

Na batida do funk: modelagem de tópicos combinando LLM, Engenharia de Prompt e BERTopic

Jesus Yopez¹, Bruno Tavares¹, Fabíola Peres², Karin Becker¹

¹ Instituto de Informática

² Instituto de Filosofia e Ciências Humanas

Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre, Brasil

{jesus.rojas, bruno.tsantos, karin.becker}@inf.ufrgs.br,

fabioladecarvalholeite@gmail.com

Abstract. *Song lyrics impose additional challenges to topic modeling, as the discourse is often implicit and must be understood within its context, using figurative and poetic language, and slangs. This paper proposes LLMusic, a new topic modeling approach that leverages Large Language Models (LLMs) to analyze lyrics, using Brazilian funk as a case study. Funk is a rich social portrait of the periphery. We use LLMs and prompting to summarize song excerpts into central themes in an iterative and unsupervised process applied to a corpus representative of the genre. These themes are grouped into a lean, coherent set of topics using BERTopic. Through zero-shot prompts, one can classify new lyrics based on these topics. We applied LLMusic to analyze the discourse in the 100 most popular funks, showing its potential for large-scale analysis.*

Resumo. *Letras de músicas impõem desafios à modelagem de tópicos, já que o discurso nem sempre é explícito, devendo ser compreendido dentro de seu contexto, sua linguagem figurativa e poética, gírias, etc. Este artigo propõe o LLMusic, uma nova abordagem de modelagem de tópicos que explora o potencial de Grandes Modelos de Linguagem (LLMs) para analisar letras de música, utilizando o funk brasileiro como estudo de caso. O Funk, muito popular no país, é um rico retrato social da periferia. LLMs e prompting são usados para resumir trechos de músicas em temas centrais, em um processo iterativo e não supervisionado aplicado a um corpus representativo do gênero. Esses temas são agrupados em um conjunto enxuto e coerente de tópicos usando BERTopic. Novos trechos de música são classificados nesses tópicos usando prompts zero-shot. Usamos LLMusic para analisar o discurso dos 100 funks mais populares, mostrando seu potencial para análise em larga escala.*

1. Introdução

O estudo de músicas proporciona uma visão multidimensional para as complexas relações entre cultura e sociedade, oferecendo uma oportunidade de desenvolvimento de análises interdisciplinares e descoberta de conhecimento. Através de ritmos e letras, músicas refletem tradições, crenças e valores de um determinado povo. Letras musicais expõem as experiências e perspectivas dos indivíduos dentro de uma sociedade, servindo como um espelho que reflete suas condições sociais. Em muitos gêneros musicais (e.g. *blues*, *folk* e *rap*), as letras refletem as realidades sociais de uma época.

O funk brasileiro é um gênero musical que nasceu nas favelas do Rio de Janeiro. O gênero incorporou elementos de gêneros estrangeiros como *soul*, *funk* americano, *R&B*

e *rap*, mesclados com características de gêneros brasileiros como o samba. As letras combinam assuntos do dia a dia nas comunidades, resiliência e críticas sociais, e incluem temas como violência, uso de álcool e drogas, sexo e objetificação das mulheres. Mesmo com resistências quanto ao valor artístico do gênero [Lopes 2011], o funk furou a bolha dos bailes das favelas cariocas e é hoje uma das maiores manifestações culturais de massa do Brasil, sendo escutado por todas classes sociais, e com projeção internacional.

Diferentes trabalhos qualitativos já mergulharam nos aspectos sociais e culturais do funk. Estes trabalhos geralmente têm como base um conjunto cuidadosamente curado de canções representativas de algum fenômeno a ser observado, tais como a expressão da masculinidade [Peres 2023], o racismo velado da sociedade brasileira [Lopes 2011] ou sua relação com a violência [Brilhante et al. 2019]. Pelas técnicas de análise empregadas, esses estudos limitam-se a pequenos conjuntos de dados. Ferramentas computacionais têm o potencial de auxiliar tais estudos qualitativos, permitindo, em maior escala, a identificação, análise, classificação e sumarização dos discursos presentes nestes textos.

A utilização de técnicas computacionais para análise em larga escala de letras de músicas foram realizadas em trabalhos abordando viés de gênero [Betti et al. 2023], letras e áudio [Calcina 2022], temas em letras [Junior et al. 2019] e classificação de sentimentos [Devi and Saharia 2020]. Boa parte desses trabalhos explora Modelagem de Tópicos (MT) como ferramenta de extração de informação em textos, sendo LDA [Blei et al. 2003] a mais popular. Por explorar a distribuição estatística de coocorrência das palavras, LDA não é capaz de captar relações semânticas dentro das letras, o que é particularmente problemático no contexto de letras musicais, onde metáforas, símiles, eufemismos e linguagens figurativas diversificam as palavras utilizadas para representação de um mesmo contexto. BERTopic [Grootendorst 2022] explora embeddings, similaridade e agrupamento por densidade para a MT, resultando em avanços na interpretação semântica de letras musicais [Calcina 2022]. Contudo, ao focar na tarefa de agrupamento de textos semanticamente semelhantes, pode gerar agrupamentos sem relevância do ponto de vista do discurso. Para ilustrar essa dificuldade, tomemos simplificações de trechos reais de funk. Trechos como “Deus, como ela é linda”, “Deus me guiou até aqui”, “Vou mandar o cara para Deus”, tendem a ser considerados similares em MT por sua evocação a Deus, mas nas respectivas letras expressam discursos totalmente diferentes, a saber mulheres, reflexões sobre a vida, e violência, respectivamente.

Grandes Modelos de Linguagem (Large Language Models - LLMs) vêm sendo aplicados com sucesso em uma diversidade de tarefas de processamento de linguagem natural (PLN). Uma das formas de se interagir com LLMs é através de *prompts*, cujo design através da Engenharia de Prompts (EP) influencia os resultados. Dada a configuração correta dos parâmetros e *prompts*, LLMs têm demonstrado capacidade de sumarização de textos comparáveis às feitas por humanos [Zhang et al. 2024]. Isto sugere que a utilização de LLMs para superar desafios de MT é uma abordagem promissora para a análise de letras musicais. LLMs conseguem captar nuances contextuais e semânticas que métodos tradicionais de MTs não alcançam. Isso ocorre pois estes modelos são treinados em vastas quantidades de texto, desenvolvendo uma compreensão mais profunda das estruturas e significados linguísticos. A presente pesquisa soma-se aos esforços para combinar o potencial de EP e LLMs para MT [Pham et al. 2024].

Este trabalho propõe LLMusic, um *framework* para extração de tópicos em um *corpus* de letras de músicas. LLMusic combina sinergicamente o poder de PE e LLMs, com técnicas de MT avançadas como o BERTopic, visando assim contribuir às limitações

de métodos de MT tradicionais para identificação de tópicos subjetivamente representados em textos. O método requer como entrada um *corpus* de letras representativas do gênero musical. Usamos a EP e LLMs de duas formas: *identificação de temas* e *atribuição de tópicos*. Primeiramente, usamos *prompts* e LLMs para extrair temas expressos nos trechos de músicas, explorando a capacidade generativa de LLMs de sumarização. Para criação de uma distribuição de temas robusta e representativa, combinamos aleatoriamente trechos de músicas de um *corpus* referência em múltiplas iterações. Usamos então o BERTopic para resumir essa distribuição em uma lista não redundante de tópicos representativos. Usamos também PE sobre LLMs para atribuir de forma não supervisionada (*prompts zero-shot*) os tópicos identificados a trechos de música para análise em larga escala. Apesar do potencial de generalização do método, nossas pesquisas focam inicialmente no funk. Usando as músicas curadas em [Peres 2023] como *corpus* referência do funk, nossos resultados mostram o desempenho superior de LLMusic comparado a BERTopic na extração de tópicos subjetivamente representados, bem como bom desempenho de classificação não supervisionada de tópicos. Ilustramos a aplicação de LLMusic como uma estratégia computacional para analisar os cem (100) funks mais tocados, ilustrando o potencial para análise em larga escala e complementar a estudos qualitativos.

O restante do texto está assim organizado. A Seção 2 sumariza os trabalhos relacionados. A Seção 3 detalha o *framework* LLMusic, e a Seção 4 detalha os resultados de sua aplicação em letras de funk. A Seção 5 ilustra sua aplicação para analisar os 100 funks mais populares. A Seção 6 apresenta conclusões e direções futuras.

2. Trabalhos relacionados

Os estudos que ligam técnicas computacionais e análise musical vêm aumentando, permitindo extrair *insights* a partir da extensa quantidade de músicas disponíveis [Oramas et al. 2018]. Eles destacam como a música tem sido cada vez mais vista como um sistema linguístico e, portanto, pode se beneficiar das técnicas de PLN de ponta, tais como LLMs. Nos trabalhos que abordam a análise de letras musicais, encontramos aplicações supervisionadas e não supervisionadas.

Técnicas de modelagem supervisionadas têm sido usadas na análises de grandes *corpus* de letras. Em [Betti et al. 2023], modelos baseados na arquitetura BERT foram ajustados usando uma base anotada para identificação de sexismo e viés de gênero. Os modelos foram posteriormente aplicados em um corpus composto por 377.808 letras de músicas em inglês para analisar a relação entre sexismo e gênero do artista. Por ser uma análise supervisionada, sua reprodução é restrita a contextos onde existam bases anotadas de referência para ajuste dos modelos.

Nos estudos de música que utilizam técnicas não supervisionadas de MT, LDA é a técnica prevalente. Em [Devi and Saharia 2020] é aplicada para classificação de sentimentos. [Junior et al. 2019] investiga como termos específicos (e.g., álcool e relacionamentos) são empregados nas letras de sertanejo, gênero musical popular no Brasil. Como o LDA extrai relações de coocorrências das palavras, a geração de agrupamentos representando tópicos está atrelada a uma recorrência dos artistas em utilizar as mesmas palavras para representar as mesmas situações.

Em [Calcina 2022] é feita uma comparação de artistas baseada no conjuntos de tópicos abordados em suas canções, onde BERTopic é utilizado para descobrir semelhanças letras de diferentes gêneros musicais (e.g., *folk* e *rap*) com base nos tópicos comuns das letras (e.g, lutas diárias, problemas sociais). O uso de BERTopic em letras

é ainda recente, mas sua capacidade de extrair informações de um *corpus* está atrelada à capacidade do modelo de linguagem utilizado ser capaz de representar corretamente, por meio de *embeddings*, as informações explicitamente contidas no *corpus*. Letras de música usam de subjetividade, linguagem figurativa e poética, metáforas, entre outros, para expressar idéias que requerem uma habilidade abstração.

Trabalhos recentes destacam o potencial generativo de LLMs em tarefas de sumarização de textos. Em [Zhang et al. 2024], é feita uma avaliação humana de dez diferentes LLMs, onde sugere-se que a configuração do *prompt* é mais importante que o tamanho do modelo em tarefas de sumarização *zero-shot*. O trabalho aponta que a qualidade de sumarização dos LLMs é equivalente à humana.

O uso de LLMs explora o paradigma de *prompt learning* para sua parametrização. A estrutura básica do aprendizado baseado em *prompt* está ilustrada na Figura 1, contendo um *input* [X], um Template onde constam o conjunto de instruções do *prompt*, e um *output* [Y]. A estratégia (e.g. *zero-shot*, *few-shot*, *chain of thought*) escolhida no processo de criação do *prompt* é fundamental, uma vez que impacta diretamente o resultado do modelo [Pengfei Liu and Neubig 2023]. Alguns dos desafios para o uso de LLMs incluem alucinações (i.e respostas sem sentido ou irrelevantes do modelo), ou simplesmente lidar com a natureza não determinística dos LLMs (isto é, o mesmo *prompt* pode gerar respostas diferentes). Essas questões requerem exploração para identificar o que melhor se adapta à tarefa em questão, como misturar diferentes tipos de *prompts*.

<p>Input (X): <i>E eu, gostava tanto de você</i> Template: [X] Sugira 4 tópicos para esse trecho de letra: [Y]. Output (Y): Amor; Fim de relacionamento; Memórias; Saudade.</p>
--

Figura 1. Exemplo de *prompt* para gerar tópicos de um trecho de música

Estudos abordaram o potencial de EP em MT de forma geral [Pengfei Liu and Neubig 2023, Pham et al. 2024]. Em [Pham et al. 2024] propõe-se um *framework* onde uma distribuição de tópicos é gerada por EP considerando artigos de Wikipédia e resumos de notas do congresso americano. Os tópicos são então refinados e finalmente atribuídos aos documentos por meio de um processo de classificação usando novamente EP. O método foi superior a modelos tradicionais de MT como LDA e BERTopic. No entanto, o refinamento dos tópicos exige intervenção manual para análise e definição dos critérios de fusão e exclusão de tópicos. Além disso, a etapa de atribuição de tópicos requer a elaboração de *prompts* contendo exemplos de cada tópico (*few shot*).

3. LLMusic: Modelagem de tópicos em músicas com Prompt Engineering e BERTopic

Existem vários desafios na análise das letras musicais, visto que os textos são ambíguos e subjetivos, com múltiplos temas e estruturas líricas entrelaçadas, tornando a identificação automática de tópicos desafiadora para *frameworks* tradicionais de MTs. No contexto do funk existem desafios adicionais, já que as letras utilizam gírias únicas que refletem a vida na periferia [Lopes 2011]. Nesse contexto propomos o LLMusic, uma abordagem *end to end* para identificação e classificação de tópicos em letras de músicas. A estrutura proposta combina a capacidade de interpretação de texto das LLMs via engenharia de *prompt* com a capacidade de identificação de tópicos do BERTopic.

A capacidade de sumarização das LLMs permite que os tópicos gerados usando LLMusic sejam mais próximos de anotações humanas que os tópicos gerados por abor-

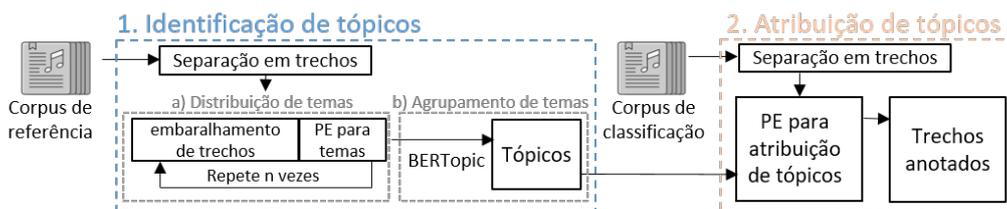


Figura 2. Visão geral do Framework LLMusic



Figura 3. Templates de PE para identificação e atribuição de tópicos

dagens tradicionais de MT. Como resultado, LLMs são capazes de derivar os principais tópicos das letras de músicas e categorizar seus trechos com uma abordagem multi-tópico. Embora LLMusic seja agnóstico em relação ao gênero musical, o desenvolvemos usando o funk como gênero musical alvo.

A Figura 2 resume o *framework* LLMusic, que aborda duas tarefas principais, *Identificação de Tópicos* e *Atribuição de Tópicos*, detalhadas no restante dessa seção. Ambas são baseadas em estruturas de *prompt zero-shot*, ilustradas na Figura 3, em um processo totalmente não supervisionado.

O *framework* assume como entrada um *corpus* não rotulado de letras representativas do fenômeno ou gênero a ser estudado, i.e. *Corpus Referência*. Uma música pode discorrer sobre múltiplos tópicos, até mesmo em um mesmo trecho, e a interpretação pode ser subjetiva e não determinística. Com o objetivo do resultado final possuir uma granularidade capaz de identificar quais partes da música correspondem a quais tópicos, cada música do *corpus* é separada em trechos.

3.1. Identificação de Tópicos

Como mostra a Figura 2, a etapa de Identificação de Tópicos é realizada em duas fases: a) criação de uma *distribuição de temas* contidos nas músicas usando PE sobre um LLM (e.g. Sabiá, GPT4), e b) *agrupamento dos temas* relevantes e frequentes em um conjunto representativo e não redundante de tópicos usando BERTopic.

Essa estratégia de LLMusic para a identificação de tópicos tem algumas vantagens. Primeiro, a mistura aleatória de trechos aumenta a diversidade de temas identificados através de um PE *zero-shot*, criando uma distribuição confiável através da repetição. Como efeito secundário, o *corpus* referência acaba sendo reformulado através dessa combinação, dando destaque a temas originalmente marginais nas letras, o que minimiza a preocupação com o seu tamanho e possível viés. Segundo, como o BERTopic é baseado em agrupamento por densidade, ele ajuda a eliminar *outliers*, priorizando os temas mais frequentes dentro da distribuição. Ao descrever o *corpus* musical a partir dos temas com maior frequência, minimizamos o problemas de alucinações e de respostas pouco significativas, comuns em aplicações de LLMs.

1) Criação de uma Distribuição de Temas: Para criar uma distribuição confiável, em-

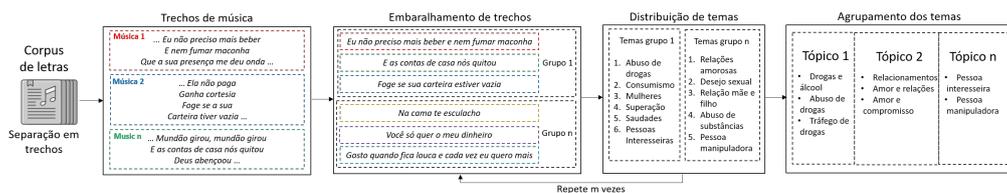


Figura 4. Ilustração de identificação de tópicos

baralhamos iterativamente os trechos para formar grupos aleatórios, e os usamos como entrada para um modelo de PE, repetindo este processo várias vezes, e agregando ao final os resultados, como ilustrado na Figura 4:

1.a) *Embaralhamento de trechos*: os trechos de todas as músicas são embaralhados e divididos aleatoriamente em grupos. A premissa é que a mistura aleatória de grupos proporcionará uma maior diversidade de temas, ao mesmo tempo que lida com alucinações e não determinismo. Além disso, tem como efeito secundário reformular o *corpus* inicial. Ainda, ao organizar os trechos em grupos menores, pode-se ajustar o tamanho de tais grupos para lidar com limitações de tamanho máximo de *tokens* do LLM. Dada a necessidade de trabalhar com menor número de *tokens*, pode-se reduzir o número de trechos por *prompt*. Por outro lado, um menor número de *prompts* diminui a quantidade de requisições à API do LLM.

1.b) *Engenharia de Prompt para extração de temas*: exploramos PE conforme o template da Figura 3.(a) para extrair vários temas para cada grupo de trecho. Combinado com o embaralhamento e a repetição, esta estratégia garante uma distribuição ampla de temas. A distribuição final de temas, que é a agregação de todos os temas ao longo das repetições, permite descrever o *corpus* musical a partir de temas recorrentes, minimizando a influência de possíveis alucinações do LLM.

2) **Identificação de Tópicos pelo Agrupamento de Temas**: o passo anterior resulta em uma distribuição de temas rica, mas possivelmente redundante. Usamos BERTopic para resumir esses temas em um conjunto menor de tópicos representativos. BERTopic garante o agrupamento de temas redundantes em um mesmo tópico baseado na similaridade semântica entre os temas, ao mesmo tempo que trata regiões pouco densas (temas infrequentes ou pouco representativos) como *outliers*.

3.2. Atribuição de Tópicos

Dado o conjunto de tópicos obtido na etapa anterior, a segunda tarefa é a *Atribuição de Tópicos* a trechos de músicas de um *corpus* a ser analisado. LLMusic adota uma estrutura de PE *zero-shot*, configurando-se como uma tarefa de classificação não-supervisionada, onde não são necessários dados rotulados para o treinamento dos classificadores multi-rótulo. Essa etapa explora uma LLM com *template* da Figura 3.(b), sendo que cada requisição compara individualmente cada tópico com cada trecho de música, avaliando o nível de relacionamento semântico entre ambos.

4. Resultados

Nesta seção avaliamos como o framework LLMusic sobre músicas do gênero funk. Mostramos como são extraídos tópicos a partir de um *corpus* referência do funk, a serem usados para classificar trechos de outras músicas do gênero. O repositório público¹ contém

¹https://github.com/yepez26/SBBD2024_FUNK

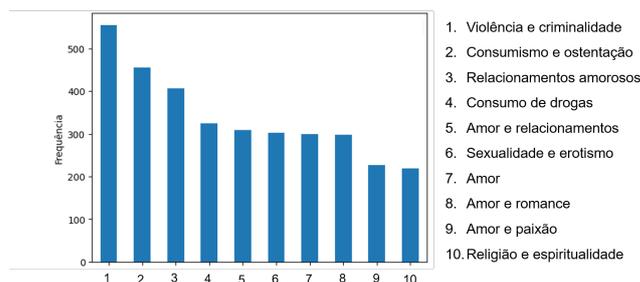


Figura 5. Distribuição de temas

detalhes do *corpus* referência, dos tópicos encontrados, e os notebooks com detalhes da implementação de cada etapa do LLMusic.

4.1. Corpus referência

Adotamos as 18 músicas analisadas em [Peres 2023], que estuda a expressão da masculinidade em letras de funk. O estudo observa a expressão do poder através da violência, ostentação, objetificação da mulher, e glorificação da figura materna, temáticas identificadas também em outros estudos (e.g. [Lopes 2011]). Extraímos as músicas do site *Letras.com.br* por *scraping*, e a divisão em trechos corresponde à separação de estrofes no site. As 18 músicas do *corpus* foram separadas em 174 trechos.

4.2. Identificação de tópicos

A estratégia e estrutura do *prompt*, bem como a LLM, foram definidas experimentalmente. Como LLM, resultados experimentais levaram à adoção do *Sabiá-2-medium* [Ramon Pires and Nogueira 2023], treinado em Português, pois para a tarefa de geração de temas mostrou-se levemente superior ao GPT4, com custo bem menor. Em termos de *prompting*, o melhor resultado foi obtido usando a estratégia *zero-shot* com o *template* da Figura 3.(a). Observamos maior consistência dos resultados quando solicitado um número específico de temas. Definimos 5 temas por *prompt*, um limiar baixo o suficiente para fugir de temas não representativos e alto suficiente para garantir um bom número de temas representativos.

Com 20 repetições, geramos 1.191 temas. Na Figura 5, que mostra os 10 temas mais frequentes, observamos redundância nos temas *relacionamentos amorosos*; *amor*; *amor e romance*; *amor e paixão*.

Para extrair tópicos a partir dessa distribuição de temas, configuramos o BERTopic com SBERT como modelo pré-treinado, redução a 5 dimensões (UMAP), e mínimo de 25 temas por cluster (HDBSCAN). Como resultado, os 1.191 temas foram reduzidos a 13 tópicos, onde 260 temas foram considerados *outliers*. Todos os tópicos foram avaliados por uma antropóloga especialista em funk e seu impacto social. A pesquisadora interpretou os agrupamentos, atribuiu semântica aos tópicos, recomendando o descarte de dois tópicos próprios ao funk (Apologia ao funk, Apresentação de MCs) pela pouca relevância ao estudo social que o gênero permite.

O resultado é sumarizado na Tabela 1. Cada tópico é descrito por um identificador, a descrição do rótulo elaborada pela especialista, os temas mais representativos dentre os agrupados no tópico e trechos de músicas que ilustram o tópico.

Tabela 1. Trechos das músicas “Perdoa mãe” (1) “Mãe de traficante” (2), “Amor de verdade”, (3), “Novinha terrível” (4, 10), “Fala mal de mim” (9), “Mais amor, menos recalque” (7, 11), “Angra dos Reis” (5, 6) e “Nunca vendeu maconh” (8).

ID Tópico	Descrição por especialista	Temas Representativos	Trechos de Exemplo
1	Arrependimento	arrependimento e pedido de perdão aos pais;	“Perdoa, mãe, por eu não ter ouvido a senhora Perdoa, mãe, a todos os conselhos que eu joguei fora Perdoa, mãe, por eu ter abandonado a escola ...”
2	Relações familiares	Relações familiares e independência financeira; Afeição e relações familiares	“Ai que saudade daquele menino correndo, sorrindo com os olhos cheio de felicidade, que decepção, meu filho traficante...”
3	Relacionamentos	Relacionamentos e indecisão; Relacionamentos e comunicação	“Na vida podem existir milhares, mas nenhuma vai ser com você...” “...o amor foi de verdade, que fez a minha vida acontecer...”
4	Aparência e comportamento das mulheres	Comportamento e aparência das mulheres em situações sociais; Estereótipos de aparência e comportamento das mulheres	“Anda bem cheirosa Porta Morena Rosa Melhor que tá tendo E quando passa eu falo nossa!”
5	Dinâmicas de sedução, sexo, desejo sexual	Comportamento sexual e romântico; Comportamento sexual e atração física	“Deixa os parça de escanteio e sobra espaço pra mais seis mulher”
6	Festas e bailes Funk	Cultura do funk e dança; Música e cultura funk	“Angra dos Reis, 40 graus, eu quero baile de funk...” “De 1100, rolé vai adiante ...”
7	Consumo e ostentação	Estilo de vida e consumo de produtos de marcas famosas; Estilo de vida e consumo de produtos de moda e beleza	“Que eu vou de Hornet, de Amarak 20 mil cruzeiros pra gastar”
8	Tráfico e consumo de substâncias	Consumo de álcool e drogas; Consumo de drogas e violência	“Quem disse que dinheiro não cresce em árvores; Nunca vendeu maconha”
9	Violência e crime	Violência e linguagem vulgar na música; Violência e linguagem vulgar na música e na sociedade	“Não olha pro lado, quem tá passando é o bonde; Se ficar de caozada, a porrada come.”
10	Vida nas favelas	Localização geográfica e referências a bairros e favelas; Representação da cultura e comportamento das favelas brasileiras	“A quebrada sabe Que ela é outro nível.”
11	Reflexões sobre a vida e superação de desafios	Trabalho e superação de dificuldades; Reflexão sobre a vida e superação de obstáculos	“Vai se acostumando que eu cheguei para ficar Porque daqui ninguém me tira”

4.3. Atribuição de tópicos via Engenharia de Prompt

Dado um *corpus* não anotado de músicas funk divididas em trechos, a relação semântica de cada trecho com cada tópico é quantificada em uma escala de 1 a 5 usando o template de *prompt* da Figura 3.(b). Para essa tarefa, o LLM GPT4 teve melhor desempenho comparado ao Sabiá, que apresentou comportamento inconsistente nas respostas do modelo.

A Figura 6 ilustra os tópicos atribuídos para dois trechos da musica “Novinha Terrível”, que descreve uma jovem da favela sob a ótica masculina. O primeiro fala sobre marcas de roupa que ela usa, e o modelo relacionou esse trecho com tópico 7 (*Consumismo e ostentação*) com nota máxima. O modelo classifica o segundo trecho com nota alta nos tópicos 4 (*Aparência e comportamento das mulheres*), 5 (*Dinâmicas de sedução, sexo, desejo sexual*) e 6 (*festas, bailes Funk*), pois critica que ela prefere ir a festas ao invés de envolver-se amorosamente. Assim, o modelo mostra-se capaz de classificar corretamente os trechos dentro dos tópicos identificados, incluindo uma certa subjetividade.

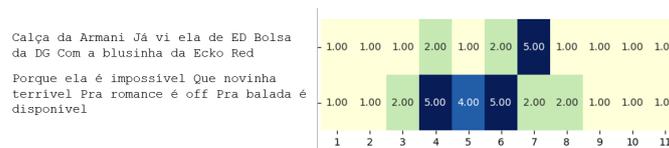


Figura 6. Classificação de trechos da música “Novinha Terrível”

4.4. Avaliação

a) Desempenho da Identificação dos Tópicos: com o objetivo de validar a qualidade dos tópicos gerados no processo descrito na Seção 4.2, fizemos uma comparação utilizando BERTopic diretamente sobre o corpus referência de 18 músicas (Seção 4.1). No BERTopic foi usado UMAP para reduzir os *embeddings* a 5 dimensões, e um mínimo de 7 documentos por cluster no HDSCAN. Esses parâmetros foram escolhidos visando maximizar a métrica de coerência c_v [Röder et al. 2015]. O valor máximo obtido para c_v foi 0,54, resultando em 8 tópicos. Os tópicos obtidos com LLMusic resultaram nesse mesmo valor de c_v . Ambos os tópicos foram avaliados por uma antropóloga especialista em funk.

A Tabela 2 apresenta as quatro (4) palavras mais representativas de cada tópico encontrado em cada abordagem. No LLMusic, os tópicos são identificados com os mes-

Tabela 2. Tópicos extraídos por LLMusic e BERTopic do *Corpus* Referência

LLMusic ID	Tópicos LLMusic	BT ID	Tópicos BERTopic
1	perdão, religião, arrependimento, pedido	A	nos, trampo, quem, minha
2	família, familiares, relações, mãe	B	quica, xerequinha, senta, xerecao,
3	relacionamentos, relacionamento, amor, comportamento	C	vai, hoje, revoada, mc
4	mulheres, comportamento, aparência, estereótipos	C	mulher, quer, patrão, ama
5	sexual, comportamento, sexualidade, relacionamentos	E	piriguete, mexe, rebola, devagar
6	comportamento, dança, festas, ambientes	F	solta, beat, espelho, apitar
7	moda, marcas, estilo, aparência	G	bala, mete, love, cheirosa
8	violência, consumo, drogas, comportamento	H	favela, baile, menor, preparado
9	música, vulgar, violência, linguagem		
10	referências, favelas, bairros, cultura		
11	vida, sucesso, superação, trabalho		
12	funk, cultura, comportamento		
13	funk, músicas, cultura, brasileiras		

mos IDs da Tabela 1, sendo que os tópicos 12 e 13 corresponde àqueles descartados por sugestão da especialista. No BERTopic, os tópicos são identificados pelas letras A .. H. Pode-se observar a enorme diferença entre os tópicos encontrados considerando as respectivas palavras representativas.

Os tópicos gerados por BERTopic são baseados no conteúdo explicitamente representado, e portanto, limitado aos trechos do *corpus* de referência, cujas músicas foram cuidadosamente selecionadas para estudar a expressão da masculinidade em músicas funk. Os tópicos envolvem prioritariamente temáticas identificadas em [Peres 2023], e as palavras representativas revelam algum nível de relação com temáticas de objetificação da mulher (B, E), violência e poder (A, C, G), além de comportamentos de bailes funk (H). Alguns tópicos são menos claros em sua semântica (D, F), e as palavras nem sempre permitem identificar o tema com clareza (e.g. G, A, C). Não há tópicos específicos para ostentação e a figura materna, assim como outros tópicos que expressam a cultura funk como um retrato social da periferia [Lopes 2011].

Em relação aos tópicos identificados pelo LLMusic, a antropóloga classificou-os como abrangentes e representativos do movimento. Cabe enfatizar que, apesar de extraídos de um *corpus* reduzido com um enfoque específico, a capacidade generativa da LLM, aliada ao embaralhamento dos trechos, permitiu a captura de temas tangenciados nessas letras. Assim, foi possível capturar temas sutilmente representados, como drogas, a relação familiar, a vida na favela, etc. Ainda, os tópicos gerados por LLMusic diferenciam-se claramente uns dos outros, não somente em relação às palavras que os representam, mas também quanto aos conteúdos cobertos. Temas como relações familiares, ostentação, relacionamentos amorosos e consumo de drogas são claramente decifráveis a partir do subconjunto de palavras. Já nos tópicos gerados pelo BERTopic, a interpretação dos tópicos gerados requer um exercício de abstração maior do avaliador, onde a repetição das palavras (e.g., ama, mulher, patrão) ao longo dos tópicos e a falta de coesão semântica no conjunto das palavras dificulta a interpretação de significados aos tópicos gerados.

Conclui-se que o LLMusic é capaz de identificar tópicos implicitamente representados, e que tem boa capacidade para lidar com linguagem figurativa, poética, e gírias. Ainda, é robusto ao tamanho, enfoque e possível viés do *corpus* referência.

b) Desempenho da Classificação de Trechos: para avaliar o desempenho da abordagem *zero-shot* de atribuição de tópicos, construímos uma base de teste. Selecionamos músicas conhecidas usando como critério conter tópicos identificados na Tabela 1. Após esclarecimentos sobre o significado de cada tópico, três anotadores foram instruídos a ler trechos de músicas, e atribuir a cada trecho ausência ou relação com os tópicos. Selecionamos somente os trechos onde pelo menos dois anotadores concordaram com a relação com o tópico, o que resultou em 131 trechos anotados com pelo menos um tópico, dos quais 82

Tabela 3. Desempenho da atribuição de tópicos

Tópico	1	2	3	4	5	6	7	8	9	10	11
Revocação ponderada	92,40%	73,93%	75,14%	77,94%	91,95%	69,37%	92,98%	89,84%	86,36%	78,92%	72,52%
Precisão ponderada	87,70%	81,35%	80,24%	78,35%	92,31%	84,45%	88,85%	87,52%	88,13%	81,64%	86,72%
F1 ponderada	81,65%	80,61%	78,84%	77,40%	92,04%	82,74%	84,53%	84,83%	87,64%	81,54%	86,87%
Acurácia	85,50%	83,97%	82,44%	78,63%	92,37%	81,68%	87,02%	86,26%	88,55%	82,44%	87,02%

estavam associados a mais de um tópico. Esses trechos passaram pelo *prompt* de cada um dos 11 tópicos, com resultados de 1 a 5. O tópico foi considerado presente usando o limiar ≥ 3 para interpretar a saída do *prompt*, e não presente se inferior. Comparamos então esse resultado com os rótulos atribuídos pelos anotadores.

Considerando os 131 trechos anotados como *ground truth*, a Tabela 3 mostra o desempenho do método *zero-shot* proposto para atribuição de tópicos. Utilizamos as métricas precisão, revocação e F1 para medir o resultado da classe positiva (presença do tópico) e negativa (ausência) para cada tópico, onde os trechos anotados com um dado tópico foram considerados positivos, e os demais, negativos. Os resultados apresentados correspondem à média ponderada das classes positiva/negativa, onde o peso foi a proporção de casos em cada classe. A acurácia média considerando todos os tópicos é de 85,1% e o F1 ponderado médio é de 83,5%. A precisão ponderada varia de 78,35% a 92,31%, enquanto que a revocação ponderada varia entre 72,52% e 92,4%. Os valores de precisão são superiores aos de revocação, salvo no tópico 1.

Observamos assim um bom desempenho, com bom compromisso entre precisão e revocação, e com equilíbrio entre as classes positiva e negativa. A análise dos resultados trecho a trecho sugerem uma certa dificuldade do *prompt* na tarefa de atribuição em trechos onde o tópico é expresso de forma mais subjetiva e indireta, o que explica a revocação mais baixa, e a precisão mais alta.

5. Estudo de caso: Análise dos 100 funks mais tocados

Para ilustrar o potencial do framework LLMusic, analisamos as 100 músicas classificadas com o gênero funk mais tocadas segundo o site *Letras.com.br*, que usa as visualizações do *Youtube* para compor o ranking. Segundo a especialista, o número de visualizações no *Youtube* é mais representativo que o ranking de outras ferramentas de *streaming*, como o *Spotify*, pouco consumido pela população da periferia. Foram considerados os tópicos contidos na Tabela 1. As 100 letras foram segmentadas em 1.113 trechos, que após remoção de duplicados, resultou em 805 trechos únicos. O repositório público detalha as músicas e os rótulos atribuídos a cada trecho.

A Figura 7 representa a distribuição de tópicos nos 11 funks mais tocados, apresentados em ordem alfabética, ilustrando a riqueza do gênero. Os pesos representam a média dos tópicos considerando todos os trechos de cada música. O valor 5 representa que todos os trechos da música contêm o tópico, enquanto que o valor 1 indica que nenhum dos trechos da música contêm o tópico. Por exemplo, a música *Amor de Verdade*, que fala do amor por uma pessoa especial, recebe o peso de 4,75 no tópico 3 (*Relacionamentos*), e 1,12 (i.e. próximo a inexistente) no tópico 4 (*Aparência e comportamento das mulheres*). O preconceito ao funk muitas vezes está centrado nas temáticas associadas aos tópicos 4-9, mas percebe-se que o tópico 3 é o mais recorrente nas músicas mais tocadas. Além disso, observa-se a importância de temas como os *Relacionamentos Familiares* (2) e *Superação* (11). Alguns trechos dessas músicas ilustram os tópicos na Tabela 1.

Analisando os tópicos atribuídos aos trechos dessas 11 músicas, a antropóloga es-



Figura 7. Distribuição de tópicos nos 11 funks mais tocados.

pecialista avaliou que, em linhas gerais, LLMusic produziu resultados consistentes para as narrativas das letras. A música *Angra dos Reis* descreve uma festa nessa região de alto padrão do Rio de Janeiro. A letra descreve o evento em termos de marcas de luxo e seu custo e da presença de mulheres (tópicos 4, 6 e 7), e da conquista que esse evento representa para alguém da periferia (tópico 10). *Novinha Terrível* e *Fala mal de mim* são em certo sentido similares, descrevendo a força, beleza e charme da mulher de periferia, a primeira pela perspectiva de um homem (objetificação) e a segunda da mulher (empoderamento). Ambas falam de aparência de mulheres e festa (tópicos 4 e 6), mas *Novinha Terrível* objetifica a mulher (tópico 5). *”Nunca vendeu maconha”* descreve diversas alucinações após o consumo de *cannabis*. Essa letra ilustra a dificuldade de perceber certas nuances, pois embora 100% dos trechos mencionem o consumo da droga, somente 8 dos 24 trechos tiveram notas maiores ou igual a 3. As gírias (e.g. *puxada*, *brisa*) que associam a droga à alucinação não são reconhecidas pelo modelo.

A Figura 8 mostra a participação de cada um dos 11 tópicos dentro das 100 músicas de funk em termos de porcentagem, utilizando dois critérios. O limiar é ≥ 3 para a presença do tópico, e ausência, em caso contrário. O primeiro critério associa a música ao tópico se ela inclui pelo menos um trecho daquele tópico. Sobre essa ótica, confirma-se que a narrativa dessas músicas envolvem sexualização, mulheres e festas. Esse resultado indica prevalência dos tópicos 6, 4 e 5, nessa ordem, em 70-80% das músicas, reforçando assim o estereótipo associado ao gênero. Além disso, aproximadamente 50% das top-100 músicas incluem a temática mais geral de relacionamentos (tópico 3). A segunda abordagem é mais restritiva, e considera que um tópico está contido dentro da música se ao menos 50% dos trechos o contêm. Isso permite abstrair de referências superficiais, concentrando nas narrativas centrais de cada letra. Essa abordagem destaca a prevalência dos tópicos 6 e 4, seguidos por 5 e 3 em iguais proporções.

As figuras 7 e 8 ilustram o potencial do *framework* para proporcionar *insights* a partir de um *corpus* de músicas. Os dados quantitativos mostram que as narrativas presentes nas letras de funk vão além dos estereótipos, incluindo relacionamentos em geral e com a família, a vida cotidiana da favela, e histórias de superação sob a ótica da falta de oportunidades nas favelas. Essa variedade revela a multiplicidade de experiências vivenciadas no dia a dia das comunidades da periferia [Peres 2023, Lopes 2011].

6. Conclusões e trabalhos futuros

Apresentamos o LLMusic, um *framework* para a identificação e classificação não supervisionada de tópicos em letras de músicas, que explora o potencial de LLMs através da

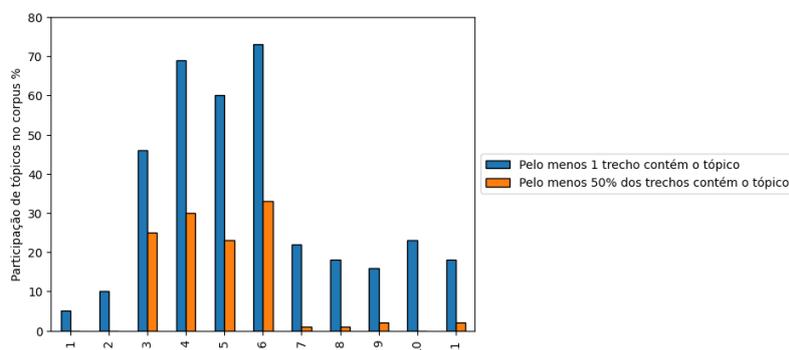


Figura 8. Participação de tópicos no corpus.

EP para vencer as limitações de métodos tradicionais de MTs. Apesar do potencial de generalização, nos limitamos ao gênero funk, cujas letras expressam um discurso do dia a dia das periferias de usando gírias, figuras de linguagem e subjetividade. Dado um *corpus* referência do gênero musical, exploramos a capacidade generativa das LLMs para interpretar trechos musicais combinados aleatoriamente, reduzindo-os a temas abordados na letras, os quais são posteriormente condensados em uma lista de tópicos coerente, não redundante, e enxuta usando BERTopic. Os tópicos assim gerados podem ser atribuídos a novos trechos de músicas, também em uma abordagem de *prompting zero-shot*.

Com uma antropóloga especialista em funk avaliamos a relevância e representatividade dos tópicos extraídos, e desenvolvemos um estudo de caso. O estudo de caso ilustrou a capacidade de análises em maior escala, confirmando que o gênero vai além dos estereótipos. A abordagem proposta pode ser utilizada como uma ferramenta de recuperação de informações, apoiando estudos qualitativos em busca de músicas representativas que abordem um tema específico. O LLMusic serve também para identificar músicas com determinadas características, aprofundar os tópicos em subtópicos (e.g. detalhar o discurso sobre as mulheres a partir das letras associadas ao tópico 4).

LLMusic elimina a necessidade de dados anotados, e destaca-se pela sua flexibilidade e adaptabilidade, pois permite configurar tamanho de corpus, critério de separação de letras e modelos LLMs. Também é robusto às características e tamanho do corpus referência. Adicionalmente, sua estrutura modular permite substituições independentes das etapas, bem como a utilização de diferentes estruturas de *prompt* tanto na geração dos temas como na atribuição dos tópicos, além da substituição do BERTopic por outro modelo de MTs no agrupamento de temas.

Utilizamos o funk como um estudo de caso para avaliar a eficiência do LLMusic em capturar nuances e tópicos que aparecem em letras. No geral, apresentou bom desempenho, mas identificamos limitações, possivelmente relacionadas à singularidade das gírias, à natureza informal das letras e à subjetividade expressa que depende do contexto das favelas. Trabalhos futuros envolvem um estudo temporal do funk para analisar mudanças no gênero, generalizações do método para outros gêneros musicais, suporte à avaliação dos tópicos, e melhoria no *framework* LLM propriamente dito.

Agradecimentos: Pesquisa parcialmente apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Projeto PETWIN (financiamento FINEP e Consórcio de LIBRA).

Referências

- Betti, L., Abrate, C., and Kaltenbrunner, A. (2023). Large scale analysis of gender bias and sexism in song lyrics. *EPJ Data Science*, 12(1):10.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Brilhante, A. V. M., Giaxa, R. R. B., Branco, J. G. d. O., and Vieira, L. J. E. d. S. (2019). Cultura do estupro e violência ostentação: uma análise a partir da artefactualidade do funk. *Interface-Comunicação, Saúde, Educação*, 23:e170621.
- Calcina, Erik e Novak, E. (2022). Measuring the similarity of song artists using topic modelling. In *Proc. of the 25th Intl. Multiconference Information Society - Data Mining and Data Warehouses (SiKDD)*, page 103–106.
- Devi, M. D. and Saharia, N. (2020). Exploiting topic modelling to classify sentiment from lyrics. In *Proc. of the 2nd Intl. Conferemce on Machine Learning, Image Processing, Network Security and Data Sciences (MIND)*, pages 411–423.
- Grootendorst, M. (2022). BERTopic: Leveraging bert and topic modeling for efficient document clustering. <https://maartengr.github.io/BERTopic>.
- Junior, J. S., Rossi, R., and Lobato, F. (2019). Uma abordagem baseada em letras para a descoberta de conhecimento da música brasileira: o sertanejo como um estudo de caso. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 949–960.
- Lopes, A. C. (2011). *Funk-se Quem Quiser: No Batidão Negro Da Cidade Carioca*. Bom Texto FAPERJ.
- Oramas, S., Espinosa-Anke, L., Gómez, F., and Serra, X. (2018). Natural language processing for music knowledge discovery. *Journal of New Music Research*, 47:365–382.
- Pengfei Liu, Weizhe Yuan, J. F. Z. J. H. H. and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACMCom-put.*, 55(9):35.
- Peres, F. C. (2023). Puta ou santa: as relações com mulheres enquanto elemento constituinte das masculinidades do funk brasileiro? In *Anais do IV Encontro Anual de Antropologia do Mercosul*.
- Pham, C. M., Hoyle, A., Sun, S., Resnik, P., and Iyyer, M. (2024). Topicgpt: A prompt-based topic modeling framework. <https://doi.org/10.48550/arXiv.2311.01449>.
- Ramon Pires, Hugo Abonizio, T. S. A. and Nogueira, R. (2023). Sabía: Portuguese large language models. *Anais da XII Brazilian Conference on Intelligent Systems*, 12(1):15.
- Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.
- Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., and Hashimoto, T. B. (2024). Benchmarking Large Language Models for News Summarization. *Transactions of the Association for Computational Linguistics*, 12:39–57.