
Análise do Posicionamento Político dos Brasileiros em relação à Guerra Irã x EUA utilizando dados do YouTube

Lavínia Barbosa Dantas de Souza



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Lavínia Barbosa Dantas de Souza

**Análise do Posicionamento Político dos
Brasileiros em relação à Guerra Irã x EUA
utilizando dados do YouTube**

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Área de concentração: Sistemas de Informação

Orientador: Profa. Dra. Fabíola Souza Fernandes Pereira

Monte Carmelo - MG

2026

*Aos meus pais, Edna e Efanazio, que sempre me deram asas para sonhar e suportaram
minha ausência ao longo dessa jornada.*

Agradecimentos

O desenvolvimento dessa monografia jamais seria possível sem ter, durante essa jornada, a ajuda de algumas pessoas. Dentre elas:

Minha orientadora, Profa. Dra. Fabíola Souza, por ter me instruído e me dado todo o suporte necessário no decorrer do trabalho;

Todos os professores que passaram não só por minha formação acadêmica, mas também por meu desenvolvimento enquanto pessoa. Dos quais, destaco: professoras Ana, Alessandra, Sara e Fernanda. Professores André, Claudiney, Diego e Victor;

Meus pais, que me deram todo suporte emocional e sempre fizeram de tudo para que esse sonho se tornasse realidade;

Por fim, meus colegas de faculdade, que se tornaram, além de amigos, parte da minha família e estiveram comigo nos melhores e nos piores dias. Em especial, Camilly, Mariana e Nicolly.

“[...]pois o próximo instante é desconhecido. O próximo instante é feito por mim? Ou se faz sozinho? Fazemo-lo juntos com a respiração. E com uma desenvoltura de toureiro na arena”
(Clarice Lispector)

Lista de ilustrações

Figura 1 – Imagem retirada do livro Network Science (BARABÁSI, 2016)	18
Figura 2 – Fluxograma das principais etapas desenvolvidas.	23
Figura 3 – Vídeos selecionados para a coleta de comentários. Classificação realizada manualmente em relação ao tipo, narrador principal, gênero e posicionamento político.	24
Figura 4 – Conceito da projeção monopartida de uma rede bipartida comentário-termo	27
Figura 5 – Gráfico com as proporções relacionadas ao posicionamento Guerra e Lado	30
Figura 6 – Quantidade de comentários de acordo o tipo de vídeo	30
Figura 7 – Quantidade média de palavras por comentário de acordo o tipo de vídeo	31
Figura 8 – Nuvens de palavras segmentadas em relação ao posicionamento na guerra.	31
Figura 9 – Rede de comentários classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.	32
Figura 10 – Rede de comentários classificados de acordo com o posicionamento em relação aos lados do conflito. Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou aos Estados Unidos.	33
Figura 11 – Rede de comentários extraídos de vídeos, classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.	34
Figura 12 – Rede de comentários extraídos de vídeos curtos (shorts), classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.	34
Figura 13 – Rede de comentários extraídos de vídeos, classificados de acordo com o posicionamento em relação aos lados do conflito. Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou aos Estados Unidos.	35

Figura 14 – Rede de comentários extraídos de vídeos curtos (shorts), classifica-	
dos de acordo com o posicionamento em relação aos lados do conflito.	
Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou	
aos Estados Unidos.	35
Figura 15 – Comunidades de Girvan-Newman geradas	37

Lista de tabelas

Tabela 1 – Tabela Comparativa de Trabalhos Relacionados. T1: (RICCI, 2020); T2: (OLIVEIRA; OLIVEIRA, 2023); T3: (SARI et al., 2025)	22
Tabela 2 – Comparação dos resultados por tipo de planilha e posicionamento . . .	26
Tabela 3 – Cinco comentários com maiores valores de centralidade de grau.	38
Tabela 4 – Cinco comentários com maiores valores de centralidade de intermediação.	38
Tabela 5 – Cinco comentários com maiores valores de centralidade de proximidade.	39
Tabela 6 – Cinco comentários com maiores valores de PageRank.	39

Lista de siglas

API Interface de Programação de Aplicações - *Application Programming Interface*

COVID-19 Doença do Coronavírus de 2019 - *Coronavirus Disease 2019*

EUA Estados Unidos da América

ISP Provedor de Serviços de Internet - *Internet Service Provider*

LLM Modelo de Linguagem de Larga Escala - *Large Language Model*

ML Aprendizado de Máquina - *Machine Learning*

PLN Processamento de Linguagem Natural

SNS Serviços de Redes Sociais - *Social Networking Services*

TF-IDF Frequência do Termo - Inverso da Frequência nos Documentos - *Term Frequency - Inverse Document Frequency*

Sumário

1	INTRODUÇÃO	11
1.1	Motivação	11
1.2	Problema	12
1.3	Hipótese	12
1.4	Objetivos	13
1.5	Contribuições	13
1.6	Organização da Monografia	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Mineração de mídias sociais	15
2.2	Mineração de textos	16
2.3	Redes complexas	17
2.3.1	Métricas de centralidade em redes complexas	18
2.3.2	Deteção de comunidades em redes complexas	19
2.4	Coleta de dados de redes sociais	19
3	TRABALHOS RELACIONADOS	21
4	METODOLOGIA PROPOSTA	23
4.1	Construção da base de dados a partir da coleta de comentários do YouTube	24
4.2	Rotulagem da base de dados	25
4.3	Modelagem das redes complexas adequadas para a tarefa de análise de posicionamento	26
4.4	Construção das redes complexas projetadas	27
4.5	Cálculo de métricas das redes propostas	28
5	ANÁLISE E DISCUSSÃO DOS RESULTADOS	29
5.1	Análise exploratória por meio de estatística	29

5.2	Caracterização Geral das Redes	31
5.2.1	Rede de Posicionamento em Relação à Guerra	32
5.2.2	Rede de Posicionamento em Relação aos Lados do Conflito	32
5.2.3	Redes por Tipo de Vídeo	33
5.3	Análise de Comunidades	36
5.4	Análise de Métricas de centralidade	37
6	CONCLUSÃO	41
6.1	Principais Contribuições	41
6.2	Trabalhos Futuros	42
	REFERÊNCIAS	44

Introdução

O grande volume de dados produzidos e compartilhados nas redes sociais tem ampliado as possibilidades de análise de fenômenos sociais e políticos. Discussões relacionadas a temas de interesse público, como política, conflitos internacionais e questões sociais, geram grandes quantidades de conteúdos que podem ser utilizados para identificar padrões de opinião e comportamento. Nesse contexto, a análise de dados provenientes das redes sociais tem ganhado relevância como ferramenta de investigação da opinião pública (SILVA, 2021).

Neste trabalho, a análise é direcionada aos comentários publicados no YouTube em vídeos relacionados à guerra entre os Estados Unidos e o Irã. Essa análise, feita a partir desses comentários, é importante para investigar como técnicas de mineração de dados e análise de redes complexas podem ser utilizadas para identificar com maior agilidade padrões de posicionamento e relações entre diferentes manifestações presentes nas discussões.

1.1 Motivação

As redes sociais constituem uma importante fonte de dados para a observação de manifestações espontâneas dos usuários. Informações extraídas dessas plataformas podem ser utilizadas em diferentes contextos, incluindo estudos de opinião pública, comunicação política e análise de debates sociais (ZAFARANI; ABBASI; LIU, 2014a).

Trabalhos anteriores já demonstraram a possibilidade de utilizar comentários e publicações em redes sociais para analisar temas de interesse social e político. O trabalho de (ROCHA et al., 2024), por exemplo, utiliza comentários do YouTube para investigar percepções relacionadas à segurança pública. Outros estudos empregaram técnicas de análise de sentimentos e mineração de dados em contextos como a pandemia de Doença do Coronavírus de 2019 (COVID-19) (PACHECO et al., 2022), as eleições presidenciais de 2022 (OLIVEIRA; OLIVEIRA, 2023) e a Reforma da Previdência de 2019 (RICCI, 2020).

Apesar dessas iniciativas, permanece a possibilidade de explorar outras formas de representação das relações entre comentários, especialmente por meio de redes complexas construídas a partir da similaridade entre os textos. Redes complexas são estruturas que apresentam nós que se ligam entre si, em um contexto real esses nós podem representar pessoas, entidades ou comentários, que, quando relacionados, se conectam através de arestas (BARABÁSI, 2016). Nesse sentido, este trabalho busca investigar se a estrutura formada pelas relações entre comentários pode contribuir para identificar padrões de posicionamento e grupos de discussão relacionados a um conflito internacional.

1.2 Problema

A grande quantidade de comentários disponíveis em plataformas como o YouTube dificulta sua análise individual e torna necessário o uso de métodos capazes de identificar padrões em grandes volumes de dados. Além disso, a simples classificação dos comentários não permite observar como diferentes manifestações estão relacionadas entre si.

Dessa forma, o problema investigado neste trabalho consiste em verificar de que maneira os comentários publicados no YouTube podem ser representados e analisados por meio de redes complexas, de forma a identificar padrões de similaridade, posicionamento e organização das discussões sobre a guerra entre os Estados Unidos e o Irã.

Essa abordagem também envolve desafios relacionados à qualidade e à disponibilidade dos dados, à classificação dos comentários e à escolha de representações capazes de preservar relações relevantes entre os textos.

1.3 Hipótese

A hipótese principal deste trabalho é que a representação de comentários de redes sociais como uma rede complexa, construída a partir de relações de similaridade lexical, pode fornecer informações úteis para identificar padrões de posicionamento e relações entre diferentes grupos de comentários.

A partir dessa hipótese, busca-se verificar se métricas estruturais, como centralidade e detecção de comunidades, permitem identificar comentários que ocupam posições de destaque e grupos de comentários com características semelhantes.

Como questão exploratória adicional, investiga-se se a origem dos comentários em diferentes formatos de conteúdo do YouTube, especificamente vídeos convencionais e *shorts* (vídeos na orientação vertical e com até 3 minutos de duração), está associada a diferenças na estrutura das redes. Essa análise é tratada como exploratória, uma vez que sua avaliação depende da quantidade e da distribuição dos comentários disponíveis em cada categoria.

1.4 Objetivos

Este trabalho tem como objetivo geral analisar como comentários publicados no YouTube sobre a guerra entre os Estados Unidos e o Irã podem ser representados por meio de redes complexas para identificar padrões de posicionamento e organização das discussões.

Tal objetivo geral desdobra-se nos seguintes objetivos específicos:

- ❑ construir uma base de dados de comentários de vídeos do YouTube relacionados à temática estudada;
- ❑ classificar os comentários de acordo com seu posicionamento em relação à guerra e aos lados envolvidos no conflito;
- ❑ modelar redes complexas a partir das relações de similaridade entre os comentários;
- ❑ analisar a estrutura das redes por meio de métricas de centralidade;
- ❑ identificar comunidades presentes nas redes;
- ❑ analisar possíveis diferenças entre redes formadas a partir de comentários de vídeos convencionais e de *shorts*.

1.5 Contribuições

Como contribuição, este trabalho apresenta uma abordagem para a utilização de comentários do YouTube como fonte de dados para a análise de fenômenos sociais e políticos por meio de redes complexas.

Outra contribuição consiste na construção de uma base de dados de comentários relacionados à temática estudada, que poderá ser ampliada e utilizada em investigações futuras. Também foram desenvolvidos códigos para as etapas de coleta, tratamento dos dados, construção das redes e realização das análises, permitindo a reprodução e o aprimoramento da metodologia proposta.

Por fim, o trabalho contribui para a aplicação conjunta de técnicas de mineração de dados, análise de redes sociais e Teoria de Redes Complexas na investigação de discussões políticas presentes em mídias sociais.

1.6 Organização da Monografia

Esta monografia está organizada em seis capítulos.

O Capítulo 2 apresenta a fundamentação teórica necessária para o desenvolvimento da pesquisa, abordando mineração de mídias sociais, mineração de textos, redes complexas e coleta de dados de redes sociais.

O Capítulo 3 apresenta os trabalhos relacionados, discutindo pesquisas que utilizaram técnicas de análise de sentimentos, mineração de dados e dados provenientes de redes sociais em diferentes contextos sociais e políticos.

O Capítulo 4 descreve a metodologia proposta, incluindo a construção da base de dados, a rotulação dos comentários, a modelagem e construção das redes e o cálculo das métricas utilizadas.

O Capítulo 5 apresenta a análise e discussão dos resultados obtidos. Inicialmente, são caracterizadas as redes construídas e, posteriormente, são analisadas as comunidades e as métricas de centralidade, incluindo os resultados obtidos para diferentes recortes dos dados.

Por fim, o Capítulo 6 apresenta as conclusões do trabalho, suas principais contribuições e as possibilidades de trabalhos futuros.

Fundamentação Teórica

Este capítulo trata dos conceitos teóricos fundamentais para o desenvolvimento deste trabalho. Com isso, busca-se explicar cada um deles, especialmente os que têm maior relação com a temática proposta.

2.1 Mineração de mídias sociais

Em (ZAFARANI; ABBASI; LIU, 2014a) são definidos conceitos relevantes sobre a mineração de dados em redes sociais e que são importantes para o entendimento desse trabalho.

Rede social, é definida, em Musso (2006,p.34) como “uma das formas de representação dos relacionamentos afetivos, interações profissionais dos seres humanos entre si ou entre seus agrupamentos de interesses mútuos.”. Também pode-se enxergar as redes como uma ferramenta que busca diminuir as distâncias do mundo físico por meio de um mundo virtual (ZAFARANI; ABBASI; LIU, 2014b).

Com isso, fica claro que a interação virtual, assim como a física, produz resultados. No primeiro caso, trata-se principalmente de um grande volume de dados que cirula nessas redes.

A área da ciência responsável por analisar esses dados é chamada *Social Media Mining*, definida como um processo de análise, representação e extração de padrões específicos dos dados que circulam pelas mídias sociais. Essa ciência foca em apresentar os principais algoritmos para investigação desses dados e discute novos métodos, interligando áreas como ciência da computação, *machine learning*, sociologia, estatística e matemática (ZAFARANI; ABBASI; LIU, 2014b).

Ainda trazendo algumas definições relevantes, é importante saber que quando se trata da análise de dados em redes sociais, alguns elementos importantes: os indivíduos (ou usuários), entidades que podem ser sites e/ou aplicativos, e as interações estabelecidas entre esses dois elementos (ZAFARANI; ABBASI; LIU, 2014b).

Por ser uma área em ascensão, existem alguns desafios ainda sem respostas. Zafarani *et al.* (2014) citam alguns como: O Paradoxo da Big Data, que mostra que o excesso de dados obtidos nessas redes sociais pode, em alguns casos, atrapalhar em vez de ajudar. Isso ocorre principalmente na geração de recomendações. Como recomendar informações relevantes quando se analisa um conjunto de dados cada vez mais genérico? Outro problema está relacionado à obtenção limitada desses dados. Como é possível analisar um recorte de um conjunto de dados e inferir informações sobre o todo? Ou seja, como medir o quanto resultados encontrados indicam, de fato, algum padrão? Além disso, os dados obtidos das redes sociais contam com uma grande quantidade de ruído, o que torna necessário um alto investimento em pré-processamento.

2.2 Mineração de textos

A utilização da Mineração de Textos (*Text Mining*) em conjunto com a Mineração de Dados (*Data Mining*) é bastante frequente na análise de dados para marketing (MATSU-MOTO *et al.*, 2017), em técnicas de recomendação, na tomada de decisão, entre outros. A mineração de textos, no entanto, cuida especificamente da análise de textos e palavras, extraídas, por exemplo, de comentários, questionários e até legendas de fotos.

Quando traz-se a mineração de texto para o contexto das redes sociais, entende-se que técnicas como a extração de opiniões a partir de palavras-chave podem ser muito úteis para detectar sentimentos, sejam eles positivos ou negativos (YI *et al.*, 2003). Diversos trabalhos (RICCI, 2020) (OLIVEIRA; OLIVEIRA, 2023) já têm feito uso dessa ciência para analisar comportamentos sociais a partir de redes como Twitter, YouTube, Instagram.

Para que a análise de sentimentos seja feita, duas etapas são muito importantes, a primeira envolve o pré-processamento, que está relacionado com a seleção das palavras-chave que representem a ideia central do texto. Para isso é feita a tokenização, que divide a frase em vários trechos/tokens, a lematização, que converte as palavras para seu radical, e a remoção de stopwords, que elimina palavras que não acrescentam significado à frase, como preposições (com, em, de). Além do pré-processamento, há o uso de técnicas de processamento de linguagem natural (Processamento de Linguagem Natural (PLN)) para treinar softwares capazes de interpretar semanticamente um texto de maneira similar aos humanos. Esses softwares também conseguem ponderar cada palavra-chave de acordo com o sentimento que ela provoca.

Alguns desafios encontrados na análise de sentimentos são: extrair do texto a parte que realmente é mais relevante e projetar algoritmos que fazem a associação entre sentimento e texto, já que é um documento composto de vários tópicos, que contêm sentimentos relacionados a todos eles (YI *et al.*, 2003).

2.3 Redes complexas

As redes complexas são também um dos pilares para compreender e trabalhar com análise de dados em redes sociais. Isso porque a representação por meio de grafos pode auxiliar na detecção de padrões e oferecer uma boa visualização das interações dos usuários (ZAFARANI; ABBASI; LIU, 2014b), como amizade, afinidade e seguidores.

Barabási (2016) define que para entender um sistema complexo, primeiro devemos entender seus elementos e como eles interagem entre si. Redes complexas são mapeadas em formato de grafos, onde tem-se os vértices e as arestas que os ligam. No entanto, o mesmo autor chama atenção para as nomenclaturas. Grafos são usados principalmente quando trata-se da representação matemática das redes. Mas para a análise de sistemas reais, as redes são mais adequadas, contendo nós (*nodes*) que são ligados (*links*) entre si.

Um exemplo é a internet. É fácil de visualizar os usuários e as relações estabelecidas entre eles. Uma rede nesse contexto possui os usuários como nós e a relação que liga dois usuários que, por exemplo, se seguem, é chamada *link* ou aresta.

Na Figura 1 tem-se três redes de sistemas diferentes representadas por um único formato de grafo (d). No primeiro exemplo, uma rede de computadores conectados entre si. Na segunda rede (b), atores que estão ligados se trabalharam juntos em algum filme. Por fim, (c) mostra proteínas ligadas quando existe alguma evidência científica de que as duas podem se ligar em uma mesma célula.

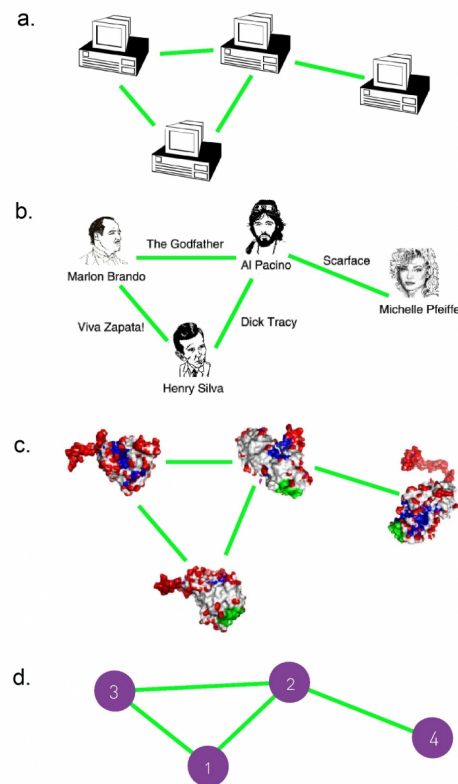


Figura 1 – Imagem retirada do livro Network Science (BARABÁSI, 2016)

As ligações em uma rede podem ser direcionadas, quando apresentam um único sentido ou não direcionadas, quando a ligação é recíproca, ou seja, de ida e volta. Por exemplo, em uma rede como o Instagram um usuário pode seguir outro e esse outro o seguir, o que seria uma ligação não direcionada.

Uma rede é chamada “dígrafo” se contém apenas ligações direcionadas e “não direta” quando suas ligações não são direcionadas. Algumas redes podem ser dos dois tipos (BARABÁSI, 2016).

2.3.1 Métricas de centralidade em redes complexas

As medidas de redes complexas proporcionam a compreensão de como os nós e arestas interagem entre si. Em especial, métricas de centralidade de nó indicam o quão central é a posição de um nó dentro da topologia da rede. A seguir, são destacadas as principais medidas de centralidade de nó utilizadas neste trabalho.

- Grau de Centralidade (*Degree Centrality*): Essa métrica mede o número de conexões de um nó na rede. Nós com um alto grau de centralidade têm muitas conexões e desempenham um papel significativo na rede.

- ❑ Centralidade de Proximidade (*Closeness Centrality*): Essa métrica avalia a proximidade de um nó em relação a outros nós na rede. Nós com alto *closeness centrality* estão próximos de muitos outros nós na rede.
- ❑ Centralidade de Intermediação (*Betweenness Centrality*): Essa métrica mede a importância de um nó como um intermediário na rede. Nós com alto *betweenness centrality* atuam como pontes entre diferentes partes da rede. A remoção de um nó com alto *betweenness* pode indicar a geração de duas componentes conexas na rede.
- ❑ PageRank: O PageRank é uma métrica que avalia a importância dos nós na rede com base em suas conexões. Nós com alto PageRank possuem vizinhos também com altos valores de PageRank.

2.3.2 Detecção de comunidades em redes complexas

O algoritmo Girvan-Newman (GIRVAN; NEWMAN, 2002) é o mais utilizado para detecção de comunidades em redes complexas. Ele separa a rede removendo progressivamente as arestas com maior centralidade de intermediação (*edge betweenness*), ou seja, mais importantes para a conexão entre diferentes partes do grafo.

A medida de Modularidade (*Modularity*) é um *score* matemático que mede a qualidade de uma divisão específica de rede comparando a densidade de arestas dentro de uma comunidade (*intra-community edge density*) em um grafo aleatório. É usada, portanto, para identificar grupos de nós na rede. Ela ajuda a compreender como os nós estão organizados em sub-redes. Quanto mais alta a modularidade de uma rede, maior a capacidade dela de se organizar em grupos bem definidos.

2.4 Coleta de dados de redes sociais

Dada a quantidade e variedade de dados disponíveis pela Internet, coletar esses dados pode ser uma tarefa trabalhosa, principalmente quando esbarra-se nas diversas questões éticas. Felizmente, existem diversas ferramentas e técnicas nessa área. Nessa Seção, são discutidas algumas delas.

Existem diversas formas de extrair dados das redes sociais. A depender dos objetivos com a coleta, pode-se usar desde entrevistas até uso de APIs e técnicas de *Web scraping*.

O uso de entrevista é a técnica mais básica, mas que, a depender da intenção da coleta, pode ser bem útil. Um exemplo trazido pelo trabalho (BENEVENUTO, 2010) é o caso de uma entrevista com usuários do Facebook, em que os pesquisadores (JOINSON, 2008) conseguiram obter diversas informações, como os motivos que levavam esses usuários a utilizar a aplicação. Essa metodologia é bastante empregada também em sistemas que precisam coletar feedback do usuário com relação à experiência com interfaces.

Outras técnicas mais rebuscadas e bastante utilizadas envolvem pontos de coleta intermediários, ou seja, coletam dados dos pontos de agregação de tráfego na rede. Isso pode ser feito a partir de dados que passam, por exemplo, pelo provedor de internet Provedor de Serviços de Internet (ISP) (BENEVENUTO, 2010).

Coletar dados a partir de servidores proxy que estão intermediando o tráfego entre clientes e servidores é uma prática comum. Alguns trabalhos (GILL et al., 2008) (SCHNEIDER et al., 2009) (ZINK et al., 2008) (AAZIZ et al., 2018) fazem uso dessa prática para coletar dados de algum servidor próximo e conseguir caracterizar e analisar tráfego e padrões de alguma localidade.

Um outro método de coleta seria a partir dos servidores que hospedam as aplicações, mas devido a questões de privacidade e segurança dos usuários, a maior parte das aplicações bloqueia e impede esse acesso. É assim que entra a técnica de *scraping*, que consiste em acessar determinada página e, por meio de um robô *scraper*, realizar a coleta de informações públicas. Como exemplo de ferramenta, o Snsrape é utilizado especificamente para fazer a raspagem em redes sociais, também chamadas Serviços de Redes Sociais (SNS) (social networking services), coletando dados como informações de perfil, hashtags, comentários, entre outros (NAIK et al., 2024).

O trabalho (NAIK et al., 2024) traz um bom exemplo de como utilizar essas ferramentas para coletar comentários do Twitter, atual X, e utiliza técnicas de *Machine Learning* para analisar e classificar esses comentários. No trabalho os autores usam a Interface de Programação de Aplicações (API) Tweepy, que permite, através da linguagem Python, a interação com a plataforma do Twitter e a coleta de posts e comentários a partir do já mencionado Snsrape.

Trabalhos Relacionados

O trabalho (RICCI, 2020) teve como objetivo analisar o sentimento dos usuários do Twitter a respeito da Reforma da Previdência no Brasil, no ano de 2019, utilizando a mineração de dados para entender o posicionamento da sociedade sobre o tema. Para isso, o trabalho coletou dados do Twitter e aplicou técnicas de Processamento de Linguagem Natural (PLN), incluindo pré-processamento (como tokenização e remoção de *stopwords*), para então realizar a análise de sentimentos. Ao final, o estudo conseguiu identificar e classificar comentários como positivos, negativos ou neutros, revelando o sentimento geral da população em relação à reforma.

A relação com este trabalho é direta, servindo como um modelo metodológico de como a mineração de dados em redes sociais pode ser utilizada para analisar o posicionamento da sociedade sobre temas polêmicos.

O artigo de B. Oliveira e L. Oliveira (2023) investigou a aplicação de técnicas de aprendizado de máquina e análise de sentimento para analisar dados de redes sociais relacionados às eleições presidenciais de 2022. A metodologia consistiu na coleta de dados de plataformas como Twitter e YouTube, seguida pela aplicação de algoritmos de *machine learning* para classificar os comentários e postagens. O resultado foi a identificação e a classificação do sentimento expresso pelos usuários, fornecendo um panorama do comportamento social e do debate político online. Tal trabalho valida a abordagem de usar a análise de sentimento para entender comportamentos sociais em redes como o YouTube e o Twitter.

Sari *et al.* (2025) discutem como a identidade nacional é expressa nos comentários de vídeos do YouTube, oferecendo um bom alinhamento com este trabalho. Os autores utilizam uma abordagem de análise de conteúdo de texto em larga escala para extrair padrões e sentimentos dos comentários, visando entender como os usuários se expressam sobre sua identidade nacional. A pesquisa resulta em uma compreensão mais profunda das interações e da manifestação cultural dentro da rede social. A importância desse trabalho reside no fato de que ele utiliza o YouTube como fonte de dados, que é mesma plataforma escolhida para este trabalho.

A Tabela 1 estabelece um comparativo geral sobre os trabalhos anteriormente mencionados.

Tabela 1 – Tabela Comparativa de Trabalhos Relacionados. T1: (RICCI, 2020); T2: (OLIVEIRA; OLIVEIRA, 2023); T3: (SARI et al., 2025)

Característica	T1	T2	T3
Objetivo	Análise de sentimento no Twitter sobre a Reforma da Previdência.	Investigação de ML e análise de sentimento em redes sociais.	Entender como identidade nacional é expressa no YouTube.
Metodologia	Mineração de dados e pré-processamento de texto.	Coleta de dados (Twitter/YouTube) e uso de algoritmos de ML.	Análise de conteúdo e extração de padrões em larga escala.
Principal Resultado	Classificação de comentários.	Panorama do comportamento social e debate político online.	Compreensão de interações e manifestação cultural em redes.
Relação com o Projeto	Modelo para mineração de dados em redes sociais.	Valida o uso de ML para análise de comportamento social.	Usa YouTube como fonte de dados e analisa comentários.

Pode-se concluir que os trabalhos apresentados nessa seção abordam temáticas que relacionam-se com a proposta de utilizar dados de redes sociais para obter informações relevantes sobre contexto político-social. No entanto, vale ressaltar que nenhum deles utilizou a técnica de análise por meio de redes complexas, ao invés disso, a maioria foca em técnicas de *machine learning*.

Metodologia Proposta

Para realizar a análise do posicionamento dos brasileiros em relação à guerra Irã x EUA utilizando dados do YouTube, é proposta uma metodologia que engloba desde a coleta e rotulagem da base de dados até a aplicação de técnicas de análise de redes complexas. A Figura 2 ilustra tal metodologia, cujas etapas são descritas nas seções subsequentes.



Figura 2 – Fluxograma das principais etapas desenvolvidas.

4.1 Construção da base de dados a partir da coleta de comentários do YouTube

A etapa de construção da base de dados foi essencial no desenvolvimento do trabalho, uma vez que todas as demais etapas dependiam dela. Nessa fase, que durou aproximadamente 1 mês, foi construída uma base de dados inédita, por meio da coleta dos comentários de vídeos do YouTube relacionados com a temática “Guerra Irã e EUA”.

A extração começou a partir da seleção arbitrária de 16 vídeos da plataforma YouTube, encontrados por meio da busca “Guerra Irã e EUA” e ordenados pelo número de visualizações. Em seguida, eliminou-se aqueles vídeos que estavam fora dos critérios estabelecidos e baseados no estudo da média geral dessa primeira amostra:

- ❑ Data de publicação menor ou igual a 1 ano
- ❑ Mínimo de 500 mil visualizações
- ❑ Mínimo de 400 comentários

Os vídeos foram assistidos e categorizados em Narrador principal (Feminino ou Masculino), Gênero (Informativo, Humorístico, Noticiário, Reportagem) e Posicionamento político do vídeo e do canal (Direita, Esquerda, Neutro). Tudo isso foi importante para perceber se a escolha dos vídeos não estava desproporcional ou tendenciosa. A Figura 3 sumariza o resultado obtido, totalizando 10 vídeos.

VIDEO_ID	TÍTULO	TIPO	NARRADOR PRINCIPAL	GÊNERO	POSICIONAMENTO POLÍTICO
UbQVMxPg-Yo	ESTADOS UNIDOS VS IRÃ O QUE ESTÁ ACONTECENDO? (Brasil vai ser impactado?)	Vídeo	Masculino	Informativo	Neutro
aTD4gNDatM8	Passo a passo: como EUA e Israel conseguiram se infiltrar no espaço aéreo iraniano	Vídeo	Feminino	Reportagem	Neutro
3D_vS0XIC1w	Entenda o conflito entre IRÃ e ESTADOS UNIDOS de uma FORMA SIMPLES	Vídeo	Masculino	Informativo	Neutro/Esquerda
OAYZlqY4Cxg	O que está por trás do ataque de EUA e Israel ao Irã	Vídeo	Feminino	Reportagem	Neutro
n3sAl2PKuKk	EUA e Israel ATACAM o Irã: entenda o que aconteceu e o que vem agora!	Vídeo	Masculino	Opinativo	Direita
Kes-d8bJBuU	Irã vs Israel. 5 minutinhos sem perder a amizade!	Shorts	Masculino	Humorístico	Neutro?
PQGF2-G9YIQ	Marinha do Irã está "completamente destruída", diz Trump LIVE CNN	Shorts	Feminino	Noticiário	Neutro
kCuZj39ONNU	Vídeo mostra míssil iraniano caindo na capital do Catar, Doha	Shorts	Sem narrador	Noticiário	Neutro
1VIT8dlcRU0	CONFLITO ENTRE IRÃ E ISRAEL ENTRA EM NOVA FASE CRÍTICA	Shorts	Sem narrador	Noticiário	Neutro
6_UbB96L2qY	Irã cria uma animação em estilo 'Lego' sobre a guerra	Shorts	Feminino	Noticiário	Neutro

Figura 3 – Vídeos selecionados para a coleta de comentários. Classificação realizada manualmente em relação ao tipo, narrador principal, gênero e posicionamento político.

Com os vídeos selecionados, implementou-se um código em Python para coletar comentários e metadados de cada vídeo, i.e: conteúdo textual de cada comentário, a data de publicação, o nome de usuário do autor do comentário, e as respostas que cada comentário gerou, além da quantidade de curtidas. Utilizou-se a API do YouTube ¹ para realizar tal coleta e foram obtidos 14.052 comentários.

¹ <https://developers.google.com/youtube/v3>

4.2 Rotulagem da base de dados

A etapa de rotulação da base envolveu a classificação de cada comentário nos seguintes rótulos:

- ❑ Em relação a guerra, se o comentário evidenciava ter um posicionamento a favor da guerra, contra a guerra ou neutro.
- ❑ Em relação ao posicionamento, se o comentário uma vez classificado a favor, se mostrava favorável a um lado: EUA ou Irã.

Para rotular toda a base, inicialmente foi feita uma classificação manual, em que os 50 primeiros comentários de cada vídeo eram classificados nos rótulos explicados acima. Uma vez tendo a amostra, ela foi utilizada como exemplo para que um Modelo de Linguagem de Larga Escala (LLM) (Modelo de Linguagem de Larga Escala) classificasse todos os outros comentários. Esse tipo de *prompt* é chamado *Few-shot Prompting* porque contém alguns poucos exemplos de entrada e saída para que o modelo aprenda primeiro e consiga então reproduzir um resultado no formato desejado (BROWN et al., 2020).

A escolha do LLM para classificação baseou-se naquele que obteve um melhor desempenho nos testes, ou seja, aquele modelo que mais acertou quando comparado com os comentários manualmente classificados.

Foi utilizado no GPT 5.5 e no Gemini 3.1 Pro o seguinte prompt:

“Em anexo um csv com comentários de vídeos relacionados à Guerra Irã e EUA, extraídos do YouTube, classifique cada comentário em relação ao posicionamento na guerra (A favor, Contra, Neutro) e quanto ao lado (a favor do Irã ou dos EUA), gerando duas novas colunas para essas duas classificações na planilha. Considere ainda que: Só há classificação em relação ao lado se o posicionamento em relação à guerra for a favor, Se houver incerteza opte por Neutro no posicionamento em relação à guerra, Deixe em branco a classificação em relação ao lado se a classificação em relação à Guerra for Neutro ou Contra. Deixo abaixo 6 exemplos:[...]”

As planilhas anexadas continham todos os metadados dos comentários e o conteúdo textual de cada um e estavam separadas por tipo de vídeo para evitar sobrecarga do modelo.

Após, foi calculado o índice de concordância Kappa, proposto por Cohen (COHEN, 1960), entre o valor retornado pelo respectivo LLM e a classificação manual. Esse índice permite avaliar a concordância entre duas classificações categóricas, considerando também a concordância esperada ao acaso. Para a interpretação dos valores obtidos, foi utilizada a classificação proposta por Landis e Koch (LANDIS; KOCH, 1977). A Tabela 2 sumariza os resultados obtidos.

Tabela 2 – Comparação dos resultados por tipo de planilha e posicionamento

Planilha	Classificação	GPT 5.5	Gemini 3.1 Pro
Shorts	Posicionamento Guerra	0.042	0.4249
Shorts	Posicionamento Lado	0.225	0.6597
Vídeos	Posicionamento Guerra	0.197	0.4081
Vídeos	Posicionamento Lado	0.137	0.5639

O LLM Gemini 3.1 Pro obteve maiores índices Kappa para as 4 tarefas de classificação (posicionamento guerra, posicionamento lado para shorts e posicionamento guerra, posicionamento lado para vídeos). Foi, portanto, o escolhido para classificar toda a base de dados.

4.3 Modelagem das redes complexas adequadas para a tarefa de análise de posicionamento

Para a análise de posicionamento, cada comentário foi modelado como um nó da rede. As relações entre os comentários foram definidas com base na similaridade lexical. Essa técnica baseia-se principalmente no conceito do Freqüência do Termo - Inverso da Freqüência nos Documentos (TF-IDF), que apesar de não ter sido efetivamente aplicado, refletiu-se no conceito geral de que comentários que compartilham termos relevantes apresentam algum grau de proximidade discursiva (SALTON; WONG; YANG, 1975).

A rede foi modelada como um grafo não direcionado e ponderado. As arestas representam relações de similaridade textual entre pares de comentários, enquanto o peso de cada aresta corresponde à quantidade de termos compartilhados entre os comentários conectados. Por se tratar de uma relação de similaridade, a conexão entre dois comentários não possui uma direção específica.

O procedimento aqui aplicado pode ser compreendido conceitualmente como a projeção monopartida de uma rede bipartida comentário-termo (LATAPY; MAGNIEN; VECCHIO, 2008). Na estrutura bipartida intermediária, um conjunto de nós representou os comentários e outro representou os termos presentes nesses textos, sendo feita uma ligação comentário-termo. Em seguida, a projeção final (monopartida) resultou em uma rede composta apenas por comentários, na qual dois nós são conectados quando possuem palavras relevantes em comum. A Figura 4 ajuda a compreender melhor a abordagem utilizada.

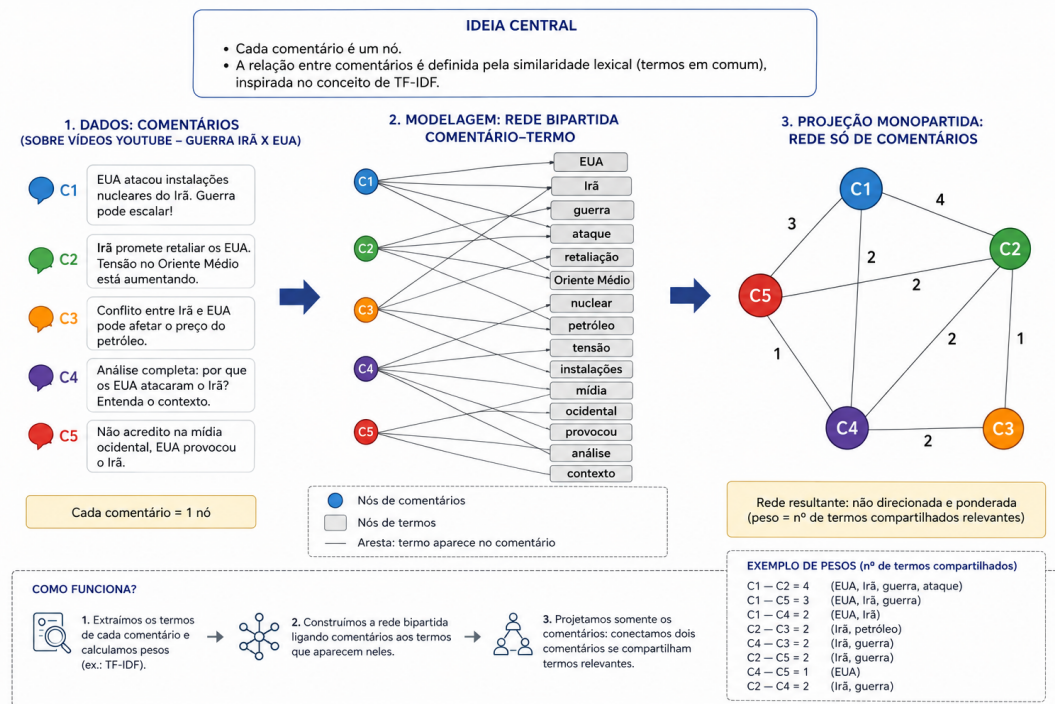


Figura 4 – Conceito da projeção monopartida de uma rede bipartida comentário-termo

Essa modelagem permite representar a organização dos comentários segundo sua proximidade vocabular. A partir da estrutura relacional obtida, torna-se possível identificar agrupamentos de comentários, elementos estruturalmente centrais e padrões associados aos posicionamentos presentes nas discussões analisadas.

4.4 Construção das redes complexas projetadas

Para realizar a construção da rede foi necessário antes o pré-processamento linguístico dos comentários classificados. Essa etapa foi essencial para preservar somente termos compostos por letras (MANNING; RAGHAVAN; SCHÜTZE, 2008). Dessa forma, os textos foram convertidos para letras minúsculas, tokenizados e filtrados. Também foram removidas as *stopwords* da língua portuguesa, palavras com até duas letras e outros termos considerados pouco informativos.

Além disso, os comentários classificados como neutros em relação ao posicionamento sobre a guerra foram excluídos da construção das redes. E para reduzir o número conexões artificiais produzidas por palavras excessivamente frequentes, foram desconsiderados os termos presentes em mais de 5% dos comentários.

Após o pré-processamento, cada comentário foi representado por um conjunto de termos relevantes. Em seguida, os comentários foram comparados entre si, sendo criada uma aresta sempre que dois deles apresentavam uma ou mais palavras em comum. O peso atri-

buído à aresta foi definido pela quantidade de termos compartilhados pelo respectivo par de comentários.

Por último, foram aplicados filtros mínimos de peso às arestas, com o objetivo de evidenciar principalmente as conexões que apresentavam maior evidência de similaridade lexical. As redes projetadas resultantes foram então utilizadas nas etapas de visualização, cálculo das métricas de centralidade e identificação de comunidades.

4.5 Cálculo de métricas das redes propostas

Na computação, assim como em diversas áreas, é necessário trabalhar com métricas e propriedades para entender melhor um sistema e também comprovar ou descartar hipóteses. Esse foi um passo importante para computar as propriedades das nossas redes complexas, tais como a análise de centralidades e comunidades. Com isso, foi possível analisar bem as relações entre os elementos das redes obtidas.

As métricas de centralidade permitem identificar vértices que ocupam posições de destaque na estrutura de uma rede. Para isso foram utilizadas as seguintes métricas de centralidades: grau, centralidade de intermediação (*betweenness*), centralidade de proximidade (*closeness*) e o *PageRank*. Em conjunto, essas métricas permitiram observar diferentes aspectos da importância estrutural dos comentários nas redes analisadas.

Por fim, para identificação das comunidades da rede foi utilizado o algoritmo de Girvan-Newman, que separa a rede removendo progressivamente as arestas com maior centralidade de intermediação (edge betweenness), ou seja, mais importantes para a conexão entre diferentes partes do grafo (GIRVAN; NEWMAN, 2002).

Análise e Discussão dos Resultados

Este Capítulo tem como objetivo discutir os resultados obtidos a partir da análise das redes complexas construídas. Dessa forma, será possível compreender não apenas como essas redes estão estruturadas, mas também a semântica associada a cada um dos grafos relacionados aos comentários publicados no YouTube sobre a guerra entre os Estados Unidos e o Irã.

Inicialmente, é apresentada uma análise exploratória com as principais características da base de dados construída. Em seguida, são apresentadas informações gerais sobre cada rede, como o número de vértices e arestas, a densidade, o grau médio e a quantidade de componentes conexos. Por fim, são analisadas medidas relacionadas à importância e ao posicionamento dos comentários nas redes, incluindo métricas de centralidade e a identificação de comunidades.

5.1 Análise exploratória por meio de estatística

A análise exploratória dos dados por meio de técnicas da estatística descritiva permitiu ter reflexões importantes sobre a base de dados e entender melhor os padrões que podem ser extraídos dela, além de visualizar possíveis relações entre as variáveis do conjunto de dados em questão.

Primeiro foi realizado o pré-processamento, que envolveu majoritariamente a tarefa de remover comentários em brancos e também inserir o marcador “nulo” para aqueles comentários que não tinham um posicionamento definido. Realizado o pré-processamento, a base, até então, com 14057 comentários, passou a ter 14052. Ou seja, foram removidas 5 instâncias que não continham conteúdo textual.

Após essa etapa foi possível obter diversas análises, como a da proporção da base de dados em relação a quantidade de comentários classificados em cada uma das categorias de “Posicionamento Guerra” e também em relação as categorias de “Posicionamento Lado”. O resultado é mostrado na Figura 5:

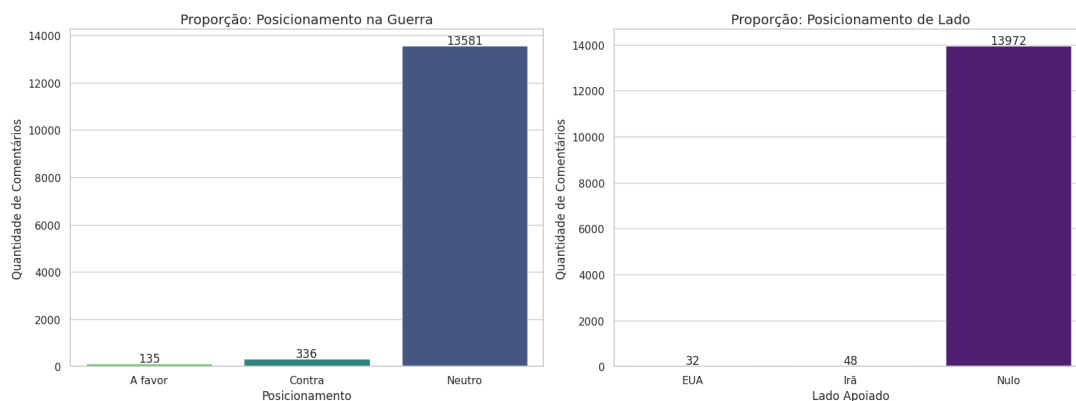


Figura 5 – Gráfico com as proporções relacionadas ao posicionamento Guerra e Lado

Após análise das proporções, constatou-se que o volume de comentários neutros era muito maior que a quantidade de comentários classificados em cada categoria. Isso pode ter ocorrido, porque muitas vezes, os comentários eram apenas de cunho informativo e declarativo, mas não opinativo o suficiente para ser classificado.

Outras duas análises envolveram a proporção do número de comentários de acordo o tipo de vídeo (Shorts ou Vídeo) (Figura 6) e a média de palavras por comentário em cada tipo de vídeo (Figura 7). Ambas evidenciaram um desbalanceamento significativo, que posteriormente impactou nos resultados das redes complexas quando se segmentava a rede por tipo de vídeo.

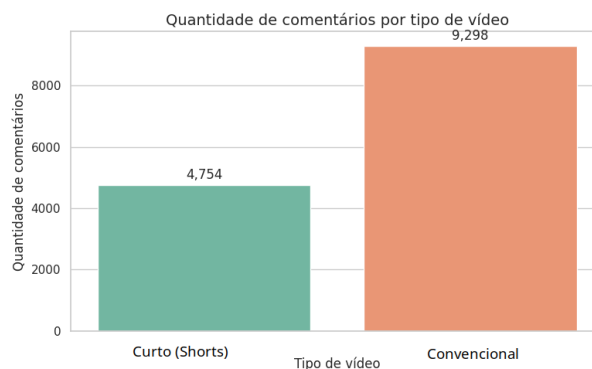


Figura 6 – Quantidade de comentários de acordo o tipo de vídeo

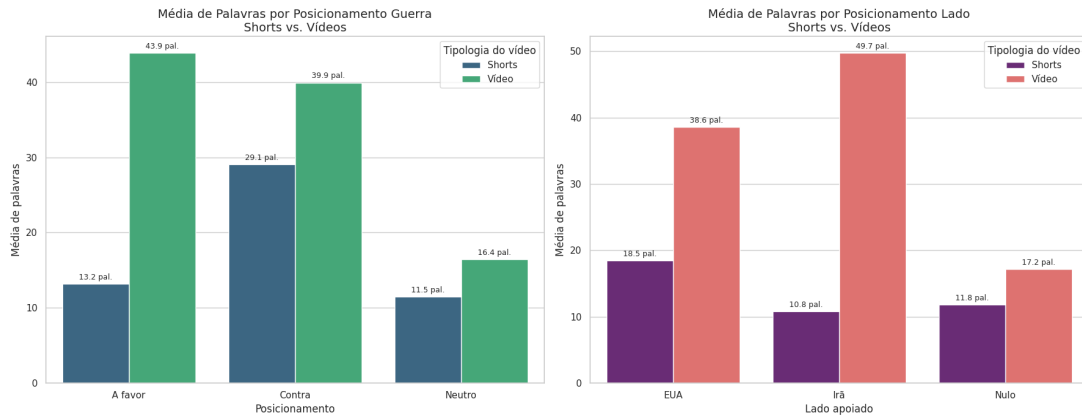


Figura 7 – Quantidade média de palavras por comentário de acordo o tipo de vídeo

Por fim, a nuvem de palavras obtida e mostrada na Figura 8 também foi importante para validar se os comentários continham palavras de fato relacionadas à temática proposta neste trabalho, o que foi confirmado. Observou-se ainda que a temática dos tópicos e das palavras variam conforme o posicionamento, a favor ou contra a guerra.



Figura 8 – Nuvens de palavras segmentadas em relação ao posicionamento na guerra.

5.2 Caracterização Geral das Redes

A análise descritiva foi iniciada pela caracterização geral das redes obtidas. Para cada rede, foram observados o número de vértices, que corresponde aos comentários coletados de vídeos e de *shorts* — vídeos curtos — do YouTube, e o número de arestas, que representa as relações de similaridade entre esses comentários. Também foram consideradas medidas estruturais que permitem avaliar o nível de conectividade e a forma de organização das discussões.

As redes apresentadas a seguir são derivadas de uma rede inicial, construída com base na similaridade entre os comentários. Essa rede inicial conectou comentários com termos relevantes em comum, atribuindo maior peso às conexões entre comentários que compartilham uma quantidade maior de palavras.

5.2.1 Rede de Posicionamento em Relação à Guerra

A rede apresentada na Figura 9 representa os posicionamentos dos comentários em relação à guerra, sua estrutura contém 187 nós e 763 arestas. Essa rede tem como objetivo analisar os comentários classificados como “a favor” ou “contra” a guerra entre os Estados Unidos e o Irã.

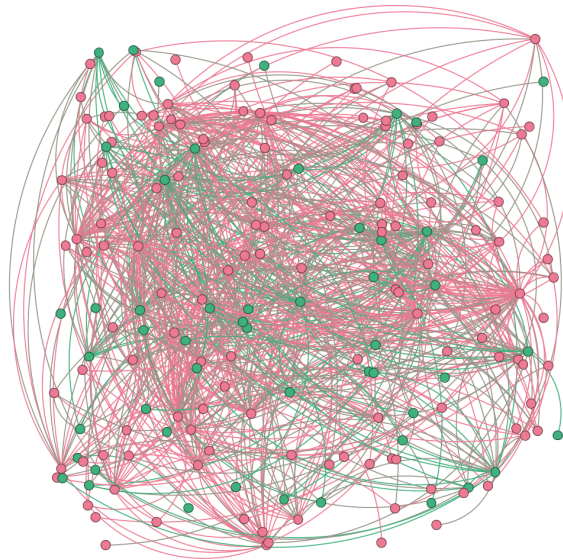


Figura 9 – Rede de comentários classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.

Os comentários neutros, ou seja, aqueles sem uma classificação definida, foram removidos da rede. Além disso, foram eliminados os nós com grau inferior a 3, isto é, os nós que possuem menos de três conexões. Esse filtro contribui para a redução do “ruído” da rede, permitindo destacar os comentários com maior participação na estrutura de similaridade.

5.2.2 Rede de Posicionamento em Relação aos Lados do Conflito

A rede apresentada na Figura 10, com 34 nós e 75 arestas, também foi obtida a partir de um recorte da rede inicial. Nesse caso, foram removidos os comentários contrários à guerra e mantidos os comentários favoráveis, posteriormente diferenciados de acordo com o lado apoiado no conflito, isto é, Irã ou Estados Unidos. O filtro de grau mencionado anteriormente também foi mantido.

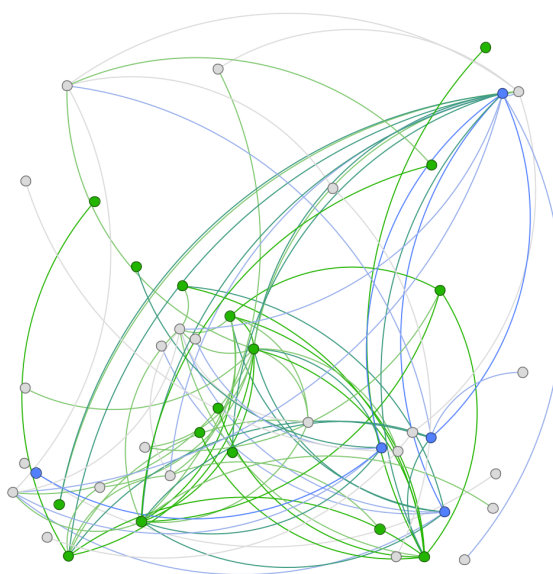


Figura 10 – Rede de comentários classificados de acordo com o posicionamento em relação aos lados do conflito. Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou aos Estados Unidos.

5.2.3 Redes por Tipo de Vídeo

Com o objetivo de verificar se o formato do conteúdo publicado no YouTube poderia influenciar a estrutura das redes, as duas redes analisadas anteriormente foram subdivididas de acordo com o tipo de vídeo de origem dos comentários. Dessa forma, foram construídas quatro novas redes: duas contendo apenas comentários provenientes de vídeos convencionais (Figuras 11 e 13) e duas contendo apenas comentários provenientes de *shorts* (Figuras 12 e 14).

A primeira subdivisão foi realizada sobre a rede de posicionamento em relação à guerra, resultando em uma rede formada por comentários de vídeos convencionais e outra formada por comentários de *shorts*.

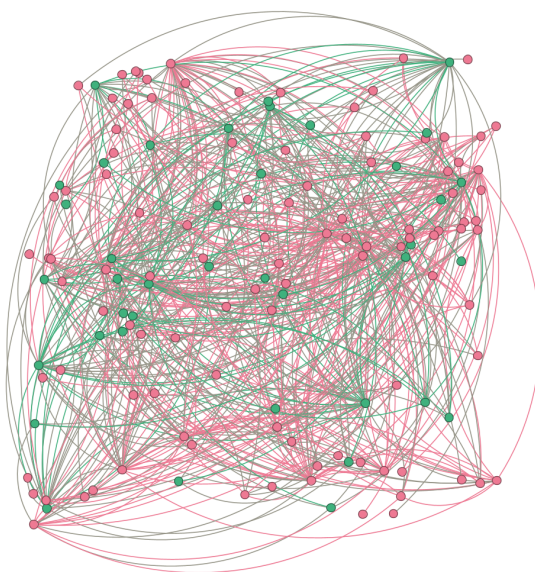


Figura 11 – Rede de comentários extraídos de vídeos, classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.

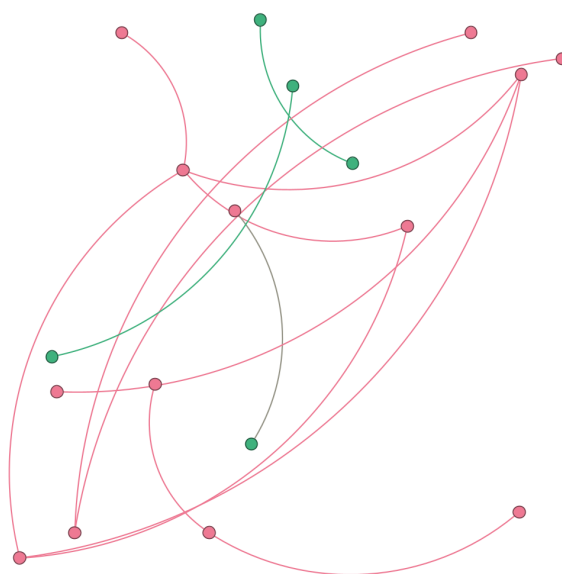


Figura 12 – Rede de comentários extraídos de vídeos curtos (shorts), classificados de acordo com o posicionamento em relação à guerra. Verde: favorável à guerra; Vermelho: Contrário à guerra.

O mesmo procedimento foi aplicado à rede de posicionamento em relação aos lados do conflito, produzindo outras duas redes. Foram mantidos, em cada recorte, os mesmos critérios de classificação e filtragem empregados nas redes apresentadas anteriormente:

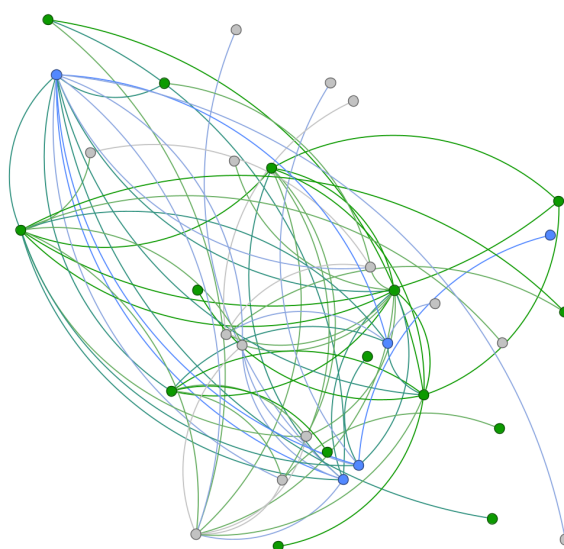


Figura 13 – Rede de comentários extraídos de vídeos, classificados de acordo com o posicionamento em relação aos lados do conflito. Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou aos Estados Unidos.

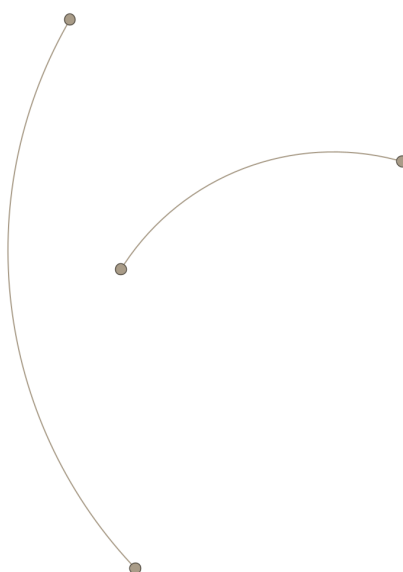


Figura 14 – Rede de comentários extraídos de vídeos curtos (shorts), classificados de acordo com o posicionamento em relação aos lados do conflito. Verde: Irã; Azul: Estados Unidos; Cinza: Sem apoio definido ao Irã ou aos Estados Unidos.

A hipótese inicial era de que o tipo de conteúdo poderia influenciar a organização dos comentários, uma vez que vídeos convencionais e *shorts* apresentam formatos, durações e formas de consumo distintas. Assim, esperava-se identificar diferenças na maneira como os comentários se agrupavam e nas relações de similaridade estabelecidas entre eles.

Entretanto, a divisão dos dados reduziu consideravelmente a quantidade de comentários disponível em cada rede, especialmente no conjunto proveniente de *shorts*. O menor volume de dados dificultou a identificação de padrões estruturais consistentes e limitou a interpretação das métricas de centralidade e das comunidades formadas. Por esse motivo, os resultados das redes de *shorts* foram considerados inconclusivos.

Embora o recorte por tipo de vídeo não tenha produzido evidências suficientes para confirmar a hipótese inicialmente proposta, a análise exploratória permitiu identificar uma limitação da base de dados. A comparação adequada entre vídeos convencionais e *shorts* exigiria uma quantidade maior e mais equilibrada de comentários em cada categoria. Portanto, não é possível concluir, a partir dos dados disponíveis, que o tipo de vídeo não influencia a estrutura das redes, mas apenas que o volume de comentários analisado não foi suficiente para demonstrar essa possível influência.

5.3 Análise de Comunidades

As comunidades da rede foram identificadas pelo algoritmo de Girvan-Newman (GIRVAN; NEWMAN, 2002). Antes da detecção das comunidades, foram considerados apenas comentários não neutros, conexões com peso mínimo igual a 3 e foram removidos nós isolados, para que a análise focasse nas relações mais consistentes da rede.

O algoritmo gera diferentes possibilidades de divisão da rede. Para escolher a melhor partição, foi calculada a modularidade, uma medida que indica o quanto os grupos formados estão bem separados entre si e bem conectados internamente (GIRVAN; NEWMAN, 2002). A partição selecionada foi aquela com maior modularidade entre as primeiras divisões avaliadas. Por esse critério, foram encontrados 10 grupos, conforme mostrado na Figura 15.

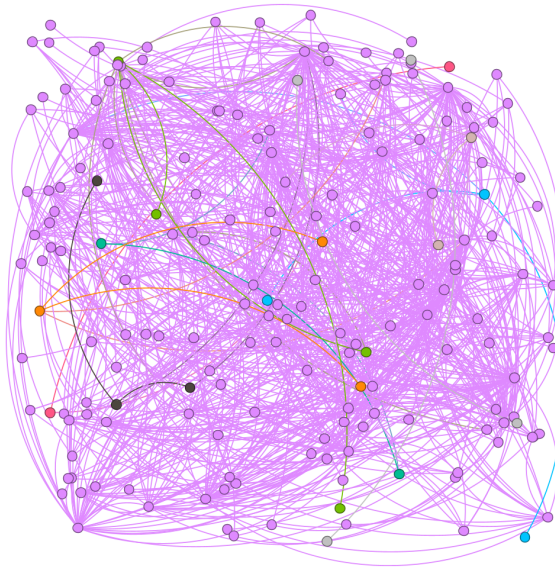


Figura 15 – Comunidades de Girvan-Newman geradas

A existência de 10 grupos significa que a estrutura das conexões entre os comentários apresentou 10 conjuntos relativamente distintos. Em termos interpretativos, esses grupos indicam blocos de comentários que compartilham vocabulário, temas ou padrões de posicionamento mais próximos entre si do que em relação ao restante da rede. No resultado gerado, a maior comunidade concentra a maioria dos comentários, enquanto os demais grupos menores indicam núcleos mais específicos ou periféricos de discussão.

5.4 Análise de Métricas de centralidade

As métricas de centralidade foram utilizadas para identificar comentários que ocupam posições de destaque na estrutura da rede. A análise foi realizada sobre a rede não direcionada e ponderada após a remoção dos comentários neutros -aqueles que não tinham posicionamento definido em relação a Guerra- e a manutenção apenas das arestas com peso maior ou igual a 3. A rede resultante possui 187 nós e 763 arestas, distribuídos em três componentes conectados. O maior componente contém 183 nós, correspondendo a aproximadamente 97,9% da rede, o que indica que quase todos os comentários pertencem a uma mesma estrutura relacional principal.

Conforme discutido no Capítulo 2, a centralidade de grau (*degree centrality*), a centralidade de intermediação (*degree betweenness*) e a centralidade de proximidade (*degree closeness*) representam diferentes formas de avaliar a posição de um vértice. Além dessas medidas, foi calculado o PageRank, que considera a importância dos vértices conectados e, neste trabalho, também o peso das relações de similaridade lexical (BRIN; PAGE, 1998). As Tabelas 3 a 6 apresentam os cinco comentários mais bem posicionados em cada métrica.

Tabela 3 – Cinco comentários com maiores valores de centralidade de grau.

Rank	Posicionamento	Lado	Grau	Trecho do comentário
1	Contra	–	0,3656	Israel é um estado genocida desde 1948 Infelizmente o brasileiro não estuda, não pesquisa, só escuta pastores e...
2	Contra	–	0,3011	Você vai ficar impressionado, mas vou citar aqui 4 afirmações que a Bíblia Sagrada faz: 1) A Bíblia indica que no...
3	Contra	–	0,2581	A GRANDE AFLIÇÃO QUE YESHUA PREVIU AOS JUDEUS. Vem o Fim, o Fim vem, despertou-se contra ti; eis que vem. "Quando,...
4	A favor	EUA	0,2473	Viram aí porque é que o Irã precisava de uma bomba atômica? Pra se defender de assassinos como Trump. Os iranianos...
5	A favor	Irã	0,2258	Pelo que entendo do conflito, os EUA acabam se envolvendo porque o Irã se tornou uma das principais forças hostis...

Tabela 4 – Cinco comentários com maiores valores de centralidade de intermediação.

Rank	Posicionamento	Lado	Intermediação	Trecho do comentário
1	Contra	–	0,1494	Israel é um estado genocida desde 1948 Infelizmente o brasileiro não estuda, não pesquisa, só escuta pastores e...
2	Contra	–	0,1378	Você vai ficar impressionado, mas vou citar aqui 4 afirmações que a Bíblia Sagrada faz: 1) A Bíblia indica que no...
3	Contra	–	0,0863	A GRANDE AFLIÇÃO QUE YESHUA PREVIU AOS JUDEUS. Vem o Fim, o Fim vem, despertou-se contra ti; eis que vem. "Quando,...
4	Contra	–	0,0703	A função de um presidente deveria ser, acima de tudo, proteger a vida, garantir a segurança do seu povo e buscar...
5	A favor	EUA	0,0659	Viram aí porque é que o Irã precisava de uma bomba atômica? Pra se defender de assassinos como Trump. Os iranianos...

Tabela 5 – Cinco comentários com maiores valores de centralidade de proximidade.

Rank	Posicionamento	Lado	Proximidade	Trecho do comentário
1	Contra	–	0,5726	Israel é um estado genocida desde 1948 Infelizmente o brasileiro não estuda, não pesquisa, só escuta pastores e...
2	Contra	–	0,5397	Você vai ficar impressionado, mas vou citar aqui 4 afirmações que a Bíblia Sagrada faz: 1) A Bíblia indica que no...
3	A favor	EUA	0,5364	Viram aí porque é que o Irã precisava de uma bomba atômica? Pra se defender de assassinos como Trump. Os iranianos...
4	Contra	–	0,5316	A GRANDE AFLIÇÃO QUE YESHUA PREVIU AOS JUDEUS. Vem o Fim, o Fim vem, despertou-se contra ti; eis que vem. "Quando,...
5	A favor	Irã	0,5192	Pelo que entendo do conflito, os EUA acabam se envolvendo porque o Irã se tornou uma das principais forças hostis...

Tabela 6 – Cinco comentários com maiores valores de PageRank.

Rank	Posicionamento	Lado	PageRank	Trecho do comentário
1	Contra	–	0,0516	Israel é um estado genocida desde 1948 Infelizmente o brasileiro não estuda, não pesquisa, só escuta pastores e...
2	Contra	–	0,0408	Você vai ficar impressionado, mas vou citar aqui 4 afirmações que a Bíblia Sagrada faz: 1) A Bíblia indica que no...
3	Contra	–	0,0301	A GRANDE AFLIÇÃO QUE YESHUA PREVIU AOS JUDEUS. Vem o Fim, o Fim vem, despertou-se contra ti; eis que vem. "Quando,...
4	A favor	EUA	0,0295	Viram aí porque é que o Irã precisava de uma bomba atômica? Pra se defender de assassinos como Trump. Os iranianos...
5	A favor	Irã	0,0271	Pelo que entendo do conflito, os EUA acabam se envolvendo porque o Irã se tornou uma das principais forças hostis...

Os resultados mostram consistência entre as diferentes medidas. O mesmo comentário ocupa a primeira posição nos quatro rankings, apresentando centralidade de grau igual a 0,3656 (que indica que esse comentário está diretamente conectado a 68 dos 186 outros nós possíveis), centralidade de intermediação igual a 0,1494, centralidade de proximidade igual a 0,5726 e PageRank igual a 0,0516. A presença desse comentário na primeira posição em todas as métricas mostra que ele reúne diferentes características estruturais: possui muitas conexões, encontra-se próximo de grande parte da rede, participa de caminhos que conectam diferentes regiões e mantém relações lexicalmente fortes com outros comentários relevantes.

Também foi observada uma elevada recorrência entre os comentários mais bem classificados. Quatro comentários aparecem entre os cinco primeiros nas quatro métricas, ainda que em posições diferentes. Esse resultado sugere a existência de um conjunto reduzido

de comentários que ocupa o núcleo estrutural da rede e concentra uma parcela importante das relações de similaridade lexical.

A distribuição da centralidade de grau também evidencia uma estrutura desigual. Enquanto o maior valor foi aproximadamente 0,366, a média foi 0,044 e a mediana foi 0,016. A diferença entre esses valores indica que poucos comentários apresentam muitas conexões, enquanto a maior parte permanece em posições mais periféricas. Assim, a rede possui um núcleo relativamente conectado e um número maior de comentários com baixa participação na estrutura de similaridade.

Em suma, esses resultados devem ser interpretados exclusivamente no contexto da rede construída. Valores elevados de centralidade não indicam maior qualidade, veracidade ou relevância política de um comentário. Eles indicam apenas que o comentário ocupa uma posição estrutural importante na rede de similaridade lexical, seja por compartilhar vocabulário com muitos outros comentários ou por conectar grupos discursivos. Como a construção das arestas depende da quantidade de termos compartilhados, comentários mais extensos ou com vocabulário mais amplo também podem apresentar mais oportunidades de conexão, aspecto que deve ser considerado na interpretação dos rankings.

Conclusão

Este trabalho teve como objetivo avaliar de que maneira dados disponíveis em redes sociais podem contribuir para a análise de aspectos sociais e políticos. Para isso, foram coletados e analisados comentários publicados no YouTube relacionados à guerra entre os Estados Unidos e o Irã, a partir da construção e análise de redes complexas baseadas na similaridade lexical entre os comentários.

A metodologia desenvolvida permitiu caracterizar a estrutura das redes e identificar padrões de organização dos comentários por meio de métricas de centralidade e de detecção de comunidades. Os resultados das métricas de centralidade permitiram identificar comentários que ocupam posições de destaque na estrutura das redes e, a partir de sua classificação, observar os respectivos posicionamentos em relação ao conflito. Dessa forma, os resultados demonstram a possibilidade de utilizar estruturas de redes para auxiliar na identificação de padrões de opinião presentes em dados provenientes de mídias sociais.

A análise exploratória também evidenciou uma quantidade significativa de comentários classificados como neutros ou sem um posicionamento claramente definido. Esse resultado indica que nem todos os comentários apresentam elementos suficientes para a identificação de uma posição em relação ao tema analisado, o que deve ser considerado em análises futuras e na interpretação dos resultados.

Apesar das limitações encontradas na base de dados, especialmente em relação à quantidade e à distribuição dos comentários entre diferentes tipos de vídeos, os objetivos propostos foram alcançados. O trabalho demonstrou que técnicas de mineração de dados e análise de redes complexas podem ser aplicadas conjuntamente para investigar a organização de opiniões expressas em redes sociais.

6.1 Principais Contribuições

As principais contribuições deste trabalho envolvem a construção de uma base de dados sobre os comentários analisados, que poderá ser ampliada e utilizada em trabalhos

futuros, e a aplicação integrada de conceitos de Teoria de Redes Complexas, Análise de Redes Sociais, análise de sentimento e mineração de dados em mídias sociais.

Também foi desenvolvido um repositório¹ contendo os códigos utilizados nas diferentes etapas do trabalho, desde a extração dos dados do YouTube até a construção e análise das redes. A disponibilização desses materiais contribui para a reprodução da metodologia e para o desenvolvimento de novas pesquisas a partir dos dados e procedimentos utilizados.

Os resultados obtidos fornecem evidências de que dados disponíveis em redes sociais podem ser utilizados para investigar padrões de opinião social e política. Entretanto, a capacidade de generalização desses resultados depende da disponibilidade de uma base de dados mais ampla e equilibrada, especialmente em relação à quantidade de comentários e à diversidade dos conteúdos analisados.

O dataset rotulado, bem como todas as redes geradas estão publicamente disponíveis via DOI: <<https://doi.org/10.5281/zenodo.21996246>>.

6.2 Trabalhos Futuros

As limitações identificadas ao longo do desenvolvimento do trabalho abrem possibilidades para novas investigações e aprimoramentos da metodologia proposta. Uma primeira possibilidade consiste em ampliar e equilibrar a quantidade de dados coletados, permitindo testar novamente se o tipo de vídeo, especialmente a diferença entre vídeos convencionais e *shorts*, influencia a estrutura e o comportamento das redes construídas.

Outra possibilidade consiste em aplicar a metodologia a diferentes temas de caráter político e social, permitindo verificar se os padrões observados neste estudo também aparecem em outros contextos e se diferentes temas apresentam níveis distintos de polarização e organização das opiniões.

Também pode ser investigada a ampliação da etapa de classificação dos comentários, buscando reduzir a quantidade de comentários sem posicionamento definido e melhorar a identificação das diferentes categorias de opinião. Da mesma forma, podem ser avaliadas outras técnicas de representação textual e de construção das redes, com o objetivo de obter relações semânticas mais abrangentes do que aquelas baseadas exclusivamente na similaridade lexical.

Além disso, o uso de séries temporais poderia permitir analisar a evolução das redes ao longo do período da guerra, verificando se sua estrutura e os posicionamentos dos comentários se modificam de acordo com os principais eventos e marcos do conflito. Essa abordagem poderia relacionar variações nas métricas de centralidade, na formação de comunidades e na distribuição dos posicionamentos a acontecimentos específicos, possibilitando identificar mudanças no comportamento das discussões ao longo do tempo.

¹ Disponível em: <<https://github.com/LaviniaBDantas/TCC.git>>. Acesso em: 19 ago. 2026.

Por fim, uma extensão natural do trabalho seria o desenvolvimento de um sistema automatizado, com uma interface de usuário, que permitisse submeter uma lista de vídeos ou uma determinada temática para realizar automaticamente a coleta dos comentários, a construção das redes e a apresentação dos principais resultados, como comunidades, métricas de centralidade e distribuição dos posicionamentos.

Referências

- AAZIZ, O. et al. Exploring and quantifying how communication behaviors in proxies relate to real applications. In: **2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)**. [S.l.: s.n.], 2018. p. 12–22. Citado na página 20.
- BARABÁSI, A.-L. **Chapter 2: Networks and Graphs**. 2016. Disponível em: <<https://networksciencebook.com/chapter/2#networks-graphs>>. Citado 3 vezes nas páginas 5, 12 e 18.
- BENEVENUTO, F. Redes sociais online: Técnicas de coleta, abordagens de medição e desafios futuros. In: **Anais dos simpósios SBSC'10, Webmedia'10, IHC'10 e SBBD'10**. [S.l.: s.n.], 2010. Citado 2 vezes nas páginas 19 e 20.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. **Computer Networks and ISDN Systems**, v. 30, n. 1–7, p. 107–117, 1998. Citado na página 37.
- BROWN, T. B. et al. **Language Models are Few-Shot Learners**. 2020. Disponível em: <<https://arxiv.org/abs/2005.14165>>. Citado na página 25.
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. Citado na página 25.
- GILL, P. et al. Characterizing user sessions on youtube. In: **IEEE Multimedia Computing and Networking (MMCN)**. [S.l.: s.n.], 2008. Citado na página 20.
- GIRVAN, M.; NEWMAN, M. E. J. Community structure in social and biological networks. **Proceedings of the National Academy of Sciences**, v. 99, n. 12, p. 7821–7826, 2002. Citado 3 vezes nas páginas 19, 28 e 36.
- JOINSON, A. N. Looking at, looking up or keeping up with people?: motives and use of facebook. In: **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)**. [S.l.]: ACM, 2008. p. 1027–1036. Citado na página 19.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159–174, 1977. Citado na página 25.

- LATAPY, M.; MAGNIEN, C.; VECCHIO, N. D. Basic notions for the analysis of large two-mode networks. **Social Networks**, v. 30, n. 1, p. 31–48, 2008. ISSN 0378-8733. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378873307000494>>. Citado na página 26.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. [S.l.]: Cambridge University Press, 2008. <<https://www-nlp.stanford.edu/IR-book/>>. Acesso em: 6 ago. 2026. Citado na página 27.
- MATSUMOTO, T. et al. Data analysis support by combining data mining and text mining. In: **2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)**. [S.l.: s.n.], 2017. p. 313–318. Citado na página 16.
- NAIK, S. L. et al. Sentimental analysis of tweets using machine learning. In: **2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)**. [S.l.: s.n.], 2024. p. 1–5. Citado na página 20.
- OLIVEIRA, B.; OLIVEIRA, L. Aprendizado de máquina e análise de sentimento em redes sociais: Um estudo de caso usando as eleições presidenciais em 2022. In: **Anais do XX Congresso Latino-Americano de Software Livre e Tecnologias Abertas**. Porto Alegre, RS, Brasil: SBC, 2023. p. 107–112. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/latinoware/article/view/26090>>. Citado 4 vezes nas páginas 7, 11, 16 e 22.
- PACHECO, M. L. et al. A holistic framework for analyzing the COVID-19 vaccine debate. In: CARPUAT, M.; MARNEFFE, M.-C. de; RUIZ, I. V. M. (Ed.). **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. Seattle, United States: Association for Computational Linguistics, 2022. p. 5821–5839. Disponível em: <<https://aclanthology.org/2022.naacl-main.427/>>. Citado na página 11.
- RICCI, R. D. **Análise de Sentimentos no Twitter sobre a Reforma da Previdência no ano de 2019**. 70 p. Trabalho de Conclusão de Curso (Graduação em Estatística) — Universidade Federal de Uberlândia, Uberlândia, 2020. Disponível em: <<https://repositorio.ufu.br/handle/123456789/30900>>. Citado 5 vezes nas páginas 7, 11, 16, 21 e 22.
- ROCHA, S. et al. Monitorando a opinião pública sobre operações policiais no brasil via comentários de vídeos no youtube. In: **Anais do XIII Brazilian Workshop on Social Network Analysis and Mining**. Porto Alegre, RS, Brasil: SBC, 2024. p. 158–171. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/29340>>. Citado na página 11.
- SALTON, G.; WONG, A.; YANG, C. S. A vector space model for automatic indexing. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 18, n. 11, p. 613–620, nov. 1975. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/361219.361220>>. Citado na página 26.
- SARI, D. K. et al. We know who we are: expressing national identity on youtube video comments. **Journal of Computational Social Science**, v. 8, n. 1, p. 85, 2025. Disponível em: <<https://doi.org/10.1007/s42001-025-00426-3>>. Citado 2 vezes nas páginas 7 e 22.

- SCHNEIDER, F. et al. Understanding online social network usage from a network perspective. In: **ACM SIGCOMM Internet Measurement Conference (IMC)**. [S.l.: s.n.], 2009. p. 35–48. Citado na página 20.
- SILVA, J. A. d. **Análise de sentimentos de comentários em redes sociais como ferramenta de apoio à gestão de eventos**. Dissertação (Mestrado) — Universidade Federal de Uberlândia, 2021. Disponível em: <<https://repositorio.ufu.br/handle/123456789/30900>>. Citado na página 11.
- YI, J. et al. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: **Third IEEE International Conference on Data Mining**. [S.l.: s.n.], 2003. p. 427–434. Citado na página 16.
- ZAFARANI, R.; ABBASI, M.; LIU, H. **Social Media Mining: An Introduction**. Cambridge University Press, 2014. Disponível em: <<https://www.socialmediamining.info/book/>>. Citado 2 vezes nas páginas 11 e 15.
- ZAFARANI, R.; ABBASI, M. A.; LIU, H. **Social Media Mining: An Introduction**. Cambridge University Press, 2014. Disponível em: <<https://www.socialmediamining.info/SMM.pdf>>. Citado 2 vezes nas páginas 15 e 17.
- ZINK, M. et al. Watch global, cache local: Youtube network traces at a campus network - measurements and implications. In: **IEEE Multimedia Computing and Networking (MMCN)**. [S.l.: s.n.], 2008. Citado na página 20.