

Influenciadores Digitais Católicos: Uma Análise Centrada em Dados

Fernanda Ferreira do Nascimento¹, Melissa Sousa Felipe
José Maria Monteiro¹, Javam Machado¹

¹Universidade Federal do Ceará (UFC)
Av. Humberto Monte, s/n, Pici - CEP 60440-593 – Fortaleza – CE – Brasil

{fernanda.nascimento, melissasousa}@alu.ufc.br

{jose.monteiro, javam.machado}@lsbd.ufc.br

Abstract. Nowadays, social media platforms are widely used to publish and consume spiritual content, reflecting the presence of a broad and diverse audience seeking to exercise their faith online. In this context, Catholic digital influencers stand out, sharing religious content using innovative language and boasting a large number of followers. This work investigates the content produced by four prominent Catholic digital influencers: Bernardo Küster, Paulo Ricardo, Patrick Fernandes, and Júlio Lancellotti. To achieve this, four datasets were constructed — one for each influencer — based on videos published on YouTube. Next, the LDA and PTM algorithms were applied to identify latent topics in each dataset. Subsequently, the LIWC tool was used to analyze the emotional, cognitive, and structural components of the corresponding texts. The findings reveal distinct discursive patterns, engagement strategies, and political-religious positioning among the influencers, highlighting opposing narratives and contrasting communicative styles

Resumo. Atualmente, as redes sociais são amplamente utilizadas para publicar e consumir mensagens de caráter espiritual, refletindo a presença de um público amplo e diverso interessado em exercitar sua fé em ambientes digitais. Nesse contexto, destacam-se os influenciadores digitais católicos, que compartilham conteúdos religiosos por meio de uma linguagem inovadora e reúnem um expressivo número de seguidores. Este trabalho investiga o conteúdo produzido por quatro influenciadores digitais católicos de destaque: Bernardo Küster, Paulo Ricardo, Patrick Fernandes e Júlio Lancellotti. Para tal, foram construídos quatro conjuntos de dados — um para cada influenciador — com base nos vídeos publicados na plataforma YouTube. Em seguida, os algoritmos LDA e PTM foram utilizados para identificar os tópicos latentes em cada conjunto analisado. Posteriormente, utilizou-se a ferramenta LIWC para examinar os componentes emocionais, cognitivos e estruturais dos textos. Os resultados revelam padrões discursivos, estratégias de engajamento e posicionamentos político-religiosos distintos, evidenciando narrativas e estilos comunicativos completamente diferentes, muitas vezes contraditórios entre si.

1. Introdução

Atualmente, milhares de pessoas utilizam as redes sociais para publicar e consumir mensagens de natureza espiritual, o que indica a existência de um público amplo e diversificado interessado em exercer sua fé em ambientes digitais. Nesse contexto, destacam-se os influenciadores digitais católicos (IDCs), que compartilham conteúdos religiosos por meio de linguagens inovadoras e acumulam um número expressivo de seguidores. Esses influenciadores têm adquirido relevância crescente na formação de opinião e de valores entre fiéis católicos, superando, em muitos casos, a própria estrutura oficial de ensino (catequese) mantida pela hierarquia da Igreja Católica [Medeiros et al. 2024]. Ademais, a atuação desses influenciadores vem se consolidando como uma tendência irreversível no campo religioso contemporâneo.

Este artigo tem como objetivo analisar o conteúdo discursivo compartilhado por quatro dos principais influenciadores digitais católicos no Brasil — Bernardo Küster, Paulo Ricardo, Patrick Fernandes e Júlio Lancellotti — por meio de vídeos publicados em seus canais no YouTube. A escolha desses nomes foi baseada no número de seguidores durante o período analisado, considerando também a diversidade de estilos e posicionamentos, conforme discutido anteriormente em [Medeiros et al. 2024].

Para viabilizar as análises, foram construídos quatro conjuntos de dados distintos, com base nas transcrições dos vídeos publicados entre janeiro e maio de 2024. Para a análise dos principais temas abordados pelos influenciadores, foram aplicados os algoritmos *Latent Dirichlet Allocation* (LDA) [Blei et al. 2003] e *Pseudo-document-based Topic Modelling* (PTM) [Zuo et al. 2016], selecionados por sua ampla utilização e eficácia em tarefas de descoberta de tópicos latentes em textos longos. Em seguida, foi realizada a extração e análise das entidades nomeadas mais frequentes, utilizando técnicas de NER (Named Entity Recognition), com o objetivo de mapear os principais personagens referenciados nos discursos. A partir das entidades identificadas, foram realizadas buscas textuais na ferramenta *ElasticSearch*, que permitiram identificar trechos representativos para a caracterização do discurso. Por fim, foi utilizada a ferramenta LIWC (*Linguistic Inquiry and Word Count*) [Tausczik and Pennebaker 2010] para identificar e comparar traços emocionais, cognitivos e estruturais nos conteúdos de cada influenciador.

A partir da análise qualitativa dos tópicos e trechos extraídos, foi possível identificar quatro tendências discursivas predominantes entre os influenciadores analisados, que aqui são interpretadas como diferentes modos de atuação. Tais categorias não foram identificadas de forma automática, mas emergiram da análise dos dados, em diálogo com a literatura.

O primeiro modo foi identificado como “extremo conservadorismo”, caracterizado por uma oposição veemente à doutrina social da Igreja e pela defesa de uma economia ultraliberal, mesmo que isso implique deslegitimar a autoridade do Papa e do Magistério [Oro 1996, McConahay 2023]. O segundo, denominado conservadorismo, busca preservar os valores e a identidade católica tradicional, rejeitando ideias consideradas modernas ou seculares [Villaga 2006, Passos 2021]. O terceiro modo, chamado “isenção programada”, foca em conteúdos espirituais e devocionais, com pouca ou nenhuma referência a questões sociopolíticas [Medeiros et al. 2024]. Por fim, o quarto modo, identificado como “ativismo progressista”, prioriza pautas sociais e prega uma Igreja mais aberta e engajada com os desafios contemporâneos [Dias 2012, Campos 2022].

Cabe ressaltar que as tendências discursivas identificadas não são absolutas nem exclusivas, mas sim aproximações analíticas baseadas em padrões observados nos dados. A metodologia utilizada, embora aplicada ao contexto brasileiro, pode ser replicada em estudos sobre influenciadores religiosos em outros contextos culturais e plataformas digitais.

O restante deste artigo está estruturado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve a metodologia utilizada e os conjuntos de dados construídos; a Seção 4 detalha as análises realizadas; e, por fim, a Seção 5 apresenta as conclusões obtidas e sugere possíveis desdobramentos futuros.

2. Trabalhos Relacionados

Vários estudos recentes têm explorado o uso de redes sociais como espaços de disputa simbólica e ideológica, especialmente em contextos polarizados ou religiosos. Nesta seção, discutem-se os principais trabalhos relacionados ao tema, com ênfase nas metodologias utilizadas e nas contribuições em relação à presente pesquisa.

Em [Ebeling et al. 2020], os autores analisaram como a polarização política influencia o comportamento de grupos com visões opostas, tomando como base o debate público sobre a COVID-19 no Brasil. Dois grupos de usuários do Twitter foram selecionados — os “Cloroquiners” e os “Quarenteners” — e submetidos a uma análise composta por quatro etapas: (i) inferência da orientação política com base em interações e conteúdo; (ii) aplicação de modelagem de tópicos para medir a coesão temática dos discursos; (iii) análise de redes para identificar comunidades e sua estrutura relacional; e (iv) exame linguístico para inferência de aspectos psicológicos. O estudo destaca a influência da identidade política na formação discursiva, mas se limita a textos curtos (*tweets*) e não contempla o discurso religioso, ao contrário do presente trabalho.

No trabalho apresentado por [Brum et al. 2020], foi realizada a caracterização de mais de 56 milhões de *tweets* em português durante 70 dias da pandemia. A análise baseou-se em frequência de termos, agrupamento por *hashtags* e identificação de temas recorrentes. As mensagens giravam majoritariamente em torno de temas políticos e sanitários, como quarentena, uso de hidroxicloroquina e aglomerações. Embora o volume de dados seja expressivo, o enfoque é quantitativo-descritivo e pouco se aprofunda em questões emocionais ou discursivas complexas, como as que abordamos por meio de ferramentas como o LIWC.

O estudo descrito em [Ebeling et al. 2023] propõe um *framework* de análise multidimensional, combinando coleta de dados, inferência de inclinação política, análise topológica de redes, modelagem híbrida de tópicos e caracterização de fontes de informação. O foco do trabalho de [Ebeling et al. 2023] é o alinhamento político, enquanto o presente artigo concentra-se no discurso religioso e suas variações ideológicas.

Em [Jacintho et al. 2020], os autores conduziram uma profunda análise das eleições presidenciais brasileiras com base em padrões temporais e geográficos. O estudo utilizou agrupamentos geolocalizados e séries temporais para traçar o comportamento do eleitorado. Apesar de abordar comportamento político, o enfoque é macrossociológico e não textual/discursivo, o que distancia sua abordagem da proposta deste artigo.

[Guimarães et al. 2022] analisaram a estrutura e o conteúdo de propagandas eleitorais antecipadas no Twitter, em períodos pré-eleitorais (2016, 2018 e 2020). Foi observada a prevalência de discursos negativos ou neutros e uma estrutura textual relativamente constante. O estudo compartilha o interesse em padrões discursivos, mas trata de mensagens explicitamente políticas e publicitárias, enquanto o presente trabalho aborda manifestações religiosas e espirituais.

Em [Silva et al. 2021], foram utilizados algoritmos de modelagem de tópicos para resumir comentários sobre projetos de lei no Congresso Nacional. A metodologia incluiu uma técnica de incorporação de sentenças para gerar representações vetoriais de textos em português, permitindo agrupamento temático automático. Esse estudo dialoga diretamente com a modelagem de tópicos presente neste trabalho, embora o corpus seja jurídico e a análise se restrinja a conteúdo racional-argumentativo.

Já o trabalho de [Melo et al. 2024] explora o uso de *stickers* (adesivos) com fins políticos em grupos públicos do WhatsApp no Brasil. A análise abarcou padrões visuais, frequência de envio e conotações ideológicas dos *stickers*, com base em uma grande amostragem. Apesar do objeto ser multimodal (visual e textual), o estudo destaca como mídias alternativas também atuam na formação de discursos, aspecto convergente com nossa investigação sobre plataformas digitais religiosas.

Em síntese, os trabalhos relacionados utilizam diversas técnicas relevantes, como modelagem de tópicos, análise de redes e inferência semântica. No entanto, poucos exploram a dimensão religiosa da comunicação em ambientes digitais, tampouco aplicam uma abordagem integrativa que combina LDA, PTM, NER, ElasticSearch e LIWC sobre um corpus extenso de transcrições. Este trabalho contribui, portanto, ao preencher essa lacuna, oferecendo uma análise sistemática do discurso religioso e ideológico em mídias sociais.

3. Materiais e Métodos

Esta seção descreve o delineamento metodológico adotado neste estudo, contemplando a seleção dos influenciadores, a coleta e transcrição dos vídeos, a preparação dos dados textuais e as análises aplicadas para identificar padrões temáticos, discursivos e psicolinguísticos. Todos os *scripts* utilizados no processo de coleta, transcrição, pré-processamento e análise estão disponíveis publicamente em nosso repositório no GitHub¹, assegurando a reprodutibilidade da pesquisa.

3.1. Seleção dos Influenciadores e Escopo do Estudo

Foram selecionados quatro influenciadores digitais católicos com ampla atuação no YouTube e diferentes perfis de comunicação: Bernardo Kuster, Paulo Ricardo, Patrick Fernandes e Júlio Lancellotti. A seleção foi baseada em dois critérios principais: (i) número de inscritos nos canais; e (ii) diversidade de posicionamentos religiosos e sociais, conforme reconhecido em estudos prévios [Medeiros et al. 2024]. O período de coleta compreende publicações entre 1º de janeiro e 5 de maio de 2024.

Vale ressaltar que o padre Fábio de Melo, embora relevante, não foi incluído por não ter publicado vídeos no período analisado. O YouTube foi escolhido como fonte por permitir discursos mais extensos e espontâneos, ideais para análise temática e emocional.

¹<https://github.com/jmmfilho/influenciadorescatolicos>

3.2. Coleta, Transcrição e Preparação dos Dados

A coleta foi realizada por meio da API oficial do YouTube², utilizando a biblioteca `googleapiclient.discovery`. Um *script* em Python foi desenvolvido para obter os metadados dos vídeos (título, data, descrição, duração, número de visualizações, curtidas e comentários). Os vídeos foram baixados com `yt-dlp` e os áudios extraídos via `ffmpeg` em MP3 (192 kbps).

A transcrição foi feita com o modelo *Whisper* da OpenAI³, via `Torch`. O modelo foi escolhido por sua alta acurácia em português. A validação foi realizada por amostragem: trechos dos áudios foram escutados manualmente e comparados com o texto transcrito, verificando-se elevada fidelidade semântica, mesmo com pequenas imperfeições de pontuação ou ortografia. As transcrições foram armazenadas em arquivos CSV, um para cada influenciador, contendo o identificador do vídeo e o texto correspondente. A Tabela 1 apresenta um resumo dos canais analisados, enquanto a Tabela 2 mostra estatísticas básicas dos textos transcritos.

Tabela 1. Características Gerais dos Canais Analisados

| Influenciador | Nº de Inscritos | Quantidade de Vídeos | Duração Total dos Vídeos |
|-------------------------------|-----------------|----------------------|--------------------------|
| Bernardo Kuster | 1.000.000 | 14 | 9h54min |
| Paulo Ricardo | 1.840.000 | 310 | 76h08min |
| Patrick Fernandes | 258.000 | 97 | 51h47min |
| Júlio Lancellotti (OArcaNoAr) | 92.600 | 197 | 39h22min |

Tabela 2. Estatísticas dos Textos Transcritos por Influenciador

| Estatística | Bernardo Kuster | Patrick Fernandes | Paulo Ricardo | Júlio Lancellotti |
|------------------------|-----------------|-------------------|---------------|-------------------|
| Média de Caracteres | 37.501 | 13.880 | 195 | 260 |
| Desvio-Padrão (Carac.) | 31.583 | 19.271 | 128 | 110 |
| Média de Tokens | 6.604 | 2.536 | 35 | 46 |
| Desvio-Padrão (Tokens) | 5.604 | 3.511 | 23 | 20 |
| Mínimo de Tokens | 1.828 | 10 | 1 | 1 |
| Máximo de Tokens | 23.080 | 12.779 | 112 | 89 |
| Caracteres Únicos | 99 | 134 | 78 | 83 |
| Total de Caracteres | 525.018 | 1.346.384 | 59.721 | 50.892 |

As diferenças de volume de dados entre os influenciadores refletem a dinâmica própria de cada canal: enquanto alguns publicam poucos vídeos longos (como Bernardo Kuster), outros preferem conteúdos curtos e frequentes (como Paulo Ricardo). Essa heterogeneidade é uma característica empírica relevante da amostra, e não uma limitação.

No contexto deste trabalho, um *token* refere-se a qualquer unidade léxica identificada durante a tokenização dos textos — geralmente uma palavra, número ou símbolo relevante. As contagens de *tokens* apresentadas na Tabela 2 consideram a transcrição integral, sem remoção de *stopwords*, visando preservar a estrutura linguística para ferramentas como o LIWC. Para modelagem de tópicos, as *stopwords* foram removidas com base em listas padrão para o português.

²<https://developers.google.com/youtube/v3>

³<https://github.com/openai/whisper>

Observa-se ainda, na Tabela 2, a ocorrência de vídeos com apenas 1 *token* nos conjuntos de Paulo Ricardo e Júlio Lancellotti. Esses casos correspondem a vídeos extremamente curtos ou com conteúdo predominantemente não verbal, como vinhetas ou transmissões com fundo musical. Foram mantidos por completude estatística, mas desconsiderados nas análises temáticas e linguísticas.

3.3. Etapas da Análise dos Dados

3.3.1. Reconhecimento de Entidades Nomeadas

Entidades nomeadas (ou *Named Entities*, em inglês) são expressões textuais que representam elementos específicos e bem definidos do mundo real, geralmente com nome próprio. O Reconhecimento de Entidades Nomeadas (ou *Named Entity Recognition* - *NER* é uma tarefa que consiste em detectar automaticamente entidades nomeadas em um determinado texto, além de classificá-las em categorias predefinidas (pessoas, organizações, lugares, valores monetários, datas e horários) [Hu et al. 2024]. Neste trabalho, a biblioteca *SpaCy*, juntamente com um modelo pré-treinado para português, foi utilizada para reconhecer entidades nomeadas [Schmitt et al. 2019]. As entidades nomeadas mais frequentes no vocabulário de um determinado influenciador podem fornecer “*insights*” sobre suas preferências e posições ideológicas.

3.3.2. Busca Textual com ElasticSearch

Um motor de busca textual é um sistema computacional que realiza operações de indexação, recuperação e ranqueamento de documentos textuais com o objetivo de responder eficientemente a consultas em linguagem natural ou formal. O Elasticsearch é um motor de busca textual distribuído, o qual foi projetado para indexar, pesquisar e analisar grandes volumes de dados em tempo real [Kim and Kim 2022]. As transcrições obtidas a partir dos vídeos publicados pelos influenciadores em seus canais do YouTube foram indexadas utilizando o ElasticSearch. Desta forma, foi possível utilizar as entidades nomeadas previamente identificadas com a finalidade de recuperar os trechos onde estas ocorrem. Esses textos foram analisados manualmente, a fim de obter maiores informações sobre o discurso de cada influenciador.

3.3.3. Análise de Polaridade

A análise de polaridade é uma técnica de Processamento de Linguagem Natural (PLN) que visa identificar automaticamente a valência emocional expressa em um texto, atribuindo-lhe uma categoria polar (positiva, negativa ou neutra) com base no conteúdo lexical, semântico ou contextual [Zhang et al. 2023]. Para realizar a análise de polaridade, os textos foram segmentados em blocos de 1.000 *tokens*, utilizando ferramentas da biblioteca NLTK. Essa estratégia teve por finalidade assegurar maior comparabilidade entre os textos analisados.

3.3.4. Modelagem de Tópicos

A Modelagem de tópicos (*Topic Modeling*) é um método estatístico de descoberta não supervisionada de temas latentes (tópicos) em um corpus textual, agrupando palavras que frequentemente ocorrem juntas e associando-as a tópicos interpretáveis. Neste trabalho, os algoritmos LDA (Latent Dirichlet Allocation) [Blei et al. 2003] e PTM (Pseudo-document-based Topic Model) [Zuo et al. 2016] foram utilizados para identificar os tópicos mais representativos de cada influenciador. O número de tópicos k variou de 2 a 30, sendo escolhida a configuração ótima com base na métrica de coerência C_V [Röder et al. 2015]. A avaliação da qualidade dos tópicos foi realizada a partir das métricas C_V, NPMI (*Normalized Pointwise Mutual Information*) [Bouma 2009], diversidade e IRBO (*Inverted Rank-Biased Overlap*) [Corsi and Urbano 2024], conforme implementado na biblioteca OCTIS [Terragni et al. 2020].

3.3.5. Análise Linguística

A ferramenta LIWC (*Linguistic Inquiry and Word Count*) é um software baseado em dicionários que analisa textos e fornece estatísticas sobre a frequência de palavras pertencentes a categorias psicológicas, sociais e linguísticas, permitindo inferências sobre o estado emocional, estilo de linguagem e traços cognitivos do autor [Carvalho et al. 2019]. Neste trabalho, a análise psicolinguística dos textos transcritos foi realizada com o dicionário LIWC 2007 adaptado ao português brasileiro [Balage Filho et al. 2013], o qual contempla 64 categorias distribuídas entre dimensões psicológicas, linguísticas e sociais. Esse tipo de análise permite identificar, por exemplo, se um influenciador utiliza mais frequentemente uma linguagem individualista ou coletiva, expressões emocionais ou racionais, termos associados à autoridade ou à empatia. Esses padrões auxiliam na interpretação do estilo comunicacional de cada influenciador analisado.

4. Análise de Dados

Esta seção apresenta os principais resultados obtidos por meio da aplicação de diferentes técnicas de análise textual sobre os conteúdos publicados por influenciadores digitais católicos no YouTube. São descritas as análises de entidades nomeadas, polaridade, tópicos e aspectos psicolinguísticos com o LIWC.

4.1. Análise de Entidades Nomeadas e de Discurso

Como mencionado anteriormente, uma entidade nomeada é uma expressão que pode ser extraída de um texto e é essencial para compreender, concluir ou ignorar um determinado contexto ou ação. O reconhecimento de entidades nomeadas tem por finalidade identificar entidades presentes em um texto, bem como classificá-las dentro de um conjunto de categorias predefinidas. A Tabela 3 ilustra as dez entidades nomeadas mais frequentes nos discursos de cada um dos influenciadores estudados.

A partir das entidades nomeadas mais frequentes, foram realizadas buscas com a ferramenta *ElasticSearch*⁴, a fim de localizar trechos com potencial de representar os discursos mais relevantes ou recorrentes. A Tabela 4 apresenta exemplos desses trechos.

⁴<https://www.elastic.co/pt/elasticsearch>

Tabela 3. Principais Entidades Nomeadas

| Influenciador | Entidades Nomeadas |
|-------------------|--|
| Paulo Ricardo | Pai do Filho e do Espírito Santo, Deus, Senhor, Jesus, Evangelho, Igreja, An-tigo Testamento, Satanás, Dez Mandamentos e Quaresma |
| Bernardo Kuster | Deus, Cristo, Brasil , Jesus, Bíblia, Lutero, Papa Francisco , Invencível, Sa-tanás e Bom Combate |
| Julio Lancellotti | Padre Júlio, Jesus, Senhor, Deus, Brasil, Lucas, Mateus, Papa Francisco , Evangelho e Nossa Senhora Aparecida |
| Patrick Fernandes | Senhor, Deus, Jesus Cristo, Pedro , Cordeiro de Deus, Osana, Maria, Paróquia e Evangelho |

Analisando a Tabela 4, pode-se observar que o influenciador Bernardo Kuster recorre frequentemente à polêmica e ao confronto para gerar engajamento, enquanto Paulo Ricardo apresenta discursos de formação, com ênfase em doutrina e dogmas. Patrick Fernandes foca em mensagens devocionais e de auto-referência, e os conteúdos de Júlio Lancellotti, no período analisado, foram majoritariamente compostos por manifestações de apoio devido à repercussão de denúncias contra ele.

A Figura 1 ilustra a distribuição da polaridade de trechos de 1.000 *tokens* das transcrições dos vídeos analisados. Observe que o influenciador Patrick Fernandes tende a ter mensagens mais positivas, já que seu conteúdo é mais devocional. Padre Ricardo apresenta um conteúdo mais neutro, uma vez que os vídeos eram majoritariamente formados por homilias e divulgação de cursos. As mensagens mais positivas ou altamente negativas, em geral, foram publicadas pelos influenciadores Bernardo Kuster e Júlio Lancellotti.

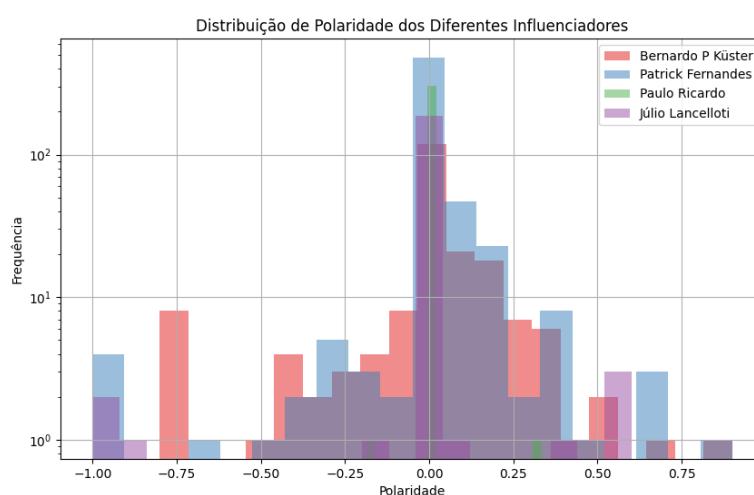


Figura 1. Distribuição da Polaridade das Transcrições dos Vídeos Analisados

Tabela 4. Principais Trechos dos Conjuntos de Dados Analisados

| Influenciador | Texto | Categoria |
|----------------------|---|------------------|
| Bernardo K. | ... as encíclicas que francisco já fez até agora que são altamente criticáveis né eu sou livre para fazer isso ... | Rompimento |
| Bernardo K. | ... por exemplo eu denunciei sozinho a articulação entre a cúpula da cnbb e a cúpula do pt ... campanha da fraternidade cheia de escândalos e coisas podemos dizer até anticatólicas ... | Rompimento |
| Bernardo K. | ... falsos profetas no mesmo modo e não tem entre vós falsos doutores como lutero que introduzirão seitas ... | Intolerância |
| Bernardo K. | ... o governo bolsonaro que aumentou a quantidade de armas reduziu o número de homicídios | Ideologia |
| Bernardo K. | ... esse padre ali o padre pedro paulo está sacrificando a verdade no altar da tolerância ... esse padre aí visita terreiros de umbanda... dom orlando se você não se converter tu vai pro inferno ... o padre júlio lancelotti que é um comunista nato | Conflito |
| Paulo R. | ... a razão de ser desse curso meus queridos é para vocês pararem de matar a vida eterna e alimentá-la essa é a razão de ser do curso ... | Curso |
| Paulo R. | ... esse curso das navegações portuguesas não é um curso de ideologia portuguesa seja ela de direita de esquerda uma exaltação da nação portuguesa ... | Curso |
| Paulo R. | ... o inferno existe gente e nós estamos fazendo esse curso aqui exatamente por quê por duas coisas uma coisa positiva e uma coisa negativa ... | Curso |
| Paulo R. | ... a igreja sempre considerou que judas estava no inferno não tem como ter a esperança de que não haja ninguém no inferno é impossível o inferno está cheio de bilhões de demônios não tem como estar vazio ... | Polêmica |
| Patrick F. | é tão bonito quando as pessoas chegam para a gente e diz assim nossa como que você transmite uma luz bonita como que você transmite uma paz como que você transmite uma alegria ... meu deus como que as pessoas notam quando a gente tem algo diferente a oferecer para elas não é verdade ... | Auto-referência |
| Patrick F. | ... coração se anima eu acho e isso porque a gente está tendo a vigília na nossa paróquia em mais dois lugares ... | Devoção |
| Júlio L. | ... a coleta nacional para a campanha da fraternidade será feita no próximo domingo e o objetivo desta coleta é apoiar projetos nas paróquias nas dioceses espalhados por todo o brasil que tenham a marca da solidariedade e da amizade social | Campanha |
| Júlio L. | ... não nos enganemos essa campanha odiosa contra o padre julio faz parte de uma articulação nacional e internacional contra a cnbb e contra o papa francisco | Apoio |
| Júlio L. | ... total apoio à luta do padre júlio que defende os excluídos os de baixo que defende os pobres ... | Apoio |

4.2. Modelagem de Tópicos

Os métodos utilizados para modelagem de tópicos foram o *Latent Dirichlet Allocation* (LDA) e o *Pseudo-document-based Topic Model* (PTM) [Zuo et al. 2016]. A avaliação considerou as métricas de coerência (C_V e NPMI) [Campagnolo et al. 2022] e diversidade (IRBO e Diversidade). A implementação do IRBO foi baseada no código descrito em [Terragni et al. 2021]. Tanto para o LDA quanto para o PTM, o número de tópicos (k) para cada conjunto de dados foi definido utilizando-se a maximização da coerência C_V, variando-se k de 2 a 30. A Tabela 5 ilustra a quantidade de tópicos identificada para cada combinação entre os métodos de modelagem de tópicos explorados e os influenciadores estudados. Já a Tabela 6 exibe os valores obtidos para as métricas de coerência e diversidade nos diferentes cenários investigados.

Tabela 5. Quantidade de Tópicos

| | Quantidade de Tópicos | |
|-------------------|-----------------------|-----|
| | LDA | PTM |
| Júlio Lancellotti | 4 | 10 |
| Patrick Fernandes | 4 | 10 |
| Paulo Ricardo | 8 | 10 |
| Bernardo Kuster | 10 | 10 |

Tabela 6. Métricas para Avaliação da Qualidade dos Tópicos

| Influenciador | LDA | | | | PTM | | | |
|---------------|--------|-------|-------|-------------|--------|-------|-------|-------------|
| | NPMI | C_V | IRBO | Diversidade | NPMI | C_V | IRBO | Diversidade |
| Paulo | -0.067 | 0.590 | 0.687 | 0.700 | -6.354 | 0.599 | 0.788 | 0.488 |
| Bernardo | -0.096 | 0.470 | 0.648 | 0.350 | -0.157 | 0.676 | 0.081 | 0.891 |
| Patrick | -0.064 | 0.383 | 0.426 | 0.650 | -0.073 | 0.558 | 0.049 | 0.755 |
| Julio | -0.239 | 0.446 | 0.715 | 0.700 | -0.268 | 0.723 | 0.118 | 0.919 |

A Tabela 7 ilustra os principais tópicos identificados utilizando-se o algoritmo LDA. Observe que os Tópicos 1 e 3 do influenciador Paulo Ricardo se concentram nos cursos por ele ofertados, enquanto os Tópicos 2 e 4 se referem às homilias dominicais. No caso de Bernardo Kuster, o Tópico 1 associa-se a discursos de castigo, o Tópico 2 remete à Páscoa, e os Tópicos 3 e 4 abordam temas ligados à fé na Igreja. Para Júlio Lancellotti, os Tópicos 1, 3 e 4 agrupam manifestações de apoio, e o Tópico 2 evoca o domingo de Páscoa. Já Patrick Fernandes apresenta tópicos mais homogêneos, com ênfase devocional.

A Tabela 8 exibe os principais tópicos identificados utilizando-se o algoritmo PTM. Observe que os Tópicos 1 e 2 de Paulo Ricardo correspondem às missas dominicais, o Tópico 3 trata de cursos e o Tópico 4 aborda temas da modernidade, como revolução francesa e aborto. Em Bernardo Kuster, o Tópico 1 aponta críticas a padres divergentes, o Tópico 2 refere-se a um curso, o Tópico 3 aborda o celibato e o marxismo, e o Tópico 4 reflete intolerância religiosa. Os Tópicos 1 e 2 de Júlio Lancellotti concentram mensagens de apoio, o Tópico 3 refere-se a missas, e o Tópico 4 foca em sua paróquia. Em Patrick Fernandes, os Tópicos 1, 2 e 4 são devocionais, e o Tópico 3 destaca elementos auto-referenciais.

Tabela 7. Tópicos Gerados pelo LDA

| Tópico | Paulo Ricardo | Bernardo Kuster | Júlio Lancellotti | Patrick Fernandes |
|----------|--|---|--|---|
| Tópico 1 | curso, mandamentos, pecado , graça, navegações , aula , peca | inferno , morte , jesus , mortos, rei, és, poderoso, poder, glória | padre, júlio, lance-lotti, senhor, anos, apoio , paulo, pessoas | senhor , deus , jesus , gente, vida, cristo, fazendo, amor |
| Tópico 2 | feliz, domingo , queridos, irmãos, alegria, respeito, evangelho | páscoa , primavera, data , terra, domingo , calendário | padre, júlio, trabalho, domingo , dia, gente, falar , jesus , senhor, apoio | deus , senhor , gente, parte, jesus , vida, fazendo, cristo, santo |
| Tópico 3 | vez, importante, deus, mandamentos, pastor , fé, história , vendo | deus, fé , igreja , gente, lá, agora, tranquilo, pessoas, mundo, fazer | júlio, padre, nome, luta, jesus , olá , anos, senhor, lance-lotti, querido | senhor , deus , lá, luz, terra, santo, vida, amor, igreja, espírito |
| Tópico 4 | evangelho , jesus , hoje , nome, irmãos, queridos, filho, pai | deus, igreja , pode, dignidade, gente, fé , pessoas, ter, agora | padre, júlio, lance-lotti, paulo, nome, trabalho , hoje, rua , miguel, olá | parte, fazemos, senhor , deus , vida, jesus , sebastião, santo, cristo |

Tabela 8. Tópicos Gerados pelo PTM

| Tópico | Paulo Ricardo | Bernardo Kuster | Júlio Lancellotti | Patrick Fernandes |
|----------|---|---|--|---|
| Tópico 1 | grande, alegria, irmãos, dia , queridos, irmãs, igreja, dias, celebramos | deus, jesus , <i>padre</i> , <i>falando</i> , <i>quer</i> , dizer, negócio, vezes, ensino | alguma, pessoas, cuidado, penitenciária, degradante , acusar , preconceitos | senhor, adorar , santo , verbo, evangelho, rogai, eleição, oração , osana, intercessão |
| Tópico 2 | primeira , tema, tempestade, celebramos , ondas, barca, pequena, assustada | invencível, formação , aula , aulas, defesa, defender, religiões, você, ataques, valor, preparar | bom , trabalho , anos, conheço, exemplo , vida, pão, pastor, cristão, testemunha, comunidades | faz , agora , dentro, coração , tempo, cruz, dor, mal, pecado, evangelho, inteiro, presença, recebe, experiência |
| Tópico 3 | portugal , história, mundo, concupiscência, fé, refletir , poema, infante , enrique | igreja, padre, casados, castidade, celibato , papa, propriedade, mulher, marxismo | professora, igreja, senhor, missa , horas, domingo , poder, unesp, vereadores, título | diante, alguém, pessoa , tão , coisas, dizer, certeza, sinto, confiança, mim |
| Tópico 4 | aborto , francesa, frança , revolução , constituição, resultando, decapitação, irá, reverter, dor, matando | paganismo , religiões , sociedade, conhecimento, teoria, antiga, doutrina, século, magia, agenda, moral | nome, miguel , paróquia , júlio, comunidade , arcanjo , participo, conheço, pertença, moca, capela | fazendo, gente, parte, projeto, mostrei-me, conversão , caminhos, realidade, comunhão , eleição |

4.3. Análise Linguística

Para examinar os aspectos emocionais, cognitivos e estruturais dos textos em cada conjunto de dados, utilizou-se a ferramenta LIWC. O LIWC emprega um dicionário de palavras desenvolvido para o português do Brasil (LIWC 2007 Português) [Balage Filho et al. 2013], que inclui 127.149 palavras, cada uma associada a uma ou mais categorias. Essas categorias representam diversas perspectivas linguísticas e psicológicas, totalizando 64 no total. A Tabela 9 exibe a quantidade de categorias com diferenças estatísticas significativas entre os pares de conjuntos de dados analisados. O elevado número de categorias com diferenças estatísticas confirma que os estilos comunicacionais dos influenciadores analisados são linguisticamente distintos.

Tabela 9. Quantidade de Categorias do LIWC com Diferenças Significativas

| | J. Lancellotti | B. Kuster | P. Fernandes | Paulo Ricardo |
|-----------------------|-----------------------|------------------|---------------------|----------------------|
| J. Lancellotti | - | 65.6% (42) | 54.7% (35) | 60.9% (39) |
| B. Kuster | 65.6% (42) | - | 90.6% (58) | 67.2% (43) |
| P. Fernandes | 54.7% (35) | 90.6% (58) | - | 70.3% (45) |
| Paulo Ricardo | 60.9% (39) | 67.2% (43) | 70.3% (45) | - |

5. Conclusões e Trabalhos Futuros

Este trabalho investigou o conteúdo produzido por quatro influenciadores digitais católicos de destaque: Bernardo Küster, Paulo Ricardo, Patrick Fernandes e Júlio Lancellotti. Para tal, foram construídos quatro conjuntos de dados — um para cada influenciador — com base nos vídeos publicados na plataforma YouTube. Em seguida, os algoritmos LDA e PTM foram utilizados para identificar os tópicos latentes em cada conjunto analisado. Posteriormente, utilizou-se a ferramenta LIWC para examinar os componentes emocionais, cognitivos e estruturais dos textos. Os resultados revelaram padrões discursivos, estratégias de engajamento e posicionamentos político-religiosos distintos, evidenciando narrativas e estilos comunicativos completamente diferentes, e por vezes contraditórios.

Uma limitação importante desta pesquisa está relacionada ao período da coleta, que abrangeu a Quaresma — um dos momentos litúrgicos mais relevantes do calendário católico. Esse recorte temporal pode ter influenciado o tom ou a natureza dos discursos analisados, tornando-os eventualmente mais moderados, introspectivos ou voltados à reflexão espiritual. Investigações futuras podem considerar diferentes períodos do ano litúrgico para avaliar possíveis variações sazonais no conteúdo e no estilo comunicacional desses influenciadores.

Como sugestões para trabalhos futuros, propõe-se: (i) a análise longitudinal dos canais, de modo a capturar transformações discursivas ao longo do tempo; (ii) a inclusão de outros influenciadores e plataformas digitais, ampliando a representatividade e diversidade dos discursos religiosos analisados; (iii) a combinação de análise textual com indicadores de engajamento (curtidas, comentários, compartilhamentos) para correlacionar estratégias discursivas e impacto junto ao público; e (iv) a utilização de métodos de análise visual e sonora para capturar dimensões multimodais da comunicação religiosa digital.

Referências

- Balage Filho, P., Pardo, T. A. S., and Aluísio, S. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(null):993–1022.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009*, volume Normalized, pages 31–40, Tübingen.
- Brum, P. V., Teixeira, M. C., Miranda, R., Vimieiro, R., Jr, W. M., and Pappa, G. L. (2020). A characterization of portuguese tweets regarding the covid-19 pandemic. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 177–184, Porto Alegre, RS, Brasil. SBC.
- Campagnolo, J. M., Duarte, D., and Dal Bianco, G. (2022). Topic coherence metrics: How sensitive are they? *Journal of Information and Data Management*, 13(4).
- Campos, B. M. (2022). *Fundamentalismos religiosos em perspectiva*. Pluralidades.
- Carvalho, F., Rodrigues, R., Santos, G., Cruz, P., Ferrari, L., and Guedes, G. (2019). Avaliação da versão em português do liwc lexicon 2015 com análise de sentimentos em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 24–34, Porto Alegre, RS, Brasil. SBC.
- Corsi, M. and Urbano, J. (2024). The treatment of ties in rank-biased overlap. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '24*, page 251–260, New York, NY, USA. Association for Computing Machinery.
- Dias, Z. M. (2012). *Os Varios Rostos do Fundamentalismo*. Pluralidades.
- Ebeling, R., Nobre, J., and Becker, K. (2023). A multi-dimensional framework to analyze group behavior based on political polarization. *Expert Syst. Appl.*, 233(C).
- Ebeling, R., Sáenz, C., Nobre, J., and Becker, K. (2020). Quarenteners vs. cloroquiners: a framework to analyze the effect of political polarization on social distance stances. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 89–96, Porto Alegre, RS, Brasil. SBC.
- Guimarães, S., Silva, M., Caetano, J., Araújo, M., Santos, J., Reis, J., Silva, A., Benevenuto, F., and Almeida, J. (2022). Análise de propagandas eleitorais antecipadas no twitter. In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*, pages 85–96, Porto Alegre, RS, Brasil. SBC.
- Hu, Z., Hou, W., and Liu, X. (2024). Deep learning for named entity recognition: a survey. *Neural Comput. Appl.*, 36(16):8995–9022.
- Jacinto, L. H., da Silva, T., Parmezan, A., and Batista, G. (2020). Brazilian presidential elections: Analysing voting patterns in time and space using a simple data science pipeline. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 217–224, Porto Alegre, RS, Brasil. SBC.

- Kim, M. and Kim, D. (2022). A suggestion on the lda-based topic modeling technique based on elasticsearch for indexing academic research results. *Applied Sciences*, 12(6).
- McConahay, M. J. (2023). *Playing God: American Catholic Bishops and the Far Right*. Melville House Publishing.
- Medeiros, F. F., Silva, A. A. d., Souza, A. R. d., Sbardelotto, M., and Gomes, V. B. (2024). *Influenciadores Digitais Católicos: Efeitos e Perspectivas*. Paulus.
- Melo, P., Couto, J. M. M., Kansaon, D., Mafra, V., Reis, J. C. S., and Benevenuto, F. (2024). A sticker is worth a thousand words: Characterizing the use of stickers in whatsapp political groups in brazil. *CoRR*, abs/2406.08429.
- Oro, I. P. (1996). *O outro é o demônio: uma análise sociológica do fundamentalismo*. Paulus.
- Passos, J. D. (2021). *No lugar de Deus: Ensaio (neo)teocráticos*. Editora Paulinas.
- 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, WSDM '15*, page 399–408, New York, NY, USA. Association for Computing Machinery.
- Schmitt, X., Kubler, S., Robert, J., Papadakis, M., and LeTraon, Y. (2019). A replicable comparison study of ner software: Stanfordnlp, nltk, opennlp, spacy, gate. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 338–343.
- Silva, N., Silva, M., Pereira, F., Tarrega, J., Beinotti, J., Fonseca, M., Andrade, F., and Carvalho, A. (2021). Evaluating topic models in portuguese political comments about bills from brazil's chamber of deputies. In *Anais da X Brazilian Conference on Intelligent Systems*, Porto Alegre, RS, Brasil. SBC.
- Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.
- Terragni, S., Fersini, E., and Messina, E. (2021). Word embedding-based topic similarity measures. In *International Conference on Applications of Natural Language to Information Systems*, pages 33–45. Springer.
- Terragni, S., Ganguly, D., and Crestani, F. (2020). Octis: Comparing and optimizing topic models is simple! In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, pages 2201–2204. ACM.
- Villaça, A. C. (2006). *O Pensamento Católico No Brasil*. Civilização Brasileira.
- Zhang, W., Li, X., Deng, Y., Bing, L., and Lam, W. (2023). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Trans. on Knowl. and Data Eng.*, 35(11):11019–11038.
- Zuo, Y., Wu, J., Zhang, H., Lin, H., Wang, F., Xu, K., and Xiong, H. (2016). Topic modeling of short texts: A pseudo-document view. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, page 2105–2114, New York, NY, USA. Association for Computing Machinery.