

Music Explorer

Daniel de Miranda Almeida
FGV EMAP
Rio de Janeiro, Brasil
danielalmemiranda@gmail.com

Pedro Santos Tokar
FGV EMAP
Rio de Janeiro, Brasil
pedrotokar2004@gmail.com

Vitor do Nascimento
FGV EMAP
Rio de Janeiro, Brasil
vitornascimento@pm.me

Abstract—A indústria musical é ampla, complexa e está envolvida com uma variedade de pessoas: artistas, ouvintes, produtores, executivos, entre outros. Apesar de existirem diversas bases de dados relacionadas à essa indústria, todas com diversas informações sobre artistas e músicas, a principal forma de visualização desses dados é por meio de textos e tabelas.

Este trabalho propõe uma ferramenta de visualização baseada em grafos para dados de artistas e de relações entre eles.

Index Terms—Visualização de Dados, Música, Grafos, Ciência de Dados, Visualização da Informação

I. INTRODUÇÃO

Música é um tema muito vasto e presente diariamente no cotidiano da população. São inúmeras as pessoas que dedicam parte de seu tempo para ouvir música, e igualmente massivo é o número de pessoas que produzem novas músicas, desde pequenos artistas independentes até grandes nomes que produzem e lançam hits anualmente.

Quando se observa a indústria musical atualmente, percebe-se que existem muitas relações entre as pessoas e grupos que a compõe. Indo desde artistas que colaboram para a produção de uma música até artistas que se juntam em um grupo para fazer novas composições, as relações são incontáveis.

Este trabalho tem como objetivo fornecer uma ferramenta (chamada de *Music Explorer*) que utilize conceitos da área de visualização de dados e permita ao usuário explorar melhor essas relações, empoderando-o para que ele possa descobrir interações entre artistas específicos e entender como um artista se posiciona nesse universo em relação aos demais.

A. Trabalhos relacionados

Diversas bases de dados sobre informações musicais podem ser encontradas na internet, como o RateYourMusic [1], além de diversos *datasets* soltos que podem ser encontrados em sites como o Kaggle. A base de dados utilizada para este projeto foi a MusicBrainz [2].

Além de simplesmente ter dados, existem serviços que também apresentam visualizações sobre o mundo musical. Como exemplo notável é possível destacar o site Last.fm [3], que tem uma seção reservada (Labs) para visualizações sobre os hábitos musicais dos usuários.

Já no campo de visualização de grafos, dentre as inúmeras que estão disponíveis na internet, destacamos o site Wiki-Graph [4], que foi a principal inspiração para o formato de grafos para dispor os dados de artistas.

II. MÉTODOS

O desenvolvimento da ferramenta pode ser dividido em duas frentes: (i) o processamento dos dados que alimentam a ferramenta e (ii) o desenvolvimento da página da ferramenta. Essas duas frentes serão detalhadas a seguir, com ênfase especial no desenvolvimento da página e nos recursos de visualização empregados nela.

A. Processamento de dados

A ferramenta de visualização se fundamenta em três conjuntos de dados: (i) uma árvore de gêneros, (ii) um grafo de artistas e (iii) estatísticas de gêneros.

Esses três conjuntos são processados com base nos dados disponíveis da base de dados MusicBrainz. Esta é uma base de acesso livre e gratuito que visa ser referência de acesso à metadados da indústria musical. O acesso às informações da base, seja via API REST pública ou download de cópias do site, é livre e não exige autenticação ou compra de planos. Como melhor detalhado abaixo, o acesso aos dados foi feito pelos dois métodos disponíveis, e todos os algoritmos foram programados na linguagem *Python*.

a) **Árvore de gêneros**: Os dados para gerar a árvore de gêneros foram obtidos através de um webscraping da plataforma do *MusicBrainz*. Dentro das páginas de gêneros do site, um scraper mapeia a hierarquia dos gêneros baseando-se nos hyperlinks de cada página e salva o resultado final no formato JSON.

Tanto a estrutura desta árvore quanto do grafo dos artistas é feita usando um esquema de chaves e conjuntos. Cada gênero tem sua chave no dicionário, e o valor é um outro dicionário que contém diferentes conjuntos, um para cada tipo de relação, além de outras chaves para armazenar informações como nomes. Cada conjunto contém as chaves para acessar as informações dos outros nós, agindo como uma lista de adjacências.

Essa estrutura se difere da que é usada por bibliotecas como o D3.js para processar grafos, já que não consiste em um conjunto de nós e um separado de arestas. A escolha desse formato se deu pela conveniência para construir os grafos e pela maior eficiência para acessar arestas, já que não é necessário percorrer um conjunto de arestas em busca daquelas que tem como uma das pontas um nó específico.

b) **Grafo de artistas**: O grafo de artistas é construído a partir da navegação do dump de dados. Existem duas motivações para pré-processar o grafo:

(i) Processar os dumps em tempo real é inviável, dado a natureza dos dados compactados. Para navegar nos dumps é necessário ler linha a linha até encontrar o objeto desejado;

(ii) Nem todas as relações presentes no grafo estão disponíveis facilmente neles. As relações relacionadas com cover dependem de um busca e lógica mais extensas para serem inferidas.

Para processar o grafo, são usadas três entidades do MusicBrainz: “artists”, que representa um artista, seja solo ou grupo; “works”, que representam músicas; e “releases”, que representam lançamentos de mídia física ou digital de uma música ou álbum.

O processo exato para realizar as inferências é extenso e não será detalhado no presente relatório. Uma descrição das principais etapas de processamento realizadas para a construção do grafo estão disponíveis no arquivo `extras/README.md` do repositório com o código fonte da ferramenta.

Assim como a árvore de gêneros, o arquivo do grafo é salvo em JSON. Devido ao arquivo ser pesado para os padrões comuns de sites web, esse arquivo é separado em diversos arquivos menores, baseados no início do ID dos artistas. Isso permite com que o site carregue dinamicamente os arquivos durante a execução e sob demanda, evitando com que o usuário tenha que esperar um longo período de tempo para começar a explorar.

c) *Estatísticas de gêneros*: As estatísticas de gêneros são computadas utilizando os dumps de dados. Conceitualmente, elas consistem em uma matriz que relaciona anos e gêneros, com suas entradas sendo quantos álbuns, singles, EPs e etc. foram lançados com um gênero em um ano específico. Na prática, esses dados são armazenados no formato JSON

O processo para calcular estas estatísticas também é melhor detalhado no arquivo `extras/README.md` do repositório. Em resumo, são realizadas contagens levando em conta os campos de gênero e ano de lançamento das entradas de uma base de dados de lançamentos.

Estes três conjuntos de dados são construídos antes da execução da página. Além deles, também são usadas informações acessadas dinamicamente da API do MusicBrainz. Essas informações incluem principalmente metadados de artistas e seus lançamentos.

B. Visualizações

A interface para a ferramenta é feita utilizando tecnologias de desenvolvimento *web*, como as linguagens *HTML*, *CSS*, *JavaScript* e o framework *Svelte*. As tecnologias web foram escolhidas por: (i) facilitarem ao acesso da ferramenta, não dependendo de softwares compilados para apenas um sistema operacional e (ii) serem a base para a biblioteca *D3.js*.

Todas as visualizações detalhadas nas próximas subseções foram feitas utilizando a biblioteca *D3.js*. O nível de versatilidade que a biblioteca foi ideal para o desenvolvimento das visualizações, especialmente do grafo.

a) *Grafo de artistas*: Nossa visualização principal se trata de um grafo que demonstra relações entre artistas e bandas.

A partir deste grafo é possível ver quais artistas fizeram regravação de alguma música de outros artistas, assim como ver quais artistas já participaram de uma banda. Tais relações são demonstradas através de uma escala de cor categórica, a escala de cor possui uma legenda interativa, que permite remover relações com o clique do mouse.

O grafo é construído de forma iterativa pelo usuário – exceto por uma banda inicial já colocada para exemplificar a interação. A construção do grafo pode ser feita pela pesquisa por uma banda ou artista através da barra de pesquisa ou pela interação em nós já visíveis para expandi-los. Quando um artista é adicionado, os nós com relação a ele são adicionados também, e as relações entre os nós visíveis é calculada.

O grafo, além de codificar as relações entre artistas, também codifica o *tipo de relação* utilizando uma escala de cores. A legenda da escala de cores é interativa, fornecendo ao usuário opções de filtragem para controlar o tamanho do grafo e valorizar relações de interesse. Também está disponível um componente para remover nós, permitindo que o usuário tenha na tela apenas os nós de interesse sem precisar recarregar a página, e para focar em nós, permitindo que nós perdidos no grafo sejam encontrados com facilidade.

b) *Barra Lateral*: Existe uma barra lateral localizada no canto direito da tela que possui uma série de visualizações interativas que se relacionam com a visualização principal. Essa barra pode ser ocultada caso o usuário deseje.

Elementos presentes na barra lateral:

- 1) Visão geral do artista, através do ano em que ele entrou na indústria musical e através de tags para os gêneros do artista. Essas tags são interativas e se relacionam com o gráfico principal – que é detalhado na próxima subseção;
- 2) Gráfico de sinal que demonstra os anos que o artista selecionado esteve ativo e a quantidade de lançamentos que o mesmo realizou em tais anos. Este gráfico é interativo e é possível realizar um zoom para melhor visualizar os valores em cada intervalo de tempo. O usuário pode se beneficiar desse gráfico para entender melhor como foi o tempo de atividade de seu artista de interesse;
- 3) Gráfico de pizza, que mostra a proporção de cada relação do artista. Esta visualização, apesar de mostrar a mesma informação que já é visível no grafo, foca na *proporção* das relações, tornando a recuperação dessa informação mais fácil. Além disso, um tooltip interativo permite ver os números exatos de relações, sem a necessidade de contar os nós;
- 4) Gráfico de barras, que mostra as músicas do artista selecionado com mais covers realizados. Este gráfico interage com a visualização principal, já que ao clicar em uma barra (que representa uma música) os artistas que fizeram cover de tal música são destacados na visualização através da redução na opacidade dos outros nós. Essa ferramenta permite explorar mais profundamente um tipo de relação específico.

c) **Dendrograma de gêneros:** A segunda visualização principal se trata de um dendrograma, desenvolvido para demonstrar a relação hierárquica entre os gêneros musicais. Esta visualização facilita a compreensão de como os gêneros musicais se originaram e a proximidade entre certos estilos musicais.

Ao passar o mouse sobre o dendrograma, o caminho hierárquico correspondente é destacado, permitindo que se entenda melhor o caminho de um gênero específico desde sua origem. Com essa visualização, o usuário não precisa usar páginas de texto com links entre si para explorar a origem dos gêneros.

Para uma análise mais aprofundada, ao clicar em um nó (gênero musical) no dendrograma é exibido um histograma interativo. Este histograma mostra o número de lançamentos para aquele gênero musical específico por ano, demonstrando sua popularidade e evolução ao longo do tempo. Para facilitar a navegação, o dendrograma também inclui uma barra de pesquisa, similar à do grafo, que permite a busca por algum gênero musical.

A integração desta visualização com o gráfico de artistas é uma característica importante: ao clicar em uma tag de gênero no grafo de artistas, o gênero correspondente será automaticamente exibido e destacado no dendrograma, e o histograma com os lançamentos anuais para aquele gênero também será atualizado. Dessa forma, o usuário consegue entender melhor que estilos influenciaram o artista/banda de interesse dele, trazendo mais uma possibilidade de análise.

III. RESULTADOS

O resultado final foi uma página hospedada no serviço GitHub Pages com todas as visualizações detalhadas na última seção. Todas as visualizações feitas tem algum grau de interatividade, que é sinalizada tanto no texto inicial quanto próximo aos gráficos.

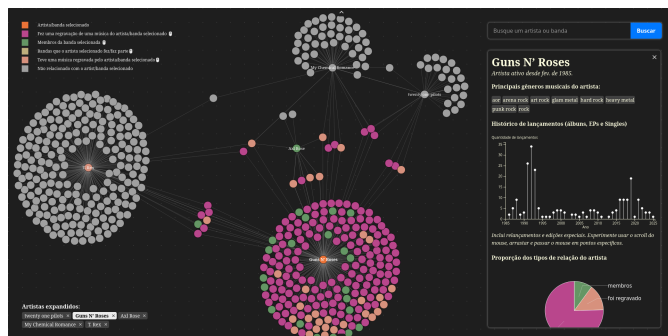


Fig. 1: Grafo de artistas interativo.

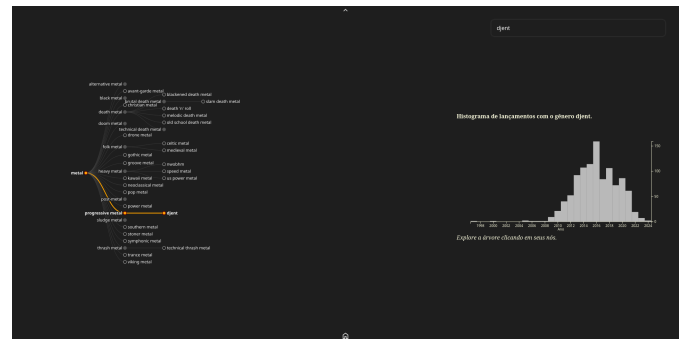


Fig. 2: Dendrograma de gêneros com histograma.

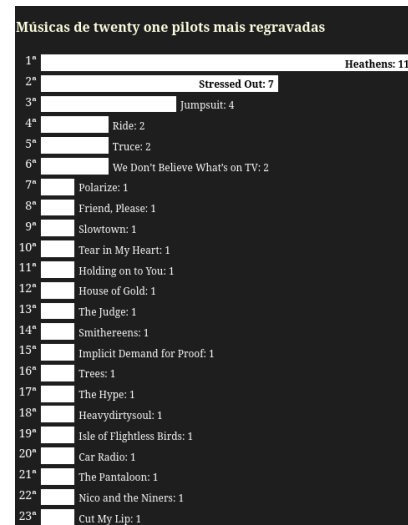


Fig. 3: Gráfico de barras de músicas mais regravadas.

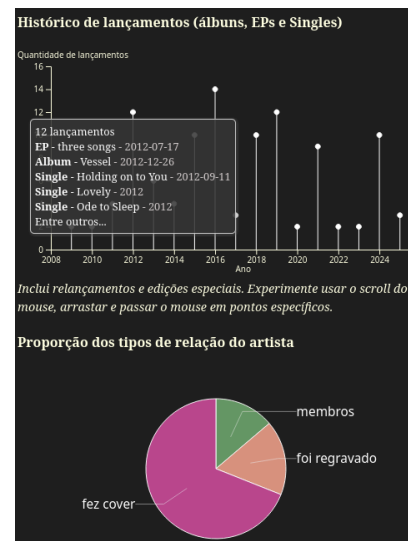


Fig. 4: Outras visualizações da barra lateral.

A. Discussão

O nosso projeto permite uma exploração interessante das relações entre artistas e gêneros musicais. Permitindo que os usuários que se interessam por alguma banda ou artista descubra relações que o mesmo ainda não conhecia.

O nosso projeto foi avaliado por alguns de nossos conhecidos e obtivemos uma boa taxa de aprovações. Muitos destes nossos colegas nos disseram ter descoberto relações do mundo da música que eles não conheciam antes, por exemplo a descoberta de novas músicas e covers.

Acreditamos que o maior benefício que o projeto oferece ao usuário é a possibilidade de fazer novas descobertas. Não ter as relações “escondidas” por trás de páginas e páginas HTML permite com que elas sejam melhor visualizadas e que sejam exploradas mais agilmente e de forma mais intuitiva.

A visualização se beneficia muito de ser relacionada à música, pois esse é um tema muito conhecido e amplo, o que torna o público alvo muito mais abrangente. Além disso, as ferramentas de interatividade permitem com que o usuário explore de acordo com suas preferências e gostos, sem estarem limitados a um subset muito pequeno de artistas. Essa característica aumenta o engajamento geral com o projeto.

B. Direções futuras

O projeto tem potencial para se expandir em algumas frentes específicas, que não foram tão exploradas devido à outras frentes que tinham prioridade no desenvolvimento:

- Criação de novas relações. Atualmente, apenas pares de tipos de relações são explorados. Com o devido planejamento e estudo da base MusicBrainz, é possível construir mais relações entre artistas, como por exemplo indicações de quais artistas fizeram colaborações com quais;
- Adição de novas visualizações de artistas. A API do MusicBrainz retorna uma variedade de informações, e nem todas foram exploradas nas visualizações feitas. Com mais estudos dos dados disponibilizados na API, é possível expandir ainda mais a gama de visualizações do site.
- Melhor integração entre o dendograma e o grafo. Atualmente, existe uma interação que parte do grafo para o dendograma, mas interações no dendograma não alteram o grafo. Com o devido planejamento de como estruturar os dados de gêneros, é possível fazer filtros que se baseiam na árvore, fornecendo uma melhor integração das visualizações.

REFERENCES

- [1] Rate Your Music, “Rate Your Music: Music and Film Community”.
- [2] MetaBrainz Foundation, “MusicBrainz”.
- [3] Last.fm, “Last.fm”.
- [4] DanTheBuilder, “DanTheBuilder's Wiki”.