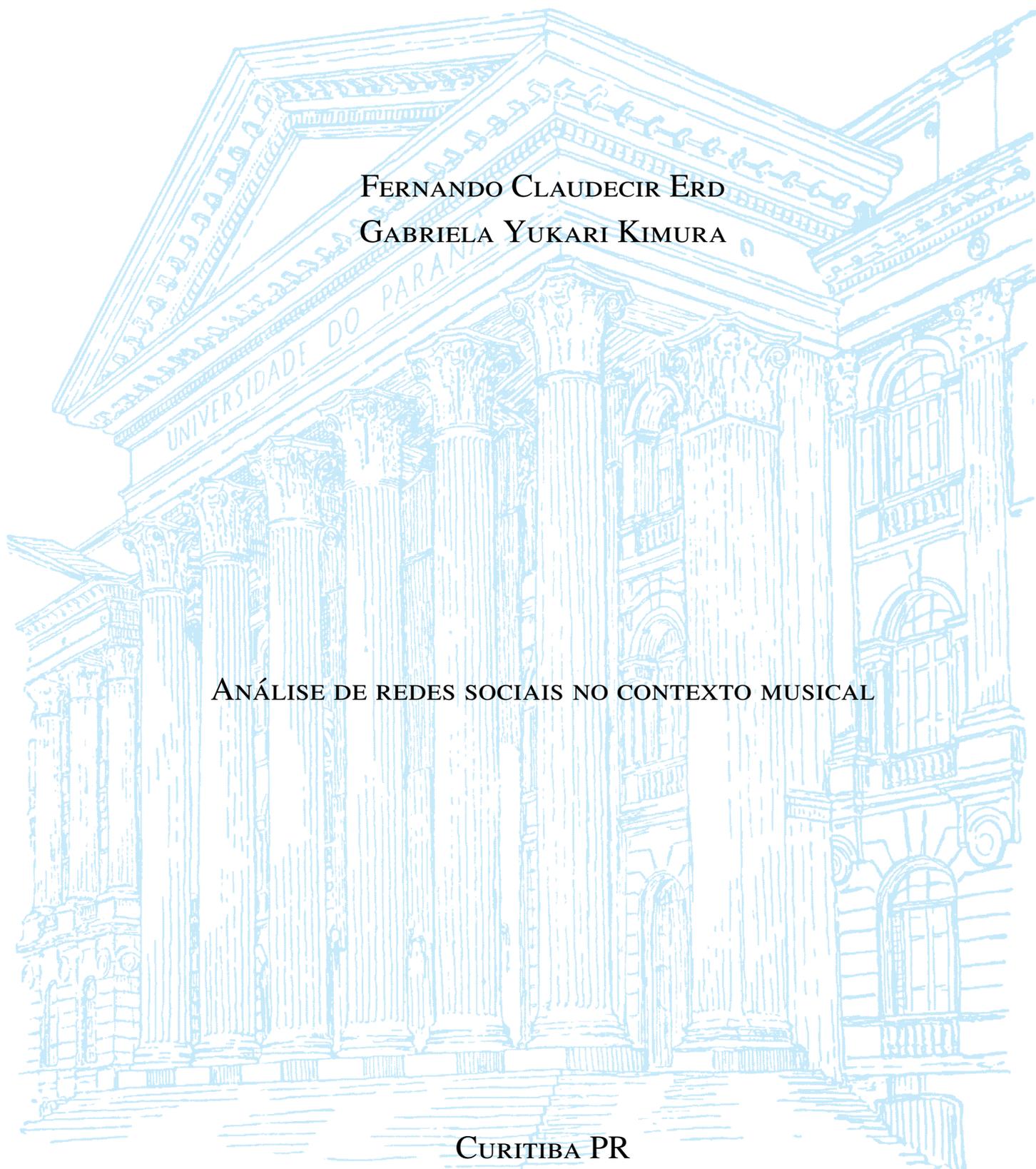


UNIVERSIDADE FEDERAL DO PARANÁ

FERNANDO CLAUDECIR ERD
GABRIELA YUKARI KIMURA

ANÁLISE DE REDES SOCIAIS NO CONTEXTO MUSICAL

CURITIBA PR
2018



FERNANDO CLAUDECIR ERD
GABRIELA YUKARI KIMURA

ANÁLISE DE REDES SOCIAIS NO CONTEXTO MUSICAL

Trabalho de Graduação em Algoritmos e Grafos apresentado ao Curso de Ciência da Computação da Universidade Federal do Paraná como requisito à obtenção do título do grau de Bacharel em Ciência da Computação.

Área de concentração: *Ciência da Computação*.

Orientador: Prof. Dr. André Luis Vignatti.

CURITIBA PR
2018

Resumo

Historicamente, a música sempre desempenhou um papel importante na formação de uma sociedade, servindo como forma de representar a arte e a maneira de pensar e agir de um povo. A evolução tecnológica modificou a maneira com que as pessoas interagem com o mundo e com tudo que o compõe. No ramo musical, por exemplo, a inserção da tecnologia permitiu que novas mídias e inteligências fossem desenvolvidas para capturar, reproduzir e distribuir músicas, possibilitando a conexão de diferentes comunidades em escala mundial. As redes sociais são uma maneira de descrever as interações de um grupo, e permitem a modelagem, em estruturas de grafos, dos relacionamentos entre artistas, produtores e compositores. Compreender os mecanismos pelos quais estes relacionamentos evoluem constitui a motivação deste projeto. Mais especificamente, a pesquisa se dedica a aplicar métricas referentes à teoria dos grafos e redes sociais a fim de traçar um comparativo em intervalos de décadas. Com base no estudo da rede social modelada, são apresentados valores que comprovam a evolução da conectividade entre os participantes que a compõem, bem como a popularidade exercida por cada um deles. Além disso, é proposta a representação de comunidades através da identificação do gênero musical predominante.

Palavras-chave: Música, Redes Sociais, Teoria dos Grafos, Fenômeno do Mundo Pequeno, Lei de Potência.

Abstract

Historically, music has always played an important role in the formation of a society, serving as a way of representing art and people's way of thinking and acting. Technological evolution has changed the way people interact with the world and with everything that surrounds it. In the musical field, for example, the insertion of technology allowed new media and intelligences to be developed to capture, reproduce and distribute music, allowing the connection of different communities on a world scale. Social networks are a way of describing the interactions of a group, and allow the modeling, in graph structures, of the relationships between artists, producers and songwriters. Understanding the mechanisms by which these relationships evolve is the motivation of this project. More specifically, the research is dedicated to applying metrics related to graph theory and social networks in order to draw a comparative in intervals of decades. Based on the study of the social network, values are presented in a way that demonstrate the evolution of the connectivity between the participants that compose it, as well as the popularity exerted by each of them. In addition, it is proposed the representation of communities through the identification of the predominant musical genre.

Keywords: Music, Social Networks, Graph Theory, Small World Phenomenon, Power Law.

Lista de Figuras

2.1	Representação gráfica de um grafo com 4 vértices e 4 arestas.	9
2.2	Exemplo de um grafo simples.	10
2.3	Exemplo de um grafo G com 4 vértices e 4 arestas.	11
2.4	Exemplo de um grafo com 4 componentes.	11
2.5	Exemplo de um grafo bipartido.	12
2.6	Rede social de colaboração de artigos com Paul Erdős (Easley e Kleinberg, 2010).	12
2.7	Identificação de comunidades em um grafo (Balakrishnan e Deo, 2006).	13
2.8	Exemplos de grafos G e H	14
2.9	Distribuição de caminhos percorridos pelas cartas do experimento de Stanley Milgram (Travers e Milgram, 1969).	16
2.10	Distribuição de lei de potência em escala linear e escala logarítmica (García-Manso e Martín-González, 2008).	16
3.1	Exemplo do grafo Artista \times Artista.	20
3.2	Exemplo do grafo Artista \times Produtor.	21
3.3	Exemplo da transformação aplicada para o grafo Artista \times Artista com compositores ou produtores em comum.	21
3.4	Resultado da coleta de dados.	22
4.1	Cinco maiores componentes apresentadas nas décadas de 1960 e 2010, sendo que cada cor representa uma componente.	24
4.2	Distribuição de graus para o grafo Artista \times Artista.	26
4.3	Maior componente das décadas de 1960 e 2010, respectivamente. Os vértices em vermelho representam os artistas e os vértices em verde representam os compositores.	28
4.4	Maior componente das décadas de 1960 e 2010, respectivamente. Os vértices em vermelho representam os artistas e os vértices em verde representam os produtores.	28
4.5	Grau médio de compositor e artista por década.	30
4.6	Grau médio de produtor e artista por década.	31
4.7	Distribuição de graus para o grafo Artista \times Compositor.	32
4.8	Distribuição de graus para o grafo Artista \times Produtor.	33
4.9	Maior componente para as décadas de 1960 e 2010. Os vértices representam os artistas e as arestas, o relacionamento entre artistas com algum compositor ou produtor em comum.	35
4.10	Distribuição de graus para o grafo Artista \times Artista com compositores ou produtores em comum.	37
4.11	Comunidades identificadas da década de 1960 do grafo Artista \times Artista com compositores ou produtores em comum.	38
4.12	Comunidades identificadas da década de 1970 do grafo Artista \times Artista com compositores ou produtores em comum.	39

4.13	Comunidades identificadas da década de 1980 do grafo Artista × Artista com compositores ou produtores em comum.	40
4.14	Comunidades identificadas da década de 1990 do grafo Artista × Artista com compositores ou produtores em comum.	41
4.15	Comunidades identificadas da década de 2000 do grafo Artista × Artista com compositores ou produtores em comum.	41
4.16	Comunidades identificadas da década de 2010 do grafo Artista × Artista com compositores ou produtores em comum.	42

Lista de Tabelas

4.1	Propriedades do grafo Artista \times Artista.	24
4.2	Dimensão da maior componente para o grafo Artista \times Artista.	25
4.3	Comparação da quantidade de artistas, compositores e produtores.	29
4.4	Propriedades do grafo Artista \times Compositor.	29
4.5	Propriedades do grafo Artista \times Produtor.	30
4.6	Dimensão da maior componente para o grafo Artista \times Compositor.	31
4.7	Dimensão da maior componente para o grafo Artista \times Produtor.	31
4.8	Propriedades dos grafos Artista \times Artista com compositores ou produtores em comum.	35
4.9	Dimensão da maior componente para o grafo Artista \times Artista com compositores ou produtores em comum.	36
4.10	Valor da <i>resolução</i> para cada década.	38
4.11	Quantidade de vértices e gêneros musicais identificados na década de 1960 com resolução de 3,0.	39
4.12	Quantidade de vértices e gêneros musicais identificados na década de 1970 com resolução de 1,5.	39
4.13	Quantidade de vértices e gêneros musicais identificados na década de 1980 com resolução de 1,6.	40
4.14	Quantidade de vértices e gêneros musicais identificados na década de 1990 com resolução de 1,5.	41
4.15	Quantidade de vértices e gêneros musicais identificados na década de 2000 com resolução de 0,6.	42
4.16	Quantidade de vértices e gêneros musicais identificados na década de 2010 com resolução de 1,2.	42

Sumário

1	Introdução	8
2	Fundamentos Teóricos	9
2.1	Grafos	9
2.1.1	Grau de um vértice	10
2.1.2	Distância	10
2.1.3	Componente	11
2.1.4	Grafo bipartido	11
2.2	Redes Sociais	12
2.2.1	Comunidades	13
2.2.2	Coefficiente de Agrupamento	13
2.2.3	Fenômeno do Mundo Pequeno	15
2.2.4	Lei de Potência	15
3	Proposta	18
3.1	Objetivo Geral	18
3.2	Objetivos Específicos	18
3.3	Metodologia	18
3.4	Modelagem	19
3.4.1	Grafo Artista \times Artista	20
3.4.2	Grafo Artista \times Produtor e Grafo Artista \times Compositor	20
3.4.3	Grafo Artista \times Artista com Compositores ou Produtores em Comum	21
3.5	Restrições	21
4	Resultados	23
4.1	Grafo Artista \times Artista	24
4.2	Grafo Artista \times Compositor e Grafo Artista \times Produtor	28
4.3	Grafo Artista \times Artista com Compositores ou Produtores em Comum	35
5	Conclusão e Trabalhos Futuros	43
	Referências	44

1 Introdução

Historicamente, a música sempre desempenhou um papel importante na formação de uma sociedade, servindo como forma de representar a arte e a maneira de pensar e agir de um povo. A evolução tecnológica, que teve início no século XX, modificou a forma com que as pessoas interagem com o mundo e com tudo que o compõe. No ramo musical, por exemplo, o aumento da popularidade do rádio permitiu que a música não ficasse limitada somente a concertos e clubes, possibilitando aos artistas a obtenção de fama nacional e, até mesmo, internacional. Da mesma forma, o público passou a estar exposto a um maior leque de opções de gêneros musicais, resultando na conexão de diferentes culturas em escala mundial.

As redes sociais podem ser representadas por estruturas matemáticas capazes de descrever as interações sociais de um grupo que compartilham valores ou objetivos comuns. Permitem identificar, através da modelagem em grafos ou em matrizes, os atores e as ligações que as constituem, bem como analisar os seus relacionamentos de forma a compreender o seu desenvolvimento. Este trabalho visa modelar uma rede social em quatro estruturas de grafos, onde os vértices representam os artistas, compositores e produtores, e as arestas, os relacionamentos entre os mesmos. Esta modelagem tem como motivação a análise dos mecanismos pelos quais os relacionamentos entre artistas, produtores e compositores evoluem.

Com o objetivo de delimitar a amostra de estudo, o presente trabalho destina analisar todas as músicas que apareceram, ao menos uma vez, no ranking Hot 100 da revista Billboard (Billboard Staff, 2011), especializada em informações da indústria musical. Mais especificamente, a pesquisa se dedica a aplicar métricas referentes à teoria dos grafos e redes sociais a fim de traçar um comparativo em intervalos de décadas, entre os anos de 1960 e 2017. A contribuição do trabalho consiste na apresentação de resultados relacionados à popularidade exercida pelos participantes que compõem o espaço musical.

Este documento foi organizado na forma de capítulos, e cada capítulo, por sua vez, é dividido em seções. O Capítulo 2 corresponde à fundamentação teórica, onde são apresentados e devidamente referenciados os princípios teóricos referentes à teoria dos grafos e redes sociais. O Capítulo 3 descreve os objetivos gerais e específicos propostos pelo trabalho, bem como as restrições e os procedimentos metodológicos e de modelagem que foram necessários para o desenvolvimento do mesmo. Em seguida, o Capítulo 4 apresenta uma análise dos grafos desenvolvidos referentes às décadas do espaço musical estudado. Por fim, o Capítulo 5 corresponde à conclusão a respeito das informações apresentadas neste documento e a sugestão de futuros trabalhos que poderiam ser realizados.

2 Fundamentos Teóricos

Este capítulo tem como objetivo apresentar os conceitos básicos necessários para um bom entendimento deste trabalho. Tais conceitos estão relacionados à teoria dos grafos e modelos de redes sociais. Apresentam-se aqui apenas os conceitos mais gerais, deixando os aspectos específicos para serem referenciados quando necessários. A Seção 2.1 apresenta os principais conceitos de teoria dos grafos e a Seção 2.2 define os conceitos de redes sociais.

2.1 Grafos

Grafos são estruturas matemáticas utilizadas para a representação de relacionamentos entre objetos. Um grafo G pode ser descrito por um conjunto de vértices ou nós $V(G) = \{v_1, v_2, \dots, v_n\}$ e um conjunto de arestas ou ligações $E(G) = \{e_1, e_2, \dots, e_m\}$, onde cada aresta é um par não ordenado de vértices de G . O número total de vértices e arestas em G são denotados por $|V(G)|$ e $|E(G)|$, respectivamente.

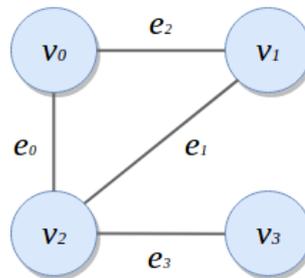


Figura 2.1: Representação gráfica de um grafo com 4 vértices e 4 arestas.

A Figura 2.1 apresenta um exemplo gráfico da composição de um grafo G , tal que:

$$\begin{aligned} V(G) &= \{v_0, v_1, v_2, v_3\} \\ E(G) &= \{e_0, e_1, e_2, e_3\} \end{aligned}$$

onde

$$\begin{aligned} e_0 &= \{v_0, v_2\} \\ e_1 &= \{v_1, v_2\} \\ e_2 &= \{v_0, v_1\} \\ e_3 &= \{v_2, v_3\} \end{aligned}$$

Além disso, um grafo H é um subgrafo de um grafo G se, e somente se, todo vértice de H é vértice de G e toda aresta de H é aresta de G , ou seja, $V(H) \subseteq V(G)$ e $E(H) \subseteq E(G)$ e, portanto, $H \subseteq G$. Assim, é dito que G contém H e que H está contido em G .

2.1.1 Grau de um vértice

Em um grafo simples G , dois vértices x e y são adjacentes (ou vizinhos) se existe uma aresta $e = \{x, y\}$ em G . Esta aresta é dita ser incidente a ambos os vértices, x e y .

Por conseguinte, o grau de um vértice v , denotado por $\delta_G(v)$, representa o número de arestas incidentes em v ou, de forma análoga, o número de vértices adjacentes à ele. Além disso, um vértice de grau 0 é denominado vértice isolado de G e denotamos como $\overline{\delta}_G$ o grau médio de G , que é calculado pela seguinte equação:

$$\overline{\delta}_G = \frac{1}{|V(G)|} \sum_{v \in V(G)} \delta_G(v)$$

Desta maneira, seja G o grafo apresentado pela Figura 2.2. O vértice v_0 apresenta duas arestas adjacentes em G e, portanto, $\delta_G(v_0) = 2$. O vértice v_1 é um vértice isolado em G , visto que $\delta_G(v_1) = 0$. E o grau médio, $\overline{\delta}_G$, é igual a 1,2.

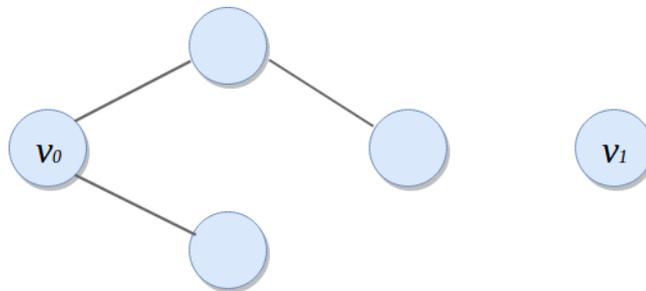


Figura 2.2: Exemplo de um grafo simples.

2.1.2 Distância

Em um grafo qualquer G , um caminho é uma sequência finita não vazia de vértices distintos $C = (v_0, v_1, \dots, v_k)$, onde $\{v_{i-1}, v_i\} \in E(G)$ para todo $1 \leq i \leq k$. Os vértices v_0 e v_k representam, respectivamente, a origem e o término do caminho C , e os vértices v_1, \dots, v_{k-1} , os vértices internos de C . O comprimento de um caminho C é denotado $|C|$, e representa o número de arestas contidas na sequência.

Dizemos ainda que a distância entre os vértices v_1 e v_2 em G , denotada por $d_G(v_1, v_2)$, é o tamanho do menor caminho de v_1 a v_2 em G , isto é:

$$d_G(v_1, v_2) = \min\{|C| \text{ tal que } C \text{ é caminho de } v_1 \text{ a } v_2 \text{ em } G\}$$

Se não existe caminho conectando dois vértices em G , ou seja, se dois vértices pertencem a diferentes componentes de G , temos que a distância entre eles é infinita, e esta distância é dita não válida. Além disso, denotaremos \overline{d}_G a distância média em um grafo G , que é calculada pela seguinte equação:

$$\overline{d}_G = \frac{1}{\binom{|V(G)|}{2}} \sum_{v \in V(G)} \sum_{w \in V(G)} d_G(v, w)$$

Para demonstrar o cálculo da distância média de um grafo, considere o grafo da Figura 2.3. A distância de cada par de vértices de G é:

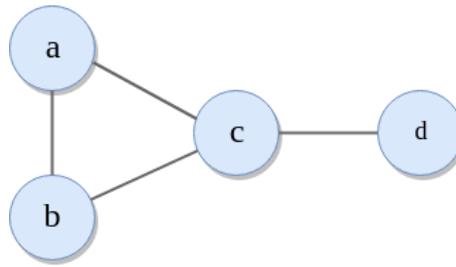


Figura 2.3: Exemplo de um grafo G com 4 vértices e 4 arestas.

$$\begin{array}{lll} d_G(a, b) = 1 & d_G(a, c) = 1 & d_G(a, d) = 2 \\ d_G(b, c) = 1 & d_G(b, d) = 2 & d_G(c, d) = 1 \end{array}$$

Logo, temos que a distância média do grafo é:

$$\begin{aligned} \overline{d_G} &= \frac{1}{6} \cdot (1 + 1 + 2 + 1 + 2 + 1) \\ &= \frac{8}{6} \\ &= 1,333 \end{aligned}$$

2.1.3 Componente

Um grafo G é dito conexo se, para todo par de vértices distintos x e y , existe em G um caminho com origem x e término y . Caso contrário, o grafo G não é conexo, e se apresenta dividido em subgrafos conexos denominados componentes de G .

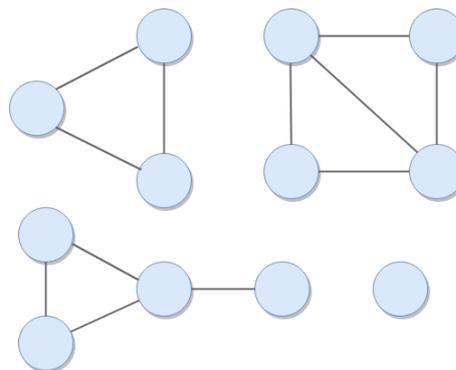


Figura 2.4: Exemplo de um grafo com 4 componentes.

2.1.4 Grafo bipartido

Um grafo G é bipartido se $V(G)$ pode ser particionado em dois subconjuntos X e Y , tal que $X \cup Y = V(G)$ e $X \cap Y = \emptyset$, de modo que toda aresta de G possui um extremo em X e outro em Y .

Vários problemas da vida real, ao serem representados por uma estrutura de grafos, são caracterizados como grafos bipartidos. Para exemplificar, suponha a seguinte situação. Sejam dois conjuntos distintos de vértices, onde no primeiro conjunto, cada vértice representa um

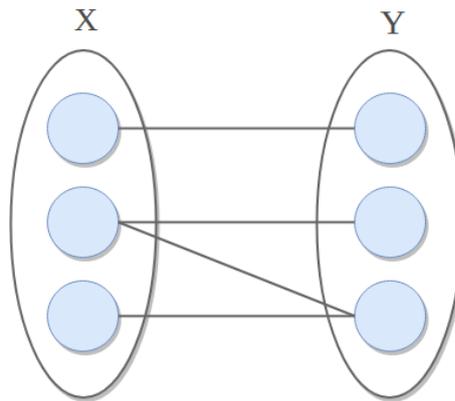


Figura 2.5: Exemplo de um grafo bipartido.

candidato à procura de emprego, e no segundo, uma vaga de emprego. Cada candidato só pode se candidatar à uma ou mais vagas de emprego. De maneira gráfica, o grafo é composto apenas por duas partições, uma delas contendo as vagas de emprego, e a outra, os candidatos, com arestas conectando ambas as partes.

2.2 Redes Sociais

As redes sociais podem ser representadas por estruturas matemáticas formadas por pessoas ou organizações que compartilham valores e objetivos comuns, e estão conectadas por diferentes tipos de relacionamentos. Permitem identificar, através da representação em grafos ou matrizes, os atores e as ligações que as compõem, bem como analisar as suas relações, de forma a compreender o seu desenvolvimento.

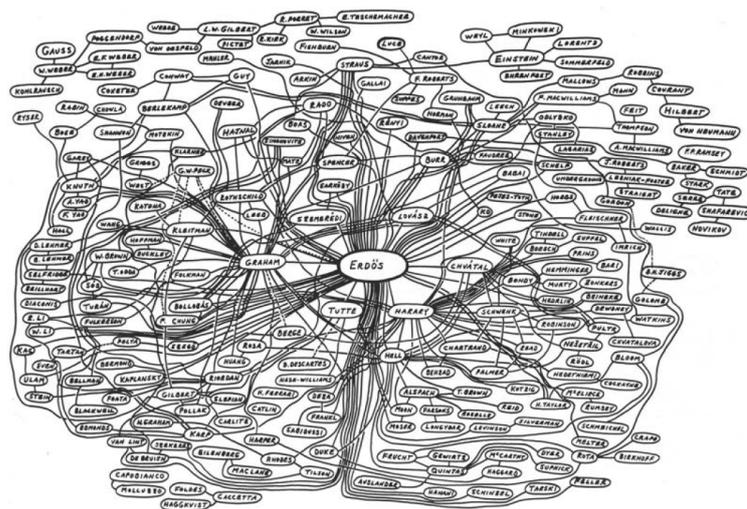


Figura 2.6: Rede social de colaboração de artigos com Paul Erdős (Easley e Kleinberg, 2010).

A importância do estudo das redes sociais está na capacidade de modelar os relacionamentos de diferentes áreas de conhecimento, existentes na natureza, na sociedade e até mesmo em um conjunto de documentos. Estas pesquisas são baseadas em características muito específicas existentes em uma rede como, por exemplo, valores de centralidade (vértice mais central) e conectividade (vértices com maior número de conexões). A partir de então, é possível analisar

e verificar o comportamento da rede social como, por exemplo, a influência de seus usuários, o potencial de alcance de seus elementos, entre outros. A Figura 2.6 ilustra o recorte de uma rede social, onde no centro está situado o famoso matemático húngaro Paul Erdős, que publicou, aproximadamente, 1500 artigos. A distância de um cientista para Erdős é conhecida como Número de Erdős.

2.2.1 Comunidades

Muitos dos problemas representados por uma estrutura de grafos apresentam comunidades, que podem ser definidas como subconjuntos de vértices (ou seja, subgrafos) cuja densidade de arestas é maior dentro de cada grupo do que entre os grupos. A Figura 2.7 apresenta um exemplo da divisão de comunidades sobre uma rede social que retrata as publicações de artigos dos pesquisadores da Santa Fé Institute.

Para identificação destes grupos, são aplicadas às redes métodos de detecção de comunidades. Por exemplo, o algoritmo proposto por Blondel et al. (2008), utilizado pela ferramenta Gephi (Gephi Website, 2018) consiste de dois passos a cada iteração. No primeiro passo, o valor da modularidade (qualidade) de uma comunidade é otimizada, permitindo apenas mudanças locais das comunidades. No segundo passo, as comunidades identificadas são agregadas para construir uma nova rede de comunidades. Estes passos são executados iterativamente até que não seja mais possível aumentar o valor de modularidade.

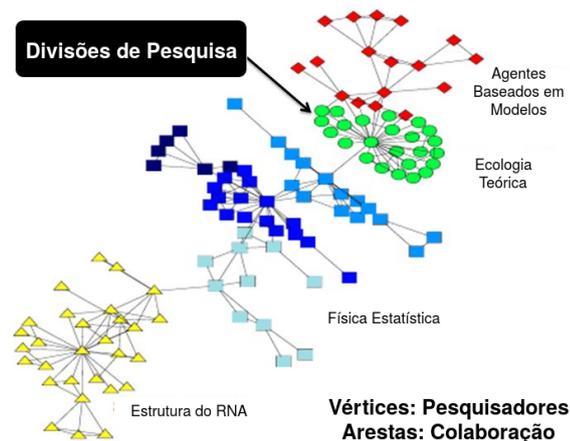


Figura 2.7: Identificação de comunidades em um grafo (Balakrishnan e Deo, 2006).

A Figura 2.7 apresenta o resultado do algoritmo de detecção de comunidades sobre a rede social que retrata as publicações de artigos dos pesquisadores da Santa Fé Institute. Cada comunidade detectada representa um campo de pesquisa.

2.2.2 Coeficiente de Agrupamento

Em teoria dos grafos, o coeficiente de agrupamento avalia o grau com que os nós de um grafo tendem a se agrupar. Em uma rede social online como a Internet, por exemplo, os indivíduos que representam os vértices, são capazes de se conectar com extrema facilidade, independentemente de suas diferenças culturais, sociais ou distância geográfica. Entretanto, as associações não tendem a ocorrer de forma aleatória, mas sim entre pessoas com afinidades semelhantes. Estas afinidades possibilitam a geração de agrupamentos.

O menor grupo possível de ser formado depende de três pessoas. Desta forma, quando duas pessoas já conectadas estão conectadas à uma terceira pessoa comum, a existência desta terceira pessoa diminui a chance destas pessoas estarem conectadas ao acaso, e torna mais forte a comunicação entre elas, uma vez que mesmo que a ligação entre elas seja rompida, ambas vão continuar recebendo as mesmas informações através do amigo comum.

O coeficiente de agrupamento é uma métrica que quantifica a transitividade destas conexões. Isto é, seja v um vértice qualquer de um grafo, o valor do coeficiente de agrupamento de v é calculado por:

$$C_v = \frac{\text{Total de arestas entre os vizinhos de } v}{\text{Número máximo possível de arestas entre os vizinhos de } v}$$

Além disso, o coeficiente de agrupamento do grafo é dado pela média do coeficiente de todos os vértices que a compõem,

$$\bar{C}_G = \frac{1}{|V(G)|} \sum_{v \in V(G)} C_v$$

Desta maneira, uma rede com alto coeficiente de agrupamento implica na existência de afinidades específicas locais, o que sugere que as conexões não tenham sido feitas aleatoriamente. Em outras palavras, se *Gabriela* e *Fernando* possuem uma amizade com *Giovanne*, é provável que *Gabriela* e *Fernando* sejam amigos entre si. As redes que apresentam o fenômeno do mundo pequeno são caracterizados por um alto valor de coeficiente de agrupamento. Entretanto em uma rede onde o grafo é bipartido, temos que $\bar{C} = 0$, pois nunca existirá arestas entre os vizinhos de v .

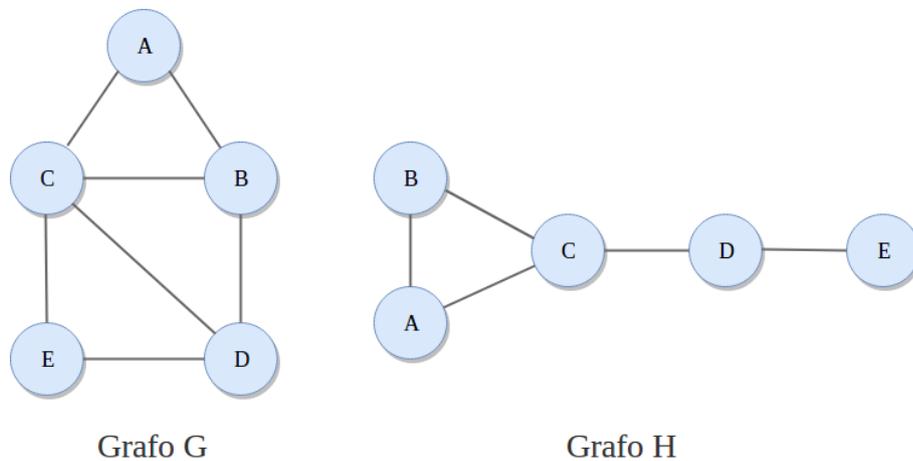


Figura 2.8: Exemplos de grafos G e H.

Para exemplificar o cálculo do coeficiente, considere os grafos da Figura 2.8. É possível notar que os vértices pertencentes ao grafo G apresentam-se mais conectados quando comparados aos vértices do grafo H , o que implica que o coeficiente de agrupamento de G é maior que o de H .

$$\begin{aligned}
\overline{C_G} &= \frac{1}{|V(G)|} \sum_{v \in V(G)} C_v \\
&= 0,20 \cdot (C_A + C_B + C_C + C_D + C_E) \\
&= 0,20 \cdot \left(\frac{1}{1} + \frac{2}{3} + \frac{3}{6} + \frac{2}{3} + \frac{1}{1} \right) \\
&= 0,20 \cdot 3,82 = 0,764
\end{aligned}$$

$$\begin{aligned}
\overline{C_H} &= \frac{1}{|V(H)|} \sum_{v \in V(H)} C_v \\
&= 0,20 \cdot (C_A + C_B + C_C + C_D + C_E) \\
&= 0,20 \cdot \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{3} + \frac{0}{1} + \frac{0}{1} \right) \\
&= 0,20 \cdot 2,33 = 0,466
\end{aligned}$$

2.2.3 Fenômeno do Mundo Pequeno

Uma rede social é dita exibir o fenômeno do mundo pequeno se, a grosso modo, quaisquer dois indivíduos da rede são conectados através de uma curta sequência de conhecidos intermediários. O conceito, proposto pelo psicólogo americano Stanley Milgram, contou com um experimento cujo objetivo era conectar pessoas desconhecidas dos Estados Unidos o mais breve possível (Milgram, 1967). Em uma das instâncias do estudo, uma pessoa do estado de Nebraska receberia uma carta que deveria entregar a outra pessoa que morava em Massachusetts. A fonte, inicialmente, receberia informações básicas sobre o alvo, incluindo seu endereço e ocupação, e seria instruída a enviar a carta a algum de seus conhecidos, de forma a alcançar o alvo de maneira mais eficaz possível. Qualquer um que recebesse subsequentemente a carta receberia as mesmas instruções, e a cadeia de comunicação continuaria até que o alvo fosse alcançado. Ao longo de vários ensaios, verificou-se que o número médio de etapas intermediárias das sequências que obtiveram sucesso, situava-se entre cinco e seis, quantidade que deu origem ao princípio de “seis graus de separação”, nome popular que denominou o fenômeno do mundo pequeno.

Este fenômeno pode ser explicado devido ao padrão estrutural apresentado pela maioria das redes sociais. A conexão de várias comunidades de diferentes dimensões, isto é, o número de vértices que as compõem, através de um número reduzido de arestas garantem à rede a maximização da conectividade ao mesmo tempo em que minimiza o número total de conexões. Consequentemente, a navegação pela rede ocorre de forma muito mais eficiente e é possível conectar elementos de diferentes comunidades muito mais rapidamente.

2.2.4 Lei de Potência

A lei de potência é uma relação matemática aplicada entre duas variáveis ou quantidades, onde a modificação de uma delas dá origem a uma alteração relativa proporcional sobre a outra. A função geral da lei de potência pode ser descrita pela seguinte equação matemática, e a sua distribuição pode ser visualizada na Figura 2.10.

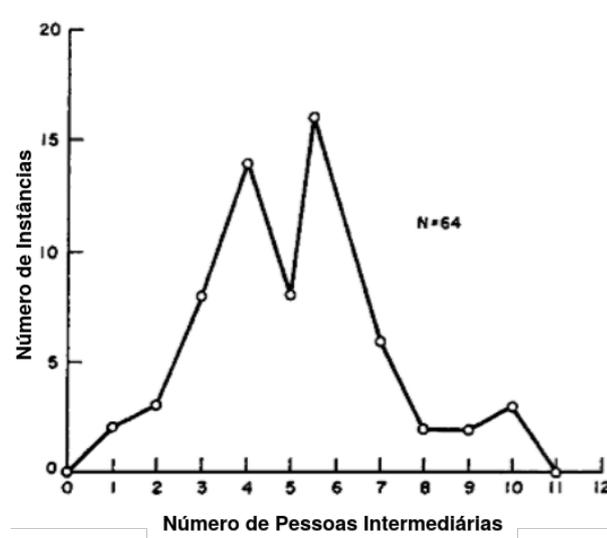


Figura 2.9: Distribuição de caminhos percorridos pelas cartas do experimento de Stanley Milgram (Travers e Milgram, 1969).

$$f(x) = ax^\alpha$$

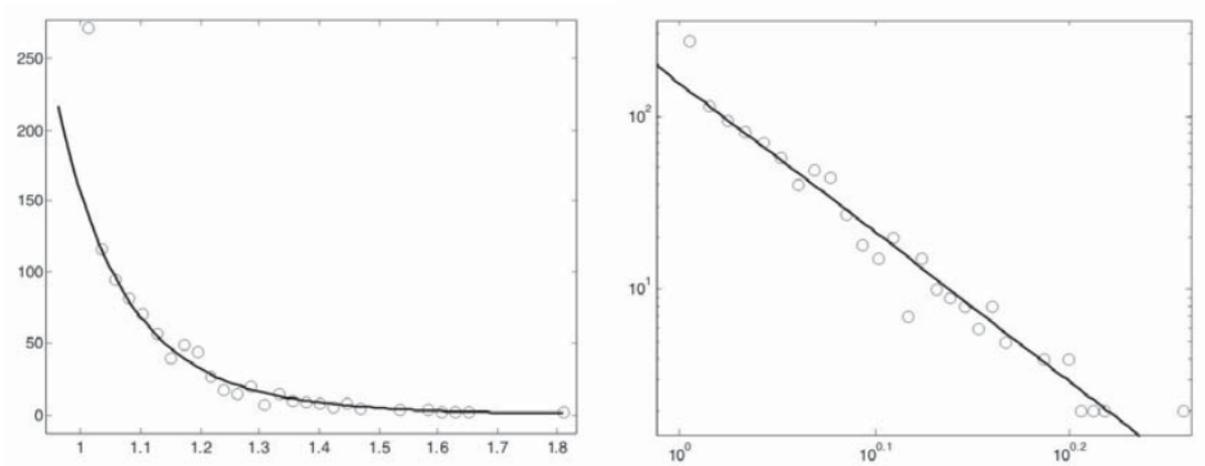


Figura 2.10: Distribuição de lei de potência em escala linear e escala logarítmica (García-Manso e Martín-González, 2008).

Esta distribuição implica que pequenas ocorrências da relação entre as variáveis analisadas são extremamente comuns, enquanto grandes quantidades são raras. Na visão de grafos, a distribuição apresenta-se na forma de poucos vértices com grau elevado, e muitos vértices com grau baixo. Outra característica da lei de potência é que ela é dita invariante em escala, isto é, se pegarmos uma amostra da distribuição, ela apresentará o mesmo formato da distribuição completa. Da mesma maneira, temos que ao multiplicar a variável da função de potência, o valor da função simplesmente aumentará ou reduzirá em um fator constante.

Ao aplicarmos o logaritmo sobre a função da lei de potência temos que:

$$\log(f(x)) = \log(c) + \alpha \cdot \log(x)$$

Logo, ao traçarmos as variáveis em um gráfico com os eixos logarítmicos, a lei de potência pode ser visualizada como uma linha reta, cuja inclinação possui valor α . Além disso, de acordo com Newman (2010), um grafo lei de potência normalmente apresenta o valor de α variando no intervalo $2 \leq \alpha \leq 3$.

3 Proposta

Este capítulo tem como finalidade apresentar os objetivos gerais e específicos propostos pelo trabalho, bem como as restrições e os procedimentos metodológicos e de modelagem que foram necessários para o desenvolvimento do mesmo.

3.1 Objetivo Geral

Este trabalho propõe-se a modelar, em uma estrutura de grafos, o universo musical dos anos de 1960 à 2017, com a finalidade de traçar um comparativo em intervalos de décadas, tendo em vista os princípios de redes sociais.

3.2 Objetivos Específicos

Para cumprir o objetivo geral proposto, fez-se necessário atingir os seguintes objetivos específicos:

- Coletar e manipular os dados fornecidos pelas ferramentas Billboard, Genius, Gephi e Spotify;
- Modelar os dados extraídos em estruturas de grafos;
- Compreender o comportamento dos grafos modelados, tendo em vista os princípios da teoria de redes sociais;
- Aplicar as métricas de teoria dos grafos e identificar as comunidades de gêneros musicais nos grafos desenvolvidos;

3.3 Metodologia

Até recentemente, o maior impedimento para a análise científica da história da música era a falta de dados. Isso mudou drasticamente com o aparecimento das tecnologias e com o surgimento de grandes coleções digitalizadas de gravações de áudio, letras e partituras. Para delimitar nossa amostra, definimos como base de estudo todas as músicas que apareceram, ao menos uma vez, no ranking Hot 100 da revista Billboard, que consiste na liberação semanal das 100 músicas mundiais mais tocadas (Billboard Staff, 2011). A escolha foi preferida devido a disponibilidade de dados desde o ano de 1958, o que possibilitaria acompanhar a evolução da música à medida que a tecnologia foi se estabelecendo.

Inicialmente, foi realizado um script em *NodeJS* para a extração anual de todos os nomes das músicas rankeadas na coluna Hot 100 da Billboard e os artistas que as interpretaram. Como

os dados extraídos não apresentavam um padrão de separação dos artistas que interpretaram músicas em conjunto, foi necessário estabelecer as strings que indicavam estas separações. Neste caso, foram consideradas as seguintes strings: ",", "&", "/", "with", "duet with", "and", "feat.", "featuring", "+" e "&&". Alguns exemplos podem ser observados abaixo.

- David Guetta **Featuring** Justin Bieber
- Flo Rida **&** 99 Percent
- Taylor Swift **Featuring** Ed Sheeran **&** Future
- Justin Bieber + BloodPop
- Kid Ink, Tyga, Wale, YG **&** Rich Homie Quan

Posteriormente, com a utilização da mídia social Genius (Genius Website, 2009), foram coletados os dados musicais como compositor, produtor, data de lançamento, álbum e gravadora de cada música encontrada na Billboard. É oferecido uma chave de autenticação, após o cadastro na plataforma Genius, para realizar consultas na API para desenvolvedores (Genius Developers, 2018). Dessa forma, cada música da Billboard foi verificada quanto a existência de seu artista no Genius. Caso o artista existisse, era feita uma requisição das músicas do artista correspondente existentes na plataforma e verificado se a música procurada pertencia à lista de músicas retornada. Por fim, era realizada uma nova requisição para a coleta das informações musicais citadas inicialmente. Contudo, os únicos atributos que apresentaram maior relevância às conclusões deste trabalho foram o de compositor e de produtor.

Realizada a extração destes dados, as músicas foram divididas em períodos de décadas, de acordo com o ano em que tiveram sua aparição na plataforma Billboard (Billboard Staff, 2011). Em seguida, os dados foram formatados em arquivos de entrada do software Gephi (Gephi Website, 2018), onde foi possível obter a representação gráfica em grafos e coletar as informações relacionadas às métricas de grau médio, distância média, coeficiente de agrupamento e tamanho de componente.

Por fim, foram obtidos os gêneros musicais de cada artista do grafo através do Spotify, serviço de música digital que oferece o acesso a milhões de músicas e outros conteúdos sobre artistas do mundo inteiro (Spotify Website, 2018). A plataforma oferece uma API para desenvolvedores (Spotify API, 2018), onde realizada uma requisição com o nome de um artista, são retornadas informações sobre o mesmo. Estes dados foram necessários para a determinação do gênero musical mais predominante em cada comunidade encontrada nos grafos apresentados pela Seção 4.3 do Capítulo 4.

Para o cálculo do valor α apresentado nos gráficos da Seção 4 que representam a distribuição de lei de potência, foi implementado um algoritmo utilizando a biblioteca *powerlaw* para a linguagem de programação *Python*. A biblioteca foi proposta por Alstott et al. (2014) e calcula métricas relacionadas a distribuição de lei de potência como, por exemplo, o valor de inclinação da reta, que foi utilizada neste trabalho.

3.4 Modelagem

Para a modelagem dos dados extraídos em estruturas de grafos, os membros da rede social, sejam eles os artistas, produtores ou compositores, foram representados por vértices, e os relacionamentos existentes entre eles, por meio de arestas. Além disso, foram propostos três tipos de estruturas de grafos, descritos nas próximas subseções.

3.4.1 Grafo Artista \times Artista

Na primeira modelagem, os dados coletados foram normalizados e as arestas indicam as colaborações realizadas entre dois ou mais artistas, isto é, as músicas que foram elaboradas em conjunto.

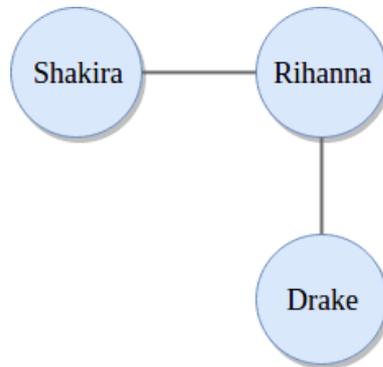


Figura 3.1: Exemplo do grafo Artista \times Artista.

A Figura 3.1 apresenta um exemplo desta relação. Neste caso, as artistas *Rihanna* e *Drake* elaboraram a música *Work* em conjunto e por isso aparecem conectadas por uma aresta. O mesmo acontece para as artistas *Rihanna* e *Shakira* na colaboração da música *Can't remember to forget you*.

3.4.2 Grafo Artista \times Produtor e Grafo Artista \times Compositor

Nesta modelagem foi utilizada a estrutura de um grafo bipartido, onde uma das partições representa os artistas e a outra, os produtores ou compositores, de acordo com grafo a ser analisado. A conexão entre as partições representa o trabalho entre artistas e compositores ou artistas e produtores. A Figura 3.2 apresenta um exemplo do trabalho entre os artistas *Drake* e *Jay-Z*, com o produtor *Kanye West*, e os artistas *Jay-Z* e *Rick Ross*, com o produtor *DJ Khaled*.

Vale ressaltar que um artista pode existir em ambas as partições, uma vez que um artista pode, além de interpretar, compor e produzir músicas. Desta forma, decidimos considerá-los indivíduos diferentes, isto é, o mesmo indivíduo aparecerá no grafo em uma das partições como artista, e na outra partição, como compositor ou produtor. A Figura 3.2 ilustra esta situação para o grafo Artista \times Produtor, onde o artista *Kanye West* aparece no grafo como produtor, uma vez que produziu músicas dos artistas *Drake* e *Jay-Z*.

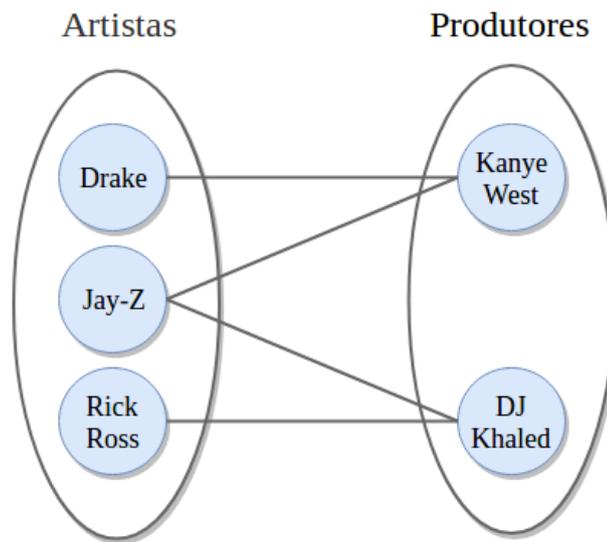


Figura 3.2: Exemplo do grafo Artista \times Produtor.

3.4.3 Grafo Artista \times Artista com Compositores ou Produtores em Comum

A terceira modelagem apresenta a relação dos artistas que trabalharam com compositores ou produtores em comum. Para exemplificar, considere as artistas *Ariana Grande* e *Avril Lavigne*. Como estas artistas trabalharam com o mesmo compositor *Max Martin*, elas aparecem conectadas no grafo Artista \times Artista com Compositores ou Produtores em Comum. A Figura 3.3 ilustra esta relação.

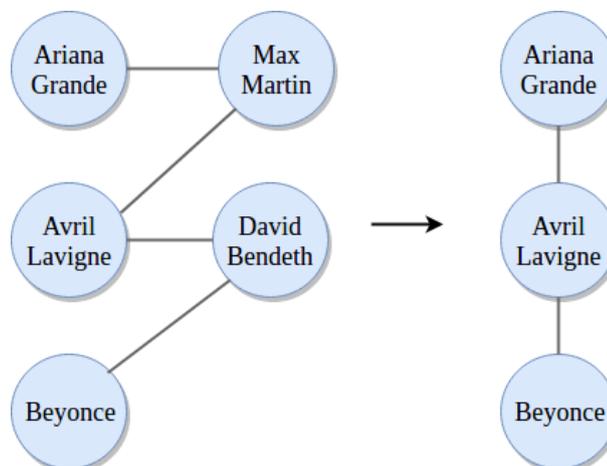


Figura 3.3: Exemplo da transformação aplicada para o grafo Artista \times Artista com compositores ou produtores em comum.

3.5 Restrições

A ferramenta Genius é uma mídia social que disponibiliza ao público uma variedade de informações relacionadas às músicas que são produzidas mundialmente. Para este trabalho, foram analisados os dados apresentados pela plataforma entre os anos de 1958 à 2017. Entretanto, como pode ser visualizado na Figura 3.4, boa parte das informações relacionadas aos compositores e produtores dos anos de 1958 e 1959 não estavam disponíveis no Genius. Desta forma, foi

decidido pela desconsideração destes anos, de forma a não comprometer as conclusões extraídas pelo trabalho.

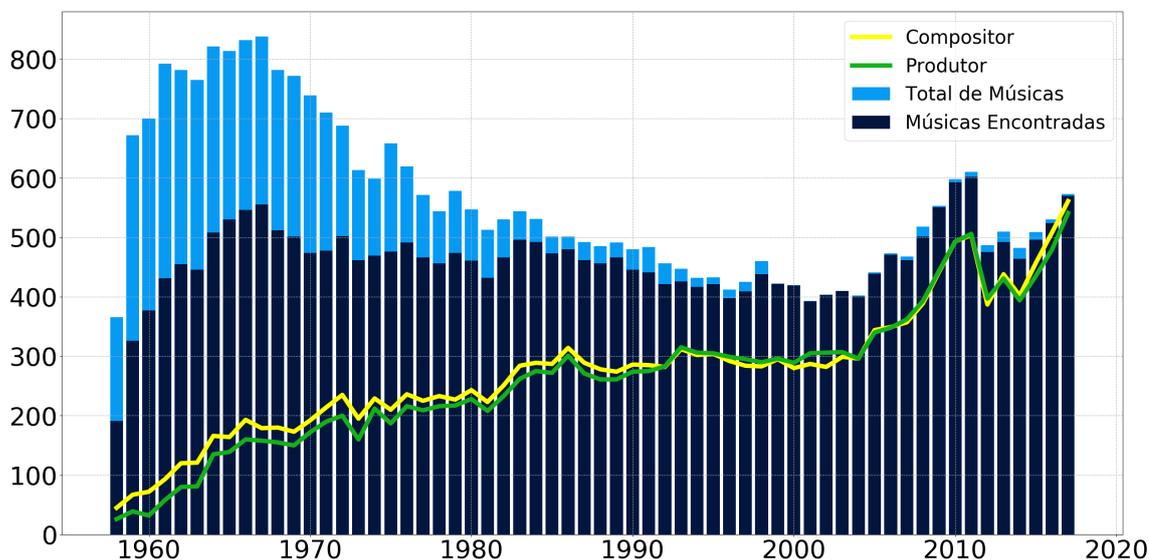


Figura 3.4: Resultado da coleta de dados.

A Figura 3.4 apresenta o desempenho do processo de extração de dados das plataformas auxiliares. A cor azul claro representa o total de músicas que apareceram na plataforma Billboard, e a azul escuro, o total de músicas cujas informações foram encontradas no Genius. Além disso, são representados pelas cores amarela e verde, respectivamente, o total de músicas em que foram encontrados os compositores e produtores, sendo estes os dados de maior relevância a pesquisa. É possível notar também que, nas décadas de 1960 e 1970, não foi possível encontrar informações suficientes em relação aos compositores e produtores, resultando somente, em cerca de 40% do total de músicas. Desta forma, os resultados obtidos podem não condizer completamente com a realidade destes anos.

Outra restrição surgiu durante o processo de identificação de comunidades dos grafos Artista \times Artista com Compositores ou Produtores em Comum. Nesta situação, foram detectadas muitas comunidades pequenas, como o caso em que um artista trabalhava apenas com um produtor/compositor durante toda a sua carreira, sendo representado por um único vértice. Como o trabalho apresenta como objetivo uma compreensão mais geral destes grafos, essas informações específicas foram evitadas, e somente as cinco maiores comunidades de cada década foram analisadas.

Ainda em relação às comunidades, um artista pode representar vários gêneros musicais diferentes. Segundo o Spotify, a banda *Coldplay*, por exemplo, possui os gêneros musicais: *permanent wave* e *pop*. Neste caso, a banda possui apenas dois gêneros musicais. Entretanto, se considerarmos agora a banda *Pink Floyd*, são apresentados sete gêneros musicais diferentes, são eles: *album rock*, *art rock*, *classic rock*, *progressive rock*, *psychedelic rock*, *rock* e *symphonic rock*. Desta forma, para a determinação do gênero musical de uma comunidade, foram considerados somente os cinco gêneros musicais mais frequentes entre os artistas que a compõem.

4 Resultados

Neste capítulo serão abordados os resultados obtidos a partir da análise dos grafos modelados de 1960 à 2017. Cada grafo foi avaliado segundo métricas obtidas da ferramenta Gephi (Gephi Website, 2018), são elas: grau de vértice, grau médio, distância média, tamanho de maior componente e coeficiente de agrupamento.

O grau de vértice, grau médio e o tamanho da maior componente permitem analisar, em média, o comportamento do relacionamento entre artistas, artistas e produtores, e artistas e compositores, e como estes estão conectados. A partir disso, é possível verificar a influência exercida por alguns produtores e compositores e a tendência de colaborações na elaboração de músicas rankeadas no Hot 100.

A distância média e o coeficiente de agrupamento permitem verificar a presença de laços fortes nos grafos do tipo Artista \times Artista, e compreender os mesmos de acordo com o fenômeno do mundo pequeno. Além disso, o agrupamento dos vértices em comunidades possibilita dividir os grafos em gêneros musicais e compará-los ao longo das décadas.

A Seção 4.1 traz uma análise sobre o grafo do tipo Artista \times Artista. A Seção 4.2 apresenta um comparativo entre os grafos Artista \times Compositor e Artista \times Produtor. Por fim, a Seção 4.3 apresenta o estudo do grafo do tipo Artista \times Artista com compositores ou produtores em comum, com ênfase nas comunidades identificadas.

4.1 Grafo Artista \times Artista

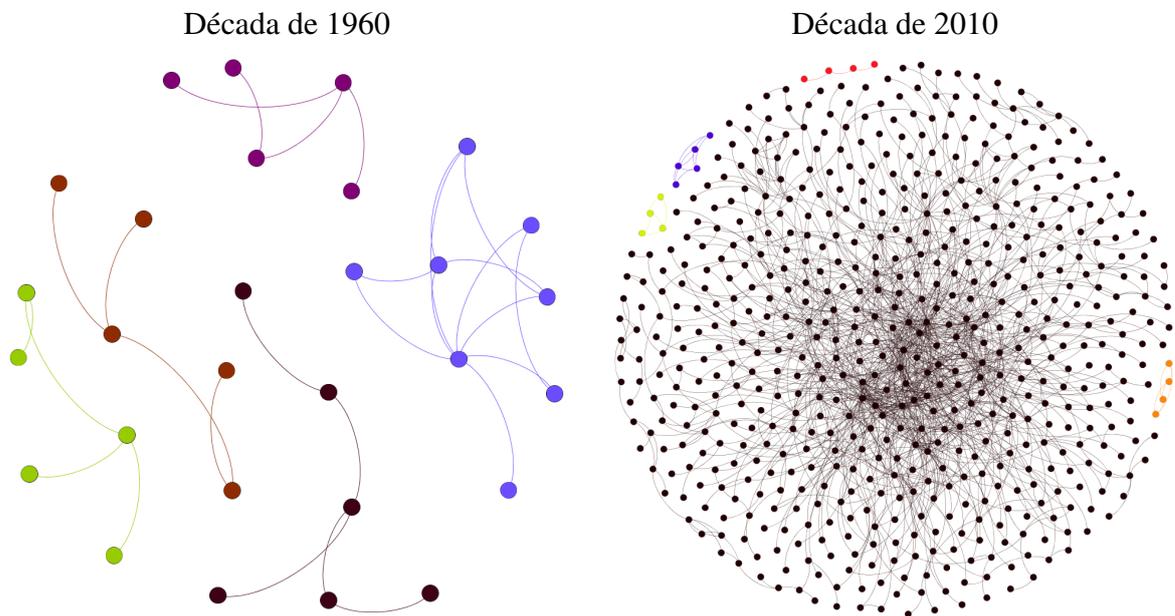


Figura 4.1: Cinco maiores componentes apresentadas nas décadas de 1960 e 2010, sendo que cada cor representa uma componente.

A inserção da tecnologia no espaço musical proporcionou aos ouvintes uma extensa variedade de opções para encontrar novas músicas, bem como a facilidade de acesso às mesmas. Desta forma, esperava-se como resultado que, com o decorrer do tempo, houvesse uma maior volatilidade nos artistas em destaque. Contudo, ao comparar os dados de vértices apresentados pela Tabela 4.1, é possível perceber que, historicamente, houve uma queda no número de novos artistas classificados no Hot 100 da Billboard. Isto demonstra a dificuldade de popularização de novos artistas, de forma que aqueles que já obtiveram sucesso tendem a manter sua popularidade por mais tempo.

Além disso, a partir da análise da Figura 4.1, torna-se evidente o crescimento da quantidade de arestas presentes nos grafos, principalmente a partir da década de 1990, conforme apresentado pela Tabela 4.1. Isto indica a tendência de músicas colaborativas, de forma que os artistas passaram a trabalhar com um conjunto maior de artistas. As métricas de grau médio e dimensão da maior componente do grafo reafirmam esta tendência.

Década	Vértices	Arestas	Grau Médio	Distância Média	Agrupamento
1960	2078	377	0,363	1,243	0,528
1970	1901	351	0,369	1,288	0,725
1980	1542	263	0,341	3,068	0,384
1990	1695	653	0,771	5,743	0,599
2000	1336	1251	1,873	4,181	0,520
2010	1161	1583	2,727	4,275	0,515

Tabela 4.1: Propriedades do grafo Artista \times Artista.

A Tabela 4.1 demonstra a evolução dos valores de grau médio dos grafos Artista \times Artista. Entre as décadas de 1960 à 1980, somente um a cada três artistas realizava alguma

colaboração, ao passo que, a partir da década de 2000, o grau médio alcança um valor em torno de 1,8, o que equivale a cada artista trabalhando com, em média, dois outros artistas. Analisando agora a Tabela 4.2, sobre a dimensão da maior componente do grafo Artista \times Artista, podemos observar que, inicialmente, as componentes eram bastante isoladas, com a maior componente da década de 1960 representando apenas 0,38% de todo o grafo. Já na década de 2010, 90% das arestas já apresentam-se concentradas na maior componente do grafo, com mais de 50% dos artistas conectados em virtude das colaborações na elaboração de músicas novas.

Uma possível suposição seria que os artistas que ganharam popularidade sejam de gêneros que tendem a favorecer colaborações como, por exemplo, *Rap* e *Pop*. Entretanto, mais dados seriam necessários para sustentar tal afirmação, sendo um tema possível para futuras pesquisas.

Década	Vértices	% de Vértices	Arestas	% de Arestas
1960	8	0,38	12	3,18
1970	9	0,47	11	3,13
1980	30	1,95	33	12,55
1990	240	14,16	402	61,56
2000	502	37,47	1065	85,13
2010	590	50,82	1454	91,85

Tabela 4.2: Dimensão da maior componente para o grafo Artista \times Artista.

A Tabela 4.1 apresenta também os valores de distância média. Nas primeiras décadas, os artistas não tinham o costume de trabalhar em conjunto, o que resultou na reduzida quantidade de arestas dos grafos correspondentes e os baixos valores de distância média, em função do aparecimento de múltiplas componentes de pequenas dimensões. A partir da década de 1990, o número de arestas aumentou significativamente, sendo possível obter valores mais precisos da distância média. Na década de 2000, por exemplo, a distância média apresentou valor igual a 4,181. Logo, sejam dois artistas distintos a_1 e a_2 , o valor de distância média indica que é possível caminhar pelo grafo partindo do artista a_1 até o artista a_2 através de apenas 4 “passos”. O restante das décadas também apresentaram valores de distância média similares, o que sugere a presença do fenômeno do mundo pequeno, que diz que uma rede social é dita exibir o fenômeno do mundo pequeno se a média das distâncias não ultrapassa um número pequeno de vértices.

Conforme apresentado na Seção 2.2.2 do Capítulo 2, o coeficiente de agrupamento demonstra o grau com que os vértices tendem a se agrupar. Quanto maior o valor desse coeficiente, maiores serão as conexões locais da rede analisada. É possível notar, segundo a Tabela 4.1, que os valores de distância média foram acompanhadas por um alto valor de coeficiente de agrupamento, o que reafirma a presença do fenômeno do mundo pequeno nos grafos Artista \times Artista.

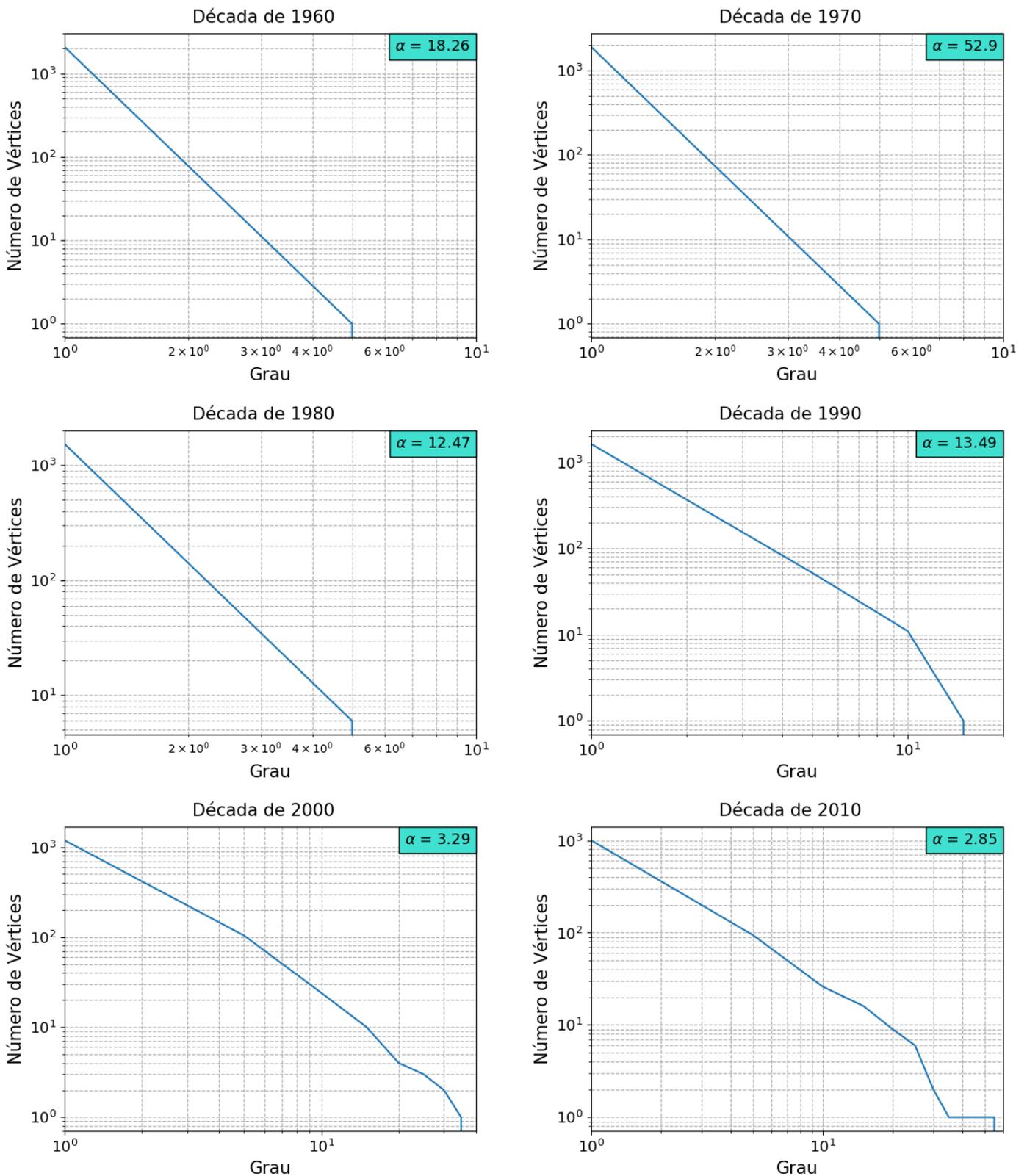


Figura 4.2: Distribuição de graus para o grafo Artista \times Artista.

Conforme definido na Seção 2.2.4 do Capítulo 2, temos que uma evidência da distribuição de lei de potência está presente na rede se o gráfico log-log apresenta um formato de reta e o nível de inclinação da reta, o valor α , se apresenta próximo do intervalo $2 \leq \alpha \leq 3$. Visualizando os gráficos da Figura 4.2, é possível perceber que até a década de 1990, apesar da apresentação em

formato de reta, os valores de α demonstraram-se muito acima do intervalo esperado, podendo ter ocorrido devido a pequena disponibilidade de arestas nestas décadas. A partir de 2000, o número de arestas presentes nos grafos aumentaram, como pode ser observado na Tabela 4.1, garantindo uma maior precisão nos resultados. Foi observado que apesar da progressão no valor de grau médio, são poucos os vértices mais influentes sobre o restante da rede. Isso pode ser confirmado através dos valores α apresentados nas décadas de 2000 e 2010 e o gráfico com formato próximo à uma reta, o que presume a presença da distribuição de lei de potência.

4.2 Grafo Artista \times Compositor e Grafo Artista \times Produtor

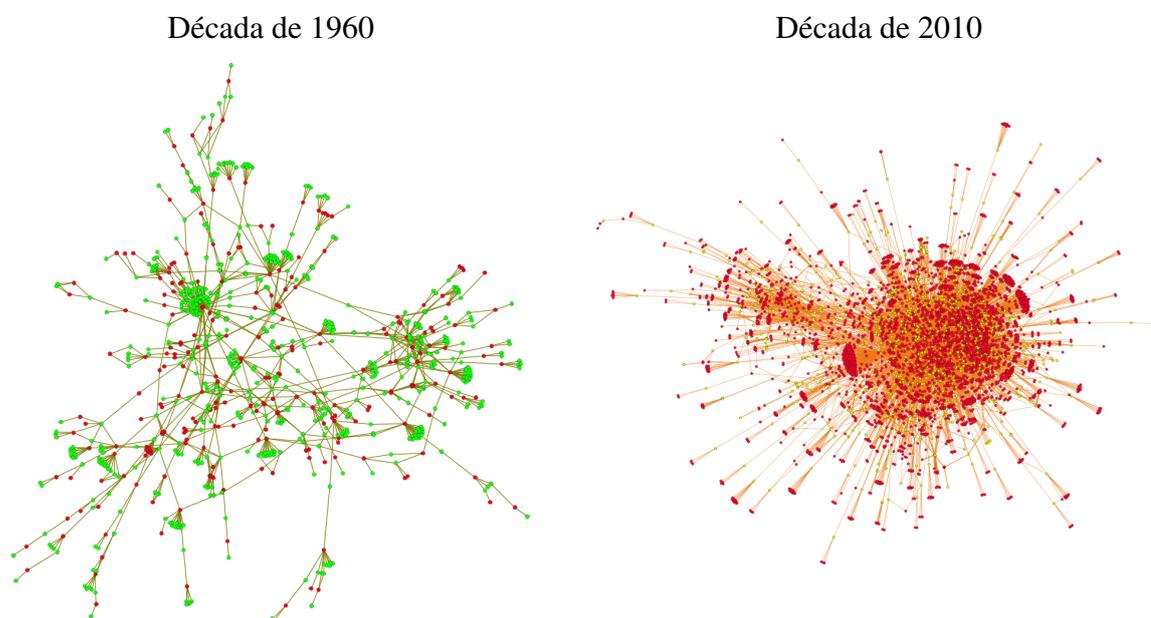


Figura 4.3: Maior componente das décadas de 1960 e 2010, respectivamente. Os vértices em vermelho representam os artistas e os vértices em verde representam os compositores.

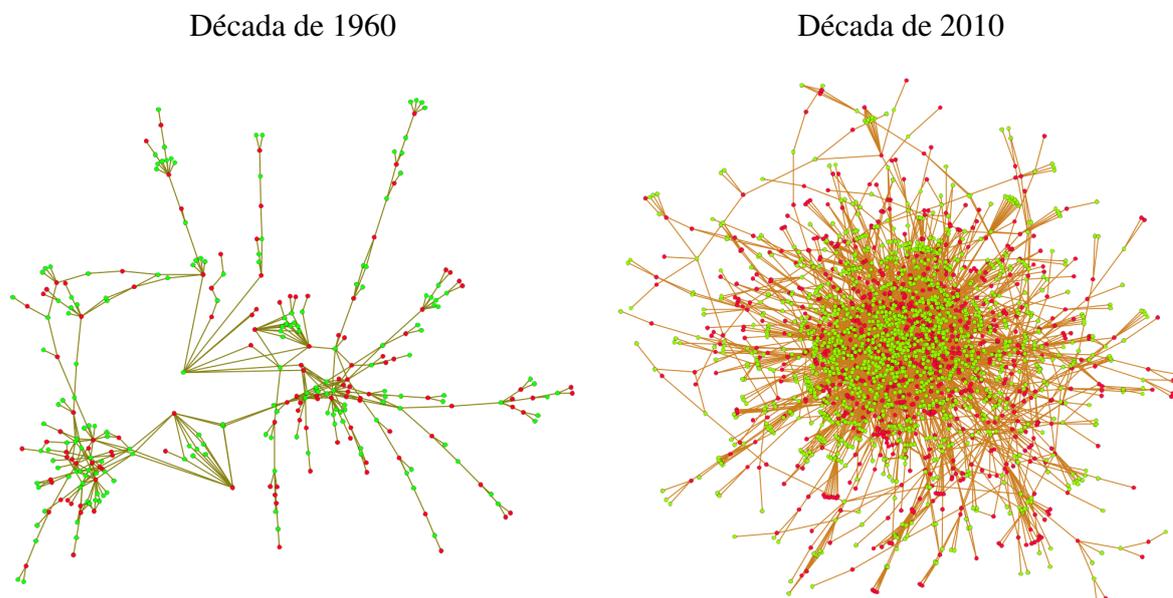


Figura 4.4: Maior componente das décadas de 1960 e 2010, respectivamente. Os vértices em vermelho representam os artistas e os vértices em verde representam os produtores.

As Figuras 4.3 e 4.4 ilustram, respectivamente, os grafos Artista \times Compositor e Artista \times Produtor, sendo os artistas representados pela cor vermelha em ambas as figuras. Visualmente, é possível notar que a densidade de arestas do grafo Artista \times Compositor é muito maior que a do grafo Artista \times Produtor na década de 1960. Isto pode ter ocorrido em decorrência da

diferença de estabilidade das duas profissões, uma vez que a profissão de produtor musical é relativamente nova, estabelecida entre as décadas de 1940 e 1950 (Gerstein, 2011), quando comparada a profissão de compositor, que apresenta registros desde a época de 200 AC (Landels, 1999).

Como já discutido na Seção 4.1, o número de artistas que conseguem se inserir no ranking Hot 100 da Billboard vem decaindo. Em contrapartida, ao analisar a Tabela 4.3, é possível perceber que a quantidade de compositores apresentou um crescimento gradativo ao longo das décadas, o que reafirma o cenário de dificuldade para que as músicas obtenham sucesso, de forma que se faz necessário uma maior participação de compositores durante a elaboração das mesmas. Realizando a mesma análise sobre o grafo Artista \times Produtor, a quantidade de produtores não demonstrou uma oscilação significativa nos períodos estudados, sendo pressuposto a característica de instabilidade apresentada pela profissão, como já mencionado anteriormente.

Década	Artistas	Compositores	Produtores
1960	2078	1018	487
1970	1901	1505	890
1980	1542	1960	1085
1990	1695	2828	1527
2000	1336	3199	1400
2010	1161	3995	1719

Tabela 4.3: Comparação da quantidade de artistas, compositores e produtores.

Uma observação mais profunda é realizada quando separamos e analisamos as partições de cada grafo, individualmente. As Figuras 4.5 e 4.6 demonstram a evolução do valor de grau médio de cada partição dos grafos Artista \times Compositor e Artista \times Produtor, nesta ordem.

De acordo com a Figura 4.5, temos que o grau médio dos compositores não demonstraram grande variação, resultando em média, em um compositor trabalhando com dois ou três artistas diferentes. Contudo, ao analisar a partição do artista, o crescimento significativo no número de compositores com quem os artistas trabalhavam é evidente, totalizando em dez compositores diferentes trabalhando com um artista, na década de 2010. Isto indica que, apesar do aumento da participação de compositores na elaboração das músicas, houve também o aumento na quantidade de novos compositores, de forma que cada compositor apresentou um número constante de colaborações.

Década	Vértices	Arestas	Grau Médio	Distância Média
1960	3096	1665	1,076	7,205
1970	3406	2174	1,277	8,647
1980	3502	2978	1,701	7,769
1990	4523	5135	2,271	7,077
2000	4535	7827	3,452	5,713
2010	5156	11893	4,613	4,826

Tabela 4.4: Propriedades do grafo Artista \times Compositor.

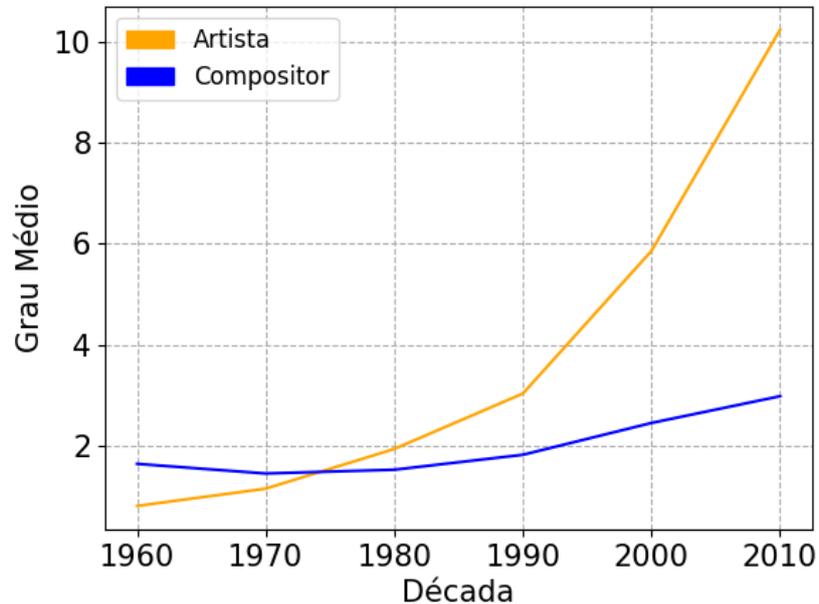


Figura 4.5: Grau médio de compositor e artista por década.

Em relação a Figura 4.6, o comportamento dos artistas e produtores se apresentou similar ao apresentado na Figura 4.5. Os produtores não apresentaram muita alteração no número de artistas com quem trabalhavam, mas quando considerada a partição dos artistas, o crescimento demonstrou-se significativo, como é o caso de um artista trabalhando com quatro ou cinco produtores na década de 2010. Entretanto, a única diferença é que a quantidade de novos produtores não demonstrou um acréscimo significativo como apresentado pelos compositores. Desta forma, devido ao aumento na participação dos produtores, isso indica que um número restrito de produtores se destacaram e aumentaram a quantidade de suas colaborações.

Década	Vértices	Arestas	Grau Médio	Distância Média
1960	2565	826	0,644	8,055
1970	2791	1437	1,030	10,226
1980	2627	1920	1,462	8,253
1990	3222	2850	1,769	7,642
2000	2736	3622	2,648	6,162
2010	2880	5192	3,606	5,319

Tabela 4.5: Propriedades do grafo Artista × Produtor.

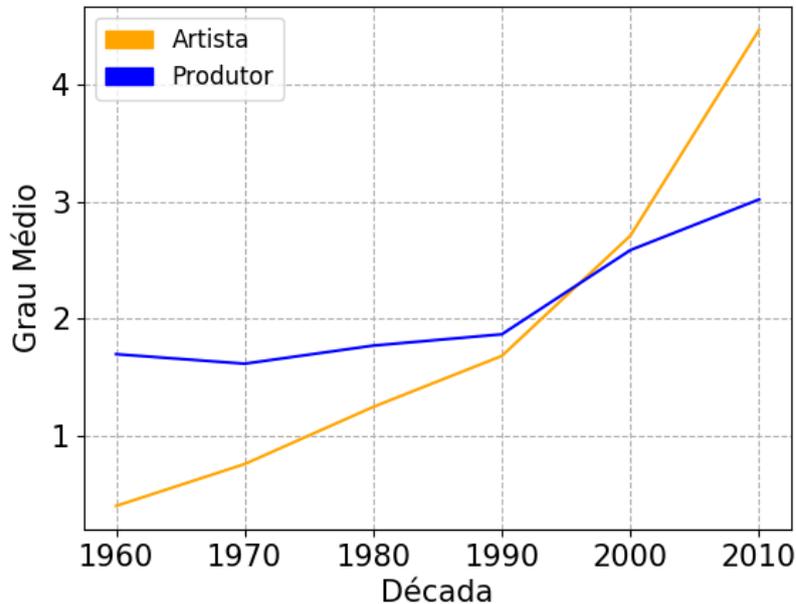


Figura 4.6: Grau médio de produtor e artista por década.

Outra métrica que confirma a maior conectividade entre artistas e compositores, e artistas e produtores, é a dimensão das maiores componentes. As Tabelas 4.6 e 4.7 demonstram este comportamento. Na década de 1960, 26% dos vértices do relacionamento entre artistas e compositores apresentavam-se concentrados na maior componente, enquanto que para o relacionamento entre artistas e produtores, eram somente 12% dos vértices. Entretanto, com o passar das décadas, esta diferença foi decaindo, a medida que os produtores foram aumentando a sua participação.

Década	Vértices	% de Vértices	Total de Arestas	% de Arestas
1960	810	26,16	1121	67,33
1970	984	28,89	1233	56,72
1980	1489	42,52	2003	67,26
1990	2504	55,36	3990	77,70
2000	3381	74,77	7293	92,03
2010	4460	86,50	11442	96,21

Tabela 4.6: Dimensão da maior componente para o grafo Artista × Compositor.

Década	Vértices	% de Vértices	Arestas	% de Arestas
1960	311	12,12	415	50,24
1970	648	23,22	755	52,54
1980	1092	41,57	1379	71,82
1990	1526	47,36	2071	72,67
2000	1894	69,23	3246	89,62
2010	2258	78,40	4883	94,05

Tabela 4.7: Dimensão da maior componente para o grafo Artista × Produtor.

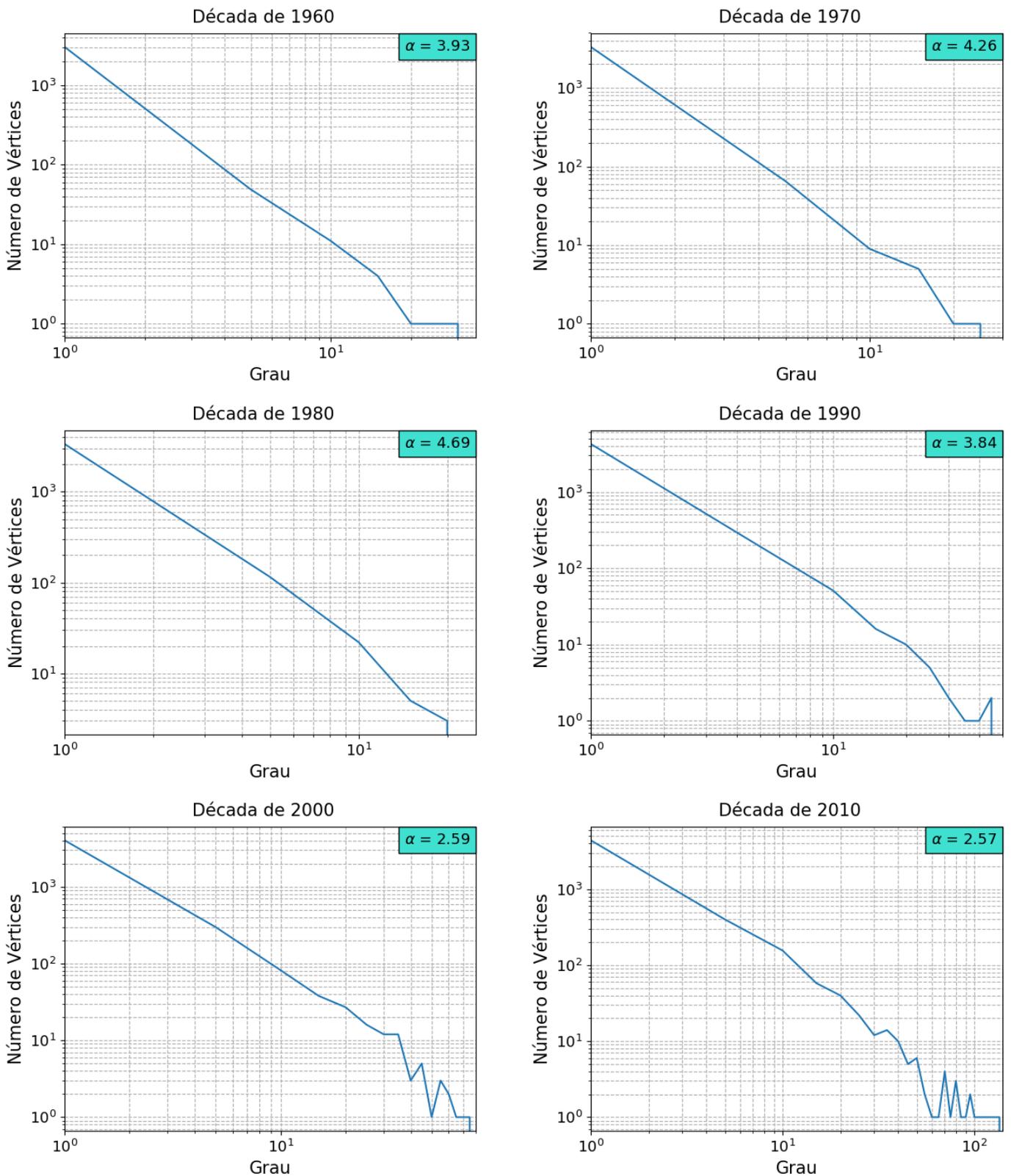


Figura 4.7: Distribuição de graus para o grafo Artista \times Compositor.

Os gráficos logarítmicos do relacionamento entre artistas e compositores são representados pela Figura 4.7, e apresentam a distribuição do número de vértices segundo o valor de grau. É possível perceber que os valores α não demonstraram uma grande variação como apresentada nos grafos da Seção 4.1. Além disso, todos os gráficos demonstraram um formato próximo de

reta e valores α muito próximos do intervalo esperado, principalmente quando analisadas as décadas de 2000 e 2010. Desta forma, é possível presumir a presença da lei de potência nos grafos Artista \times Compositor.

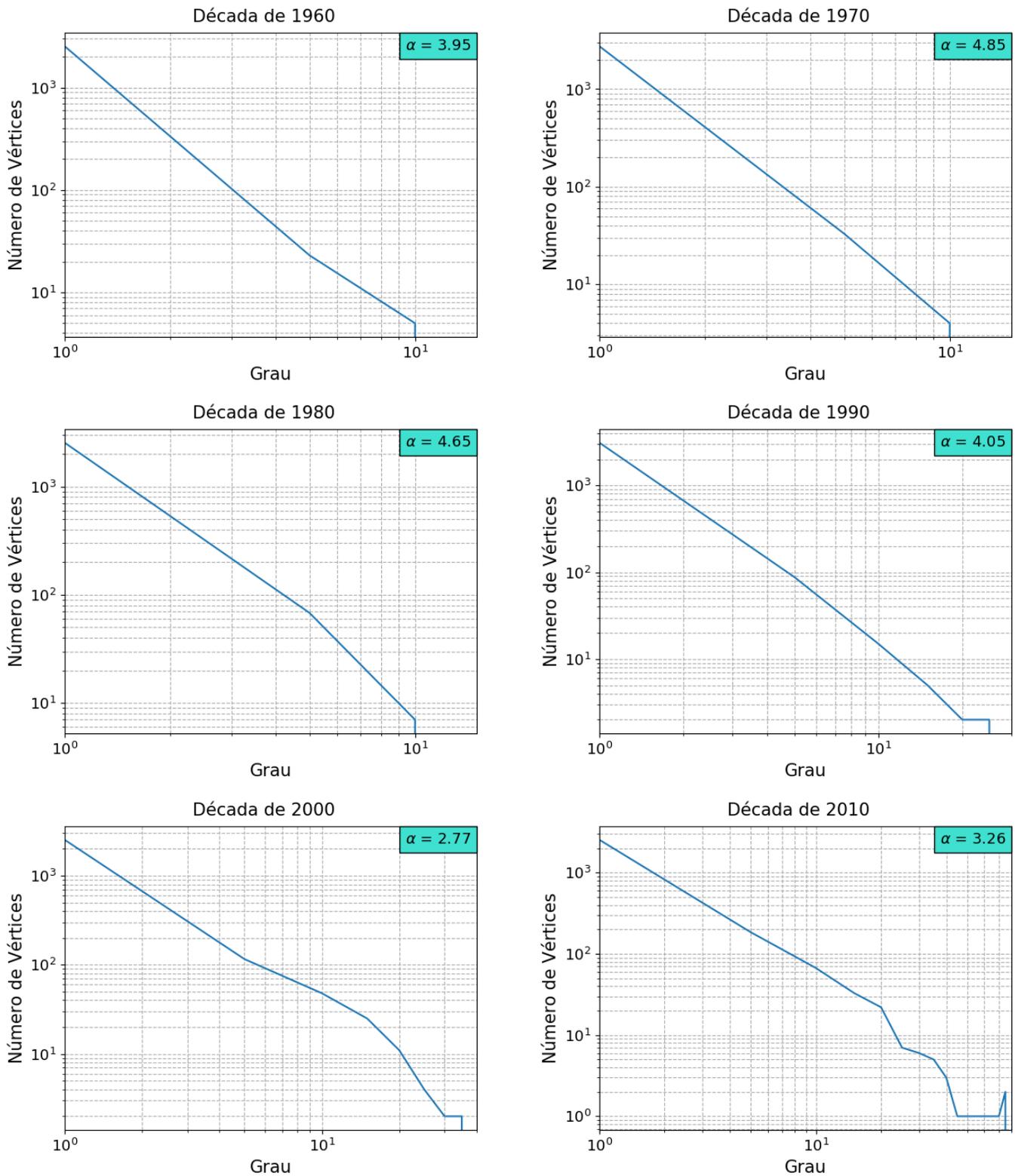


Figura 4.8: Distribuição de graus para o grafo Artista \times Produtor.

Os gráficos logarítmicos do relacionamento entre artistas e produtores são representados pela Figura 4.8, e apresentam a distribuição do número de vértices segundo o valor de grau. É possível perceber que os valores α apresentam um comportamento semelhante ao apresentado nos gráficos anteriores, no relacionamento entre artistas e compositores. Todos os gráficos demonstraram um formato próximo de reta e valores α próximos do intervalo esperado, principalmente quando analisadas as décadas de 2000 e 2010. Desta forma, é possível presumir a presença da lei de potência nos grafos Artista \times Produtor. Isto conclui que em ambos os relacionamentos, seja entre artistas e compositores, ou artistas e produtores, somente um grupo restrito de compositores e produtores exercem uma influência significativa na elaboração de músicas rankeadas no Hot 100 da Billboard.

4.3 Grafo Artista \times Artista com Compositores ou Produtores em Comum

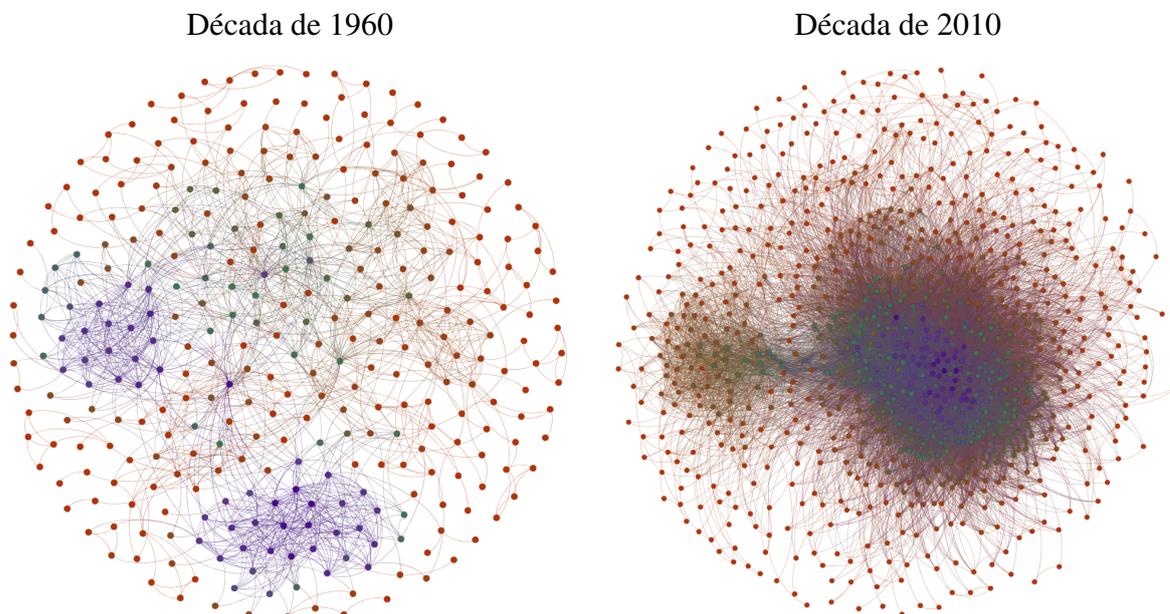


Figura 4.9: Maior componente para as décadas de 1960 e 2010. Os vértices representam os artistas e as arestas, o relacionamento entre artistas com algum compositor ou produtor em comum.

A Figura 4.9 apresenta, respectivamente, os grafos das décadas de 1960 e 2010, e representam os relacionamentos entre os artistas que trabalharam com algum compositor ou produtor em comum. Informações adicionais sobre os grafos podem ser acompanhadas na Tabela 4.9. É possível perceber que, apesar da quantidade de vértices permanecer a mesma em relação a Tabela 4.1 da Seção 4.1, a quantidade de arestas demonstrou um crescimento expressivo em todas as décadas analisadas, exibindo uma porcentagem de 90% de concentração nas componentes dos grafos e totalizando em 24 mil unidades somente nos anos entre 2010 e 2017. Isto indica que a maioria dos artistas que obtiveram sucesso trabalharam com os mesmos produtores ou compositores, e que estes compõem um grupo restrito bastante influente na elaboração de músicas rankeadas no Hot 100 da Billboard.

Década	Vértices	Arestas	Grau Médio	Distância Média	Agrupamento
1960	2078	1482	1,426	3,746	0,715
1970	1901	1824	1,919	3,875	0,710
1980	1542	3157	4,095	3,449	0,668
1990	1695	7227	8,527	3,363	0,688
2000	1336	18606	27,853	2,697	0,699
2010	1161	24584	42,350	2,381	0,732

Tabela 4.8: Propriedades dos grafos Artista \times Artista com compositores ou produtores em comum.

Agora analisando a Tabela 4.8, é possível notar que os valores de distância média não sofreram muitas alterações, iniciando com o valor 3,746 na década de 1960 e finalizando com 2,381 na década de 2010. Uma média de 3,251 de distância média entre as décadas. Em outras palavras, são necessárias, em média, 3,251 arestas intermediárias a fim de conectar quaisquer dois vértices pertencentes à rede. Além disso, todas as décadas foram acompanhadas por um alto valor de coeficiente de agrupamento, o que garante que estas conexões não tenham ocorrido de forma aleatória e que, segundo a Seção 2.2.3 do Capítulo 2, a rede apresenta as propriedades pertencentes ao fenômeno do mundo pequeno.

Década	Vértices	% de Vértices	Arestas	% de Arestas
1960	332	15,98	1437	96,96
1970	444	23,36	1697	93,04
1980	564	36,58	3076	97,43
1990	820	48,38	7118	98,49
2000	913	68,34	18573	99,82
2010	906	78,04	24549	99,86

Tabela 4.9: Dimensão da maior componente para o grafo Artista \times Artista com compositores ou produtores em comum.

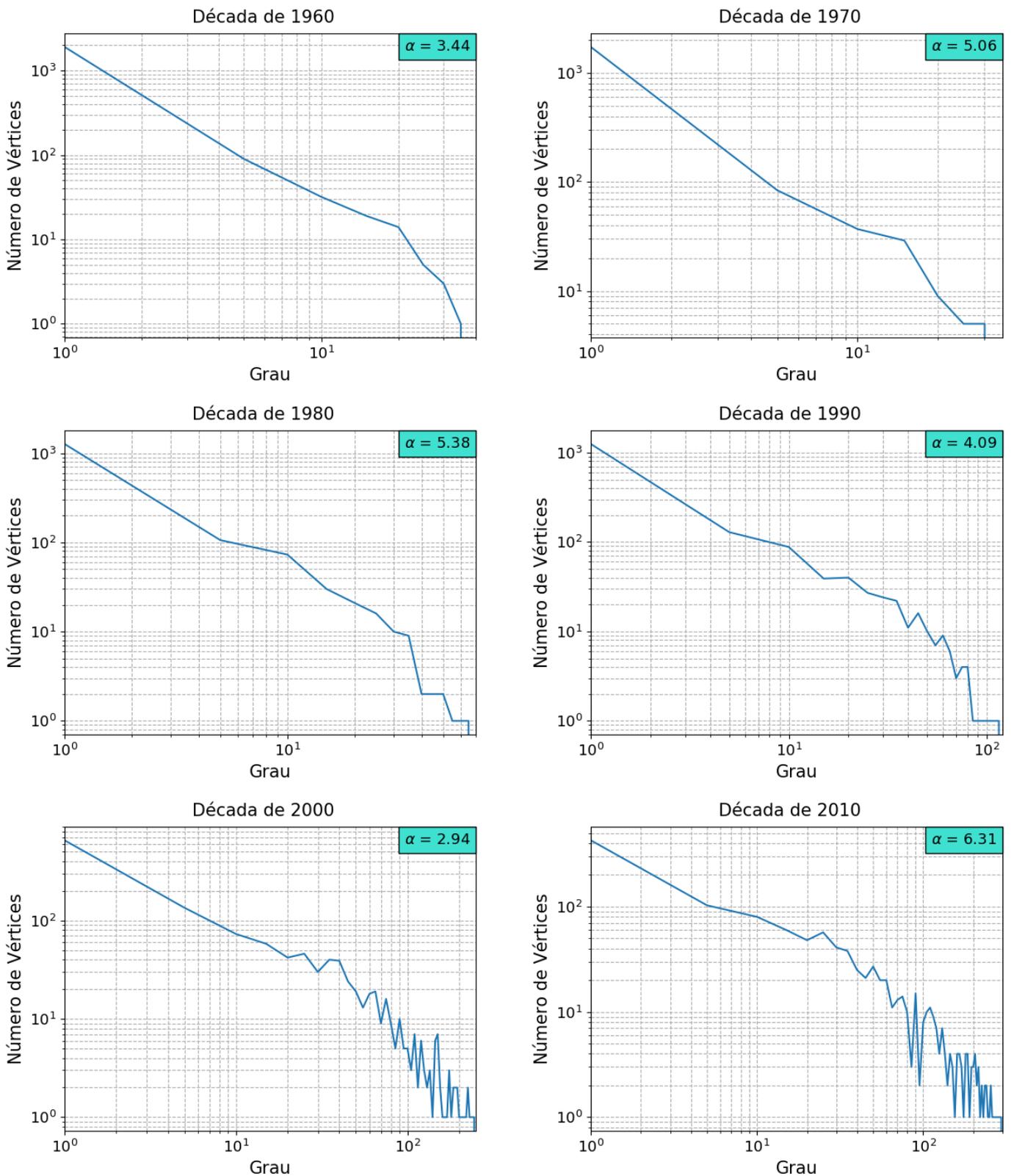


Figura 4.10: Distribuição de graus para o grafo Artista \times Artista com compositores ou produtores em comum.

Ao interpretar os gráficos gerados para a detecção de evidências da distribuição de lei de potência, exibido pela Figura 4.10, constata-se que na década de 1960 ocorre a presença de vértices com grau maior que 10, o que não ocorre nas outras modelagens de grafos apresentados

nesse trabalho. Isso indica que o grafo está altamente conectado nos anos iniciais, com um alto volume de arestas. Ao interpretar o valor da variável α , percebe-se que apenas a década de 2000 está dentro do intervalo especificado na Seção 2.2.4, o que sugere evidências da distribuição de lei de potência para esta década.

Além disso, o crescimento contínuo dos valores de grau médio, que pode ser visualizado na Tabela 4.8, demonstra a forte conexão existente entre os artistas através de seus relacionamentos com os compositores ou produtores com quem trabalharam. Este dado motivou o questionamento em relação ao impacto que a escolha de um compositor ou produtor possui sobre o estilo da música que será produzida, de forma que fosse possível dividir a rede social em grupos de artistas altamente conectados, e identificá-los segundo um gênero musical específico.

Conforme descrito na Seção 3.4, o processo de detecção de comunidades foi realizada através da ferramenta Gephi. O algoritmo disponível fornece um parâmetro denominado *resolução*, inicialmente pré-definido em 1,0, que interfere no tamanho e na quantidade de comunidades que serão identificadas. Quanto maior o valor deste parâmetro, menor será a quantidade e maior será o tamanho destas comunidades. Foi necessário ajustar o valor deste parâmetro com o propósito de limitar o número de comunidades, assim como agrupá-las com a maior quantidade de gêneros similares. A Tabela 4.10 apresenta os valores escolhidos a serem aplicados nos grafos analisados.

Década	<i>Resolução</i>
1960	3,0
1970	1,5
1980	1,6
1990	1,6
2000	0,6
2010	1,2

Tabela 4.10: Valor da *resolução* para cada década.

As próximas figuras representam as comunidades identificadas nos grafos, em décadas, que apresentaram dimensão maior que 50 vértices. Cada figura acompanha uma tabela contendo a quantidade de vértices ou artistas que compõem cada comunidade encontrada, e os cinco gêneros musicais mais frequentes entre estes artistas.

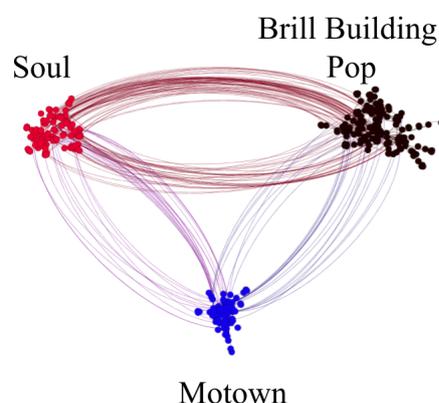


Figura 4.11: Comunidades identificadas da década de 1960 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
143	brill building pop	folk rock	bubblegum pop	adult standards	christmas
90	soul	southern soul	classic soul	folk rock	brill building pop
61	motown	soul	classic soul	brill building pop	funk

Tabela 4.11: Quantidade de vértices e gêneros musicais identificados na década de 1960 com resolução de 3,0.

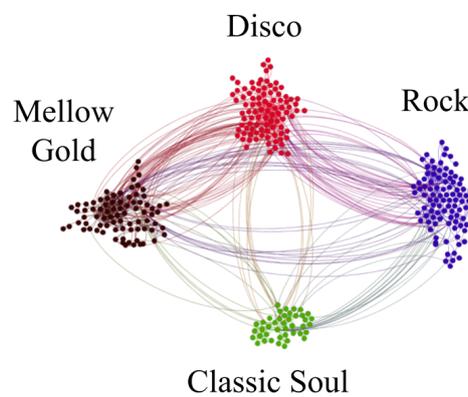


Figura 4.12: Comunidades identificadas da década de 1970 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
93	rock	folk rock	mellow gold	soft rock	classic rock
81	disco	motown	classic soul	funk	soul
79	mellow gold	soft rock	rock	album rock	classic rock
62	classic soul	motown	soul	funk	quiet storm

Tabela 4.12: Quantidade de vértices e gêneros musicais identificados na década de 1970 com resolução de 1,5.

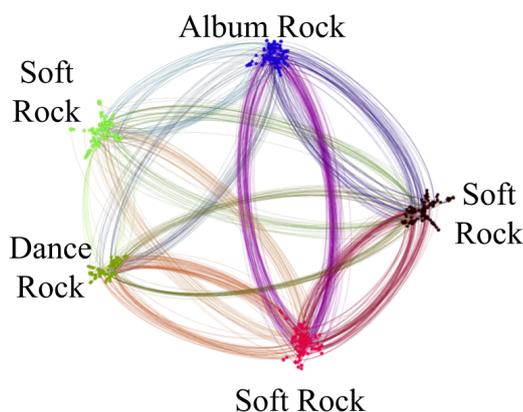


Figura 4.13: Comunidades identificadas da década de 1980 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
100	soft rock	new wave pop	mellow gold	rock	album rock
94	soft rock	new wave pop	quiet storm	dance rock	mellow gold
92	album rock	rock	soft rock	classic rock	new wave pop
92	soft rock	mellow gold	quiet storm	rock	album rock
68	dance rock	new wave	rock	quiet storm	new wave pop

Tabela 4.13: Quantidade de vértices e gêneros musicais identificados na década de 1980 com resolução de 1,6.

As primeiras décadas, que incluem os anos entre 1960 e 1989, foram marcadas pela limitação da disponibilidade de dados relacionados aos artistas, compositores e produtores. Isto pode ser visualizado pela baixa quantidade de vértices e arestas que compõem os grafos destas décadas quando comparados ao restante do intervalo analisado. Desta forma, era esperado que a falta de dados gerassem alguma interferência no processo de simplificação e identificação de um gênero musical específico. Entretanto, ao verificar a Tabela 4.11 é possível observar que as comunidades já se dividiam em comunidades com um gênero musical predominante, apenas com um ou outro gênero musical irregular, como é o caso da maior comunidade da década de 1960, contendo 143 artistas, classificado por três gêneros musicais relacionados ao gênero musical pop, e dois gêneros musicais irregulares: *rock* e *christmas*.

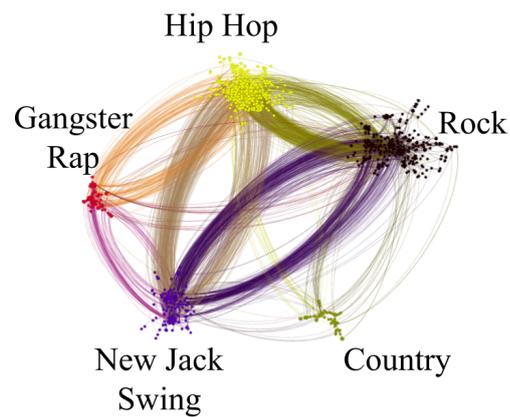


Figura 4.14: Comunidades identificadas da década de 1990 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
265	rock	pop rock	dance pop	urban contemporary	soft rock
251	hip hop	rap	hardcore hip hop	gangster rap	hip pop
111	new jack swing	urban contemporary	quiet storm	r&b	rap
69	gangster rap	west coast rap	hip hop	rap	g funk
67	country	contemporary country	country road	country dawn	country rock

Tabela 4.14: Quantidade de vértices e gêneros musicais identificados na década de 1990 com resolução de 1,5.

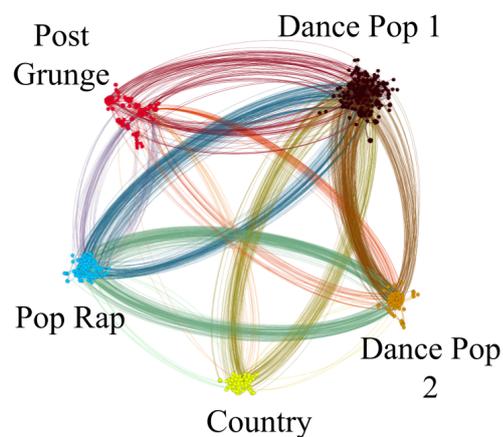


Figura 4.15: Comunidades identificadas da década de 2000 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
166	dance pop	post-teen pop	pop	pop rock	neo mellow
93	post-grunge	alternative metal	nu metal	rock	pop rock
76	pop rap	hip hop	rap	gangster rap	dirty south rap
69	country	contemporary country	country road	modern country rock	country dawn
55	dance pop	hip pop	pop rap	r&b	rap

Tabela 4.15: Quantidade de vértices e gêneros musicais identificados na década de 2000 com resolução de 0,6.

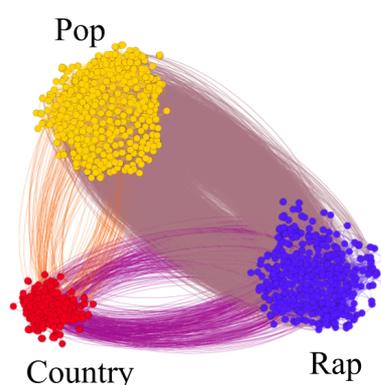


Figura 4.16: Comunidades identificadas da década de 2010 do grafo Artista \times Artista com compositores ou produtores em comum.

Vértices	1º	2º	3º	4º	5º
393	pop	dance pop	post-teen pop	pop rap	r&b
346	rap	pop rap	pop	trap music	hip hop
140	contemporary country	country road	country	modern country rock	country dawn

Tabela 4.16: Quantidade de vértices e gêneros musicais identificados na década de 2010 com resolução de 1,2.

A partir de 1990, uma maior concentração de informações relacionadas aos artistas, compositores e produtores eram acessíveis, o que contribuiu na elaboração de uma conclusão mais precisa. É o caso da primeira maior comunidade da década de 2000, contendo 166 artistas, e os cinco gêneros musicais identificados relacionados ao gênero musical pop. Entretanto, ainda houveram algumas comunidades que continham gêneros musicais que diferenciavam dos demais, fato ocasionado devido a manipulação manual dos valores de *resolução* da ferramenta Gephi, a fim de identificar comunidades de maiores dimensões mas em menores quantidades. Desta forma, alguns artistas acabaram sendo alocados em comunidades que não correspondiam ao seu estilo musical, mas que de alguma forma melhor a representavam, segundo o algoritmo utilizado pelo Gephi.

5 Conclusão e Trabalhos Futuros

Até recentemente, o maior impedimento para um relato científico da história da música estava na falta de informações. Isso mudou drasticamente com a evolução tecnológica e o surgimento de grandes coleções digitalizadas de gravações de áudio, partituras e letras. Além disso, o desenvolvimento de novas mídias e inteligências permitiram a reprodução e distribuição de músicas, possibilitando a conexão de diferentes culturas em escala mundial.

A grande importância do estudo das redes sociais reside na sua capacidade de modelar os relacionamentos de diferentes áreas de conhecimento, e analisar o comportamento dos membros que as compõem. O estudo desenvolvido compreendeu a aplicação de métricas referentes à teoria dos grafos e redes sociais a fim de traçar um comparativo, em intervalos de décadas, dos relacionamentos entre artistas, compositores e produtores classificados pela Billboard entre os anos de 1960 e 2017. A contribuição do trabalho consistiu na apresentação de resultados referentes à popularidade exercida por alguns dos artistas, compositores e produtores estudados.

As informações obtidas através da modelagem em grafos dos dados musicais demonstraram a evolução da conectividade entre os artistas, produtores e compositores, de forma que foi possível detectar a presença do fenômeno do mundo pequeno em boa parte das décadas, principalmente quando analisadas as últimas três décadas. Isto demonstrou a tendência das colaborações entre artistas na elaboração de músicas de sucesso.

Além disso, foi possível observar que, a partir da década de 1990, apesar do aumento significativo da quantidade de arestas nos relacionamentos entre artistas e compositores, e artistas e produtores, a maioria delas se concentrava em um número pequeno de vértices, o que demonstrou a grande influência exercida por um grupo restrito de compositores e produtores na elaboração de músicas rankeadas no Hot 100 da Billboard. Este fato acompanhado dos valores α e pela disposição em reta dos gráficos correspondentes sugeriram a presença da lei de potência.

Por fim, o trabalho propôs a detecção de comunidades através da identificação do gênero musical predominante entre os cinco gêneros musicais mais frequentes. Nas primeiras décadas, devido à pouca quantidade de informações disponíveis, era esperado que esta identificação não ocorresse de forma eficiente. Contudo, os resultados mostraram o contrário. Desde as décadas de 1960 e 1970, os trabalhos dos compositores e produtores já podiam ser divididos e identificados por meio dos gêneros musicais.

Um dos maiores desafios enfrentados durante o desenvolvimento deste projeto foi a realização da coleta dos dados musicais, visto que a maioria dos sites não apresentava todas as informações pertinentes à pesquisa, sucedendo a utilização de diferentes sites. Além disso, as informações encontradas não se apresentavam normalizadas, sendo exigido um tratamento especial sobre estes dados.

Para trabalhos futuros, é sugerido o questionamento sobre a popularidade de artistas influenciada por gêneros musicais que tendem a favorecer a realização de colaborações. Outra sugestão seria a realização de uma análise nacional, a partir da extração dos dados musicais do Billboard Brasil (Billboard Brasil Website, 2018), verificando se os resultados apresentados por este trabalho seguem um mesmo padrão.

Referências

- Alstott, J., Bullmore, E. e Plenz, D. (2014). powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(1): e85777e.
- Balakrishnan, H. e Deo, N. (2006). Detecting communities using bibliographic metrics.
- Billboard Brasil Website (2018). Billboard brasil - sobre. <https://billboard.uol.com.br/sobre>. Acessado em 09/11/2018.
- Billboard Staff (2011). Billboard about us page. <https://www.billboard.com/articles/events/467859/about-us>. Acessado em 13/10/2018.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. e Lefevre, E. (2008). Fast unfolding of communities in large networks.
- Easley, D. e Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, página 38. Cambridge University Press.
- García-Manso, J. e Martín-González, J. (2008). Leis de potência ou escala: sua aplicação ao fenômeno esportivo. *Fit Perf J.* 2008 mai-jun;7(3):195-202.
- Genius Developers (2018). Genius developers. <https://genius.com/developers>. Acessado em 03/12/2018.
- Genius Website (2009). Genius - about us. <https://genius.com/Genius-about-genius-annotated>. Acessado em 13/10/2018.
- Gephi Website (2018). Gephi about us. <https://gephi.org/about/>. Acessado em 13/10/2018.
- Gerstein, Y. (2011). Music production - a short historical overview.
- Landels, J. G. (1999). *Music in Ancient Greece and Rome*, página 252. Routledge.
- Milgram, S. (1967). The small world problem. *Psychology Today*, vol. 1, no. 1.
- Newman, M. (2010). *Networks An Introduction*, página 248. Oxford University Press Inc.
- Spotify API (2018). Web api spotify. <https://developer.spotify.com/documentation/web-api/>. Acessado em 03/12/2018.
- Spotify Website (2018). What is spotify? https://support.spotify.com/is/using_spotify/the_basics/what-is-spotify/. Acessado em 13/10/2018.
- Travers, J. e Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 32(4):425-443.