

Trabalho 2 - Visualização de Informação

Arthur Briganti Gini, João Gabriel, Vinicius Passos Florencio

Faculdade de Tecnologia – Unicamp

Departamento de Sistemas de Informação

Faculdade de Tecnologia (UNICAMP) – Limeira, SP – Brazil

a213253@dac.unicamp.br, j199617@dac.unicamp.br, v188237@dac.unicamp.br

Resumo. *Este relatório descreve os procedimentos para analisar dados das músicas mais ouvidas do grupo Racionais MCs e do Emicida a partir da ótica da visualização de informação utilizando as técnicas de Word Clouds e Phrase Nets, descrevemos brevemente sobre os dados e demonstramos os gráficos comparando as fontes obtendo análises pertinentes sobre os dados.*

1. Introdução

No presente trabalho apresentaremos uma análise das 5 músicas mais ouvidas do grupo Racionais Mcs e do Emicida, a partir da ótica da visualização da informação utilizando técnicas de criação de gráficos utilizando como base a letra das canções.

O objetivo é realizar uma análise gráfica a fim de identificar as principais semelhanças e diferenças nas letras de dois dos principais expoentes do rap nacional.

2. Sobre os dados estudados

Para a coleta dos dados foi realizada primeiramente uma pesquisa das músicas mais ouvidas do grupo Racionais Mcs e o rapper Emicida. Foram selecionadas as 5 músicas mais famosas de cada artista, toda a coleta foi manual no site <https://www.letras.mus.br/>.

Na coleta dos dados foi criado um arquivo com a letra de cada música, enumerando os Racionais de 1 a 5 e o Emicida de 6 a 10, assim obtendo a base bruta dos dados.

Para a realização das análises utilizando as técnicas de visualização de informação cada arquivo foi tratado separadamente e preparado de uma forma específica para cada tipo de gráfico gerado.

3. Análises realizadas

Nesta seção abordaremos como os dados foram tratados para a geração dos gráficos e realizaremos as análises inferidas a partir de cada representação.

3.1. Preparação para o Nuvens de Palavras

Com os dados separados por música de cada artista, foi realizado um processamento automático com um script na linguagem python, o objetivo é retirar as stop-words e tokenizar as palavras, separando cada uma por linha.

Após o processamento foi realizada uma vistoria manual dos dados retirando possíveis falhas do algoritmo, comumente foram retiradas pontuações extras e números erroneamente extraídos.

3.2. Nuvens de Palavras(word clouds)

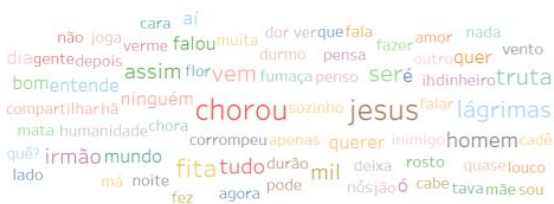
Nesta sessão abordaremos os gráficos gerados das 10 músicas, apresentando as semelhanças e diferenças visíveis sob cada artista.

Os gráficos foram gerados no software Tableau e para cada palavra foi gerada a frequência, ou seja $f(w,t)$ em que 'w' é a palavra e 't' é o texto, assim identificando quanto cada palavra se repete em cada texto.

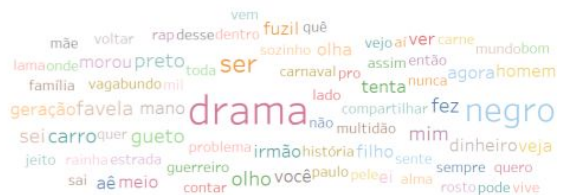
Na criação dos gráficos foram filtradas as palavras que aparecem apenas uma vez naquele texto. Para as análises a seguir foi retirado os gráficos das músicas mais ouvidas de cada artista, estas serão analisadas separadamente posteriormente.

Sobre 4 musicas do Racionais:

2- Jesus chorou



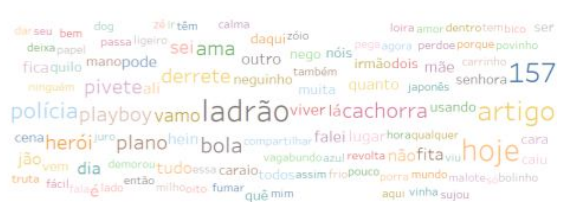
3- Nego Drama



4- Da ponte pra cá



5- Artigo 157

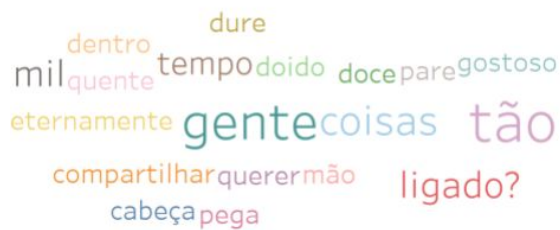


Podemos observar nos gráficos de cima as palavras que mais aparecem (maior tamanho) são chorou, jesus, drama e negro; essas que sintetizam a vivência presente nas letras dos artistas, refletindo assim nos gráficos de nuvens de palavras um pouco da realidade da periferia, podemos ver também o uso de gírias como fita, banca, rapa entre outros elementos desse dialeto.

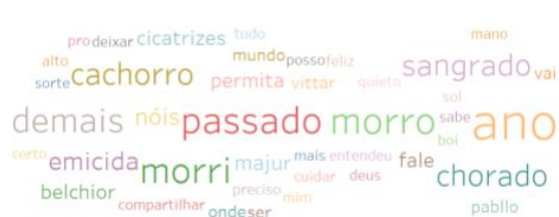
Nos gráficos de baixo na música “Eu sou 157” as palavras evidenciadas são: polícia, ladrão, artigo, 157; mostrando a realidade da visão de um ladrão e os reflexos da sua atitude nas palavras como herói e pivete. Na faixa “Da ponte pra cá” é observado uma visão mais geral da periferia e suas faces, como: querer, mundo, ser, diferente e adianta; mostrando o poder da ação a partir do dialeto da periferia com o refrão “não adianta querer tem que ser tem que pá, o mundo é diferente da ponte pra cá”.

Sobre 4 musicas do Emicida:

7- Mil coisas



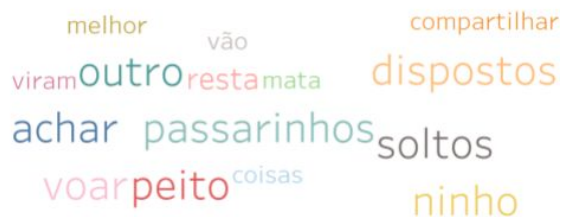
8- Amarelo



9- A cada vento



10- Passarinhos



Podemos observar que as palavras repetem mais e variam menos comparado ao artista anterior. Nos gráficos de cima observamos a palavra compartilhar como mais relevante nesse contexto mais moderno, e as palavras do gráfico seguinte levam a uma alusão ao passado do artista como: passado, morro, morri e chorado, criando uma ponte de um passado mais difícil a uma realidade melhor hoje em dia.

Nos gráficos de baixo podemos observar uma passagem na música “A cada vento” evidenciando palavras como encontrar, manhã, soprar e perder. Na música seguinte “Passarinhos” é notável uma mudança nas palavras evidenciadas, sendo dispostos, achar, peito, passarinhos e soltos.

Todas as mudanças nos gráficos são refletidas no contexto que foi escrita e lançada a obra, sendo as mais antigas evidenciando palavras sobre um futuro melhor e as mais recentes dizendo mais sobre paz, pássaros e uma vida melhor no geral.

As musica mais ouvida de cada Artista:

A análise dos gráficos a seguir foram realizadas logo após os 2 gráficos.

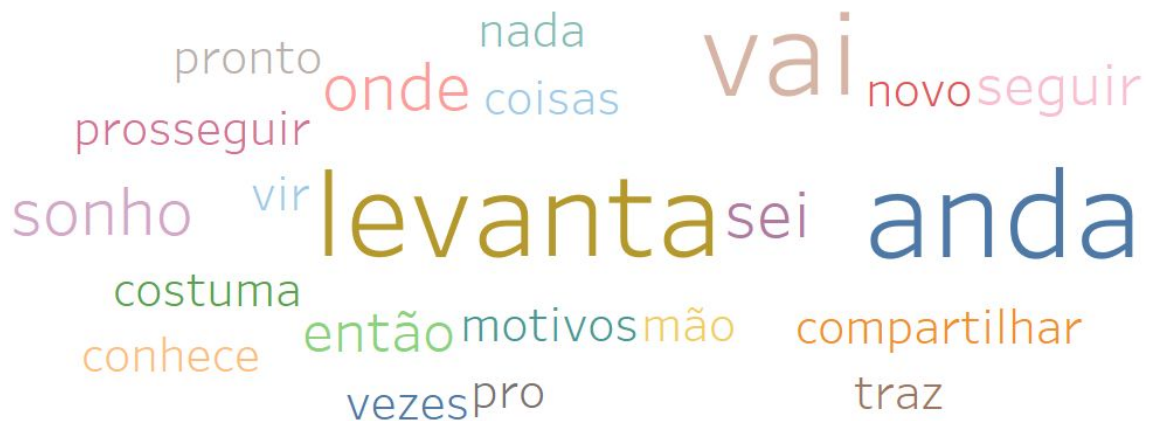
Racionais:

1- A vida é desafio



Emicida:

6- Levanta e Anda



É notável que as palavras evidenciadas nas músicas de maior sucesso de cada artista abordado são de motivação, tais como: acreditar, sempre, sonho, levanta e anda. Isso mostra a partir das técnicas de Visualização de Informação o tema com maior alcance dentro desses estilos de música.

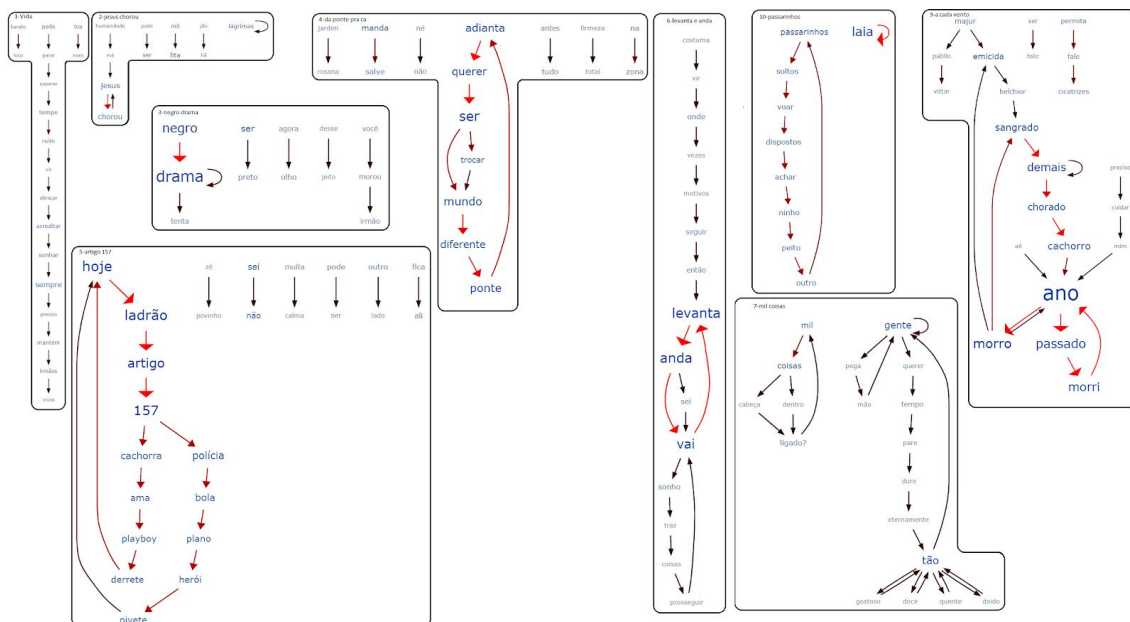
Podemos observar também uma grande mudança no estilo e tamanho das letras, sendo as letras do primeiro artista muito mais carregadas, com várias palavras repetindo mais de uma vez, enquanto do segundo as letras são mais enxutas, trazendo o mesmo contexto só que de uma forma diferente.

3.3. Preparação para o Redes de Palavras

Utilizando as bases produzidas em python para a nuvem de palavras como entrada, não era preciso pra usamos uma lista de stopwords e fizemos um algoritmo Java no qual passamos as palavras para a memória com a filtragem de stopwords, assim fazemos a contagem de palavras que se repetem em dupla.

3.4. Redes de palavras(Phrase Nets)

As imagens foram gerados no site [Graphviz](http://graphviz.org), usando da saída csv do programa em Java desenvolvido pelo grupo, o arquivo csv resultante contém as duplas vértices que se repetem e a quantidade de repetições, a partir daí os grafos são gerados, usam as repetições para definir a cor e a espessura da conexão (seta), e a frequência como cor e tamanho dos vértices(palavras), e o resultado obtido foi esse:



Por se tratar de músicas é possível perceber que muitos dos vértices com mais repetições pertence aos refrões, e muitas das vezes estão ligados fechando um ciclo quando o mesmo refrão se repete 2x e uma lista de ligações quando isso não acontece.

4. Considerações finais:

Portanto, de acordo com as análises realizadas constata-se que é possível, a partir de diferentes massas de dados, neste caso textos (letras de músicas), estabelecer uma relação entre essas massas, e posteriormente, aliados às técnicas de Visualização da Informação propostas para este trabalho conceber resultados gráficos úteis para o domínio dos dados escolhidos. É notável que as técnicas de visualização de informação ajudam muito no entendimento de diferentes bases de dados, permitindo identificar a distância entre os dados proporcionados, detectar padrões e apresentar estes resultados da melhor forma possível.

Referencias:

LETRAS. **Musicas**. Disponível em: <https://www.lettras.mus.br/>. Acesso em: 4 jul. 2020.

GITHUB. **Portuguese stop words**. Disponível em: <https://gist.github.com/alopes/5358189>. Acesso em: 3 jul. 2020.

KBOING. **As melhores musicas do Emicida**. Disponível em: <https://www.kboing.com.br/noticias/Planeta-Rock-as-melhores-musicas-do-Emicida+19080711322305.html>. Acesso em: 4 jul. 2020.

LETRAS. **Musicas**. Disponível em: <https://www.lettras.mus.br/>. Acesso em: 4 jul. 2020.

YOUTUBE. **As musicas mais tocadas dos Racionais**. Disponível em: <https://www.youtube.com/watch?v=bhcFT0AJwiE>. Acesso em: 4 jul. 2020.

GITHUB. **Trabalho2_SI803A**. Disponível em: https://github.com/ArthurGini/Trabalho2_SI803A. Acesso em: 08 jul 2020.