

MAIS DO MESMO? ESTUDO SOBRE A CORRELAÇÃO ENTRE A QUANTIDADE DE MÚSICAS PRODUZIDAS E A DIVERSIDADE TEMÁTICA EM LETRAS MUSICAIS

Palavras-Chave: PRODUÇÃO CULTURAL, MESMICE, LETRA MUSICAL

Autores/as:

BEATRIZ CAVALLI FERRÃO [FEEC/UNICAMP]

Prof. Dr. TIAGO FERNANDES TAVARES (orientador/a) [FEEC/UNICAMP]

INTRODUÇÃO:

Com o avanço da quantidade de conteúdo disponível na Internet e a facilidade de acesso a esses dados através de redes móveis nas últimas décadas, o modo de ouvir, compartilhar e de produzir músicas mudou drasticamente. Isso é comumente referido como um crescimento sem precedentes da diversidade cultural [8], embora esta afirmação em geral parta do pressuposto de que um aumento na quantidade de músicas produzidas implica necessariamente em uma grande variedade de tipos de composições. Neste trabalho, investigamos estatisticamente a correlação entre o número de objetos de mídia produzidos e sua diversidade temática.

METODOLOGIA:

Neste trabalho, investigamos a relação entre o aumento na quantidade de músicas produzidas e o aumento na variedade de temas abordados por essas mesmas músicas. Nossa hipótese é a de que o aumento na quantidade de músicas criadas em uma época não necessariamente leva à criação de músicas que abordam temas diferentes. Em outras palavras, há um volume maior de criação musical, mas não necessariamente um aumento na diversidade cultural criada em cada época.

Com o intuito de averiguarmos nossa hipótese, tivemos que encontrar uma forma de não somente analisar a complexidade do vocabulário utilizado na produção musical, mas também verificar o quanto uma letra musical se assemelha a outra. Para isso, inicialmente, selecionamos uma base de dados para fundamentar nossa pesquisa, descrita na subseção A. Posteriormente, documentamos a entropia e o coeficiente de mesmice das letras, discutidos nas subseções B e C, respectivamente.

A. Dataset

Neste projeto, utilizamos o dataset do site de letras musicais Metrolyrics [3], que contém 362236 músicas. As letras do dataset estão classificadas de acordo com seu gênero musical. A classificação do dataset usa os seguintes gêneros: country, pop, hip-hop, rock, indie, jazz, eletrônica, R&B e metal. No entanto, investigamos somente os quatro primeiros, tendo em vista que possuíam os maiores números de músicas.

B. Entropia

Entropia é um índice que representa a desordem de um sistema [2,6]. Neste projeto, utilizamos tal índice com o intuito de analisar a complexidade do vocabulário utilizado nas letras musicais. A entropia pode ser calculada da seguinte forma:

$$S = -\sum(p * \log(p)), \quad (1)$$

onde S é o índice que indica a entropia. Este pode assumir um valor de 0 a 1. Quanto mais próximo de 1, mais diversificado é o vocabulário da letra. Na Equação 1, p é a distribuição de frequência da letra de música. Em outras palavras p representa a contagem de ocorrências de cada palavra em uma música dividida pelo total de palavras na música. Em seguida, para analisarmos o comportamento do vocabulário em um determinado período, computamos a média de entropias de todas as músicas por década por gênero.

No entanto, necessitamos de um modo de relacionar o valor encontrado da complexidade do vocabulário com a quantidade de produção cultural. Para tanto, primeiramente, enumeramos o total de músicas e de artistas de cada década e gênero. Posteriormente, com o intuito de comparar as curvas de entropia com o número de artistas, utilizamos um índice de correlação, denominado Correlação linear de Pearson. Este mede o grau de relacionamento entre duas variáveis (X, Y) [7]. Neste caso, seriam a entropia (X) e a quantidade de músicas e artistas (Y), nos quais realizamos uma correlação para cada fator. Seu valor pode ser obtido através da Equação 2:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}, \quad (2)$$

onde n é a quantidade de valores computados de ambas as variáveis. O valor de “ r ” está em um intervalo de -1 a 1. Um sinal positivo implica que as variáveis são diretamente proporcionais. Já o sinal negativo, as variáveis são inversamente proporcionais. Para este caso, um valor positivo significa que um valor alto de entropia equivale a um alto número de conteúdo produzido, enquanto que um valor negativo significa que um alto valor de entropia equivale a uma baixa quantidade de conteúdo produzido, ou vice-versa.

C. Coeficiente de mesmice

O coeficiente de mesmice de uma música é definido como a probabilidade de uma palavra contida em uma música A também estar contida em uma música B, como mostra a Equação 3.

$$m = \frac{k}{t} \quad (3)$$

onde m é a probabilidade de uma palavra estar contida simultaneamente em ambas as músicas A e B. Este coeficiente pode assumir valores entre 0 e 1. Quanto mais próximo de 1, maior a quantidade de palavras que estão contidas simultaneamente em A e B, e, portanto, maior é a semelhança temática entre as músicas. Na Equação 3, k é a quantidade de palavras contidas no conjunto da intersecção do conjunto de palavras que estão nas músicas A e B; e t é o número de palavras contidas no conjunto da união de ambas as letras.

Em seguida, para analisarmos o comportamento do coeficiente em um determinado período, computamos a média das probabilidades por década por gênero.

Para este tópico, também necessitamos de uma forma de relacionar o coeficiente de mesmice com a quantidade de conteúdo produzido. Para isso, também realizamos a correlação entre as curvas. No entanto, para este caso, um valor positivo do índice de correlação significa que tanto o coeficiente de mesmice e a quantidade de músicas e artistas naquela década são altos. Em outras palavras, um valor positivo alto indica que não há variedade no conteúdo sendo produzido naquela época dentro daquele gênero específico. O contrário também é verdadeiro, ou seja, um valor negativo no índice de correlação significaria que o coeficiente de mesmice é baixo, enquanto que o número de músicas e artistas é alto, e vice versa, isto é, haveria mais diversidade no conteúdo produzido.

RESULTADOS E DISCUSSÃO:

Os resultados obtidos através da implementação dos procedimentos descritos acima, estão apresentados abaixo:

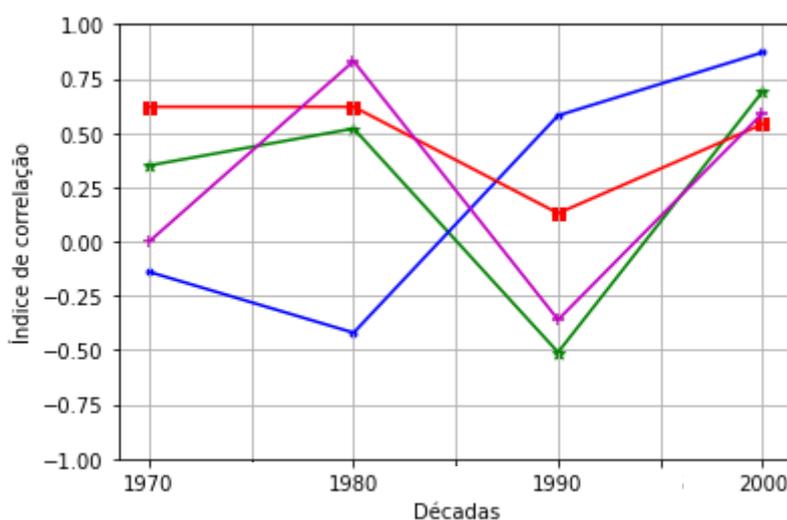


Figura 01 - Correlação anual entre quantidade de músicas produzidas e entropia das letras de música por gênero por década

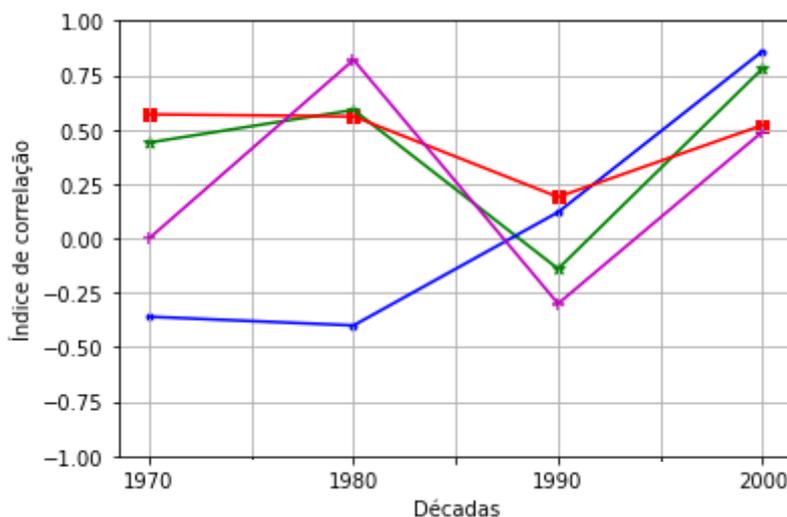


Figura 02 - Correlação anual entre número de artistas e entropia das letras de música por gênero por década

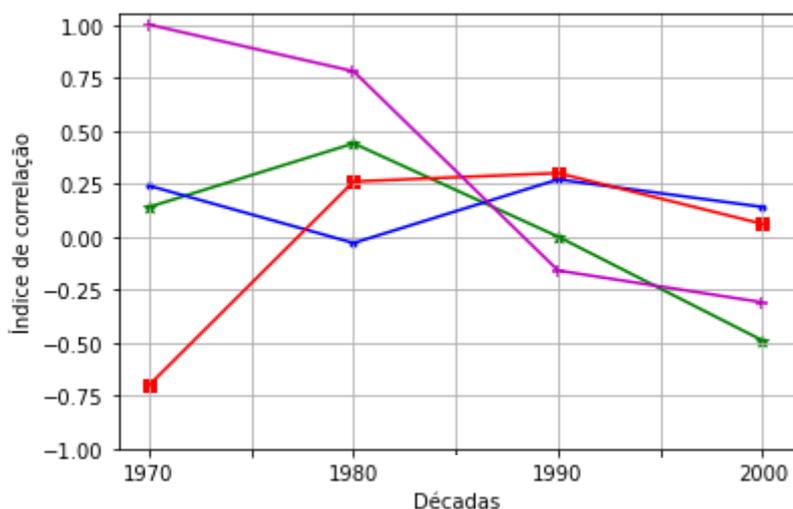


Figura 03 - Correlação anual entre o índice de mesmice e a quantidade de músicas por gênero por década

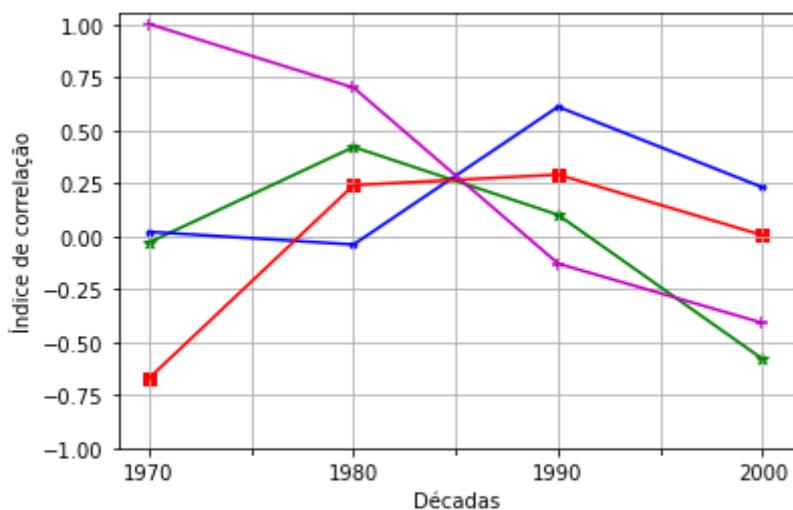


Figura 04 - Correlação anual entre o índice de mesmice e a quantidade de artistas por gênero por década

Percebemos que, primeiramente, nos anos 2000, temos um alto índice de correlação entre os números de artistas e músicas e a entropia. Isso significa que o conteúdo produzido hoje tem uma maior diversidade no vocabulário quando comparado a décadas anteriores.

No entanto, ao compararmos os dois últimos gráficos, a correlação é alta entre o coeficiente de mesmice e o número de artistas e músicas, quando analisamos os gêneros pop e rock. Isto sugere que, mesmo com um número maior de produções atualmente, não há pluralidade no conteúdo produzido, ou seja, não há uma correspondência com um possível senso comum, onde se assume que com um número maior de músicos haveria uma maior variedade de conteúdo.

Há algumas possíveis explicações para este fenômeno. Primeiramente, há uma perspectiva que o pop não é considerado um gênero musical, mas sim, uma combinação de múltiplos gêneros [1]. Seria considerado como um gênero de negócios, que visa altas posições nas paradas de popularidade musical. Sendo assim, como os músicos já teriam descoberto uma maneira de

alcançar tal feito, as músicas tenderiam a seguir uma mesma fórmula. Consequentemente, artistas que produzem exclusivamente músicas pop são tipicamente criações da indústria.

Além disso, é observado um movimento similar com o rock. Há uma teoria que pop e rock formam um único gênero (pop-rock) [5], por usarem uma instrumentação parecida. Isto poderia justificar a semelhança entre seus índices ao longo dos anos, como podemos notar no gráfico apresentado na figura 03.

No entanto, este fenômeno não é observado com os gêneros hip-hop e country. Isto é devido ao fato de que tais gêneros não seguem a mesma direção que pop-rock. Ou seja, aparecem no topo das paradas com menor frequência. O hip-hop, por exemplo, mesmo com sua expansão no cenário da indústria musical, ainda é considerado um gênero “da rua” [1].

CONCLUSÕES:

A correlação entre a quantidade de músicas produzida e o índice de mesmice é positiva para pop e rock, o que mostra que embora estejamos criando mais conteúdo, ele tende a se repetir. Esse comportamento não é observado em hip-hop e country, nos quais quanto maior a quantidade de conteúdo criado, menor a repetição temática. Portanto, aumentar a quantidade de músicas e artistas não implica em uma maior diversidade musical.

BIBLIOGRAFIA

- [1] LENNA, Jennifer C.. “**Banding Together**: how communities create genres in popular music”. Princeton: Princeton University Press, 2012.
- [2] **Entropy**. Disponível em: <https://byjus.com/jee/entropy/>. Acesso em: 18 ago. 2021.
- [3] **Metrolyrics**. Disponível em: metrolyrics.com. Acesso em: 18 ago. 2021
- [4] DALMORA, André e TAVARES, Tiago. “**Identifying Narrative Contexts in Brazilian Popular Music Lyrics Using Sparse Topic Models**: A Comparison Between Human-Based and Machine-Based Classification”. Universidade Estadual de Campinas, 2019.
- [5] REGEV, Motti. “**Pop-Rock Music**”. Cambridge, Polity Press, 2013.
- [6] SHANNON, Claude E.. “**A Mathematical Theory of Communication**”. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
- [7] LIMA FILHO, Luiz Medeiros de Araujo. “**Correlação e Regressão**”. Universidade Federal da Paraíba, Departamento de Estatística.
- [8] PINTO, Waldir de Amorim. “**Evolução da Acessibilidade aos Meios de Gravação e Produção Musical**”. Universidade Estadual de Campinas, 2012.