

Análise de Comportamento dos Usuários do Fórum Clube Ceticismo

Rafael Gontijo Sabino Neves, João Pedro Braga Ennes

Departamento de Ciências da Computação – Universidade Federal de Minas Gerais
(UFMG)

Caixa Postal 702 – 31270-901 – Belo Horizonte – MG – Brazil

{rafaelneves130, joaoennes}@ufmg.br

Abstract. *This meta-paper explores the dynamics of online forums, focusing on the Clube Ceticismo forum. With 281 users, 89 active ones, and 702 topics, the study analyzes discussions ranging from critical social issues to scientifically unsubstantiated speculations. The research aims to comprehend patterns of behavior and expressed sentiments by delving into the interaction network and employing sentiment analysis to identify emotional nuances.*

Resumo. *Este meta-artigo explora o dinamismo dos fóruns online, focalizando o fórum do Clube Ceticismo. Com 281 usuários, 89 ativos e 702 tópicos, o estudo analisa discussões que vão desde questões sociais cruciais até especulações sem embasamento científico. A pesquisa busca entender os padrões de comportamento e sentimentos expressos, explorando a rede de interações e utilizando análise de sentimentos para identificar nuances emocionais.*

1. Introdução

No que se refere ao dinamismo das mídias sociais, os fóruns emergem como pólos de interação, agregando comunidades diversas e propiciando debates que muitas vezes fogem do convencional e são mais aprofundadas do que em outras plataformas, como o twitter. O fórum do Clube Ceticismo, criado em 2020, representa um microcosmo fascinante neste contexto, oferecendo um espaço para diálogos que transcendem as fronteiras do senso comum e adentram o território das teorias não respaldadas pela ciência.

Com uma comunidade composta por 281 usuários, sendo 89 ativamente engajados nos fóruns, e um repertório de 702 tópicos variados abertos ao público, esse fórum traz à tona discussões que abraçam tanto questões urgentes e relevantes para a sociedade contemporânea, como racismo, desigualdade e homofobia, quanto temas conjecturais, como a alegada associação entre a universidade pública e a corrupção, ou a especulação sobre a origem terrestre de fenômenos UFOs. A figura 1 exemplifica assuntos presentes na plataforma.



Figura 1. Exemplos de discussões

A estrutura do fórum categoriza os tópicos em diversas categorias sendo as mais frequentadas "Ciência, Tecnologia e Saúde", "Laicismo, Política e Economia", "Religiões, Crenças e Mitos" e "História, Sociedade, Comportamento e Filosofia", como mostra a figura 2. Essas categorias fornecem um arcabouço de diversos subgrupos de discussões, sendo muitos deles assuntos que geram bastante discussões e tendem a causar polarização na sociedade.

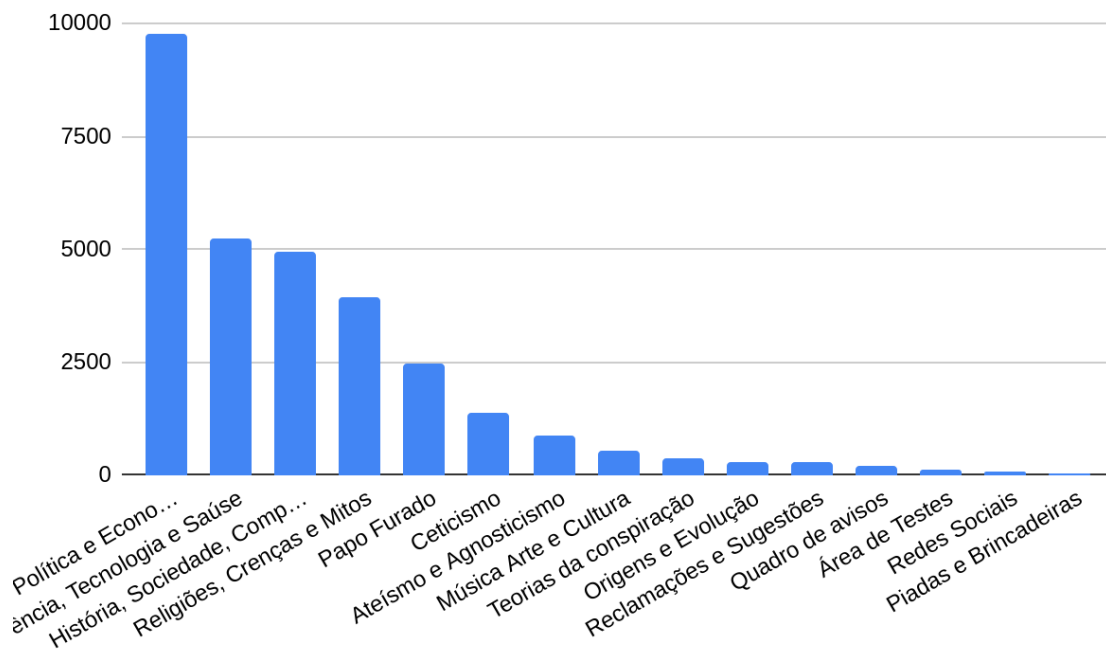


Figura 2. posts por assuntos

Este estudo busca investigar não somente nos tópicos discutidos e os padrões de comportamento dos usuários, mas também na análise minuciosa dos sentimentos expressos em cada contexto temático. Com o intuito de mapear padrões, tendências e interações, esta investigação visa analisar as dinâmicas sociais, os contornos dos pensamentos e a interconexão entre os diversos temas, contribuindo para uma compreensão aprofundada das informações presentes no Clube Ceticismo no panorama das mídias sociais e suas implicações na esfera social e cultural contemporânea.

2. Trabalhos Relacionados

Diversos estudos e pesquisas têm explorado o campo da análise de sentimentos e dinâmica de interações em comunidades online, proporcionando insights valiosos sobre diferentes contextos e grupos de usuários. Dois exemplos notáveis são os artigos "Mining Social Media Data for Understanding Students' Learning Experiences" e "Sentiment Analysis and Dynamics in an Online Cancer Survivor Community".

O estudo "Mining Social Media Data for Understanding Students' Learning Experiences" foca na compreensão das experiências de aprendizado dos estudantes por meio da análise de dados de mídias sociais. Este trabalho examina interações no Twitter, utilizando técnicas de análise de sentimentos para mapear e compreender as emoções, desafios e pontos-chave associados ao aprendizado dos alunos. Ao identificar e avaliar sentimentos expressos em discussões educacionais online, o estudo oferece insights valiosos sobre a eficácia das estratégias de ensino, áreas de interesse dos alunos e possíveis dificuldades enfrentadas durante o aprendizado.

Por outro lado, o artigo "Sentiment Analysis and Dynamics in an Online Cancer Survivor Community" direciona seu foco para uma comunidade específica: sobreviventes de câncer. Este estudo examina fóruns online nos quais pacientes com câncer compartilham informações, experiências e apoio mútuo. Ao utilizar técnicas de análise de sentimentos, o trabalho explora as emoções expressas pelos participantes, identificando tendências emocionais, necessidades não atendidas e áreas de preocupação. Isso proporciona uma visão mais profunda das experiências emocionais dos sobreviventes de câncer na comunidade online, contribuindo para a compreensão das necessidades de suporte e interações sociais em contextos sensíveis.

Ambos os estudos exemplificam como a análise de sentimentos em comunidades online pode ser aplicada a diferentes contextos e grupos de usuários. Esses trabalhos oferecem um modelo valioso para a aplicação de análises similares em outros domínios, evidenciando o potencial dessa abordagem para melhorar a compreensão e a eficácia das interações sociais online em diversas comunidades.

3. Conjunto de Dados

3.1. Coleta de Dados

A obtenção de dados em ambientes virtuais, especialmente em fóruns e plataformas menos difundidas, demanda abordagens adaptativas e personalizadas. No caso do fórum do Clube Ceticismo, a coleta de informações foi realizada por meio de um método meticuloso, empregando um crawler desenvolvido utilizando Python e Selenium.

O processo de coleta se estendeu por todas as seções do fórum, visitando cada tópico e capturando um total de 30.529 linhas de registro. Cada uma dessas linhas corresponde a uma postagem, ou seja, o conteúdo do post original ou as respostas subsequentes. As informações essenciais foram extraídas, incluindo o título da postagem, categoria, URL associada, autor, número de mensagens escritas pelo autor, data e o conteúdo do post.

Essa abordagem minuciosa permitiu a criação de um dataset abrangente disponibilizado em CSV, oferecendo uma visão detalhada e abrangente das interações e discussões dentro do fórum. Esse conjunto de dados enriquecido, contendo informações tão variadas e cruciais para compreender os padrões de comportamento, temas mais discutidos e a dinâmica dos usuários, está disponível publicamente para análise e exploração.

Para facilitar o acesso e a utilização deste valioso recurso, o crawler desenvolvido e o dataset resultante foram disponibilizados de forma aberta e gratuita. Ambos estão acessíveis no repositório público do GitHub, no endereço: <https://github.com/EnnesJp/python-crawler>.

É importante ressaltar que a ausência de barreiras significativas para a coleta de dados é um aspecto favorável neste contexto. Isso permitiu a realização do processo de coleta sem grandes entraves, viabilizando a obtenção de um corpus robusto e

representativo do fórum, apesar de sua relativa desconhecimento e menor relevância no cenário nacional. Esta disponibilidade de dados pode potencializar análises, estudos comparativos e contribuições para pesquisas acadêmicas e investigações em áreas diversas.

A transparência e a acessibilidade dos dados coletados proporcionam uma oportunidade única para explorar e compreender a dinâmica e os temas predominantes no fórum do Clube Ceticismo, ampliando assim o entendimento sobre as interações sociais e intelectuais neste contexto específico.

3.2. Grafo Gerado

A construção de um grafo que representa a rede de interações fórum do Clube Ceticismo é uma abordagem interessante para visualizar a dinâmica das discussões online. Neste contexto, cada nó no grafo representa um tópico de discussão, enquanto os links entre os nós são estabelecidos quando um mesmo usuário participa de diferentes tópicos.

Para criar esse grafo, utilizamos os dados coletados sobre as interações dos usuários em diversos tópicos. Cada usuário torna-se um ponto central de conexão, e os tópicos de discussão aos quais eles contribuem são representados como nós interligados por arestas. Quando um usuário participa de mais de um tópico, uma aresta é formada entre os nós correspondentes, indicando uma relação entre esses temas. Ao final, obtivemos um grafo com 702 nodos e 165280 arestas.

Essa representação gráfica fornece uma visão holística da interconectividade dos tópicos na rede social. Padrões emergentes, comunidades de interesse e influenciadores podem ser identificados com mais facilidade. Além disso, a análise da densidade e distribuição dos links pode revelar a coesão da comunidade e a intensidade das interações.

A construção desse grafo não apenas visualiza as interações da rede social, mas também oferece insights valiosos sobre como os usuários se envolvem em diversas discussões. Essa abordagem, baseada em grafos, é uma ferramenta poderosa para compreender a dinâmica complexa das interações sociais online.

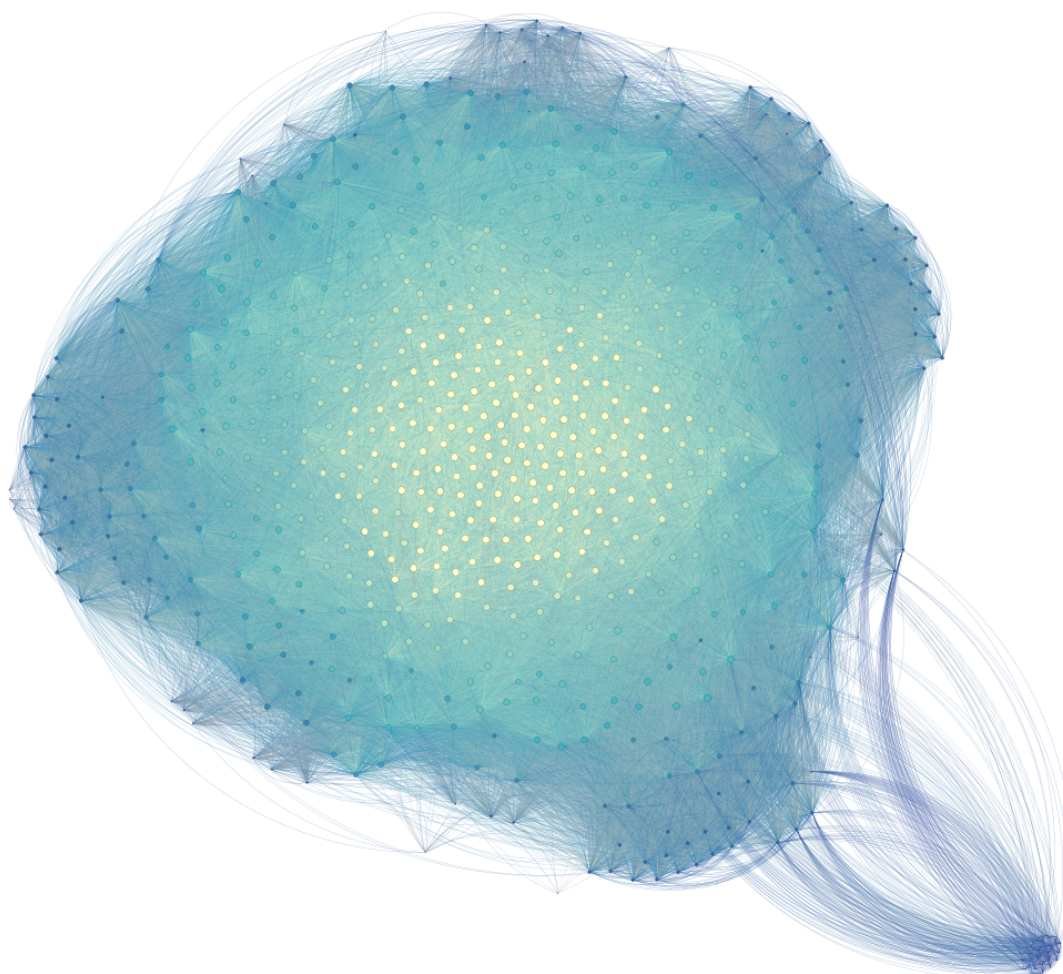


Figura 3. grafo da rede

A alta densidade do grafo é gerada em razão do baixo número de usuários na rede e do elevado número de interações por tópico. Isso revela aspectos interessantes sobre a intensidade e a coesão das interações nesse ambiente virtual.

Quando a densidade do grafo é alta, significa que uma proporção significativa dos usuários está interagindo em vários tópicos de discussão. Isso sugere uma participação ativa e diversificada da comunidade, onde os membros não se restringem a um único assunto, mas engajam-se em diversas conversas. A multiplicidade de interações por tópico contribui para a formação de um emaranhado robusto de conexões entre os nós, indicando uma teia complexa de relacionamentos e interesses compartilhados.

A baixa quantidade de usuários na rede pode resultar em uma maior probabilidade de interações entre eles, uma vez que a comunidade é mais restrita. Isso

intensifica a densidade do grafo, pois os usuários tendem a se envolver em várias discussões, promovendo uma interligação mais próxima entre os diferentes tópicos. A coesão gerada por essas interações frequentes contribui para a formação de clusters ou comunidades mais coesas, onde membros compartilham interesses comuns.

Essa alta densidade pode indicar um ambiente propício para a troca intensiva de ideias, favorecendo a formação de uma comunidade mais unida e colaborativa. No entanto, é importante considerar a diversidade e a qualidade das interações, além da quantidade. Uma análise mais aprofundada pode revelar a natureza específica dessas interações, permitindo uma compreensão mais completa do papel desempenhado pela densidade do grafo na dinâmica da rede social.

3.3. Coleta de dados de sentimentos

A análise de sentimentos é um pilar crucial na compreensão das interações humanas, especialmente em ambientes virtuais como o fórum do Clube Ceticismo. Para adentrar nesse aspecto subjetivo das postagens, foi desenvolvido um script em Python disponível em <https://github.com/EnnesJp/python-crawler/blob/master/getToxicit.py>, que percorreu todas as mensagens do dataset, extraindo o texto, removendo formatações HTML e citações para melhorar a precisão da análise.

Utilizando um modelo de inteligência artificial baseado no BERT, especificamente o modelo `distilbert-base-multilingual-cased-sentiments-student` disponível no Hugging Face's transformers, foi possível classificar as mensagens em categorias de sentimento: positivo, negativo ou neutro.

É válido mencionar algumas limitações enfrentadas durante o processo, como o desafio imposto pelo tamanho dos textos inputados, dificultando a classificação de postagens mais extensas. Contudo, apesar dessas limitações, o modelo permitiu uma análise panorâmica dos sentimentos expressos pelos autores em cada mensagem, fornecendo insights valiosos sobre o tom geral das interações.

Ao término desse processo, foram adicionados quatro novos campos ao dataset. Um deles informa o teor da mensagem, classificando-a como positiva, negativa ou neutra. Os outros três campos complementares detalham a probabilidade de determinada postagem ser positiva, negativa ou neutra, fornecendo uma visão mais granular dos resultados da análise de sentimentos.

Esta etapa de análise ampliou significativamente a compreensão do conteúdo emocional subjacente às interações no fórum do Clube Ceticismo. Ao associar os sentimentos expressos pelos autores às discussões e aos tópicos abordados, esse processo proporciona uma camada adicional de entendimento sobre a dinâmica e o tom das conversas no ambiente digital, enriquecendo ainda mais o conjunto de dados disponíveis para análises mais aprofundadas e contextualizadas.

4. Resultados

4.1. Análises dos dados coletados

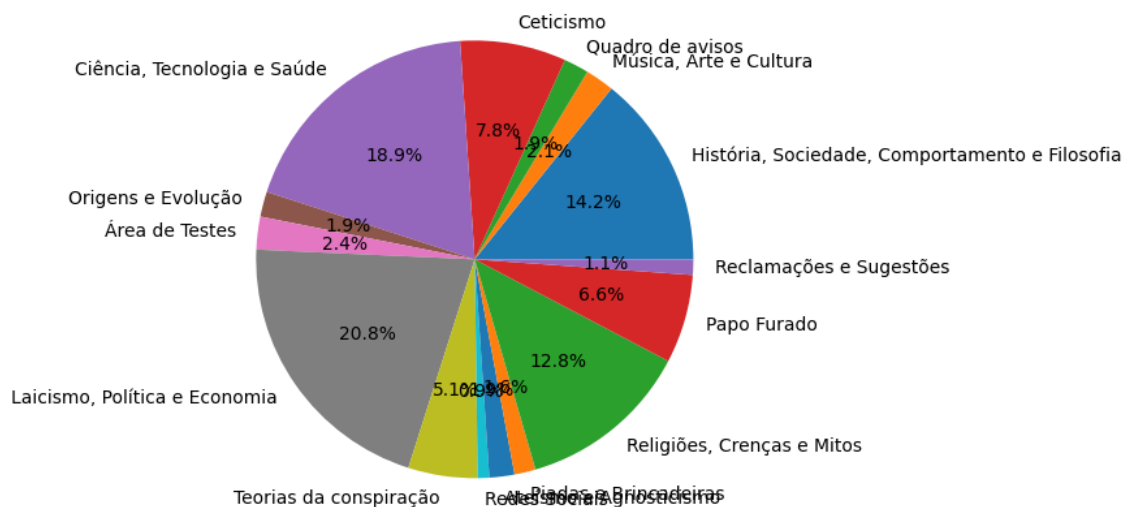


Figura 4. gráfico de interações por fórum

É possível notar uma concentração significativa de interações em tópicos sensíveis, como "Laicismo, Política e Economia", "Ciência, Tecnologia e Saúde" (especialmente relacionado à COVID-19), "História, Sociedade, Comportamento e Filosofia" e "Religiões, Crenças e Mitos", reflete a dinâmica e as prioridades da comunidade envolvida na rede social. Essa concentração pode ser influenciada por fatores externos, como eventos atuais, mudanças sociais e crises globais, que impactam diretamente os interesses e discussões dos usuários.

A alta concentração no tópico "Laicismo, Política e Economia" se deve em grande parte a debates relacionados à política, principalmente devido às últimas eleições brasileiras, onde houve uma polarização muito grande que gerou diversas discussões em ambos os lados. Essa concentração pode refletir um desejo de compreender e debater os aspectos fundamentais que moldam a sociedade, incluindo a relação entre Estado e religião. Além disso, esses espaços acabam se tornando um local propício para divulgação e propagação de "notícias" e "fatos" que não necessariamente refletem a realidade, com intuito de denegrir a imagem de um dos lados ou exaltar algum partido ou candidato específico.

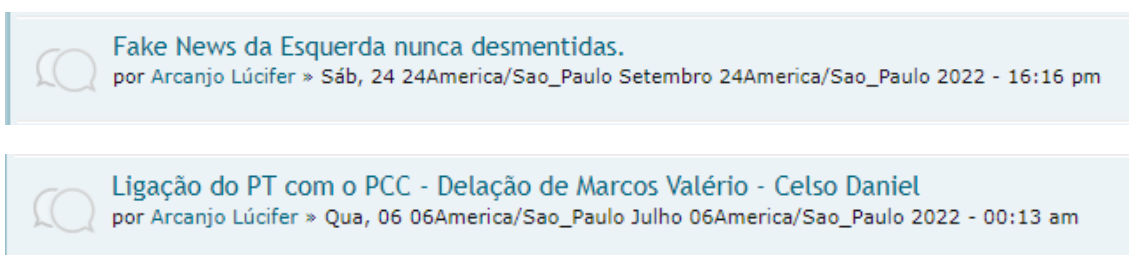


Figura 5. exemplos de tópicos de discussões políticas

A pandemia de COVID-19 tem desencadeado discussões intensas e preocupações globais, o que acabou aumentando o número de interações nos tópicos do fórum “Ciência, Tecnologia e Saúde”. A concentração neste tema reflete a sensibilidade do momento, com usuários buscando informações, compartilhando opiniões e discutindo as implicações da pandemia na ciência, tecnologia e saúde. Assim como na política, esses tópicos de discussão acabam se tornando ótimos ambientes para propagação de inverdades e discursos anti ciência, que foram amplamente divulgados nos últimos anos, pondo em dúvida a eficácia das vacinas ou até a própria seriedade com a qual o vírus deveria ser tratado.

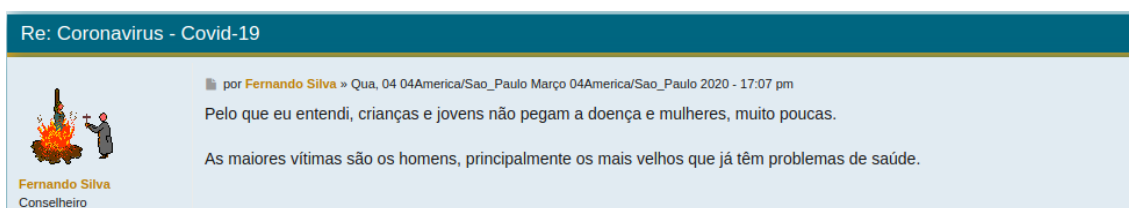


Figura 6. exemplo de comentário em discussões sobre COVID

O fórum de “História, Sociedade, Comportamento e Filosofia”, abrange uma variedade de temas amplos, mas é na parte de discussões sobre a sociedade, onde está a maior parte das interações. Dentre os temas abordados, encontramos diversos assuntos sensíveis onde moram muitos preconceitos e que através de portais como estes, podem ser discutidos de forma aberta e livre, dando espaço para disseminação e perpetuação de discursos de ódio e preconceitos presentes na sociedade.

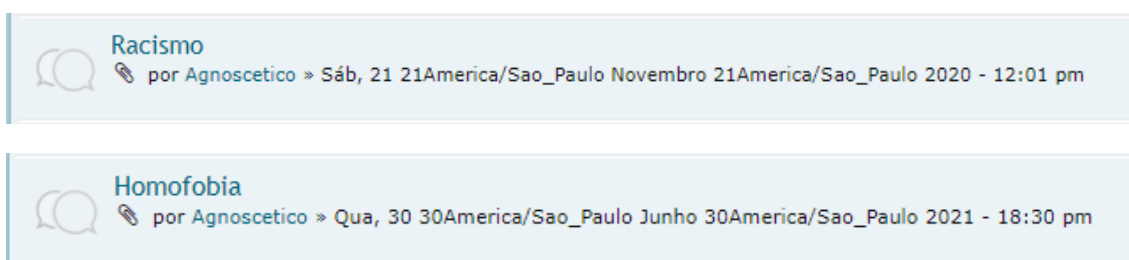


Figura 7. exemplo de tópicos relacionados a sociedade

Essa clara concentração notada nos tópicos citados, pode ser um reflexo das preocupações e interesses predominantes na comunidade, bem como da influência de eventos contemporâneos. No entanto, é crucial analisar como essa concentração afeta a diversidade de opiniões e a inclusividade da comunidade. Uma análise mais aprofundada desses tópicos pode revelar padrões comportamentais, áreas de consenso e potenciais desafios na promoção de uma troca de ideias aberta e equitativa.

4.2. Análises de sentimentos

A análise dos resultados consolidados ofereceu uma visão elucidativa sobre os padrões de sentimentos expressos em diferentes categorias dentro do fórum do Clube Ceticismo. Inicialmente, ao calcular a média de sentimentos em cada categoria, foram reveladas descobertas intrigantes.

Categoria	positivo	negativo	neutro
Área de Testes	71,7%	25,2%	3,1%
Ateísmo e Agnosticismo	21,9%	74,3%	3,9%
Ceticismo	34,0%	62,0%	4,0%
Ciência, Tecnologia e Saúde	28,3%	68,7%	3,0%
História, Sociedade, Comportamento e Filosofia	17,1%	80,4%	2,5%
Laicismo, Política e Economia	21,9%	75,7%	2,3%
Música Arte e Cultura	57,7%	40,4%	1,8%
Origens e Evolução	29,1%	64,9%	6,0%
Papo Furado	48,8%	47,3%	3,9%
Piadas e Brincadeiras	54,2%	42,4%	3,4%
Quadro de avisos	39,9%	55,9%	4,3%
Reclamações e Sugestões	32,2%	58,0%	9,8%
Religiões, Crenças e Mitos	22,6%	74,2%	3,3%
Redes Sociais	32,8%	62,5%	4,7%
Teorias da conspiração	28,0%	69,2%	2,8%

Figura 4. Análise de sentimentos

Excluindo a categoria de testes, os resultados destacaram que os assuntos com uma análise mais positiva foram "Música Arte e Cultura", "Piadas e Brincadeiras" e "Papo Furado". Esses temas parecem evocar uma atmosfera mais amigável e descontraída, onde os participantes tendem a ser mais gentis em suas interações. Eles representam áreas de discussão menos controversas e com menor propensão a gerar polarizações ou confrontos de ideias.

Por outro lado, chamou atenção que os assuntos com maior negatividade foram "História, Sociedade, Comportamento e Filosofia", "Laicismo, Política e Economia" e "Ateísmo e Agnosticismo". Esses temas mais complexos e propensos a polarizar opiniões parecem provocar uma resposta mais negativa por parte dos usuários. Isso sugere a presença de uma certa intolerância ou tensão em discussões que abordam questões sociais, políticas, filosóficas e religiosas.

Esses resultados não apenas oferecem uma visão abrangente do tom geral das conversas em cada categoria, mas também fornecem insights valiosos sobre como diferentes temas podem influenciar as emoções e atitudes dos participantes. A compreensão dessas nuances emocionais é fundamental para entender as dinâmicas

sociais e as interações humanas em ambientes online, fornecendo uma base sólida para reflexões mais profundas e estratégias para promover uma comunicação mais construtiva e empática.

5. Conclusão

O estudo do fórum do Clube Ceticismo oferece uma visão fascinante sobre a dinâmica das mídias sociais, especialmente no contexto dos fóruns online. Com 281 usuários, 89 deles ativamente envolvidos, e 702 tópicos variados, o fórum representa um microcosmo de interações que transcendem o convencional, explorando desde questões sociais urgentes até especulações mais conjecturais.

A análise do grafo da rede social revela uma alta densidade, resultado do baixo número de usuários e da intensa interação por tópico. Essa característica sugere uma comunidade ativa e diversificada, onde os membros participam de várias discussões. A coesão gerada por essas interações intensivas pode promover uma comunidade mais unida, mas é essencial avaliar a diversidade e qualidade dessas interações.

A concentração temática em "Laicismo, Política e Economia", "Ciência, Tecnologia e Saúde," "História, Sociedade, Comportamento e Filosofia," e "Religiões, Crenças e Mitos" revela a sensibilidade da comunidade para assuntos relevantes, impactados por eventos atuais como as eleições brasileiras e a pandemia de COVID-19. Essa concentração evidencia não apenas a diversidade de tópicos, mas também a propensão a debates polarizados e a disseminação de informações controversas.

A análise de sentimentos complementa a compreensão, destacando que tópicos menos controversos, como "Música Arte e Cultura," tendem a gerar interações mais positivas, enquanto discussões sobre "História, Sociedade, Comportamento e Filosofia," e "Laicismo, Política e Economia" evocam mais negatividade. Essa descoberta aponta para desafios em debates sobre temas complexos, indicando a presença de tensão e intolerância.

Em síntese, a pesquisa contribui para a compreensão aprofundada da dinâmica do Clube Ceticismo, fornecendo insights sobre padrões de comportamento, temas dominantes e o contexto emocional das interações. Essa compreensão é crucial para promover uma comunicação mais construtiva e empática em ambientes online, reconhecendo a importância de considerar não apenas o conteúdo, mas também o tom e a qualidade das interações.

6. Referencias

Qiu, Baojun, et al. "Get online support, feel better--sentiment analysis and dynamics in an online cancer survivor community." 2011 IEEE third international conference on

privacy, security, risk and trust and 2011 IEEE third international conference on social computing. IEEE, 2011.

Chen, Xin, Mihaela Vorvoreanu, and Krishna Madhavan. "Mining social media data for understanding students' learning experiences." *IEEE Transactions on learning technologies* 7.3 (2014): 246-259.

xyuan/distilbert-base-multilingual-cased-sentiments-student · Hugging Face. (2023, May 23). Huggingface.co.
<https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student>

Ennes, J. P. (2023, October 28). EnnesJp/python-crawler. GitHub.
<https://github.com/EnnesJp/python-crawler>

Clube Ceticismo - Índice (19 oct 2023). <https://clubeceticismo.com.br/>, [accessed on Dec 4].