.; JIN, N. Complex network from time series based on phase space reconstruction. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP, v. 19, n. 3, p. 033137, 2009.

GJERDINGEN, R. O.; PERROTT, D. Scanning the dial: The rapid recognition of music genres. **Journal of New Music Research**, Taylor &93–100, 2008.

GOLDBERGER, A. L. et al. Fractal dynamics in physiology: alterations with disease and aging. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 99, n. suppl 1, p. 2466–2472, 2002.

GOULART, A. J. H. **Classificação automática de gênero musical baseada em entropia e fractais**. Tese (Doutorado) — Universidade de São Paulo, 2012.

HALL, M. et al. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, ACM, v. 11, n. 1, p. 10–18, 2009.

HOLZAPFEL, A.; STYLIANOU, Y. Musical genre classification using nonnegative matrix factorization-based features. **IEEE Transactions on Audio, Speech, and Language Processing**, IEEE, v. 16, n. 2, p. 424–434, 2008.

HOPKINS, B.; WILSON, R. The truth about kõnigsberg. **The College Mathematics Journal**, Mathematical Association of America, v. 35, n. 3, p. 198, 2004.

ITZKOVITZ, S. et al. Recurring harmonic walks and network motifs in western music. **Advances in Complex Systems**, World Scientific, v. 9, n. 01n02, p. 121–132, 2006.

JACOBSON, K.; SANDLER, M. B.; FIELDS, B. Using audio analysis and network structure to identify communities in on-line social networks of artists. In: **ISMIR**. [S.l.: s.n.], 2008. p. 269–274.

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, p. 585–594, 2004.

JENSEN, K.; ANDERSEN, T. H. Beat estimation on the beat. In: IEEE. **Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.** [S.l.], 2003. p. 87–90.

JR, C. N. S.; KAESTNER, C. A.; KOERICH, A. L. Classificaç ao automática de gêneros musicais utilizando métodos de bagging e boosting. 2005.

JR., C. N. S.; KAESTNER, C. A.; KOERICH, A. L. Automatic genre classification of latin music using ensemble of classifiers. In: **Proceedings of the 33rd Integrated Software and Hardware Seminar**. [S.l.: s.n.], 2006. p. 47–53.

JR, C. N. S.; KAESTNER, C. A.; KOERICH, A. L. Automatic music genre classification using ensemble of classifiers. In: IEEE. **Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on**. [S.l.], 2007. p. 1687–1692.

KANTELHARDT, J. W. et al. Multifractal detrended fluctuation analysis of nonstationary time series. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 316, n. 1, p. 87–114, 2002.

KANTZ, H.; SCHREIBER, T. Tisean nonlinear time series analysis. **URL= http://www. mpipksdresden. mpg. de/ tisean/, Last access July 2017**.

KAREGOWDA, A. G.; MANJUNATH, A.; JAYARAM, M. Comparative study of attribute selection using gain ratio and correlation based feature selection. **International Journal of Information Technology and Knowledge Management**, v. 2, n. 2, p. 271–277, 2010.

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.

LACASA, L. et al. The visibility graph: a new method for estimating the hurst exponent of fractional brownian motion. **EPL (Europhysics Letters)**, IOP Publishing, v. 86, n. 3, p. 30001, 2009.

LAMBIOTTE, R.; DELVENNE, J.-C.; BARAHONA, M. Laplacian dynamics and multiscale modular structure in networks. **arXiv preprint arXiv:0812.1770**, 2008.

LERCH, A. **An introduction to audio content analysis: Applications in signal processing and music informatics**. [S.l.]: John Wiley & Sons, 2012.

LI, T.; OGIHARA, M.; LI, Q. A comparative study on content-based music genre classification. In: ACM. **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**. [S.l.], 2003. p. 282–289.

LIDY, T.; RAUBER, A. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: **ISMIR**. [S.l.: s.n.], 2005. p. 34–41.

LIDY, T. et al. Combining audio and symbolic descriptors for music classification from audio. **Music Information Retrieval Information Exchange (MIREX)**, 2007.

LIU, X. F.; CHI, K. T.; SMALL, M. Complex network structure of musical compositions: Algorithmic generation of appealing music. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 389, n. 1, p. 126–132, 2010.

LYKARTSIS, A.; LERCH, A. Beat histogram features for rhythm-based musical genre classification using multiple novelty functions. In: **Proceedings of the 16th ISMIR Conference,(JANUARY)**. [S.l.: s.n.], 2015. p. 434–440.

MANDELBROT, B. B.; PIGNONI, R. **The fractal geometry of nature**. [S.l.]: WH freeman New York, 1983.

MARTINS, L. A.; PÁDUA, F. L.; ALMEIDA, P. E. de. Aplicação de redes som e técnicas não paramétricas para inspeção visual automática de defeitos em aços laminados.

MASSEN, C. P.; DOYE, J. P. Identifying communities within energy landscapes. **Physical Review E**, APS, v. 71, n. 4, p. 046101, 2005.

MELO, D. d. F. P. **Estudo de Flutuações de Sinais de Áudio Classificados por Gênero Musical**. Dissertação (Mestrado) — Faculdade de Tecnologia Senai Cimatec, 2012.

MELO, D. d. F. P.; FADIGAS, I. de 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. 32, 2017.

MELO, D. F. P. Análise de flutuações de variância em sinais de áudio agrupados por gênero musical. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 1, n. 1, 2013.

MELO, D. F. P.; FADIGAS, I. S.; PEREIRA, H. B. d. B. Community detection in visibility networks: an approach to categorize percussive influence on audio musical signals. **Studies in Computational Intelligence – Complex Networks & Their Applications V**, Springer, v. 693, n. 1, p. 321–334, 2016.

MUFF, S.; RAO, F.; CAFLISCH, A. Local modularity measure for network clusterizations. **Physical Review E**, APS, v. 72, n. 5, p. 056107, 2005.

MÜLLER, M. **Information retrieval for music and motion**. [S.l.]: Springer Science & Business Media, 2007.

MÜLLER, M.; GROSCHE, P.; JIANG, N. A segment-based fitness measure for capturing repetitive structures of music recordings. In: **ISMIR**. [S.l.: s.n.], 2011. p. 615–620.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. **The structure and dynamics of networks**. [S.l.]: Princeton University Press, 2006.

NEWMAN, M.; BARABASI, A.-L.; WATTS, D. J. **The structure and dynamics of networks**. [S.l.]: Princeton University Press, 2011.

NEWMAN, M. E. Modularity and community structure in networks. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 103, n. 23, p. 8577–8582, 2006.

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

NUNEZ, A. et al. Detecting series periodicity with horizontal visibility graphs. **International Journal of Bifurcation and Chaos**, World Scientific, v. 22, n. 07, p. 1250160, 2012.

PACHET, F.; CAZALY, D. A taxonomy of musical genres. In: LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE. **Content-Based Multimedia Information Access-Volume 2**. [S.l.], 2000. p. 1238–1245.

PAMPALK, E.; RAUBER, A.; MERKL, D. Content-based organization and visualization of music archives. In: ACM. **Proceedings of the Tenth ACM international Conference on Multimedia**. [S.l.], 2002. p. 570–579.

PANAGAKIS, I.; BENETOS, E.; KOTROPOULOS, C. Music genre classification: A multilinear approach. In: **ISMIR**. [S.l.: s.n.], 2008. p. 583–588.

PANAGAKIS, Y.; KOTROPOULOS, C.; ARCE, G. R. Music genre classification via sparse representations of auditory temporal modulations. In: IEEE. **Signal Processing Conference, 2009 17th European**. [S.l.], 2009. p. 1–5.

PARK, D. et al. Topology and evolution of the network of western classical music composers. **EPJ Data Science**, Springer Berlin Heidelberg, v. 4, n. 1, p. 1, 2015.

PARK, S.-H.; FÜRNKRANZ, J. Efficient pairwise classification. **Machine Learning: ECML 2007**, Springer, p. 658–665, 2007.

PENG, C. e. a. Mosaic organization of dna nucleotides. **Physical Review**, n. E49, p. 1685–16895, 1994.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, n. 1, p. 81–106, 1986.

QUINLAN, J. R. Constructing decision tree. **C4**, v. 5, p. 17–26, 1993.

REFENES, A. N.; ZAPRANIS, A.; FRANCIS, G. Stock performance modeling using neural networks: a comparative study with regression models. **Neural networks**, Elsevier, v. 7, n. 2, p. 375–388, 1994.

SCHEDL, M.; GÓMEZ, E.; URBANO, J. Music information retrieval: recent developments and applications. **Foundations and Trends in Information Retrieval**, Now Publishers, v. 8, n. 2-3, p. 127–261, 2014.

SEYERLEHNER, K.; WIDMER, G.; KNEES, P. A comparison of human, automatic and collaborative music genre classification and user centric evaluation of genre classification systems. In: SPRINGER. **International Workshop on Adaptive Multimedia Retrieval**. [S.l.], 2010. p. 118–131.

SILVA, D. F. **Classificação de séries temporais por similaridade e extração de atributos com aplicação na identificação automática de insetos**. Tese (Doutorado) — Universidade de São Paulo, 2014.

SMALL, M.; ZHANG, J.; XU, X. Transforming time series into complex networks. **Complex sciences**, Springer, p. 2078–2089, 2009.

SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. **Bulletin of Mathematical Biology**, Springer, v. 13, n. 2, p. 107–117, 1951.

STEPHEN, M.; GU, C.; YANG, H. Visibility graph based time series analysis. **PloS one**, Public Library of Science, v. 10, n. 11, p. e0143015, 2015.

STEVENS, S.; VOLKMANN, J.; NEWMAN, E. The mel scale equates the magnitude of perceived differences in pitch at different frequencies. **J. Acoust. Soc. Am**, v. 8, n. 3, p. 185–190, 1937.

STEVENS, S. S. On the psychophysical law. **Psychological review**, American Psychological Association, v. 64, n. 3, p. 153, 1957.

STREICH, S.; HERRERA, P. Detrended fluctuation analysis of music signals: Danceability estimation and further semantic characterization. In: **Proceedings of the 118th AES Convention**. [S.l.: s.n.], 2005.

STREICH, S. et al. **Music complexity: a multi-faceted description of audio content**. [S.l.]: Universitat Pompeu Fabra, 2006.

STURM, B. L. An analysis of the gtzan music genre dataset. In: ACM. **Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies**. [S.l.], 2012. p. 7–12.

STURM, B. L. The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use. **arXiv preprint arXiv:1306.1461**, 2013.

TSE, C.; LIU, X.; SMALL, M. Analyzing and composing music with complex networks: finding structures in bach, chopin and mozart. p. 5–8, 2008.

TSUNOO, E. et al. Audio genre classification using percussive pattern clustering combined with timbral features. In: IEEE. **Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on**. [S.l.], 2009. p. 382–385.

TZANETAKIS, G.; COOK, P. Marsyas: A framework for audio analysis. **Organised sound**, Cambridge University Press, v. 4, n. 3, p. 169–175, 2000.

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. **IEEE Transactions on Speech and Audio Processing**, IEEE, v. 10, n. 5, p. 293–302, 2002.

VICKERS, E. Automatic long-term loudness and dynamics matching. In: AUDIO ENGINEERING SOCIETY. **Audio Engineering Society Convention 111**. [S.l.], 2001.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of small-world networks. **nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998.

WEI, W. W. et al. **Time series analysis: univariate and multivariate methods**. [S.l.]: Pearson Addison Wesley, 2006.

WITTEN, I. H. et al. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2016.

WITTEN, I. H. et al. Weka: Practical machine learning tools and techniques with java implementations. 1999.

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

ZHANG, J.; SMALL, M. Complex network from pseudoperiodic time series: Topology versus dynamics. **Physical review letters**, APS, v. 96, n. 23, p. 238701, 2006.