

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Thalia Sara Rodrigues Ribeiro

**Contribuição da computação na análise de
dados de redes sociais no contexto da
pandemia da COVID-19**

Uberlândia, Brasil

2021

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Thalia Sara Rodrigues Ribeiro

Contribuição da computação na análise de dados de redes sociais no contexto da pandemia da COVID-19

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.

Orientadora: Maria Camila Nardini Barioni

Universidade Federal de Uberlândia – UFU

Faculdade de Ciência da Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2021

Thalia Sara Rodrigues Ribeiro

Contribuição da computação na análise de dados de redes sociais no contexto da pandemia da COVID-19

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.

Trabalho aprovado. Uberlândia, Brasil, 26 de outubro de 2021:

Dra. Maria Camila Nardini Barioni
Orientadora

Dra. Christiane Regina Soares Brasil
Professora

Dra. Elaine Ribeiro de Faria
Professora

Uberlândia, Brasil
2021

Dedico este trabalho de pesquisa aos meus pais e a toda minha família, aos professores da FACOM e também aos meus amigos. Aos meus pais e à minha família pelo carinho, afeto, dedicação, cuidado e apoio durante toda a minha vida e jornada acadêmica, acima de todas as dificuldades. Aos meus professores por todo o conhecimento, experiência, incentivo e prática acadêmica, especialmente à minha orientadora Maria Camila que foi o meu norte na realização da monografia. Aos meus amigos pela amizade, companheirismo e ajuda.

Agradecimentos

Agradeço aos meus pais e à minha família, por todo o apoio e ajuda que me deram durante toda minha vida, por me darem acesso à educação e me incentivarem a estudar e a cursar uma faculdade. Sinto-me muito orgulhosa de mim por todo o esforço que coloquei no trabalho e espero que todos que o leiam se sintam da mesma forma.

Agradeço à minha orientadora, Maria Camila, por aceitar a me orientar neste projeto. Sua ajuda e paciência foram essenciais para a aprendizagem de novas metodologias, todos os artigos recomendados para leitura me ajudaram muito a aprender coisas muito interessantes sobre a área da Análise de Dados.

Expresso minha gratidão a todos os professores da FACOM do curso de Sistemas de Informação da Universidade Federal de Uberlândia por todo o apoio que me deram ao longo da realização do curso e bases necessárias para a realização deste trabalho.

“Em meio à dificuldade encontra-se a oportunidade.” - Albert Einstein.

Resumo

As Redes Sociais são plataformas que permitem a conectividade e interação de diferentes pessoas e culturas, gerando milhares de dados a partir dessa interação. A Análise de Dados de Redes Sociais permite que tais dados sejam analisados de forma a obter informações relevantes sobre o que está sendo discutido ou como as pessoas estão se sentindo em um determinado contexto. O trabalho de conclusão de curso apresentado aqui tem como objetivo geral a realização de uma Revisão Sistemática da Literatura sobre a contribuição da computação para a Análise de Dados de Rede Sociais, especialmente no contexto da pandemia da COVID-19. A obtenção de uma visão geral dos métodos e estratégias produzidos pela computação para analisar os dados de redes sociais neste contexto, pode ajudar os pesquisadores da área a lidar da melhor maneira possível em relação a possíveis emergências de saúde pública que possam surgir. A metodologia proposta para esse trabalho consistiu na realização de uma Revisão Mista Sequencial Exploratória que se constituiu de duas fases, Análise Quantitativa e Análise Qualitativa. Foram coletados 784 artigos e após a aplicação dos critérios de inclusão e exclusão, foram selecionados 140 artigos. Analisando os artigos selecionados foi possível identificar as principais técnicas, os métodos e as ferramentas empregados na análise dos dados. Além disso, também foi possível observar quais categorias de métodos foram mais exploradas, como está sendo a distribuição de artigos, quais países contribuíram, como a computação contribuiu para as diferentes áreas da ciência, quais foram as ferramentas mais exploradas e qual foi o período de coleta dos dados dos artigos analisados.

Palavras-chave: Revisão Sistemática, Revisão Mista Sequencial Exploratória, Análise de Dados de Redes Sociais, COVID-19.

Lista de ilustrações

Figura 1 – Revisões Convencionais.	16
Figura 2 – Revisões Mistas.	17
Figura 3 – Exemplo de postagem no <i>Twitter</i>	19
Figura 4 – Exemplo de perfil no <i>Twitter</i>	19
Figura 5 – Exemplo de resultados de uma consulta no <i>Facebook</i>	20
Figura 6 – Exemplo de modelagem de tópicos.	21
Figura 7 – Exemplos para as abordagens da Análise de Sentimentos.	22
Figura 8 – <i>Dashboard</i> do <i>Facebook Business</i>	24
Figura 9 – Processo da Revisão Sistemática.	25
Figura 10 – Categoria de métodos mais utilizadas.	32
Figura 11 – Quantidade de artigos x Curva de casos de COVID-19.	33
Figura 12 – Datas de publicação dos artigos.	34
Figura 13 – Distribuição de conferências.	34
Figura 14 – Distribuição de periódicos.	35
Figura 15 – Países que mais contribuíram com pesquisas.	37
Figura 16 – Visualização gráfica de países que mais contribuíram com pesquisas.	38
Figura 17 – Áreas que contribuíram para as pesquisas.	40
Figura 18 – Redes sociais mais exploradas.	40
Figura 19 – Fontes de dados mais exploradas para o desenvolvimento dos artigos.	41
Figura 20 – Ferramentas mais utilizadas para a extração dos dados.	42
Figura 21 – Linguagens mais utilizadas.	43
Figura 22 – Períodos das coletas de dados.	44
Figura 23 – Processo de Análise de Dados de Redes Sociais.	45
Figura 24 – <i>Dataset</i> AraCOVID19-MFH.	47
Figura 25 – Fases do pré-processamento e processamento.	50
Figura 26 – Tipos de processamento mais utilizados.	51
Figura 27 – Técnicas de análise mais utilizadas.	56
Figura 28 – <i>Word Cloud</i> utilizando o título dos artigos analisados.	63

Lista de tabelas

Tabela 1 – Consultas realizadas em cada base.	28
Tabela 2 – Critérios de inclusão.	29
Tabela 3 – Critérios de exclusão.	29
Tabela 4 – Quantidade de artigos por bases.	30
Tabela 5 – <i>Ranking</i> de periódicos e conferências.	36
Tabela 6 – Quantidade de artigos em cada categoria de motivação.	47
Tabela 7 – Tipos de dados mais utilizados nas pesquisas.	48
Tabela 8 – Tipos de processamento mais utilizados.	51
Tabela 9 – Tabela com as técnicas mais usadas.	56

Lista de abreviaturas e siglas

ACM	<i>Association for Computing Machinery</i>
API	<i>Application Programming Interface</i>
ARS	Análise de Dados de Redes Sociais
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DBSCAN	<i>Density-Based Spatial Clustering of Applications with Noise</i>
IBM	<i>International Business Machines</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
JMIR	<i>Journal of Medical Internet Research</i>
LDA	<i>Latent Dirichlet Allocation</i>
NER	<i>Named-entity Recognition</i>
NLTK	<i>Natural Language Toolkit</i>
OMS	Organização Mundial da Saúde
PLN	Processamento de Linguagem Natural
RNN	<i>Recurrent Neural Network</i>
RSL	Revisão Sistemática da Literatura
SBC	Sociedade Brasileira de Computação
SIR	<i>Susceptible Infected Recovered</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
VADER	<i>Valence Aware Dictionary for Sentiment Reasoning</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

Sumário

1	INTRODUÇÃO	12
1.1	Objetivos Gerais	13
1.2	Objetivos Específicos	13
1.3	Organização do Trabalho	13
2	REFERENCIAL TEÓRICO	15
2.1	Revisão sistemática da literatura	15
2.1.1	Tipos de Revisões Sistemáticas	16
2.1.2	Desenvolvimento da Revisão Sistemática	17
2.2	Análise de Dados de Redes Sociais	18
2.2.1	Tipos de Dados em Redes Sociais	18
2.2.2	Análise de tópicos	19
2.2.3	Análise de Sentimento	21
2.2.4	Análise de Séries Temporais	22
2.2.5	Análise de Rede	24
3	METODOLOGIA	25
3.1	Metodologia da Revisão Sistemática	26
3.2	Bases de Dados	26
3.3	Palavras-chaves e Strings de Busca	27
3.3.1	Palavras-chaves	27
3.3.2	Strings de Busca	27
3.4	Ferramenta para Condução	27
3.5	Critérios de Inclusão e Exclusão	28
3.5.1	Critérios de Inclusão	28
3.5.2	Critérios de Exclusão	29
3.6	Seleção dos Estudos Primários	30
4	ANÁLISE QUANTITATIVA	31
4.1	Questões de Pesquisa	31
4.2	Categorias de Métodos	31
4.3	Distribuição de Artigos	33
4.4	Países e Contribuição	36
4.5	Interdisciplinaridade	38
4.6	Ferramental Técnico	39
4.7	Período de Coleta de Dados	42

5	ANÁLISE QUALITATIVA	45
5.1	Processo de Análise de Dados de Redes Sociais	45
5.2	Determinar a Rede Social	45
5.3	Extração de Dados	46
5.4	Processamento dos dados	50
5.5	Técnicas de Análise	55
5.6	Apresentação dos Resultados	62
6	CONCLUSÃO	64
6.1	Trabalhos Futuros	64
	REFERÊNCIAS	66

1 Introdução

As Redes Sociais estão cada vez mais presentes na vida das pessoas. A facilidade de conectividade e interação de diferentes pessoas e culturas, faz com que o número de usuários cresça diariamente de maneira exponencial e com isso a geração de milhares de dados e informações. Mundialmente, cerca de 4,6 bilhões de pessoas usam a *internet* e as redes sociais ultrapassam 4,2 bilhões de usuários. No Brasil, há 160 milhões de pessoas com acesso à *internet* e 150 milhões de usuários em redes sociais ([We Are Social, Hootsuit, 2021](#)).

Atualmente temos vários tipos de redes sociais, *Facebook*, *Twitter*, *Instagram*, *LinkedIn*, e muitas outras. Cada uma dessas redes têm objetivos diferentes e alguns desses objetivos são descritos a seguir.

1. *Facebook*, *Twitter* e *Instagram*: O objetivo principal é conectar pessoas que buscam relações pessoais. Nessas redes sociais há muita interação e compartilhamento de dados entre usuários, desde acontecimentos do cotidiano até debates políticos ([Facebook, 2004](#); [Twitter, 2006](#); [Instagram, 2010](#));
2. *LinkedIn*: O objetivo principal é conectar pessoas que buscam boas relações e contatos profissionais, procuram aperfeiçoamento em sua carreira e oportunidades de emprego. Nessa rede há um intenso compartilhamento de conhecimento, novas tecnologias e técnicas entre os usuários ([LinkedIn, 2003](#)).

A análise de dados de redes sociais nos permite entender o comportamento, as necessidades, gostos e interesses de determinado público, possibilitando utilizar estratégias eficazes para atender, conquistar ou até mesmo entendê-lo. É possível usar como exemplo as grandes empresas que investem muito na análise de dados das suas redes sociais, desde compartilhamentos feitos até comentários em postagens, para que possam aprimorar a área de *marketing* e produção, para conquistar o sucesso através do *marketing* viral ([Leskovec; Adamic; Huberman, 2007](#)).

Entretanto, a análise de dados de redes sociais não é limitada a esse objetivo. Foi possível presenciar ao longo das últimas décadas algumas crises e a análise de dados foi importante para entendermos certos aspectos comportamentais. Em um contexto de saúde pública e doenças virais, em 2014, tivemos alguns casos suspeitos de ebola no Brasil e a repercussão aconteceu de forma rápida pelo país, principalmente no meio virtual. Analisar os dados de redes sociais na época, ajudou a identificar que os casos de racismo tiveram um aumento significativo ([Danfá, 2015](#)).

No contexto específico da pandemia da COVID-19, analisar tais dados tem uma imensa relevância. É possível obter informações valiosas, por exemplo, como as pessoas estão se sentindo e agindo perante ao isolamento social, qual está sendo o posicionamento político e também verificar se houve aumento de casos relacionados à xenofobia. Ter em mãos as ferramentas produzidas pela computação para analisar os dados de redes sociais neste contexto, ajudará a lidar da melhor maneira possível em relação a possíveis emergências de saúde pública que possam surgir.

1.1 Objetivos Gerais

O objetivo do trabalho descrito aqui consiste em fazer uma revisão sistemática da literatura, levantando os trabalhos de Análise de Dados de Redes Sociais realizados no contexto da pandemia da COVID-19. A revisão sistemática da literatura é uma modalidade de pesquisa que tem o objetivo de reunir estudos semelhantes, avaliá-los acerca de sua metodologia e por fim, reuni-los em uma análise estatística.

A importância de um trabalho como este está relacionada ao fato de evitar a duplicação de pesquisas e permitir o reaproveitamento e a aplicação de pesquisas em diferentes contextos, deixando evidente as limitações de cada artigo analisado, bem como as limitações da própria revisão (Galvão; Ricarte, 2019). Este trabalho é importante, pois possibilitou a identificação das abordagens existentes no contexto da pandemia da COVID-19 para apoio à Análise de Dados de Redes Sociais.

1.2 Objetivos Específicos

Considerando as quatro categorias de métodos de análise de dados de redes sociais, sendo estas, Análise de Tópicos, Análise de Sentimentos, Análise Temporal e Análise de Rede, os objetivos específicos do trabalho consistiram em:

1. Realizar uma análise quantitativa para apresentar as tendências de pesquisa na área.
2. Realizar uma análise qualitativa, ou seja, enumerar as técnicas usadas em cada categoria de métodos e dentro de cada categoria as usadas em cada fase principal específica de cada método. Além disso, o trabalho também pretendeu destacar as técnicas propostas;

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte forma:

1. O Capítulo 2 apresenta o Referencial Teórico utilizado para realizar o trabalho;

2. O Capítulo 3 apresenta a Metodologia proposta para realizar a revisão;
3. O Capítulo 4 apresenta a Análise Quantitativa dos artigos selecionados;
4. O Capítulo 5 apresenta a Análise Qualitativa dos artigos selecionados;
5. O Capítulo 6 apresenta a Conclusão do trabalho.

2 Referencial Teórico

Este capítulo apresenta o referencial teórico do trabalho descrito nessa monografia, discutindo os principais conceitos e tipos de Revisão Sistemática e as técnicas de Análise de dados de Redes Sociais, bem como os trabalhos correlatos encontrados nos últimos anos, abordando os objetivos e metodologias de cada um.

2.1 Revisão sistemática da literatura

A revisão sistemática consiste em uma modalidade de pesquisa que segue protocolos específicos para sua execução. Ela busca entender todos os dados provindos de um conjunto de documentos (estudos primários), verificando o que funciona, ou não, em um contexto. Dessa forma, a revisão sistemática responde a uma questão bem definida sendo caracterizada por ser sua metodologia abrangente, transparente e replicável, diferindo-a de uma revisão informal (Galvão; Ricarte, 2019).

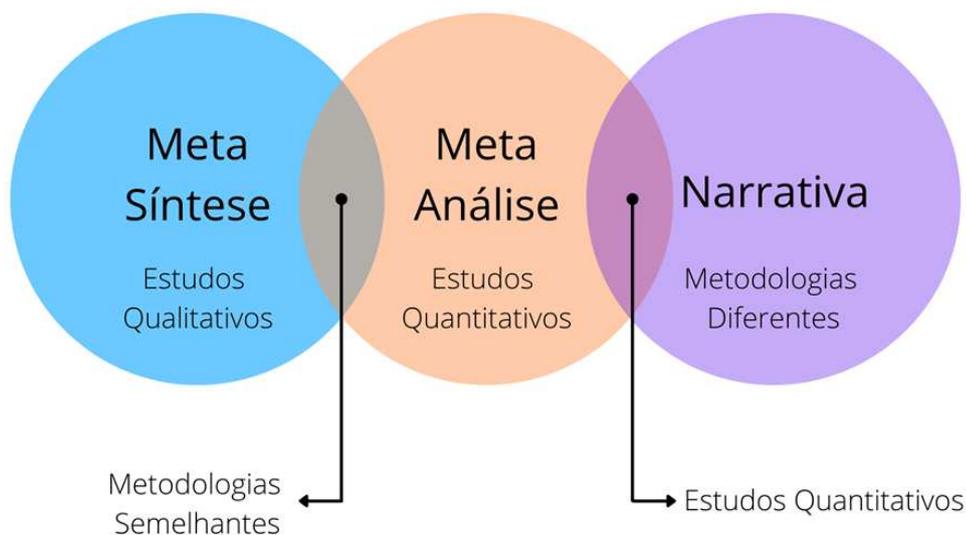
Na área de Ciência da Computação, a revisão sistemática tem sido utilizada e apoiada em diversas subáreas. Ela é uma área ampla e as pesquisas nela exigem um certo grau de interdisciplinaridade (Silva et al., 2015). No trabalho apresentado em referência está sendo abordada a área de Análise de Dados (Mineração de Dados). Nessa área há diversas técnicas e algoritmos aplicados nos dados em questão. A revisão sistemática permite uma síntese dos artigos onde são abordadas essas técnicas e algoritmos, assim é possível analisar quais estão sendo as principais técnicas, quais pontos ainda não foram abordados e como esses pontos podem ser explorados.

Além da área da Análise de dados, é possível usar como exemplo também a área de Engenharia de Software. Essa área exige pesquisas constantes que permitem a evolução de tecnologias voltadas ao *software*. Das pesquisas mencionadas, grande parte é feita de maneira informal e embora pesquisas informais sejam importantes, não é adequada devido à falta de critérios científicos presentes na revisão sistemática. A grande diversidade de material e suas estruturas de apresentação, dificultam o acesso aos conceitos básicos para pessoas interessadas em aprender a fazer revisões sistemáticas na área de Engenharia de Software (ES). Devido a isso, pesquisadores vêm abordando o conceito de revisão sistemática na área de ES, visando que mais pessoas sejam aptas a desenvolver revisões sistemáticas e com isso alavancar a evolução e qualidade das pesquisas desenvolvidas (Scannavino et al., 2017).

2.1.1 Tipos de Revisões Sistemáticas

As revisões sistemáticas podem ser classificadas em revisões sistemáticas com meta-análise, revisões sistemáticas narrativas e revisões sistemáticas com meta-síntese (Siddaway; Wood; Hedges, 2019). Uma revisão com meta-análise é adequada quando se busca relatar resultados quantitativos ou reunir estudos em relação as suas metodologias semelhantes. Ela lida com a estimativa e relato dos resultados dos estudos selecionados. Assim como a revisão com meta-análise, a revisão narrativa tem o objetivo de fazer uma revisão em relação a outros estudos quantitativos, porém onde foram utilizadas diferentes metodologias. Ela pode ser utilizada para formular uma nova teoria baseada nesses trabalhos, por exemplo. Já quando uma revisão necessita da integração de uma pesquisa qualitativa, é utilizada a revisão com meta-síntese. Nessa categoria de revisão o objetivo é sintetizar os estudos qualitativos sobre um assunto fornecendo melhor conceituação a determinado tema ou conceito (Siddaway; Wood; Hedges, 2019). A Figura 1 faz a representação das revisões convencionais.

Figura 1 – Revisões Convencionais.



Fonte: Feito pela autora com o auxílio do *Canva* (Canva, 2012).

Os estudos qualitativos nos fornecem grande quantidade de conceitos, explicações e teorias, enquanto estudos quantitativos nos fornecem muitos dados quantitativos. Muitas vezes, os dados quantitativos têm a necessidade de um complemento, para serem compreendidos da melhor forma possível e diante disso é possível ver a complexidade das pesquisas. Diante disso, surge outra abordagem, onde o foco é a integração das diferentes ciências e suas metodologias (Creswell; Clark, 2017). As revisões mistas podem ser classificadas em revisão mista de convergência, revisão mista de convergência qualitativa, revisão mista sequencial exploratória e revisão mista sequencial explanatória (Galvão; Pluye; Ricarte, 2017).

A revisão mista de convergência quantitativa é apropriada quando há um número grande de amostras, onde os resultados dos estudos qualitativos são convertidos em variáveis quantitativas. Nesse tipo de revisão, é realizada a análise de conteúdo, uma técnica de simplificação que reduz um grande número de dados em um número relativamente pequeno de categorias. Em contrapartida, a revisão mista de convergência qualitativa é utilizada quando os estudos analisados possuem amostras pequenas, onde os resultados dos estudos são transformados em variáveis qualitativas (temas). Nesse tipo de revisão é realizada a análise temática que pode ser usada para encontrar padrões (Galvão; Pluye; Ricarte, 2017).

A revisão mista sequencial exploratória é composta por duas fases, na primeira fase os resultados dos estudos selecionados são transformados em dados qualitativos usando a análise temática mencionada anteriormente e na segunda fase os resultados quantitativos de estudos semelhantes são comparados. Nesse tipo de revisão o objetivo é interpretar o resultado das fases e analisar se há alguma falha ou se falta alguma informação a ser acrescentada. Já a revisão mista sequencial explanatória é utilizada quando se quer medir os efeitos de ações e explicar as diferenças em seus efeitos. Nesse tipo de revisão a integração acontece entre a fase quantitativa e a qualitativa, onde uma fornecerá subsídios para a outra (Galvão; Pluye; Ricarte, 2017). A Figura 2 faz a representação das revisões mistas.

Figura 2 – Revisões Mistas.



Fonte: Feito pela autora com o auxílio do *Canva* (Canva, 2012).

2.1.2 Desenvolvimento da Revisão Sistemática

O desenvolvimento de uma revisão sistemática pode ser classificado em três fases, sendo elas a fase de planejamento, condução e apresentação da revisão (Petersen; Vakka-

lanka; Kuzniarz, 2015). Na fase de planejamento é identificada a necessidade da revisão e desenvolvido um protocolo de revisão, onde são abordados e definidos em profundidade o objetivo da revisão, as fontes de pesquisa, os critérios de inclusão e exclusão, e a sintetização dos dados.

Na fase de condução é feita uma busca exploratória pelos estudos primários, visando encontrar estudos já existentes e avaliar o volume de estudos que sejam interessantes. Nessa etapa, é importante considerar a importância de cada resultado e documentar todas as buscas realizadas. A partir do momento que se tem o resultado, é feita a seleção dos estudos através do título e resumo e é posteriormente feita a análise de qualidade dos estudos, permitindo assim excluir estudos não relevantes.

Na fase de apresentação da revisão é apresentada a síntese dos dados, ou seja, o agrupamento de dados extraídos, o planejamento do trabalho e a escrita da redação do trabalho.

2.2 Análise de Dados de Redes Sociais

As redes sociais têm a vantagem de produzir muitos dados, em alta velocidade e em grande variedade. Entretanto, a desvantagem é que são dados sem credibilidade, pois qualquer pessoa pode postar as informações que circulam nas redes sociais. Um grande exemplo é a disseminação de *fake news*. Perante o exposto, pesquisadores têm explorado métodos de análise para superar esses desafios que as redes sociais proporcionam (Hou; Han; Cai, 2020). Os quatro métodos principais empregados na análise de dados de redes sociais são: Análise de Tópicos, Análise de Sentimento, Análise de Séries Temporais e Análise de Rede.

2.2.1 Tipos de Dados em Redes Sociais

Nas redes sociais é possível compartilhar, curtir e postar diversos conteúdos como textos, *emojicons*, imagens ou vídeos. Além desses dados é possível extrair vários metadados, cada rede social tem seus próprios metadados de postagens e usuários.

No *Twitter* é possível acessar vários tipos de metadados através de sua API, é possível ter acesso a metadados da postagem como data de criação do *tweet*, *retweets*, *id* de quem fez a postagem, entre outros. Além dos metadados da postagem, é possível ter acesso aos metadados do usuário como o nome, data de criação da conta, número de seguidores, entre outros (Twitter, 2021).

A Figura 3 mostra um exemplo de postagem do *Twitter* com os principais tipos de dados e metadados encontrados em postagens. A Figura 4 mostra um exemplo de perfil no *Twitter* com os principais tipos de dados e metadados encontrados em perfis. Em outras

redes sociais as postagens seguem um padrão semelhante com poucas diferenças em suas nomenclaturas ou formatos, um exemplo é o *Facebook* que ao invés de *retweets* é possível encontrar compartilhamentos de postagens.

Figura 3 – Exemplo de postagem no *Twitter*.



Fonte: Captura de tela da autora.

Figura 4 – Exemplo de perfil no *Twitter*.



Fonte: Captura de tela da autora.

2.2.2 Análise de tópicos

As redes sociais se tornaram uma grande fonte de dados e com milhares de informações sendo postadas por seus membros o tempo todo foi percebida a necessidade de

realizar uma síntese dessas informações. A partir desse fato, diversos pesquisadores procuram desenvolver métodos para fazer a síntese da informação. Uma síntese que permita dar significado e coerência à informação. (Hou; Han; Cai, 2020). Algumas das soluções propostas abordadas pelos pesquisadores da área são:

- *Hashtags*: As *hashtags* são geralmente palavras acompanhadas por um símbolo “#” e são utilizadas para assuntos em tendência no momento. No *Twitter* e *Weibo*, por exemplo, há um *ranking* com as *Hashtags* mais mencionadas e utilizadas pelos usuários. Elas podem ser usadas para categorizar as postagens e também disseminar um assunto em questão. Também podem ser utilizadas como ferramenta de identificação e categorização de tópicos.
- Consultas: Toda rede social tem uma ferramenta de busca. Essa ferramenta se trata de um campo que permite que assuntos/temas sejam inseridos e buscados. Por exemplo, é possível extrair tópicos e dados a partir da consulta “vacinação COVID-19 em Uberlândia” no *Facebook*. Com essa pesquisa teremos informações como páginas, grupos e postagens recentes relacionados à Uberlândia. A Figura 5 mostra um exemplo de consulta, é possível ver que nem sempre os resultados retornados correspondem ao solicitado devido à ambiguidade na interpretação dos termos pela própria plataforma. Logo é necessário fazer classificações dessas postagens, por exemplo, classificar de acordo com o local da postagem.

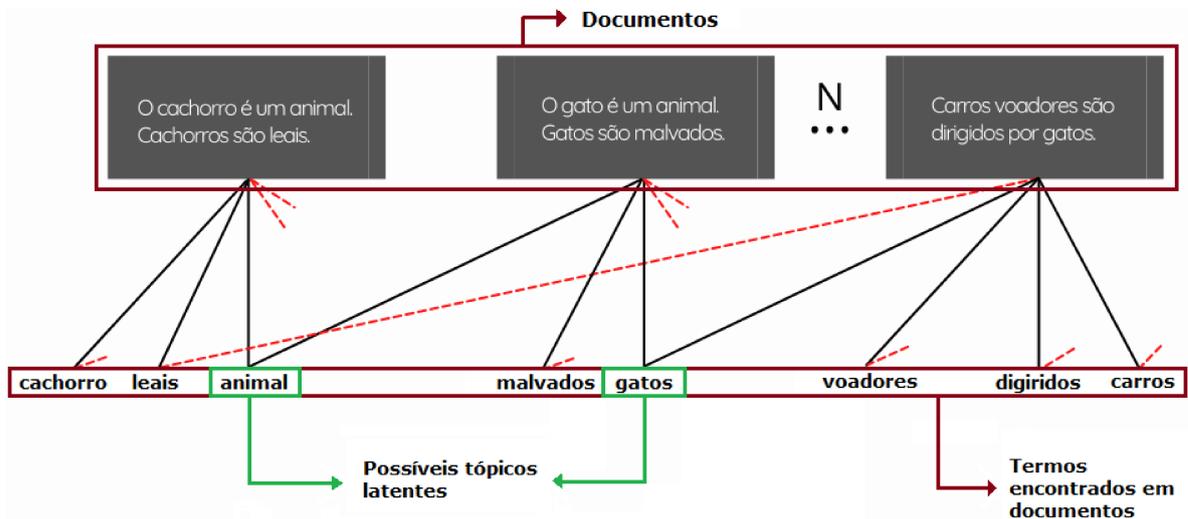
Figura 5 – Exemplo de resultados de uma consulta no *Facebook*.



Fonte: Captura de tela da autora.

- Modelagem de tópicos: A modelagem de tópicos se trata da identificação de tópicos ocultos em um ou mais documentos. Como exemplo temos a Alocação Latente de Dirichlet, uma técnica de modelagem de tópicos que permite classificar o documento e determinar sua relevância (Blei, 2012). Ela é usada como uma ferramenta de mineração para descobrir estruturas semânticas ocultas em determinado texto, permitindo assim descobrir tópicos não definidos em uma coleção de documentos. A Figura 6 mostra um exemplo com documentos fictícios, onde a partir do conteúdo dos documentos são encontrados termos frequentes em todos os documentos, tornando-os assim possíveis tópicos latentes/comuns. Dessa forma é possível definir os tópicos latentes como termos frequentes em todos os documentos e podem ser usados para conectar termos semelhantes entre si.

Figura 6 – Exemplo de modelagem de tópicos.



Fonte: Feito pela autora com o auxílio do *Canva* (Canva, 2012).

2.2.3 Análise de Sentimento

As redes sociais permitem que seus usuários deem opiniões sobre diversos assuntos, desde debate político até produtos cosméticos. A partir disso, a análise de sentimento tornou-se essencial para compreender os sentimentos e emoções a partir dessas opiniões compartilhadas. Ela visa determinar a polaridade do sentimento em positivo, negativo e em alguns casos, neutro. Há duas abordagens principais para a análise de sentimento, abordagem baseada no léxico e na aprendizagem (Hou; Han; Cai, 2020). Essas duas abordagens são descritas a seguir.

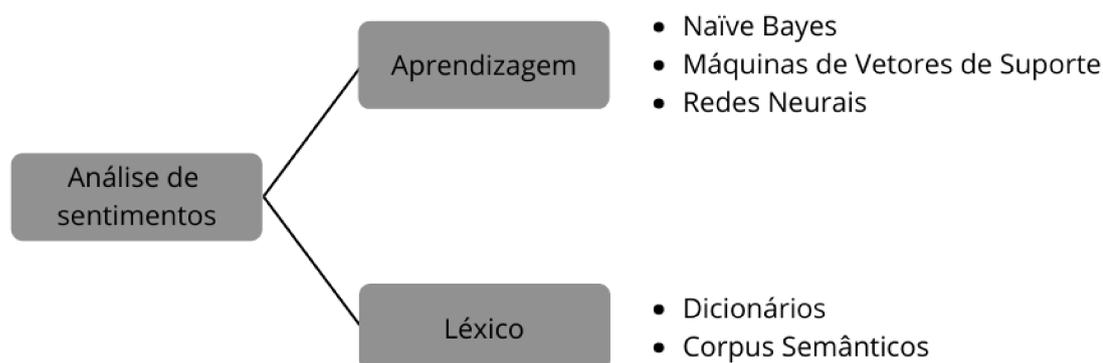
- Abordagem baseada no léxico: A abordagem léxica se caracteriza por utilizar dicionários de sentimentos, um dicionário de sentimentos é um recurso que oferece informações sobre relações semânticas e polaridade das palavras. O *WordNet* é um

dicionário de sentimentos que possui um conjunto de termos agrupados a partir de seu significado semântico (Miller, 1995). Uma das técnicas mais comuns dessa abordagem é a de co-ocorrência entre o termo e o sentimento, onde o termo é o alvo e o sentimento é o que aquele termo expressa. Nessa técnica, não é considerado a ordem ou relações dos termos e para classificar o sentimento é necessário apenas de uma palavra de sentimento, onde sua polaridade será dada pelo dicionário de sentimentos;

- Abordagem baseada na aprendizagem: A abordagem baseada na aprendizagem se tornou muito popular. Ela extrai a polaridade do sentimento baseado na característica do texto de forma automática utilizando um grande conjunto de dados como base. Porém, um dos problemas dessa abordagem está relacionado com a necessidade de dados previamente rotulados, ou seja, quando os dados não possuem uma marca/significado. Quando os dados não são rotulados com antecedência, não há qualidade no desempenho.

A Figura 7 mostra exemplos para as abordagens da Análise de Sentimentos. Na abordagem baseada no léxico, são utilizados dicionários de sentimentos ou *corpus* semânticos. Na abordagem baseada em aprendizagem são utilizados algoritmos de classificação, geralmente são utilizados algoritmos de aprendizagem supervisionado, ou seja, algoritmos voltados para aprendizado de máquina.

Figura 7 – Exemplos para as abordagens da Análise de Sentimentos.



Fonte: Feito pela autora com o auxílio do *Canva* (Canva, 2012).

2.2.4 Análise de Séries Temporais

A análise de séries temporais está relacionada com o estudo da linha temporal de determinados eventos ou acontecimentos interligados entre si. Ela se tornou outra técnica importante a ser aplicada nas redes sociais. Existem três etapas para realizar a

análise de série temporal, sendo elas a seleção de dados e pré-processamento, extração e transformação de dados, por fim, análise de série temporal (Hou; Han; Cai, 2020).

Na primeira etapa os dados são filtrados com base nas pesquisas feitas por usuários, na segunda etapa esses dados são processados de acordo com sua frequência/cronologia e na etapa final, podem ser aplicadas diferentes técnicas para a análise. Dentre essas técnicas estão as descritas a seguir.

- **Análise exploratória:** A análise exploratória é uma abordagem que usa métodos visuais para resumir suas principais características. Essa análise é feita com ajuda de ferramentas que possibilitam a visualização dos resultados obtidos. Um exemplo de ferramenta é o *TweetDeck*. O *TweetDeck* é uma ferramenta que permite o usuário criar várias colunas com base em perfis, tendências e outros critérios definidos pelos usuários do sistema e como resultado são obtidas as possíveis tendências em tempo real (Twitter, 2021);
- **Técnicas de suavização:** A técnica de suavização é utilizada para suavizar sinais irregulares e possibilitar a leitura de um sinal mais claro nas séries temporais. Um exemplo é o modelo Linear de Holt, onde a partir de um gráfico de resultados é possível realizar a suavização a partir de constantes de suavização (Queiroz et al., 2012);
- **Agrupamento:** O agrupamento temporal é feito a partir do particionamento dos dados em grupos com base na semelhança de modo que sigam os padrões de série do grupo. Um exemplo são os métodos particionais, onde os dados são classificados em grupos e cada grupo vai conter um objeto de forma que esse objeto deve pertencer exatamente a um grupo (Rodrigues, 2009);
- **Detecção de pico:** A detecção de pico é útil para identificar picos considerados relevantes para determinado contexto, pois servem como alerta para algum assunto emergente nas redes sociais. Um exemplo de método é o *Z-score*, onde assume-se que um conjunto de dados sem variação siga uma distribuição padrão e caso o dado saia desse padrão ele é considerado um ponto fora da curva. Dessa forma é possível detectar um pico em uma série temporal (Cousineau; Chartier, 2010);
- **Previsão:** A previsão se trata de prever acontecimentos se baseando no histórico de observações de acontecimentos passados, de modo a descobrir possíveis tendências ou assuntos. Um exemplo é o método de Winters onde são usados dados de quantidade, tendência e tempo para gerar previsões (Queiroz et al., 2012).

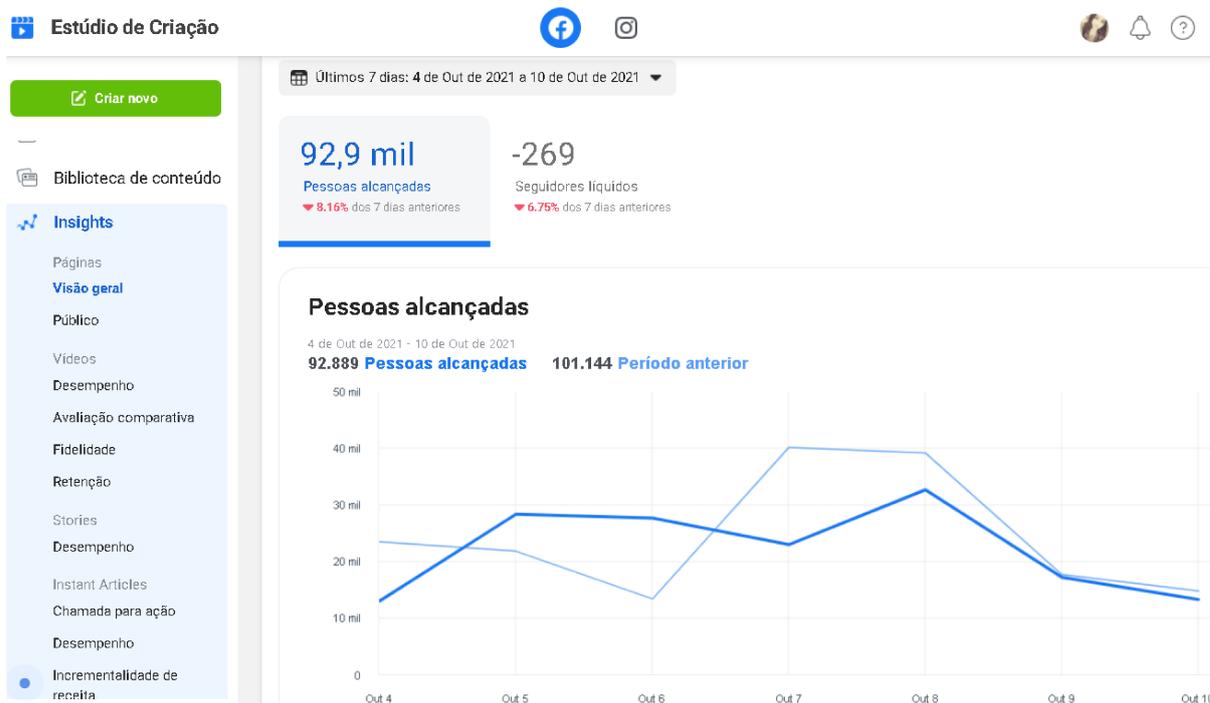
2.2.5 Análise de Rede

Nas redes sociais há diversos meios de se criar vínculos, a criação de grupos, páginas e comunidades permite que os usuários criem um vínculo entre si de acordo com seus gostos e costumes. As formas mais comuns de interação entre esses usuários são feitas através de cinco tipos (Hou; Han; Cai, 2020), são eles: marcar, comentar, compartilhar, reagir e seguir.

O objetivo da análise de rede é compreender como são construídos os relacionamentos da comunidade virtual citada anteriormente, como as informações fluem e quem contribui para isso ser possível. A *dashboard* de uma rede social fornece informações como número de visualizações, número de compartilhamentos e outras informações, a partir dela é possível ver o nível de engajamento e alcance das postagens.

A Figura 8 mostra a *dashboard* do *Facebook Business*. Essa ferramenta é muito utilizada por empresas e influenciadores, já que torna possível a extração de muita informação a respeito de seu público e também informações sobre o desempenho da empresa/influenciador.

Figura 8 – *Dashboard* do *Facebook Business*.



Fonte: Captura de tela da autora.

Através da Análise de Rede é possível utilizar algoritmos computacionais que possibilitam a identificação de possíveis influenciadores. Um exemplo é a utilização de redes complexas, onde são utilizados identificadores para a postagem, postagem que o membro respondeu e quantidade de respostas daquela postagem. Dessa forma é possível construir um grafo e identificar os possíveis influenciadores (Silva; Machado; Maciel, 2020).

3 Metodologia

A metodologia utilizada na realização desse trabalho consistiu em uma revisão sistemática da literatura. Diferente de uma revisão informal a revisão sistemática necessita de um protocolo de execução, assim foram seguidas três etapas: planejamento, condução da revisão e documentação.

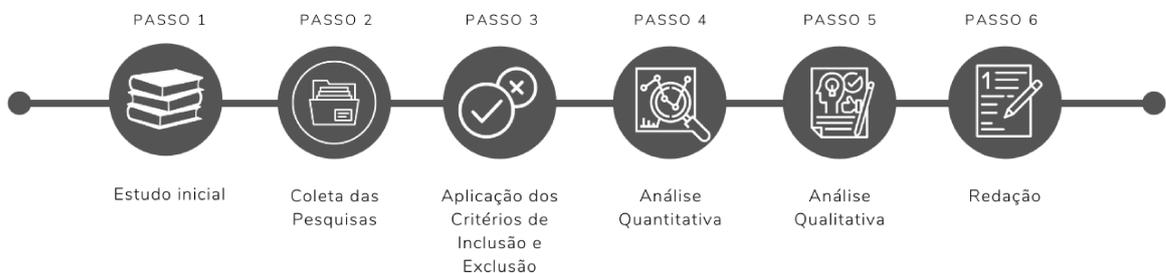
O planejamento é a etapa inicial, foi feito um estudo geral acerca das metodologias de Revisão Sistemática e da Análise de Redes Sociais. Nessa etapa, além do estudo inicial, foram definidas as questões da pesquisa e o desenvolvimento do protocolo, assim como a validação do mesmo.

Na etapa da condução o protocolo definido foi colocado em prática, assim puderam ser realizadas as buscas e seleção dos estudos primários. Os estudos primários selecionados tiveram sua qualidade e quantidade avaliadas a partir dos critérios de inclusão e exclusão definidos. Dessa forma os dados puderam ser extraídos e sintetizados, permitindo a extração das respostas para as questões de pesquisa e identificação dos pontos em aberto.

Na etapa de documentação foi feita a redação final do trabalho, documentando as *strings* de busca e ferramentas utilizadas, os critérios de qualidade envolvidos e a viabilidade de cada ponto em aberto.

A Figura 9 mostra o processo geral da revisão sistemática abordada neste trabalho. O passo 1 é a etapa de planejamento, os passos 2 a 5 fazem parte da etapa de condução e o passo 6 é a etapa de documentação.

Figura 9 – Processo da Revisão Sistemática.



Fonte: Gráfico elaborado pela autora através do auxílio do Canva ([Canva, 2012](#)).

3.1 Metodologia da Revisão Sistemática

No trabalho descrito aqui, a metodologia de revisão sistemática utilizada foi a revisão mista sequencial exploratória. A primeira fase consistiu em realizar uma análise quantitativa dos estudos selecionados, seguida por uma fase de análise qualitativa. Os resultados quantitativos e qualitativos foram utilizados para realizar interpretações acerca das tendências de pesquisa na área e também das técnicas propostas nos estudos coletados.

3.2 Bases de Dados

Nessa seção, serão apresentadas as bases de dados utilizadas na coleta dos estudos primários. As bases de dados são uma importante fonte de pesquisa, principalmente na realização de uma revisão sistemática. Elas reúnem e organizam os artigos e permitem que sejam realizadas consultas com diversas palavras-chaves, operadores e filtros (Graziosi; Liebano; Nahas, 2012). Para o trabalho descrito aqui, foram consultadas oito bases de dados, são elas:

- *Science Direct*: Biblioteca digital onde podem ser encontrados artigos de diversas áreas: Ciências Físicas e Engenharia, Ciências Biológicas, Ciências da Saúde e Ciências Sociais e Humanas (ScienceDirect, 1997).
- *Springer*: Editora que publica e disponibiliza artigos online das áreas de Ciências, Tecnologia, Matemática e Medicina (Springer, 1996).
- ACM: Base de dados que inclui periódicos científicos, revistas e anais de conferências e congressos com artigos de pesquisa nas áreas de Ciência da Computação, Engenharia de Computação, Tecnologia da Informação (ACM, 2020).
- *Scielo*: Portal eletrônico cooperativo de periódicos científicos da América Latina e também de diversas partes do mundo, como Espanha, África do Sul, entre outros (SciELO, 2002).
- IEEE: Banco de pesquisa de artigos de periódicos, anais de conferências, normas técnicas e materiais relacionados a Ciência da Computação, Engenharia Elétrica e Eletrônica e campos afins (IEEE, 2000).
- *JMIR Publications*: Portal eletrônico que organiza conferências e publica artigos online de forma gratuita, está relacionada a área médica mas abrange diversas outras áreas, como Ciência, Tecnologia, entre outras (JMIR, 1999).
- *SBC OpenLib*: Biblioteca digital mantida pela Sociedade Brasileira de Computação, SBC, que é composta por anais de eventos, revistas e livros (SBC, 2019).

- *Google Acadêmico*: Mecanismo de pesquisa que organiza e lista textos completos ou metadados da literatura acadêmica (Google, 2004).

3.3 Palavras-chaves e Strings de Busca

Nessa seção, serão apresentadas as *strings* de busca utilizadas para a obtenção dos artigos nas bases de dados. Cada base de dados oferece uma forma única de se fazer uma consulta, algumas aceitam palavras-chaves, operadores e filtros, enquanto outras aceitam apenas palavras-chaves.

3.3.1 Palavras-chaves

As palavras-chaves utilizadas para a criação das strings de busca/consulta por termos foram: *data analysis*, *data mining*, *social media*, *covid-19 pandemic*, *social networks*, análise de dados, mineração de dados, covid, análise de redes sociais e pandemia.

3.3.2 Strings de Busca

As *strings* de busca foram definidas a partir das palavras-chaves de modo à obter a máxima quantidade de artigos relacionados a análise de dados de redes sociais no contexto da pandemia COVID-19. A *string* principal foi definida da seguinte forma para obter artigos em inglês: (*Data Analysis* OR *Data Mining*) AND *Social Media* AND *Covid-19 pandemic*. Para obter os artigos em português, a string principal foi definida da seguinte forma: (Análise de Dados OR Mineração de Dados) AND Redes Sociais AND (Covid-19 OR Pandemia).

Devido a particularidade de cada base de dados, as consultas foram montadas de acordo com o escopo de cada uma, visando a obtenção mais precisa dos artigos. As consultas definidas para cada base são apresentadas na Tabela 1.

3.4 Ferramenta para Condução

Para conduzir a revisão proposta, foi utilizada a ferramenta *Excel* do *Microsoft Office*. Ela foi necessária para aplicar os critérios de inclusão e exclusão apresentados na seção 3.5. Através de recursos como a formatação condicional, por exemplo, foi possível fazer a identificação e remoção de artigos duplicados.

Tabela 1 – Consultas realizadas em cada base.

Base	Consulta	Filtros
<i>Science Direct</i>	<i>data analysis, data mining, social media, covid-19 pandemic</i>	Data
<i>Springer</i>	<i>(Data Analysis OR Data Mining) AND Social Media AND Covid-19 pandemic</i>	Tipo de documento e Área
ACM	<i>[[Abstract: data analysis] OR [Abstract: data mining]] AND [Abstract: social media] AND [Abstract: covid] AND [Publication Date: (01/01/2020 to 12/31/2021)]</i>	Data
<i>Scielo</i>	<i>((Análise de dados OR Mineração de dados) AND Social Media AND Covid) OR ((Análise de Redes Sociais OR Análise de Dados de Redes Sociais) AND (Pandemia OR Covid))</i>	NDA
IEEE	<i>(Data Analysis OR Data Mining) AND Social Media AND Covid-19 pandemic</i>	Data
<i>JMIR Publications</i>	<i>(Data Analysis OR Data Mining) AND Social Media AND Covid-19 pandemic</i>	NDA
<i>SBC OpenLib</i>	Utilizando a ferramenta de busca em página (F3), foram consultados os termos <i>social networks, social media, redes sociais, covid-19</i> nos seguintes periódicos/conferências/workshops: <i>Kdmile, Sbbd, Webmedia, Brasnam, Eniac e WAIHCWS</i>	NDA
<i>Google Acadêmico</i>	<i>((Análise de Dados OR Mineração de Dados) AND Redes Sociais AND (Covid-19 OR Pandemia))</i>	Data
	<i>((Data Analysis OR Data Mining) AND Social Media AND Covid-19 Pandemic)</i>	Data
Legendas para os filtros: Data: Artigos entre 2020 e 2021. Área: Artigos relacionados a computação. Tipo de documento: Artigos. NDA: Nenhum filtro aplicado.		

3.5 Critérios de Inclusão e Exclusão

Nessa seção serão apresentados os critérios de inclusão e exclusão para os artigos coletados, dessa forma foi possível filtrar os artigos de forma que o conjunto resultante fosse de artigos necessários para a revisão.

3.5.1 Critérios de Inclusão

A Tabela 2 contém todos os critérios utilizados para incluir os estudos primários na revisão.

Tabela 2 – Critérios de inclusão.

Índice do critério	Descrição
C1	Foram incluídos artigos nacionais e internacionais
C2	Foram incluídos artigos escritos em português e em inglês
C3	Foram incluídos artigos que foram disponibilizados online de forma gratuita
C4	Foram incluídos artigos relacionados exclusivamente à análise de dados de redes sociais
C5	Foram incluídos artigos no contexto da pandemia COVID-19

3.5.2 Critérios de Exclusão

A Tabela 3 contém todos os critérios utilizados para excluir os estudos primários da revisão.

Tabela 3 – Critérios de exclusão.

Índice do critério	Descrição
C1	Foram excluídos artigos não revisados por pares
C2	Foram excluídos artigos pagos
C3	Foram excluídos artigos que fizeram quaisquer outros tipos de análise a não ser de redes sociais
C4	Foram excluídos artigos que não estão no contexto da pandemia COVID-19
C5	Foram excluídos artigos que não foram escritos em português ou inglês
C6	Foram excluídos artigos que não foram escritos entre 2020 e 2021
C7	Foram excluídos artigos sem abstract

3.6 Seleção dos Estudos Primários

Nesta seção, serão apresentados os dados quantitativos referentes à coleta e filtragem dos estudos primários. Entre 14/07/2021 e 17/08/2021, foram coletados 784 artigos nas bases de dados selecionadas. A tabela 4 mostra a quantidade de artigos coletados em cada base.

Tabela 4 – Quantidade de artigos por bases.

Base	Quantidade de artigos
Science Direct	104
Springer	149
ACM	113
Scielo	17
IEEE	73
JMIR <i>Publications</i>	100
SBC <i>OpenLib</i>	29
Google Acadêmico	199

Após a coleta, foi feita uma leitura dos resumos e *abstracts* e a partir disso foi possível realizar a filtragem seguindo os critérios de inclusão e exclusão apresentados nas Seções 3.5.1 e 3.5.2. Sendo assim, 140 artigos foram selecionados como aptos à revisão.

4 Análise Quantitativa

Este capítulo apresenta as questões de pesquisa definidas e a análise quantitativa realizada considerando essas questões. Serão apresentados os gráficos gerados a partir da coleta de informações dos artigos selecionados e uma análise para cada questão de pesquisa.

4.1 Questões de Pesquisa

Para realizar a fase de análise quantitativa da revisão a partir dos estudos primários selecionados, foi importante definir as questões de pesquisa para melhor compreensão de como esses estudos estão distribuídos, como se relacionam, o que foi explorado, entre outros. A seguir, temos as questões definidas para este trabalho:

1. Quais categorias de métodos de análise de dados de redes sociais foram mais exploradas no contexto da pandemia da Covid-19?
2. Como a distribuição de artigos ao longo do curso da pandemia se parece?
3. É possível identificar quais os países que mais contribuíram com pesquisas nesse contexto?
4. Como diferentes áreas da ciência contribuíram para essas pesquisas? Existe interdisciplinaridade?
5. Quais foram as linguagens e ferramentas mais exploradas no desenvolvimento de cada trabalho? É possível fazer alguma observação em relação às redes sociais?
6. Qual foi o período de coleta dos dados dos artigos?

As próximas seções apresentam uma análise para cada uma das questões considerando os dados coletados dos 140 artigos selecionados. Esses dados incluem informações como periódicos/conferências em que os artigos foram publicados, métodos e ferramentas utilizadas, áreas relacionadas, entre outros.

4.2 Categorias de Métodos

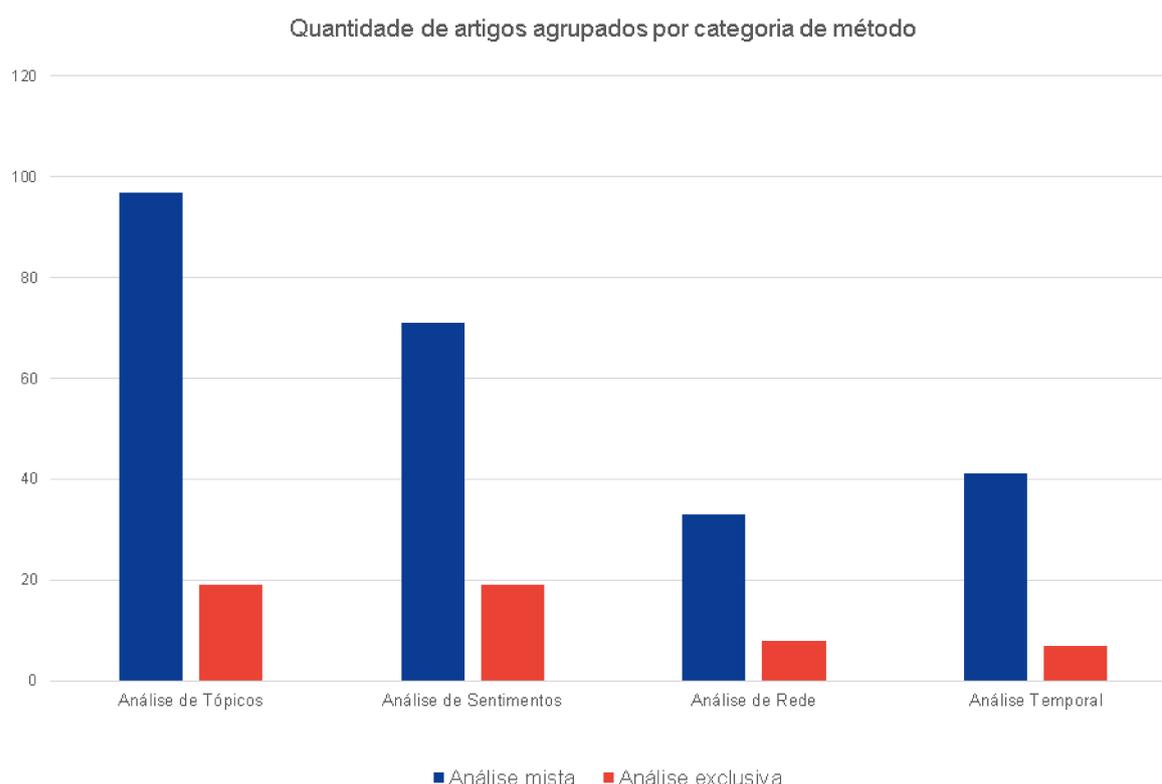
As categorias de métodos de ARS foram apresentadas na Seção 2.2 do Capítulo 2 deste trabalho, sendo elas Análise de Sentimentos, Análise de Rede, Análise Temporal e Análise de Tópicos. A partir da leitura dos artigos, foi possível detectar as categorias

de métodos utilizados em cada um deles e com isso obter variáveis quantitativas que permitiram determinar quais foram as categorias de métodos mais exploradas.

No contexto da pandemia da COVID-19, as Redes Sociais foram exploradas pelos pesquisadores com diversas finalidades. Foi possível perceber o grande foco em analisar os textos coletados das redes sociais a fim de extrair informações como assuntos mais comentados, tendências de discussões, categorizar os textos com um termo específico de escolha dos pesquisadores, entre outros.

Além disso, foi possível identificar um grande número de artigos em que os pesquisadores se preocuparam em fazer a análise de sentimentos a partir dos textos coletados das redes sociais. Estes tinham como objetivo conseguir identificar, por exemplo, como a população se sentiu em relação ao isolamento social ou também como a pandemia e o isolamento afetou o comportamento das pessoas e interação social.

Figura 10 – Categoria de métodos mais utilizadas.



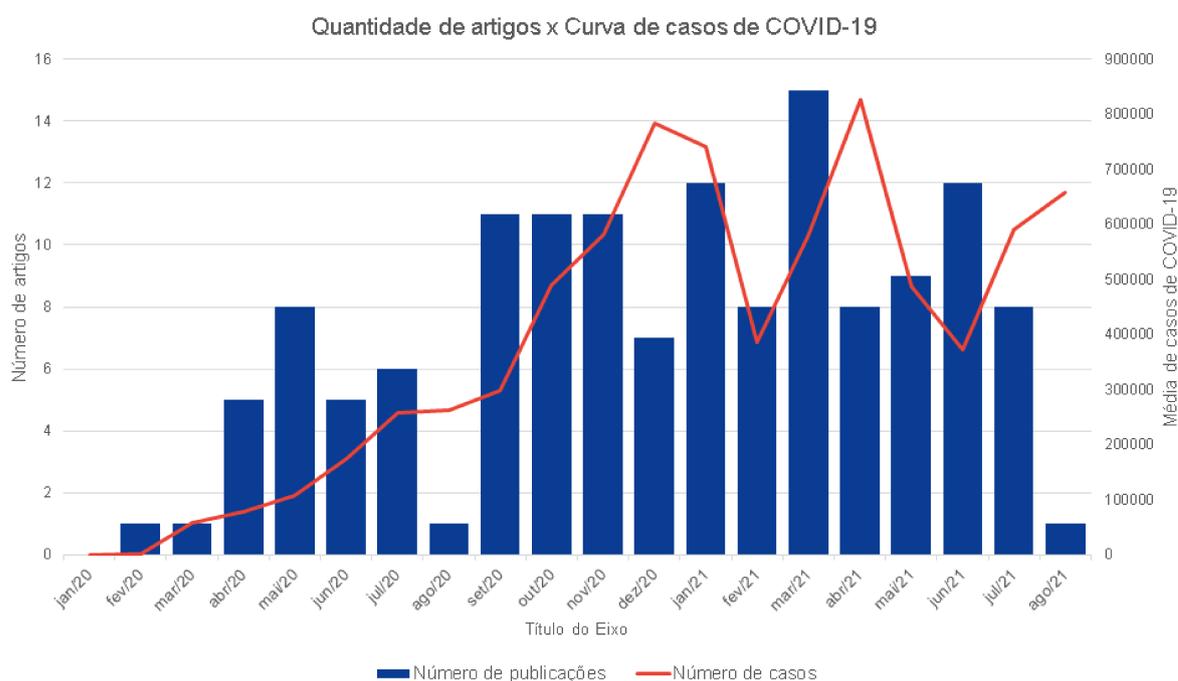
Fonte: Gráficos elaborado pela autora através do Excel.

Na Figura 10, é perceptível que a categoria de método mais utilizada foi a Análise de Tópicos, além disso é possível perceber o grande destaque na utilização da Análise de Sentimentos para a Análise de Dados de Redes Sociais. Na metodologia e desenvolvimento dos artigos selecionados foi possível perceber que os pesquisadores utilizaram mais de uma categoria de método, então é comum, por exemplo, os pesquisadores misturarem Análise de Sentimentos com Análise de Tópicos, Análise Temporal com Análise de Tópicos, entre outros.

4.3 Distribuição de Artigos

Ao longo do curso da pandemia da COVID-19 inúmeras publicações de artigos relacionados com a análise de dados de redes sociais no contexto da pandemia foram feitas em periódicos e conferências. Analisando os trabalhos selecionados, é possível perceber que o objetivo geral dos pesquisadores foi contribuir para a Ciência, apresentando as técnicas utilizadas, as descobertas feitas e os resultados de suas pesquisas tendo como base o contexto da pandemia.

Figura 11 – Quantidade de artigos x Curva de casos de COVID-19.

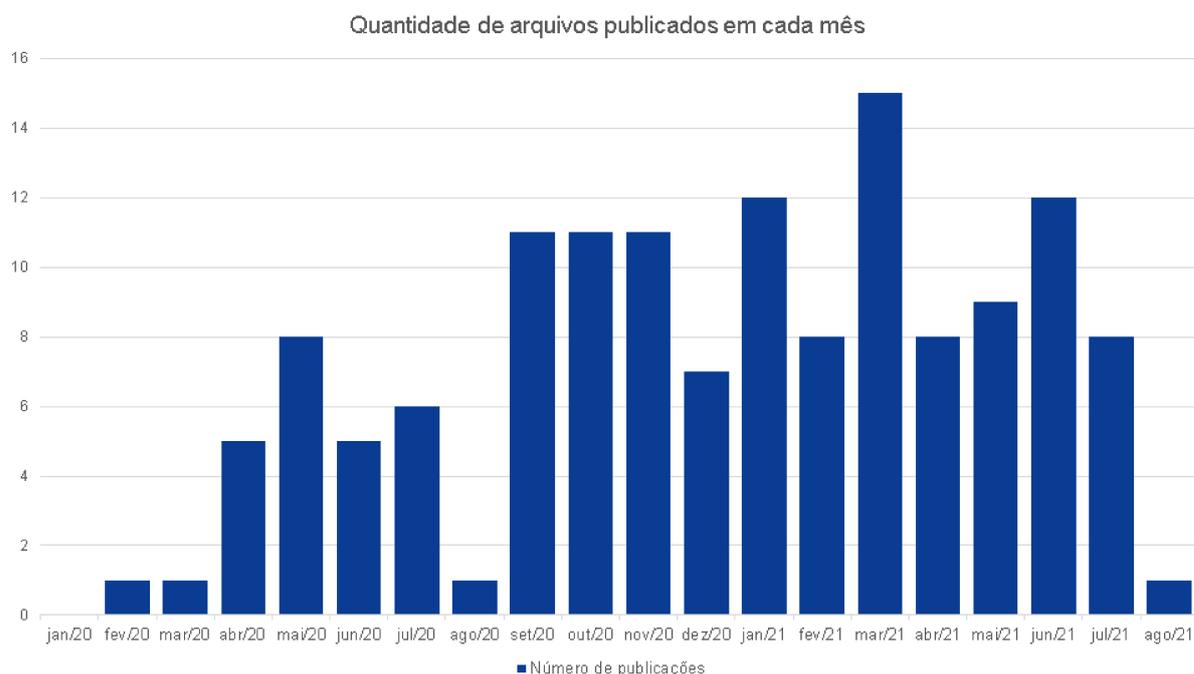


Fonte: Gráfico elaborado pela autora através do *Excel*.

A Figura 11 apresenta o gráfico com o número de artigos publicados em cada mês, e a curva do número médio de casos mundiais de COVID-19 ao longo do tempo. Desde os primeiros rumores em relação a contaminação por COVID-19, é possível observar o aumento de publicações de artigos relacionados à ARS desenvolvidos justamente nesse contexto. Os dados quantitativos para fazer a curva de casos da Figura 11 foram retirados do rastreador de casos do *The New York Times* (NYT, 2020).

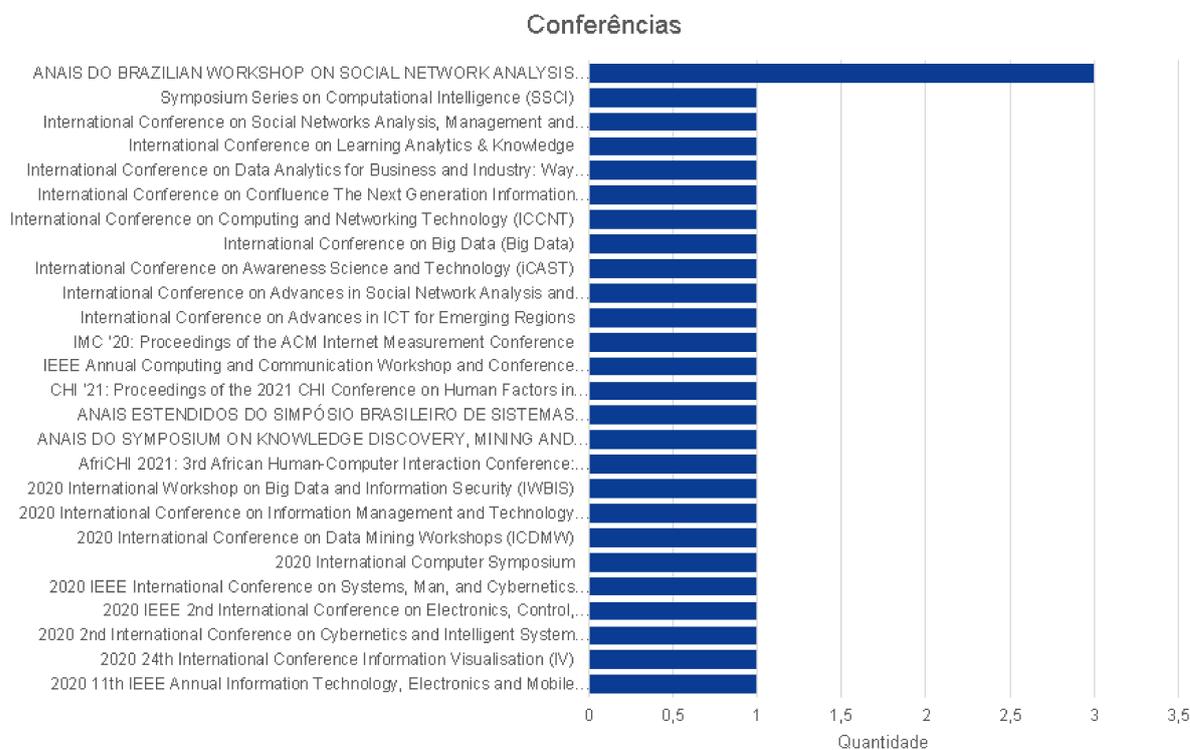
Na Figura 12 os picos podem indicar datas onde houveram conferências, por exemplo; também podem indicar um mês onde foi feito algum comunicado oficial da OMS ou atualização acerca da pandemia da COVID-19. Podem ser citados como exemplos a declaração oficial da pandemia por COVID-19 pela OMS em 11 de março de 2020 e o início da comercialização de vacinas no final de 2020 (G1, 2020; BBC, 2020).

Figura 12 – Datas de publicação dos artigos.



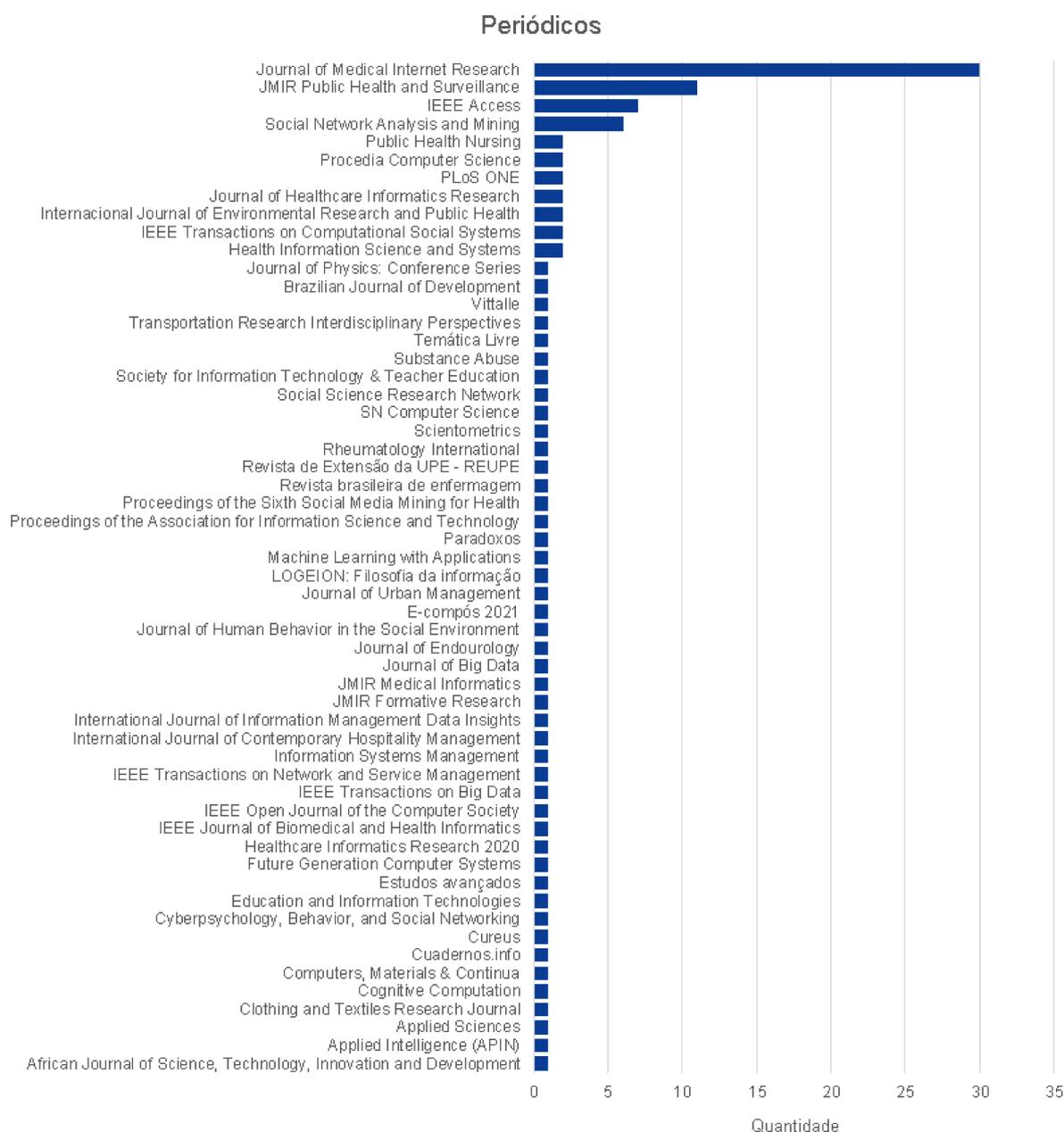
Fonte: Gráfico elaborado pela autora através do *Excel*.

Figura 13 – Distribuição de conferências.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

Figura 14 – Distribuição de periódicos.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

Nas Figuras 14 e 13, é possível ter uma visão geral de quais periódicos e conferências contiveram os artigos selecionados e quais tiveram mais artigos relacionados publicados, sendo 113 artigos de periódicos e 28 de conferências. A revista *Journal Medical Internet Research* foi a que mais conteve artigos selecionados, juntamente com a *JMIR Public Health and Surveillance* e *IEEE Access*. A Tabela 5 mostra um *ranking* dos artigos usando como base o *Qualis* na área de ciência de computação dos periódicos e conferências encontradas na plataforma da PPGCC/PUCRS (Programa de Pós-graduação em Ciência da Computação da PUCRS) (PUCRS, 2021).

O *Qualis* é importante para avaliar a produção científica desenvolvida pelos pro-

gramas de pós-graduação brasileiros, dessa forma as conferências e periódicos bem classificados ganham credibilidade e visibilidade entre os pesquisadores (Silva, 2009). Diversos critérios são definidos para uma boa avaliação, como a qualidade dos artigos das revistas e periódicos, origens dos trabalhos, periodicidade, entre outros. Revistas com baixo fator de impacto e que possuem artigos com poucas citações, tendem a não serem classificadas.

Tabela 5 – *Ranking* de periódicos e conferências.

Classificação	Quantidade de artigos
A1	6
A2	31
A4	1
B1	2
B2	1
B3	3
Sem classificação	96

A *Journal Medical Internet Research*, onde houveram mais artigos selecionados. É uma revista da *JMIR Publications*, ela tem o objetivo de publicar e disponibilizar gratuitamente artigos relacionados a medicina e saúde digital, levando em consideração a era tecnológica. Por ter esse objetivo, ela publicou muitos artigos relacionados a Análise de dados de Redes Sociais no contexto da pandemia da COVID-19 e a tendência é que ela continue publicando regularmente artigos.

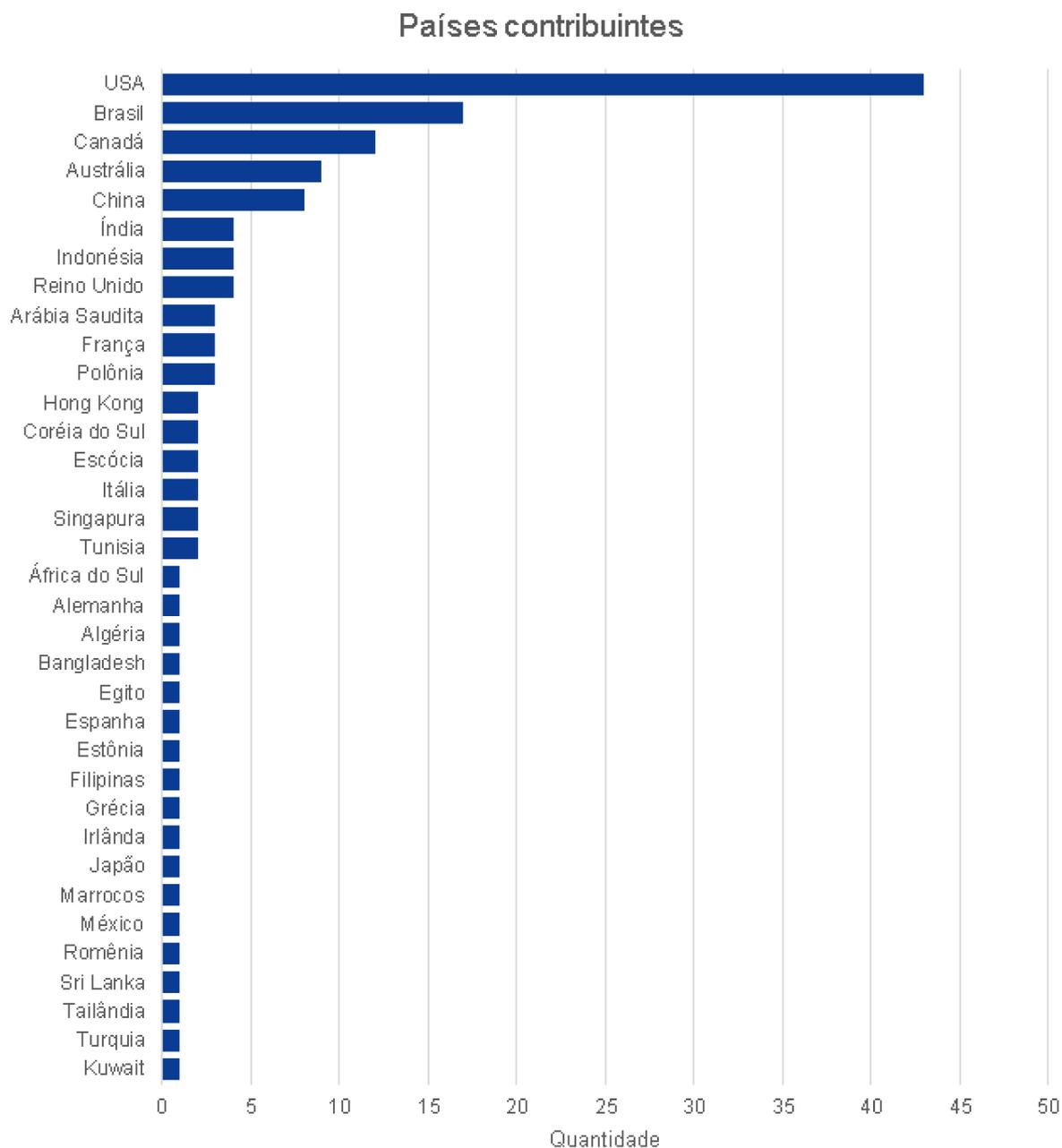
4.4 Países e Contribuição

A contribuição de diferentes países para a Ciência é muito importante, uma vez que são abordados diferentes pontos de vista, objetivos e realidades. A partir dos artigos selecionados, foi possível identificar a contribuição de diversos países ao longo da pandemia da COVID-19.

Para identificar a nacionalidade desses artigos, foi considerada a filiação mais recente da maioria dos autores, ou seja, a instituição de pesquisa a qual eles pertencem. Na Figura 15, é possível ver que houve uma grande diversidade de países que contribuíram

com pesquisas relacionadas a Análise de Dados de Redes Sociais no contexto da pandemia da COVID-19, países da América, África, Ásia e vários outros continentes.

Figura 15 – Países que mais contribuíram com pesquisas.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

Os Estados Unidos foi o país que mais colaborou com pesquisas, mas houve destaque de outros países como Brasil, Canadá, Austrália e China. Além desses destaques houveram contribuições de vários continentes, mostrando que em vários lugares do mundo há pessoas preocupadas, que procuram soluções e que têm o desejo de contribuir com seu conhecimento para que no futuro situações de emergência como essas sejam tratadas de uma forma eficaz e que possam ser resolvidas antes de se tornar uma pandemia.

Figura 16 – Visualização gráfica de países que mais contribuíram com pesquisas.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

A Figura 16, contém uma visualização gráfica dos países que mais contribuíram com pesquisas, em que permite que se tenha uma melhor percepção de como as pesquisas foram desenvolvidas ao redor do mundo. Em um tom mais escuro é possível ver os países que mais contribuíram, em um tom médio são os países que contribuíram com um número razoável e em um tom mais claro, são os países que contribuíram com um número menor de artigos.

4.5 Interdisciplinaridade

A interdisciplinaridade é um conceito que busca relacionar diferentes áreas do conhecimento, permitindo assim uma visão ampla e alto entendimento de um determinado tema ou assunto como um todo (Fazenda et al., 2008). Os artigos selecionados para a revisão sistemática mostraram a contribuição da área de computação em diversas áreas do conhecimento, de forma que foi possível fazer a identificação dos problemas relacionados ao comportamento do ser humano frente ao isolamento ou também identificação de sintomas através de *tweets*.

No trabalho descrito aqui foram definidos seis campos de conhecimento para possibilitar a categorização dos artigos em áreas baseando-se na tabela de áreas do CNPq (CNPq; CAPES; FINEP, 2005). Sendo elas Segurança e Política, Saúde e Psicologia, Questões Sociais, Informação e Comunicação, Economia e Educação.

- Saúde e Psicologia: A computação contribuiu para esse campo buscando abordar questões de saúde física e mental. Foram colocados em pauta assuntos como mu-

dança de comportamento e estresse psicológico em relação ao isolamento social e a pandemia de forma geral, uso das redes sociais para a vigilância em saúde, detectar sintomas através de *tweets*, debates em relação ao uso de máscara e recomendações médicas, entre outros.

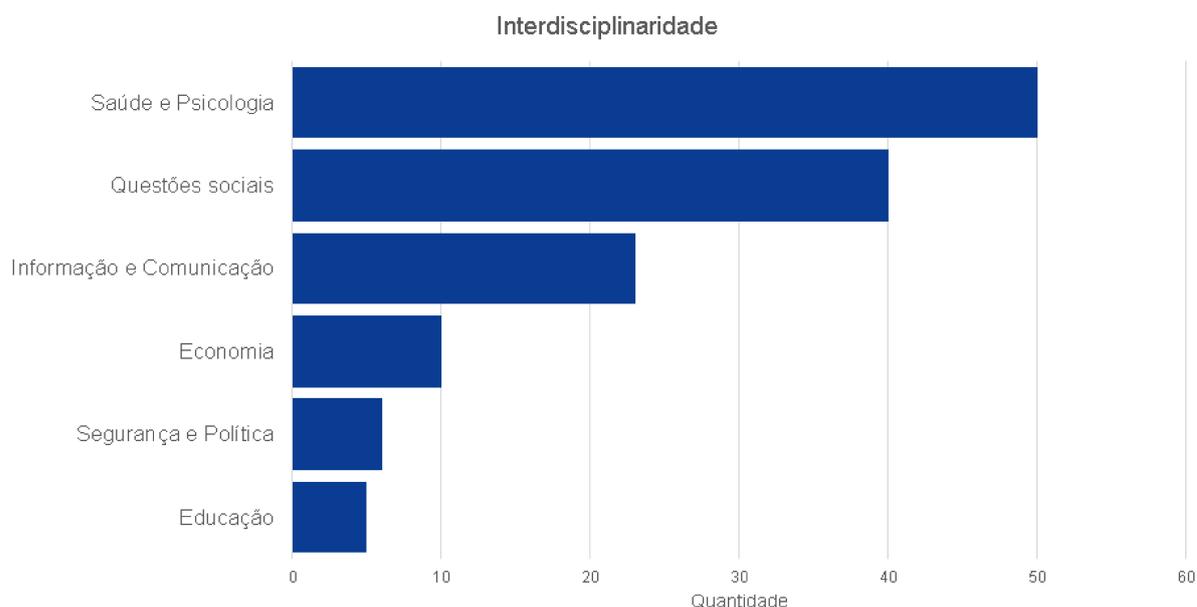
- **Questões Sociais:** A computação contribuiu para esse campo abordando assuntos relacionados às questões da sociedade. Então, foram trabalhados assuntos como aumento do uso de *internet* durante a pandemia, interação virtual, impacto da pandemia na população, quais são os assuntos mais discutidos em redes sociais, entre outros.
- **Informação e Comunicação:** A computação contribuiu para esse campo apontando os problemas relacionados a desinformação. Alguns dos temas mais abordados foram *fake news*, redes sociais como uma transportadora de informações, questões relacionadas ao *marketing*, imagem de um país durante a pandemia, entre outros.
- **Economia:** A computação contribuiu para esse campo trabalhando assuntos relacionados ao impacto econômico da pandemia. Foram trabalhados assuntos relacionados com o mercado de trabalho, adequação ao *Home Office*, comportamento de compra durante a pandemia, entre outros.
- **Segurança e Política:** A computação contribuiu para assuntos relacionados a segurança e política. Foram trabalhados assuntos como tomadas de decisão para prevenir contaminação da população, melhores maneiras de um governo gerenciar situações de emergência, violência doméstica, entre outros.
- **Educação:** A computação contribuiu para esse campo com assuntos relacionados com a educação. Então foram trabalhados assuntos como educação em saúde utilizando redes sociais, aprendizado *online* durante a pandemia, interação de alunos e instituições de ensino durante a pandemia, entre outros.

Na Figura 17, é possível ver que as áreas que mais colaboraram foram Saúde e Psicologia, juntamente com área de Questões Sociais. Isso indica que grande parte dos pesquisadores tiveram como preocupações principais a vigilância em saúde e as questões sociais de um país através da Análise de Dados de Redes Sociais.

4.6 Ferramental Técnico

Nesta seção serão apresentadas as Redes Sociais mais exploradas, as Fontes de Dados utilizadas, as Ferramentas de Extração utilizadas para extrair os dados e as Linguagens de Programação mais utilizadas para o processamentos dos dados.

Figura 17 – Áreas que contribuíram para as pesquisas.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

Os pesquisadores utilizaram diferentes Redes Sociais para desenvolverem suas pesquisas, foram elas: *Twitter*, *Facebook*, *Reddit*, *Instagram* e *Weibo*. Na Figura 18 é possível ver que o *Twitter* foi a rede social mais explorada pelos pesquisadores, outras redes como *Weibo* e *Facebook* também foram bem exploradas.

Figura 18 – Redes sociais mais exploradas.

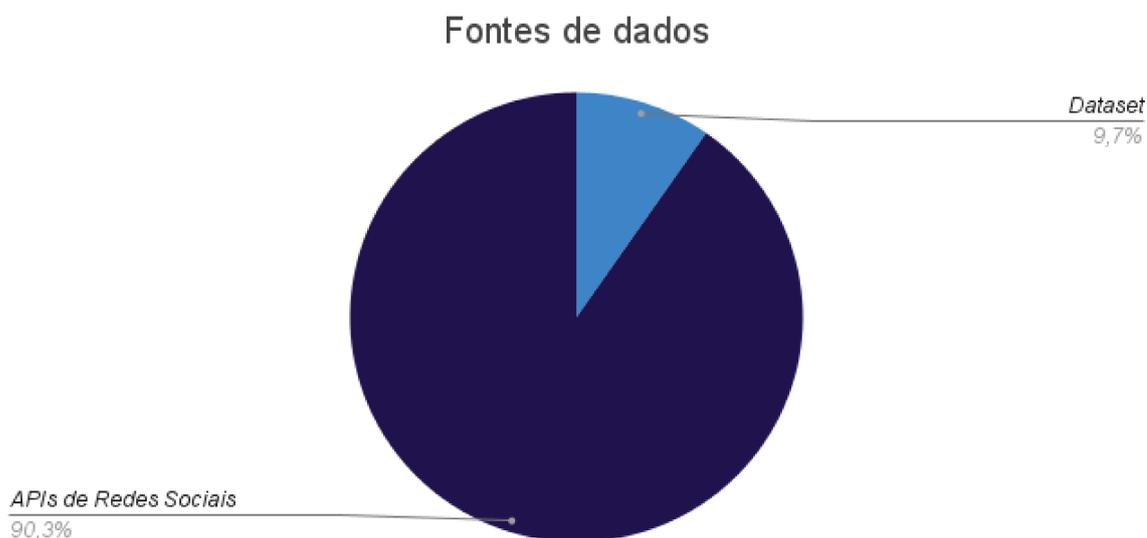


Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

As Fontes de Dados foram os meios que os pesquisadores utilizaram para acessar

as informações desejadas e posteriormente conseguir extraí-las, esse acesso pode ser feito através de uma API (*Application Programming Interface*), uma consulta utilizando a ferramenta de busca da Rede Social, um *dataset* ou a *dashboard* da Rede Social em questão.

Figura 19 – Fontes de dados mais exploradas para o desenvolvimento dos artigos.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

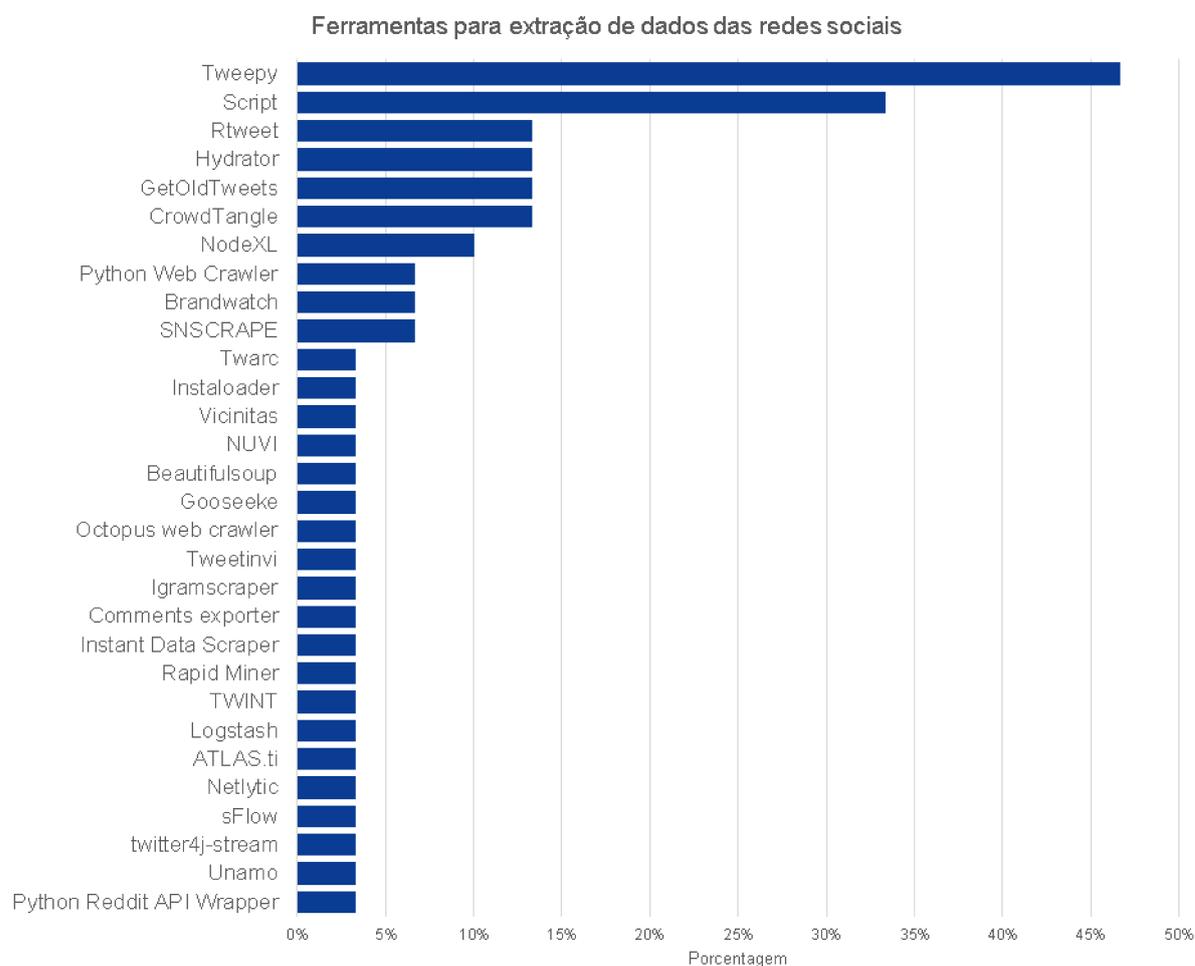
Na Figura 19 é possível ver as Fontes de Dados utilizadas pelos pesquisadores e quais foram as mais utilizadas. A fonte mais utilizada foram as próprias redes sociais, através de suas APIs. O *Twitter* fornece uma API gratuita e bem completa para os desenvolvedores, além do texto de um *tweet* é possível também obter a localização, *hashtags* utilizadas e muitos outros metadados.

Após os autores terem definido a fonte de dados a ser utilizada na pesquisa, eles precisaram fazer a extração dos dados dessas fontes. Para realizar a extração desses dados os autores utilizaram várias ferramentas, por exemplo, *softwares* de extração, *scripts* ou bibliotecas/pacotes de alguma linguagem de programação.

Na Figura 20 é possível ver quais foram as ferramentas mais utilizadas para realizar a extração de dados. A ferramenta que mais se destacou foi a biblioteca *Tweepy* do *Python*. Também é possível ver outras ferramentas bem utilizadas, por exemplo, *scripts* desenvolvidos em diferentes linguagens, o pacote *Rtweet* da Linguagem *R*, entre outros.

No desenvolvimento de cada pesquisa, foram utilizadas várias linguagens de programação com objetivos específicos, sendo eles a extração de dados (*scripts*), implementação de algoritmos para o processamento de dados e implementação das técnicas de ARS. Na Figura 21 é possível ver quais foram as linguagens mais utilizadas para essas finalidades, sendo elas *Python*, Linguagem *R*, Java e C#.

Figura 20 – Ferramentas mais utilizadas para a extração dos dados.



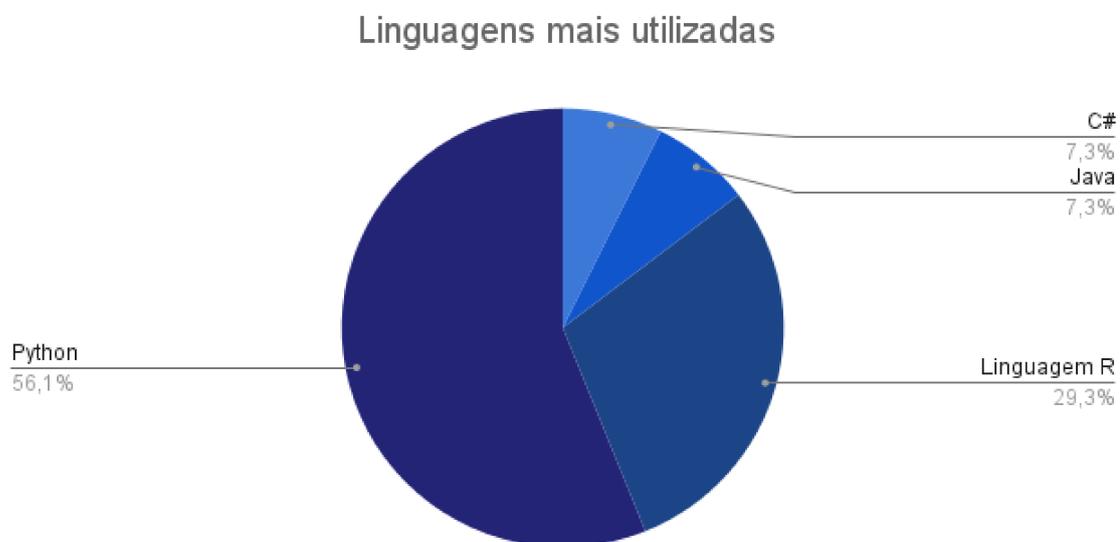
Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

A linguagem *Python* foi a mais utilizada para extração de postagens de Redes Sociais, processamento de dados e ARS, além de *Python* foi possível observar uma grande quantidade de artigos onde foi utilizada a Linguagem R. Há muitas bibliotecas em *Python* que facilitam a integração com a API do *Twitter* e também na realização da ARS, na Figura 20, é possível ver que os pesquisadores utilizaram *Tweepy*, *GetOldTweets*, *Twarc* e muitas outras bibliotecas de *Python*.

4.7 Período de Coleta de Dados

Para realizar a ARS, os pesquisadores coletaram os dados em diferentes períodos ao longo da pandemia da COVID-19. Na Figura 22 é possível ver que os dados foram coletados em um intervalo bem amplo com início em Janeiro de 2019 até o início de Agosto de 2021. Alguns pesquisadores procuraram fazer comparações entre pré-pandemia e pós-pandemia, então buscaram dados de um ano antes do primeiro caso confirmado de COVID-19 e compararam com o período pós-pandemia, com o objetivo de mostrar as

Figura 21 – Linguagens mais utilizadas.

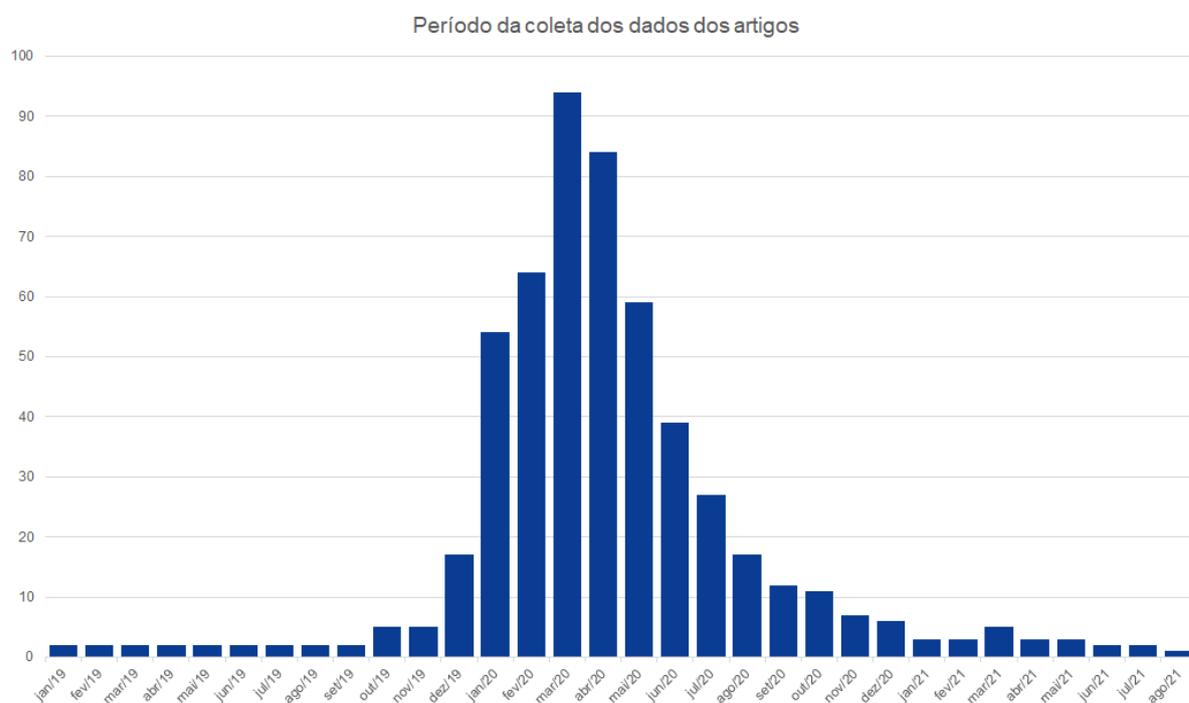


Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

principais diferenças entre eles.

Na Figura 22 é possível ver que houve um pico nos períodos de coleta, sendo esse no mês de março de 2020, mês em que foi declarado oficialmente a COVID-19 como uma pandemia pela OMS (G1, 2020). Nesse mês houveram muitas postagens em redes sociais e muito levantamento de *hashtags* no *Twitter*, dessa forma os pesquisadores utilizaram meses próximos a Março para realizarem suas pesquisas.

Figura 22 – Períodos das coletas de dados.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

5 Análise Qualitativa

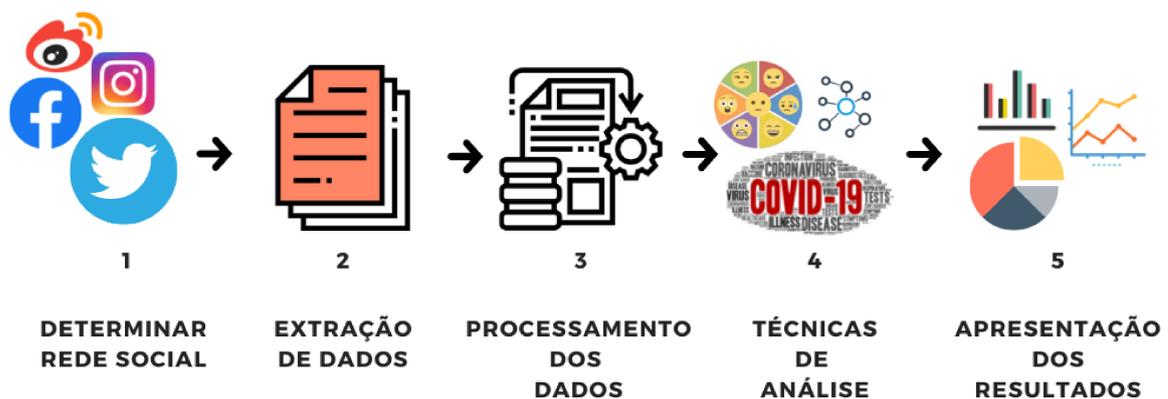
Este capítulo apresenta a análise qualitativa realizada a partir do processo de Análise de Dados de Redes sociais. Foi possível perceber quais foram as abordagens mais utilizadas e o que faltou ser abordado.

5.1 Processo de Análise de Dados de Redes Sociais

A Análise de Dados de Redes Sociais consiste em um conjunto de processos e métricas adotadas, facilitando assim o tratamento dos dados e estabelecendo um protocolo de análise. Os dados gerados por usuários em Redes Sociais possuem ruídos, os quais podem ser entendidos como informações que não são interessantes para a análise, *stopwords* ou qualquer dado que não consegue ser interpretado por máquina (Chen; Cheng; Chen, 2018).

A Figura 23 representa o Processo de ARS definido para realizar a etapa de Análise Qualitativa deste trabalho. As principais etapas envolvidas nesse processo são determinar a rede social a ser explorada, a extração de dados da rede em questão, o processamento dos dados, aplicação de alguma técnica de análise e apresentação/validação dos resultados obtidos a partir das análises.

Figura 23 – Processo de Análise de Dados de Redes Sociais.



Fonte: Imagem elaborada pela autora através do auxílio do *Canva* (Canva, 2012).

5.2 Determinar a Rede Social

A primeira etapa do processo de análise é determinar qual rede social será utilizada. Nos trabalhos analisados foram observadas cinco categorias principais para a motivação

do uso de determinada rede social. Essas categorias são: a Popularidade, Interatividade, Comunicabilidade, Acessibilidade e Especificidade. A Popularidade diz respeito a quantidade de pessoas que utilizam uma rede social e com qual frequência isso ocorre. Essa quantidade pode variar de acordo com o país ou o continente. Por exemplo, segundo o *Statista* o *Facebook* é a rede social mais utilizada no mundo com 2.8 bilhões de usuários ativos, é uma rede extremamente popular em países asiáticos, principalmente na Índia e Indonésia (We Are Social, Hootsuit, 2021; Statista, 2021).

A Interatividade é a possibilidade de membros da Rede Social conseguirem se conectar, formar comunidades e interagir na internet de forma natural e cotidiana. A Comunicabilidade é a forma em que a informação é disseminada e quais as proporções que ela toma, um exemplo são as *fake news*, elas são espalhadas rapidamente e têm um alcance muito alto.

A Acessibilidade é a facilidade em obter os dados da Rede Social em questão. O fácil acesso aos dados de uma Rede Social pode ser decisivo na motivação para a escolha por parte do pesquisador. A Especificidade se trata do quão uma Rede Social proporciona segregação em relação a alguma comunidade, um exemplo é *Reddit* que permite a criação de *subreddits* para comunidades específicas.

A Tabela 6 categoriza as principais motivações identificadas nos trabalhos analisados. Na fase da Análise Qualitativa foi possível observar o grande uso do *Twitter* para a ARS. O fato do *Twitter* ter sido escolhido como fonte de informação na maioria das pesquisas pode ser justificado pelo *Twitter* ser uma Rede Social popular e também por ter os dados facilmente acessíveis, como foi explicado na Seção 4.6 do Capítulo 4.

O *Twitter* se tornou uma rede social mundial e milhares de pessoas interagem das mais diversas maneiras, os dados gerados pelo *Twitter* vão além de textos, existem metadados como localização, número de seguidores, data de criação do *tweet*, entre outros. Além disso, o *Twitter* possui uma API para os desenvolvedores e ela é atualizada frequentemente, contribuindo cada vez mais para a Análise de Dados de Redes Sociais.

5.3 Extração de Dados

Após determinar a Rede Social a ser utilizada na pesquisa, os pesquisadores precisam determinar quais dados serão utilizados na sua pesquisa. Existem diversos tipos de dados que podem ser extraídos das Redes Sociais, desde o corpo da postagem contendo o texto até os metadados, como por exemplo, número de *likes*, localização e informações demográficas de onde a postagem foi feita. Nas pesquisas selecionadas para esta revisão, foram observados sete tipos principais de dados, sendo eles o texto, *hashtags*, mídia, perfis de usuário, *emoticons*, localização e data.

Tabela 6 – Quantidade de artigos em cada categoria de motivação.

Motivação	Exemplo	Artigos
Popularidade	”O <i>Weibo</i> é a Rede Social mais utilizada na China.”	39
Interatividade	”O <i>Facebook</i> é uma Rede Social que permite que seus membros comentem, postem conteúdos diversos e deem sua opinião.”	24
Comunicabilidade	”O <i>Twitter</i> é uma Rede Social onde os dados são disseminados com muita facilidade.”	19
Acessibilidade	”O <i>Twitter</i> fornece uma API onde os dados podem ser acessados de maneira simples pelos desenvolvedores.”	56
Especificidade	”O <i>Reddit</i> é uma Rede Social que permite a criação de vários <i>subreddits</i> , segregando os membros por assuntos de interesse.”	2

Algumas pesquisas disponibilizaram *datasets* com os dados coletados relacionados a COVID-19 para que outros pesquisadores os utilizem. Como exemplo é possível citar o *dataset* AraCOVID19-MFH que foi disponibilizado pelos pesquisadores Mohamed Seghir e Hassina Aliane. Esse *dataset* contém *tweets* em árabe que foram rotulados para serem utilizados para detecção de *fake news* ou discurso de ódio (Hadj Ameer; Aliane, 2021).

Figura 24 – *Dataset* AraCOVID19-MFH.

N°	id do Tweet	Texto do Tweet	Anotações do Tweet
1	1261897641455161345	# GermanynewCorona583, eu tenho uma receita que salva Alemanha do corona e com sucesso vai restaurar a vida cotidiana. Vamos fazer o acordo inicial. Estou pronto para entregá-la aos responsáveis. Cientistas médicos, estou pronto para explicá-la e para verificar sua eficácia. A prescrição 972522878900 não é um composto e não é um produto químico que está apenas nas mãos dos portadores da doença de Crohn.	Contém ódio: Não Contém um aviso: Não Notícia ou opinião: Opinião Culpa ou discurso negativo: Não Vale a pena verificar os fatos: Sim Fala sobre uma cura: Sim Levanta questões morais: Não Dialeto: Árabe padrão moderno Fatual: Sim Contém informação falsa: Sim
2	1265791347308539904	Deus amaldiçoe a China, Deus amaldiçoe o morcego, Deus amaldiçoe o Corona, Deus amaldiçoe o morcego, Deus amaldiçoe a pedra, amaldiçoe o pecado.	Contém ódio: Sim Contém um aviso: Não Notícia ou opinião: Opinião Culpa ou discurso negativo: Sim Vale a pena verificar os fatos: S/D Fala sobre uma cura: Não Levanta questões morais: Não Dialeto: Árabe padrão moderno Fatual: Não Contém informação falsa: S/D
3	1273965696108695552	Algumas pessoas não seguiram 100% das instruções. Proteja desse inferno e as mantenha 100% obedientes às instruções de proteção contra o Corona.	Contém ódio: Não Contém um aviso: Não Notícia ou opinião: Opinião Culpa ou discurso negativo: Não Vale a pena verificar os fatos: S/D Fala sobre uma cura: Não Levanta questões morais: Não Dialeto: Oriente médio Fatual: Não Contém informação falsa: S/D
4	1276183562560114688	A agência Al-thawra recomenda o uso de # remdesivir para o virus #Corona https://t.co/njtizkQeO8 # Covid-19 # COVID2019 https://t.co/kzupinUEI3	Contém ódio: Não Contém um aviso: Sim Notícia ou opinião: Notícia Culpa ou discurso negativo: Não Vale a pena verificar os fatos: Sim Fala sobre uma cura: Sim Levanta questões morais: Não Dialeto: Árabe padrão moderno Fatual: Sim Contém informação falsa: Não

Fonte: Adaptação do exemplo de Hadj Ameer e Aliane (2021).

A Figura 24 mostra como os pesquisadores rotularam os *tweets*, também é possível ver os dados que o *dataset* fornece além dos rótulos, sendo eles o *id* e o texto do *tweet*. Dentre os artigos analisados, aqueles que disponibilizaram os conjuntos de dados coletados

Tabela 7 – Tipos de dados mais utilizados nas pesquisas.

Tipo de dado	Descrição	Artigos
Texto	Corpo da postagem.	133
<i>Hashtags</i>	As <i>hashtags</i> são geralmente palavras acompanhadas por um símbolo “#” que são utilizadas para assuntos que estão em tendência no momento.	16
Mídia	As imagens ou vídeos, podendo estar ou não relacionados às postagens.	11
Perfis	Fornecer as informações de cada usuário da Rede Social. Pode ser o <i>username</i> da pessoa, quantidade de postagens ou outros metadados, como por exemplo, número de seguidores da conta ou <i>id</i> .	13
Data	Data em que a postagem foi feita.	31
Localização	Local de onde a postagem foi feita.	32
<i>Emoticons</i>	Símbolos que expressam uma emoção ou ação. Por exemplo, 😊 📱 😡 😞 🙌 🙏.	6
Outros	Outros metadados, como menções, número de <i>likes</i> , respostas, entre outros.	17

são Hadj Ameur e Aliane (2021), Yang, Zhou e Zafarani (2021), Chang, Monselise e Yang (2021), Lee et al. (2021), Ito e Chakraborty (2020), Kaur, Kaul e Zadeh (2020), Slavik et al. (2021), Shen et al. (2020), Cresswell et al. (2021).

Na Tabela 7 é possível ver que na maioria das pesquisas foi utilizado o corpo da postagem para realizar o processamento e posteriormente a análise, isso é devido às possibilidades de análise serem maiores com o texto e também há uma vasta quantidade de bibliotecas disponibilizadas para tratamento e manipulação de texto. Um exemplo de biblioteca muito utilizada para tratamento de texto é a NLTK do *Python*, com ela é possível extrair radicais de palavras ou também obter informações como palavras em comum nos documentos (NLTK, 2021). Além do corpo da postagem (texto) temos outros tipos de dados que podem ser importantes para a complementação de uma análise, como data ou *emoticons*. Nas pesquisas analisadas foi possível perceber o uso conjunto desses dados, um exemplo é o uso conjunto de texto e *emoticon* para a Análise de Sentimentos (Mohamed Ridhwan; Hargreaves, 2021; Malagoli et al., 2021). Os dados mencionados foram pouco explorados isoladamente no contexto da pandemia da COVID-19. A lista com os artigos referentes aos diferentes tipos de dados utilizados em conjunto é apresentada a seguir.

1. *Localização*: Kaur, Kaul e Zadeh (2020), Elgazzar, Spurlock e Bogart (2021), Liu et al. (2021), Chong e Park (2021), Liu et al. (2021), Hasni e Faiz (2021), Böttger, Ibrahim e Vallis (2020), Kaur, Kaul e Zadeh (2020), Balasubramaniam, Nayak e Bashar (2020), Bahja e Safdar (2020), Al-Shargabi e Selmi (2021), Tao, Miao e Ng (2020), Mackey et al. (2020a), Li et al. (2020), Xu et al. (2020), Wang et al. (2020), Cresswell et al. (2021), Tao et al. (2020), Xu, Dredze e Broniatowski (2020), Doogan et al. (2020), Huang et al. (2020), Maia, Oliveira e Gallegos (2021), Paula e Mello (2020), Miranda-Escalada et al. (2021), Massaad e Cherfan (2020), Brum et al. (2020), Guo et al. (2020), Wahbeh et al. (2020), Shen et al. (2020), Klein et al. (2021), Shen et al. (2021), Hussain et al. (2021), Alanazi et al. (2020);
2. *Emoticon*: Mohamed Ridhwan e Hargreaves (2021), Bhatnagar e Choubey (2021), Romeiro Paulino e Pires Ventura (2021), Malagoli et al. (2021), Morshed et al. (2021), Sari e Ruldeviyani (2020) ;
3. *Hashtag*: Mohamed Ridhwan e Hargreaves (2021), Al-Rawi et al. (2021), Singh, Jakhar e Pandey (2021), Almosa (2021), Singh, Jakhar e Pandey (2021), Forte e Pires (2020), Mourad et al. (2020), Santis, Martino e Rizzi (2020), Jang et al. (2021), Brum et al. (2020), Chen, Lerman e Ferrara (2020), Paula e Mello (2020), Leão et al. (2020), Xue et al. (2020), Trust et al. (2020), Permana et al. (2021), Slavik et al. (2021);
4. *Data*: Kaur, Kaul e Zadeh (2020), Elgazzar, Spurlock e Bogart (2021), Domalewska (2021), Liu et al. (2021), Chong e Park (2021), Yang, Zhou e Zafarani (2021), Zhou et al. (2021), Ito e Chakraborty (2020), Chen et al. (2021), Santis, Martino e Rizzi (2020), Tao, Miao e Ng (2020), Xu et al. (2020), Nurlaila et al. (2020), Boon-Itt e Skunkan (2020), Wang et al. (2020), Cresswell et al. (2021), Malagoli et al. (2021), Lyu e Luli (2021), Wrycza e Maślankowski (2020), Ruffer, Knitza e Krusche (2020), Mackey et al. (2020b), Shen et al. (2020), Adikari et al. (2021), Gerts et al. (2021), Saha et al. (2020), Chen, Lerman e Ferrara (2020), Liao et al. (2020), Kwok, Vadde e Wang (2021), Mackey et al. (2020b), Wahbeh et al. (2020), Guo et al. (2020), Alshehabi et al. (2020);
5. *Mídia*: Yang, Zhou e Zafarani (2021), Lee et al. (2021), Romeiro Paulino e Pires Ventura (2021), Forte e Pires (2020), Pran et al. (2020), Amanatidis et al. (2021), Slavik et al. (2021), Rafi et al. (2020), Jiang et al. (2021);
6. *Perfis*: Elgazzar, Spurlock e Bogart (2021), Domalewska (2021), Romeiro Paulino e Pires Ventura (2021), Mourad et al. (2020), Nurlaila et al. (2020), Abd-Alrazaq et al. (2020), Park, Park e Chong (2020), Zhang et al. (2021), Santos, Carmargo e Menossi (2020), Permana et al. (2021), Lima et al. (2021), Recuero e Soares (2021), Recuero, Soares e Zago (2021);

7. Outros metadados: Al-Rawi et al. (2021), Lee et al. (2021), Mourad et al. (2020), Yang et al. (2020), Maakoul et al. (2020), Rafi et al. (2020), Wang e Qian (2021), Jang et al. (2021), Malagoli et al. (2021), Slavik et al. (2021), Ruffer, Knitza e Krusche (2020), Amanatidis et al. (2021), Xue et al. (2020), Yang e Han (2021), Gonçalves et al. (2021), Lima et al. (2021), Paula e Mello (2020).

5.4 Processamento dos dados

Após a coleta dos dados é necessário fazer o processamento desses dados. O processamento é feito em duas etapas, sendo elas o pré-processamento e o processamento final. Nessa etapa do pré-processamento é feita a filtragem dos dados, como por exemplo, extração da localização ou remoção de *stopwords*.

Na etapa do processamento os dados já começam a ser analisados, então são utilizadas metodologias que permitam a classificação do texto, como por exemplo, algoritmos específicos para classificação ou *softwares* que disponibilizem ferramentas para classificação do texto. A Figura 25 mostra uma abstração das principais fases de pré-processamento e processamento.

Figura 25 – Fases do pré-processamento e processamento.



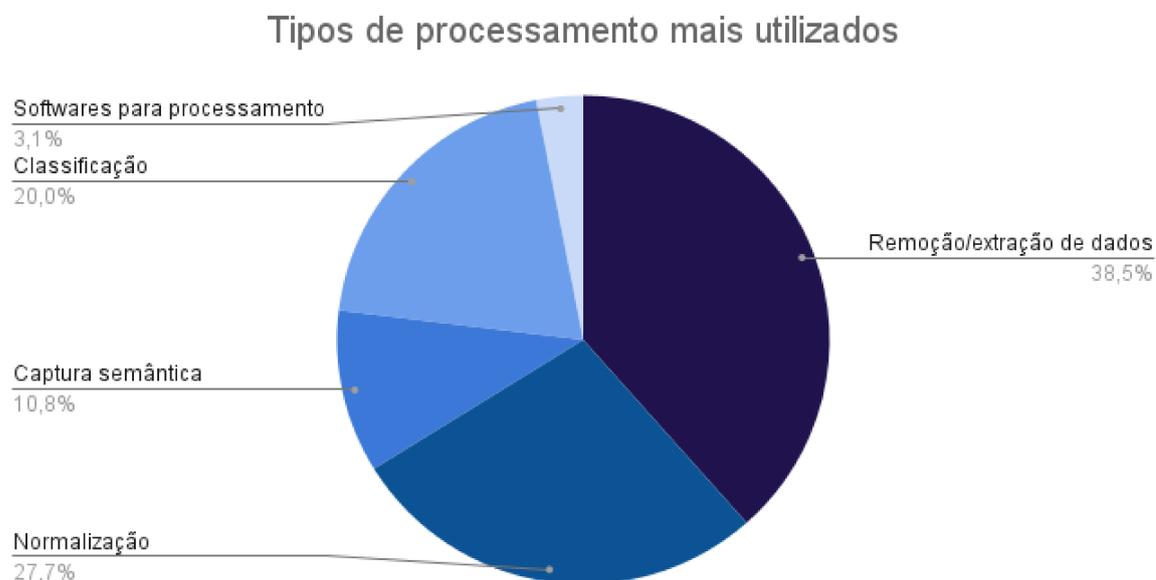
Fonte: Imagem elaborada pela autora através do auxílio do *Canva* (Canva, 2012).

Nos trabalhos coletados, foram identificados cinco tipos de pré-processamentos e processamentos utilizados nas pesquisas, sendo eles a remoção de dados irrelevantes para a análise, normalização, captura semântica, classificação e *softwares* para o processamento do texto. A Tabela 8 e a Figura 26 dão uma visão geral de quais tipos de processamento foram utilizados nas pesquisas que descreveram o método de trabalho empregado. Analisando os trabalhos foi possível observar grande concentração de uso da normalização, classificação de texto e remoção de dados a partir da construção de algoritmos baseados em Processamento de Linguagem Natural.

Tabela 8 – Tipos de processamento mais utilizados.

Tipo de processamento	Descrição	Métodos/Algoritmos comuns	Artigos
Remoção/extração de dados para a análise	Remoção de dados que não serão relevantes para a posterior análise, por exemplo, remoção de palavras repetidas, espaços em branco, entre outros. É realizada uma limpeza dos dados de forma geral. Além disso é possível extrair informações específicas, como apenas a localização de onde certa postagem foi feita.	Algoritmos baseados em Processamento de Linguagem Natural	75
Normalização	Tokenização, extração de radicais e agrupar palavras de sentido semelhante e remoção de <i>stopwords</i> .	NLTK Penn Treebank TreeTagger Porter Stemmer Gensim Methods	54
Captura semântica	Mapear as palavras de uma sentença em um vetor, é possível capturar uma palavra central.	BERT - <i>Embeddings</i> TF-IDF <i>models</i> Word2Vec FastText	21
Classificação	Etiqueta as palavras com algum tipo de classificação, por exemplo, artigo, preposição, entre outros.	Support Vector Machine Naive Bayes Random Forest Recurrent Neural Network	39
Softwares para processamento	Ferramentas para fazer o processamento dos dados.	MAXQDA ATLAS.ti CoreNLP RapidMiner	6

Figura 26 – Tipos de processamento mais utilizados.



Fonte: Gráfico elaborado pela autora através do auxílio do *Excel*.

Vários trabalhos realizaram a extração de dados específicos ou limpeza dos dados através de remoção de dados irrelevantes para a uma posterior análise ou normalização, dessa forma foram removidos dados como palavras repetidas, espaços em branco, men-

ções, *retweets*, entre outros. Os trabalhos que realizaram extração/limpeza inicial dos dados foram Rusho, Ahmed e Sadri (2020), Kaur, Kaul e Zadeh (2020), Hadj Ameer e Aliane (2021), Al-Rawi et al. (2021), Mohamed Ridhwan e Hargreaves (2021), Domalewska (2021), Liu et al. (2021), Chong e Park (2021), Hasni e Faiz (2021), Yigitcanlar et al. (2020), Amara, Taieb e Aouicha (2021), Ceron et al. (2021), Lee et al. (2021), Böttger, Ibrahim e Vallis (2020), Raheja e Asthana (2021), Duong et al. (2020), Xavier et al. (2020), Caldera, Meedin e Perera (2020), Kaur, Kaul e Zadeh (2020), Balasubramaniam, Nayak e Bashar (2020), Shan e Li (2020), Chen et al. (2021), Nurbakova, Ermakova e Ovchinnikova (2020), Nguyen e Chaturvedi (2020), Wibowo, Mahardika e Kusri (2020), Al-Shargabi e Selmi (2021), Viviani et al. (2021), Alzamzami e Saddik (2021), Yang et al. (2020), Rafi et al. (2020), Cotfas et al. (2021), Nurlaila et al. (2020), Boon-Itt e Skunkan (2020), Valdez et al. (2020), Li et al. (2020), Gao, Xie e Li (2021), Xu et al. (2020), Park, Park e Chong (2020), Alanazi et al. (2020), Zhang et al. (2021), Cresswell et al. (2021), Xu, Dredze e Broniatowski (2020), Xue et al. (2020), Huang et al. (2020), Liao et al. (2020), Huang et al. (2020), Saha et al. (2020), Ahmed et al. (2020), Chen et al. (2021), Shen et al. (2020), Shen et al. (2021), Slavik et al. (2021), Wahbeh et al. (2020), Lee et al. (2020), Ruffer, Knitza e Krusche (2020), Rosenberg (2020), Kamiński, Muth e Bogdański (2020), Shim et al. (2021), Lyu e Luli (2021), Osakwe et al. (2020), Morshed et al. (2021), Recuero, Soares e Zago (2021), Jiang et al. (2021), Miranda-Escalada et al. (2021), Xue et al. (2020), Das e Dutta (2021), Karami e Anderson (2020), Lima et al. (2021), Recuero e Soares (2021), Paula e Mello (2020), Leão et al. (2020), Brum et al. (2020), Melo (2021), Pinto et al. (2020), Santos, Carmargo e Menossi (2020), Wang e Qian (2021). Foi possível observar a utilização de softwares como MAXQDA, ATLAS.ti e *CoreNLP* para a limpeza e processamento dos dados pelos pesquisadores. A lista com os artigos referentes é apresentada a seguir.

1. MAXQDA: Almosa (2021), Osakwe et al. (2020);
2. ATLAS.ti: Romeiro Paulino e Pires Ventura (2021);
3. *CoreNLP*: Duong et al. (2020), Miranda-Escalada et al. (2021);
4. *RapidMiner*: Pastor (2020).

A normalização foi uma técnica muito utilizada nas pesquisas. Essa técnica compreende o uso de um conjunto de processos que fazem a limpeza e organização dos dados. O uso dessa técnica foi observado principalmente em relação a remoção de *stopwords* e radicais, lematização das sentenças e tokenização das palavras.

Considerando o aspecto relacionado aos softwares usados nos trabalhos descritos nos artigos selecionados foi possível observar um grande uso de *Python* nas pesquisas.

Python é uma linguagem que disponibiliza muitas bibliotecas e ferramentas para processamento de dados, uma das mais comuns e que pode ser observada nas pesquisas é a *NLTK - Natural Language Toolkit*. Também foi possível observar o uso de *Penn Treebank*, *TreeTagger*, *Porter Stemmer*, *Gensim Methods*, além de outros métodos desenvolvidos/adaptados pelos próprios pesquisadores. A lista com os artigos referentes é apresentada a seguir.

1. *NLTK - Natural Language Toolkit*: Um conjunto de bibliotecas e programas para processamento de linguagem natural, escrito em *Python*. Utilizado em [Mohamed Ridhwan e Hargreaves \(2021\)](#), [Bhatnagar e Choubey \(2021\)](#), [Jelodar et al. \(2020\)](#), [Nguyen e Chaturvedi \(2020\)](#), [Gupta et al. \(2020\)](#), [Liu et al. \(2021\)](#), [Long, Alharthi e Saddik \(2020\)](#), [Maakoul et al. \(2020\)](#), [Gupta et al. \(2020\)](#), [Xue et al. \(2020\)](#), [Doogan et al. \(2020\)](#), [Wicke e Bolognesi \(2020\)](#), [Wang et al. \(2021\)](#), [Chandrasekaran et al. \(2020\)](#), [Massaad e Cherfan \(2020\)](#), [Ruffer, Knitza e Krusche \(2020\)](#), [Li, Chaudhary e Zhang \(2020\)](#), [Liu, Xia e Lang \(2021\)](#), [Abd-Alrazaq et al. \(2020\)](#);
2. *Penn Treebank*: Corpus de texto previamente analisado que anota a estrutura sintática ou semântica das frases. Utilizado em [Agarwal et al. \(2020\)](#);
3. *TreeTagger*: É uma ferramenta para anotar o texto com informações em relação às classes gramaticais e que possuem algum significado semântico. Utilizado em [Santis, Martino e Rizzi \(2020\)](#);
4. *Porter Stemmer*: Se trata de um processo para remover as terminações morfológicas mais comuns das palavras. É muito utilizado no processo de normalização de termos em sistemas de recuperação de informação. Utilizado em [Yang et al. \(2020\)](#), [Schück et al. \(2021\)](#);
5. *Gensim Methods*: Um pacote utilizado para realizar o processamento de linguagem natural, escrito em *Python*. É utilizado para fazer a extração semântica dos documentos de uma base, por exemplo. Utilizado em [Hasni e Faiz \(2021\)](#);
6. Métodos desenvolvidos/adaptados pelos próprios pesquisadores: [Rusho, Ahmed e Sadri \(2020\)](#), [Hadj Ameer e Aliane \(2021\)](#), [Elgazzar, Spurlock e Bogart \(2021\)](#), [Domalewska \(2021\)](#), [Valle Cruz et al. \(2021\)](#), [Mourad et al. \(2020\)](#), [Abdelminaam et al. \(2021\)](#), [Gottipati et al. \(2021\)](#), [Ito e Chakraborty \(2020\)](#), [Boon-Itt e Skunkan \(2020\)](#), [Al-Ramahi et al. \(2021\)](#), [Mackey et al. \(2020a\)](#), [Jang et al. \(2021\)](#), [Tao et al. \(2020\)](#), [Chen, Lerman e Ferrara \(2020\)](#), [Xue et al. \(2020\)](#), [Daughton et al. \(2021\)](#), [Klein et al. \(2021\)](#), [Adikari et al. \(2021\)](#), [Gerts et al. \(2021\)](#), [Kwok, Vadde e Wang \(2021\)](#), [Hussain et al. \(2021\)](#), [Mackey et al. \(2020b\)](#), [Permana et al. \(2021\)](#), [Guo et al. \(2020\)](#), [Wrycza e Maślankowski \(2020\)](#), [Mutanga e Abayomi \(2020\)](#), [Yang e Han \(2021\)](#), [Saire e Cruz \(2020\)](#), [Maia, Oliveira e Gallegos \(2021\)](#), [Pinto et al. \(2020\)](#).

A captura semântica é um tipo de processamento que a partir de um vetor de palavras, é possível detectar uma palavra que é frequentemente utilizada em uma coleção de documentos ou uma palavra central em uma sentença. Para isso, alguns dos métodos propostos que realizam esse tipo de processamento são: BERT - *Embeddings*, TF-IDF *models*, *Word2Vec* e *FastText*. A lista com os artigos referentes é apresentada a seguir.

1. BERT - *Embeddings*: Se trata da construção de um modelo de linguagem com um corpo de texto. Esse algoritmo permite o entendimento das palavras, frases e conteúdos inteiros que compõem um corpo de documento a partir da criação de nós. Utilizado em [Singh, Jakhar e Pandey \(2021\)](#), [Ito e Chakraborty \(2020\)](#), [Agarwal et al. \(2020\)](#);
2. TF-IDF *models*: São modelos que visam indicar a importância de certo termo presente em um documento em relação a uma coleção de documentos. Utilizado em [Al-Rawi et al. \(2021\)](#), [Bhatnagar e Choubey \(2021\)](#), [Forte e Pires \(2020\)](#), [Xavier et al. \(2020\)](#), [Duong et al. \(2020\)](#), [Naseem et al. \(2021\)](#), [Alzamzami e Saddik \(2021\)](#), [Boon-Itt e Skunkan \(2020\)](#), [Agarwal et al. \(2020\)](#), [Wibowo, Mahardika e Kusri \(2020\)](#), [Naseem et al. \(2021\)](#), [Cotfas et al. \(2021\)](#), [Wang et al. \(2021\)](#), [Amanatidis et al. \(2021\)](#), [Li, Chaudhary e Zhang \(2020\)](#);
3. *Word2Vec*: Um algoritmo baseado em processamento natural de linguagem que realiza associações de palavras a partir de um grande corpo de texto. Utilizado em [Liu et al. \(2021\)](#), [Hasni e Faiz \(2021\)](#), [Ito e Chakraborty \(2020\)](#);
4. *FastText*: Uma biblioteca para aprendizagem de incorporação de palavras e classificação de texto, cujo objetivo é a obtenção de representações vetoriais de palavras. Utilizado em [Hasni e Faiz \(2021\)](#), [Naseem et al. \(2021\)](#), [Miranda-Escalada et al. \(2021\)](#).

No contexto da pandemia da COVID-19, além de *Python* e NLTK, foi possível observar também interesse dos pesquisadores por *Deep Learning* e *Machine Learning* na classificação do texto: [Jelodar et al. \(2020\)](#), [Long, Alharthi e Saddik \(2020\)](#), [Pran et al. \(2020\)](#), [Abdelminaam et al. \(2021\)](#), [Shan e Li \(2020\)](#), [Chen et al. \(2021\)](#), [Wibowo, Mahardika e Kusri \(2020\)](#), [Naseem et al. \(2021\)](#), [Sari e Ruldeviyani \(2020\)](#), [Maakoul et al. \(2020\)](#), [Cotfas et al. \(2021\)](#), [Gupta et al. \(2020\)](#), [Zhang et al. \(2021\)](#), [Cresswell et al. \(2021\)](#), [Wang et al. \(2021\)](#), [Xue et al. \(2020\)](#), [Chen et al. \(2021\)](#), [Daughton et al. \(2021\)](#), [Shen et al. \(2020\)](#), [Shen et al. \(2021\)](#), [Ruffer, Knitza e Krusche \(2020\)](#), [Amanatidis et al. \(2021\)](#), [Li, Chaudhary e Zhang \(2020\)](#), [Shim et al. \(2021\)](#). Algoritmos como *Naïve Bayes*, *Random Forest*, *Support Vector Machine* e *Recurrent Neural Network* tiveram destaque. As pesquisas desenvolvidas nesse âmbito contribuíram para a área da Análise de Dados e

dessa forma abrem portas para a expansão do uso dessas técnicas em pesquisas futuras. A lista com os artigos referentes é apresentada a seguir.

1. *Naïve Bayes*: O classificador *Naïve Bayes* é um dos modelos mais populares no aprendizado de máquina. Esse modelo tem como base a suposição de independência entre as variáveis de um problema, onde realiza uma classificação probabilística de observações, caracterizando-as em classes pré-definidas. Utilizado em [Mohamed Ridhwan e Hargreaves \(2021\)](#), [Bhatnagar e Choubey \(2021\)](#), [Nguyen e Chaturvedi \(2020\)](#), [Chandrasekaran et al. \(2020\)](#), [Permana et al. \(2021\)](#);
2. *Random Forest*: Esse é um método de aprendizagem para classificação, regressão e outras tarefas que operam através da construção de árvores de decisão no momento do treinamento. Utilizado em [Yigitcanlar et al. \(2020\)](#), [Duong et al. \(2020\)](#), [Gerts et al. \(2021\)](#);
4. *Support Vector Machine*: Um conjunto de métodos de aprendizado supervisionado de máquina que analisam os dados e também reconhecem padrões, ele é usado para classificação e análise de regressão. Utilizado em [Forte e Pires \(2020\)](#);
5. *Recurrent Neural Network*: É uma classe de redes neurais em que as conexões entre os nós formam um gráfico direcionado ao longo de uma sequência. Isso permite realizar a classificação dos termos baseando-se em termos recorrentes. Utilizada em [Jelodar et al. \(2020\)](#), [Mohamed Ridhwan e Hargreaves \(2021\)](#), [Yang, Zhou e Zafarani \(2021\)](#), [Hasni e Faiz \(2021\)](#), [Alzamzami e Saddik \(2021\)](#), [Abdelminaam et al. \(2021\)](#), [Pran et al. \(2020\)](#), [Yang et al. \(2020\)](#), [Rafi et al. \(2020\)](#), [Mackey et al. \(2020b\)](#), [Shim et al. \(2021\)](#).

5.5 Técnicas de Análise

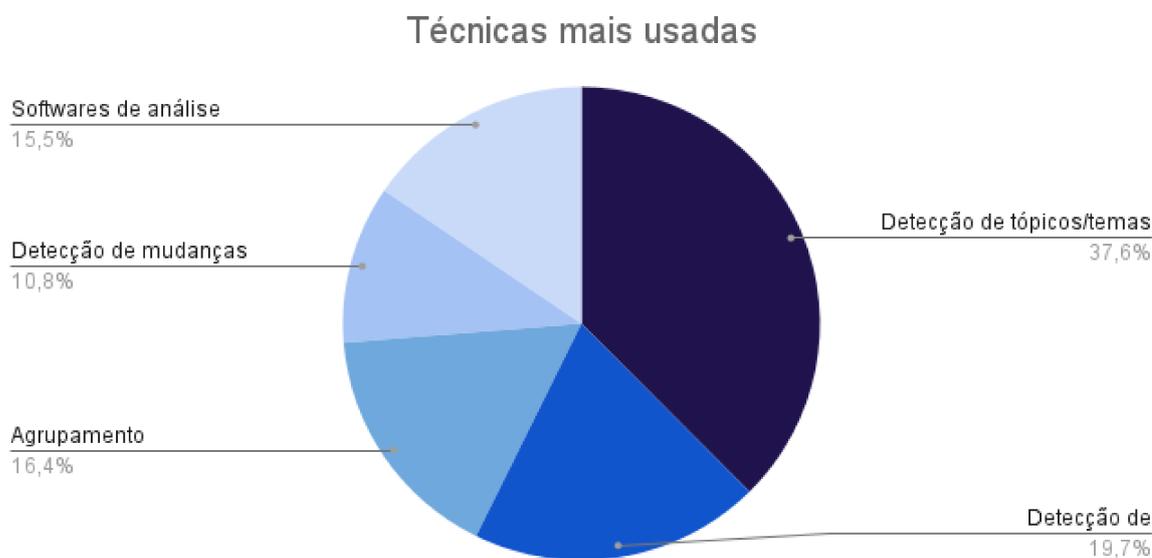
Após ser feito o Processamento dos Dados é possível utilizar alguma Técnica de Análise, dessa forma o pesquisador consegue elaborar uma resposta à questão de pesquisa ou solucionar problemas propostos em relação ao tema de pesquisa. Nessa etapa são utilizados algoritmos, métodos e modelos capazes de gerar resultados que atendam aos tipos de análise desejados. A partir da leitura dos artigos selecionados, foram identificadas cinco técnicas de análise, sendo elas a detecção de tópicos, detecção de sentimentos e polaridade, agrupamento, detecção de mudanças temporais e *softwares* que realizam análises.

Na Tabela 9 é possível ver os algoritmos e métodos mais comuns de cada técnica. Para realizar a detecção de tópicos foram utilizadas várias abordagens, através de algoritmos que calcularam frequências de certo termo ou utilizando outras abordagens supervisionadas ou não supervisionadas de aprendizado. A Figura 27 demonstra graficamente as técnicas mais utilizadas.

Tabela 9 – Tabela com as técnicas mais usadas.

Técnica	Descrição	Algoritmos/modelos mais utilizados	Artigos
Detecção de tópicos ou temas	Para realizar a detecção de tópicos foram utilizadas várias abordagens, através de algoritmos que calcularam frequências de certo termo ou utilizando outras abordagens supervisionadas ou não supervisionadas de aprendizado.	Latent Dirichlet Allocation Word Frequency Algorithm BERT - Models Identify Microblog Topics Method Non-negative Matrix Factorization Named Entity Recognition Bigrams Detection Algorithm	80
Detecção de sentimentos e determinação de polaridade	Para realizar a detecção de sentimentos e determinação de polaridade foram utilizadas várias abordagens, tanto através de algoritmos que identificaram os sentimentos envolvidos nas publicações quanto abordagens supervisionadas de aprendizado.	Valence Aware Dictionary for Bag-Of-Words Models BERT - Sentiment Analyst Approach SentiStrength Algorithm	42
Agrupamento	Para realizar agrupamentos e detecção de comunidades foram utilizados diversos algoritmos que permitissem análises voltadas para rede.	Lowvain Clustering Algorithm FluidC Algorithm PageRank Algorithm Fruchterman Reingold Algorithm Community Detection Methods	35
Detecção de mudanças temporais	Para realizar a detecção de mudanças temporais, vários algoritmos e métodos foram utilizados para prever algum evento ou tendência.	Augmented Dickey-Fuller Algorithm Smoothing Algorithms Altmetric Attention Scores Algorithm Box-Jenkins Forecasting Algorithm Topic Evolution Detection Algorithms Traffic Changes Methods Prediction Methods Time Series Variation Algorithm	23
Softwares de análise	Alguns softwares foram utilizados para realizar vários tipos de análises, principalmente análises de rede e sentimentos.	WEKA SenticNet NodeXL QDA Miner WordStat 8 Syuzhet Tool TextBlob Tool Netlytic IBM Analyze	33

Figura 27 – Técnicas de análise mais utilizadas.



Fonte: Gráfico elaborado pela autora através do auxílio do Excel.

Dentre os artigos analisados, aqueles que utilizaram essa técnica foram [Mohamed Ridhwan e Hargreaves \(2021\)](#), [Almossa \(2021\)](#), [Bhatnagar e Choubey \(2021\)](#), [Kirabo, Namara e Mcneese \(2021\)](#), [Forte e Pires \(2020\)](#), [Mourad et al. \(2020\)](#), [Santis, Martino e Rizzi \(2020\)](#), [Zhang et al. \(2021\)](#), [Tao et al. \(2020\)](#), [Chen, Lerman e Ferrara \(2020\)](#), [Liao et al. \(2020\)](#), [Ahmed et al. \(2020\)](#), [Daughton et al. \(2021\)](#), [Shen et al. \(2021\)](#), [Massaad e Cherfan \(2020\)](#), [Wrycza e Maślankowski \(2020\)](#), [Rosenberg \(2020\)](#), [Li, Chaudhary e Zhang \(2020\)](#), [Alshehabi et al. \(2020\)](#), [Saire e Cruz \(2020\)](#), [Recuero e Soares \(2021\)](#), [Pinto et al. \(2020\)](#), [Chang, Monselise e Yang \(2021\)](#), [Lee et al. \(2021\)](#), [Mourad et al. \(2020\)](#), [Alanazi et al. \(2020\)](#), [Miranda-Escalada et al. \(2021\)](#). Alguns métodos como o *Latent Dirichlet Allocation*, *Word Frequency Algorithm*, *BERT - Models*, *Identify Microblog Topics Method*, *Non-negative Matrix Factorization* e *Bigrams Detection Algorithm* tiveram destaque. A lista com os artigos referentes a cada método é apresentada a seguir.

1. *Latent Dirichlet Allocation*: Se trata de uma abordagem generativa, o algoritmo gera um modelo probabilístico que é usado para identificar grupos de tópicos. Ele permite que conjuntos de documentos sejam explicados por variáveis latentes que explicam por que algumas partes dos dados são semelhantes. Utilizado em [Mohamed Ridhwan e Hargreaves \(2021\)](#), [Rusho, Ahmed e Sadri \(2020\)](#), [Viviani et al. \(2021\)](#), [Liu et al. \(2021\)](#), [Amara, Taieb e Aouicha \(2021\)](#), [Goel e Sharma \(2021\)](#), [Duong et al. \(2020\)](#), [Jelodar et al. \(2020\)](#), [Gottipati et al. \(2021\)](#), [Ito e Chakraborty \(2020\)](#), [Balasubramaniam, Nayak e Bashar \(2020\)](#), [Agarwal et al. \(2020\)](#), [Naseem et al. \(2021\)](#), [Bahja e Safdar \(2020\)](#), [Alzamzami e Saddik \(2021\)](#), [Tao, Miao e Ng \(2020\)](#), [Boon-Itt e Skunkan \(2020\)](#), [Valdez et al. \(2020\)](#), [Gao, Xie e Li \(2021\)](#), [Al-Ramahi et al. \(2021\)](#), [Xue et al. \(2020\)](#), [Wang et al. \(2020\)](#), [Jang et al. \(2021\)](#), [Doogan et al. \(2020\)](#), [Xue et al. \(2020\)](#), [Adikari et al. \(2021\)](#), [Chandrasekaran et al. \(2020\)](#), [Kwok, Vadde e Wang \(2021\)](#), [Mutanga e Abayomi \(2020\)](#), [Shim et al. \(2021\)](#), [Lyu e Luli \(2021\)](#), [Xue et al. \(2020\)](#), [Yang e Han \(2021\)](#), [Das e Dutta \(2021\)](#), [Wicke e Bolognesi \(2020\)](#), [Liu et al. \(2021\)](#), [Maia, Oliveira e Gallegos \(2021\)](#), [Melo \(2021\)](#), [Karami e Anderson \(2020\)](#), [Melo \(2021\)](#), [Pinto et al. \(2020\)](#), [Abd-Alrazaq et al. \(2020\)](#);
2. *Word Frequency Algorithms*: Se trata de modelos que utilizam as frequências de palavras encontradas em um corpo de documentos para fazer o levantamento de possíveis tópicos latentes. Utilizado em [Rusho, Ahmed e Sadri \(2020\)](#), [Chen et al. \(2021\)](#), [Lee et al. \(2020\)](#), [Jiang et al. \(2021\)](#), [Abd-Alrazaq et al. \(2020\)](#);
3. *BERT - Models*: Se trata da construção de um modelo de linguagem com um corpo de texto. Essa abordagem do algoritmo permite a identificação de palavras (possíveis tópicos) relevantes em um corpo textual. Utilizado em [Hadj Ameer e Aliane \(2021\)](#), [Liu et al. \(2021\)](#), [Viviani et al. \(2021\)](#), [Cresswell et al. \(2021\)](#);

4. *Identify Microblog Topics Method*: Se trata de um método que produz tópicos semânticos a partir de coleções de postagens de redes sociais. Os possíveis tópicos são determinados pelo processamento de termos vinculados a uma coleção de postagens, sendo atribuídos a um gráfico de co-ocorrência dos termos. Utilizado em [Yang, Zhou e Zafarani \(2021\)](#), [Xu et al. \(2020\)](#), [Ruffer, Knitza e Krusche \(2020\)](#), [Brum et al. \(2020\)](#), [Wang e Qian \(2021\)](#);
5. *Non-negative Matrix Factorization*: é uma técnica não supervisionada. A forma como funciona é que, o NMF fatoriza vetores de dimensões elevadas em uma representação de dimensões inferiores. Esses vetores de dimensão inferior não são negativos, o que também significa que seus coeficientes não são negativos. Com essa técnica é possível descobrir tópicos ocultos em um conjunto de documentos. Utilizada em [Chang, Monselise e Yang \(2021\)](#), [Balasubramaniam, Nayak e Bashar \(2020\)](#), [Liu, Xia e Lang \(2021\)](#);
6. *Bigrams Detection Algorithm*: É uma sequência de dois elementos adjacentes, por exemplo, palavras de um vetor. A distribuição de frequência de cada bigrama em uma *string* é comumente usada para análises simples de texto em muitas aplicações, incluindo detecção e correlação de termos. Utilizada em [Bahja e Safdar \(2020\)](#), [Mackey et al. \(2020a\)](#), [Gao, Xie e Li \(2021\)](#), [Al-Ramahi et al. \(2021\)](#), [Xue et al. \(2020\)](#), [Schück et al. \(2021\)](#), [Gerts et al. \(2021\)](#), [Mackey et al. \(2020b\)](#), [Guo et al. \(2020\)](#).

Para realizar a detecção de sentimentos e determinação de polaridade foram utilizadas várias abordagens, tanto através de algoritmos que identificaram os sentimentos envolvidos nas publicações quanto abordagens supervisionadas de aprendizado. Dentre os artigos analisados, aqueles que utilizaram essa técnica foram [Domalewska \(2021\)](#), [Almossa \(2021\)](#), [Bhatnagar e Choubey \(2021\)](#), [Forte e Pires \(2020\)](#), [Raheja e Asthana \(2021\)](#), [Pran et al. \(2020\)](#), [Yang et al. \(2020\)](#), [Boon-Itt e Skunkan \(2020\)](#), [Xu et al. \(2020\)](#), [Saha et al. \(2020\)](#), [Shen et al. \(2021\)](#), [Kamiński, Muth e Bogdański \(2020\)](#), [Morshed et al. \(2021\)](#), [Jiang et al. \(2021\)](#), [Xue et al. \(2020\)](#), [Das e Dutta \(2021\)](#). Alguns métodos como o *Valence Aware Dictionary for Sentiment Reasoning*, *Bag-Of-Words Models*, *BERT - Sentiment Analyst Approach* e *SentiStrength Algorithm* tiveram destaque. A lista com os artigos referentes a cada método é apresentada a seguir.

1. *Valence Aware Dictionary for Sentiment Reasoning*: É um modelo usado para análise de sentimento de um texto, sendo sensível tanto à polaridade positiva ou negativa; também quanto à intensidade da emoção. Utilizado em [Chen et al. \(2021\)](#), [Bahja e Safdar \(2020\)](#), [Chandrasekaran et al. \(2020\)](#), [Valdez et al. \(2020\)](#), [Zhou et al. \(2021\)](#), [Al-Rawi et al. \(2021\)](#), [Kirabo, Namara e Mcneese \(2021\)](#), [Viviani et al.](#)

- (2021), Gottipati et al. (2021), Gupta et al. (2020), Gao, Xie e Li (2021), Zhang et al. (2021), Cresswell et al. (2021), Hussain et al. (2021), Wrycza e Maślankowski (2020), Maia, Oliveira e Gallegos (2021), Xue et al. (2020), Wang et al. (2020), Adikari et al. (2021);
2. *Bag-Of-Words Models*: Neste modelo, o texto é representado como um conjunto de suas palavras, desconsiderando a estrutura gramatical e até mesmo a ordenação delas, mas mantendo sua multiplicidade. utilizado em Alzamzami e Saddik (2021), Zhang et al. (2021), Amanatidis et al. (2021);
 3. BERT - *Sentiment Analyst Approach*: A partir da identificação de palavras que essa abordagem proporciona, é possível realizar a identificação de polaridade e sentimento em sentenças encontradas em documentos de texto. Utilizada em Singh, Jakhar e Pandey (2021), Duong et al. (2020), Hussain et al. (2021);
 4. *SentiStrength Algorithm*: O *SentiStrength* é um classificador de sentimentos baseado no léxico, então usa regras linguísticas para fazer a detecção do sentimento em textos informais. Existe nele uma escala de 1 a 5, onde 1 é o sentimento em menor força e 5 o sentimento em maior força. Pode ser tanto positivo quanto negativo. Utilizado em Jelodar et al. (2020), Trust et al. (2020), Pinto et al. (2020).

As técnicas de agrupamento e detecção de comunidade geralmente são utilizadas para realizar análises voltadas para rede. Dentre os artigos analisados, aqueles que utilizaram essa técnica foram Nurlaila et al. (2020), Mackey et al. (2020a), Tao et al. (2020), Xu, Dredze e Broniatowski (2020), Liao et al. (2020), Slavik et al. (2021), Permana et al. (2021), Lima et al. (2021), Recuero e Soares (2021), Pinto et al. (2020), Ito e Chakraborty (2020), Elgazzar, Spurlock e Bogart (2021), Ceron et al. (2021), Mackey et al. (2020a), Mourad et al. (2020), Park, Park e Chong (2020), Ahmed et al. (2020), Chen et al. (2021), Glowacki, Wilcox e Glowacki (2021), Massaad e Cherfan (2020), Wrycza e Maślankowski (2020), Alshehabi et al. (2020), Kamiński, Muth e Bogdański (2020), Recuero, Soares e Zago (2021). Alguns métodos como o *Louvian Clustering*, *FluidC Algorithm*, *PageRank Algorithm*, *Occupation Detection Algorithm*, *Fruchterman Reingold Algorithm* e *Community Detection Methods* tiveram destaque. A lista com os artigos referentes é apresentada a seguir.

1. *Louvain Clustering Algorithm*: É um método para extrair comunidades de grandes redes. Ele maximiza a modularidade para cada comunidade, onde a modularidade quantifica a qualidade de uma atribuição de nós às comunidades. Utilizado em Lee et al. (2021), Kirabo, Namara e Mcneese (2021), Elgazzar, Spurlock e Bogart (2021);
2. *FluidC Algorithm*: É um novo algoritmo de detecção de comunidade baseado na propagação de termos, melhorando a qualidade do conjunto de dados com base

- no *feedback* positivo e determinando a ordem de atualização do nó. Utilizado em [Bhatnagar e Choubey \(2021\)](#);
3. *PageRank Algorithm*: É um algoritmo capaz de detectar várias comunidades onde os resultados são fornecidos por meio de uma análise teórica de uma possível rede. Utilizado em [Santis, Martino e Rizzi \(2020\)](#), [Goel e Sharma \(2021\)](#), [Alzamzami e Saddik \(2021\)](#);
 4. *Fruchterman Reingold Algorithm*: É um algoritmo que agrupa os vértices de um grafo de acordo com a distância de forma a reduzir as dimensões. O grafo em duas ou mais dimensões fornece uma detecção em relação aos vértices estarem conectados. Utilizado em [Paula e Mello \(2020\)](#);
 5. *Community Detection Methods*: Existem duas abordagens para os métodos de detecção da comunidade, métodos aglomerativos e métodos divisivos. Nos métodos aglomerativos, as arestas são adicionadas uma a uma a um grafo que contém apenas nós. Nos métodos divisivos acontece o oposto, as arestas são removidas uma a uma de um gráfico completo. Utiliza em ([Goel; Sharma, 2021](#); [Al-Shargabi; Selmi, 2021](#); [Chong; Park, 2021](#); [Yang; Han, 2021](#); [Melo, 2021](#); [Maia; Oliveira; Gallegos, 2021](#); [Tao et al., 2020](#)).

Outra técnica utilizada foi a de detecção de mudanças temporais, ela foi utilizada para prever eventos ou tendências. Dentre os artigos analisados, aqueles que utilizaram essa técnica foram [Li et al. \(2020\)](#), [Mackey et al. \(2020a\)](#), [Wang et al. \(2021\)](#), [Lee et al. \(2020\)](#), [Li, Chaudhary e Zhang \(2020\)](#), [Valdez et al. \(2020\)](#), [Liao et al. \(2020\)](#), [Daughton et al. \(2021\)](#), [Adikari et al. \(2021\)](#), [Hasni e Faiz \(2021\)](#), [Wahbeh et al. \(2020\)](#), [Rosenberg \(2020\)](#), [Morshed et al. \(2021\)](#), [Jiang et al. \(2021\)](#). Alguns dos métodos utilizados para realizar essa técnica foram *Augmented Dickey-Fuller Algorithm*, *Smoothing Algorithm*, *Altmetric Attention Scores Algorithm*, *Topic Evolution Detection Algorithm*, *Traffic Changes Methods*, *Prediction Methods* e *Time Series Variation Algorithm*. A lista com os artigos referentes é apresentada a seguir:

1. *Augmented Dickey-Fuller Algorithm*: É um algoritmo aplicado em séries temporais. A estatística Dickey-Fuller se trata de um número negativo. Quanto mais negativo for, mais forte será a rejeição da hipótese que há uma raiz unitária. Possibilitando identificar se uma série está estacionada ou não. Utilizado em [Caldera, Meedin e Perera \(2020\)](#), [Xu et al. \(2020\)](#);
2. *Smoothing Algorithms*: Os algoritmos relacionados a suavização, são utilizados para suavizar sinais irregulares e possibilitar a leitura de um sinal mais claro nas séries temporais. Utilizado em [Elgazzar, Spurlock e Bogart \(2021\)](#);

3. *Altmetric Attention Scores Algorithm*: É um algoritmo que a partir da contagem ponderada, calcula a quantidade de atenção que certo assunto recebeu. Utilizado em [Chong e Park \(2021\)](#);
4. *Box-Jenkins Forecasting Algorithm*: É um algoritmo de previsão que usa estudos de regressão em dados de séries temporais. Ele é baseado na hipótese de que ocorrências passadas influenciam as futuras. Utilizado em [Caldera, Meedin e Perera \(2020\)](#);
5. *Topic Evolution Detection Algorithms*: São algoritmos relacionados a detectar as mudanças temporais de assuntos específicos. Dessa forma é possível detectar as tendências. Utilizado em [Chang, Monselise e Yang \(2021\)](#);
6. *Traffic Changes Methods*: São métodos utilizados para detectar as mudanças de tráfego, dessa forma é possível observar o que está sendo mais acessado ou mais buscado. É possível observar as tendências a partir dessas mudanças notáveis. Utilizado em [Böttger, Ibrahim e Vallis \(2020\)](#);
7. *Prediction Methods*: São métodos que visam fazer uma declaração sobre um evento ou dados futuros. Se trata de fazer uma estimativa se um evento pode ou não acontecer. Utilizado em [Shan e Li \(2020\)](#), [Li et al. \(2020\)](#), [Shen et al. \(2020\)](#);
8. *Time Series Variation Algorithm*: É um algoritmo que visa identificar a variação em uma série temporal, possibilitando a identificação de possíveis tendências. Utilizado em [\(Saha et al., 2020\)](#).

Além dos métodos e técnicas, foram utilizados também *softwares/ferramentas* para realizar as análises. Dentre os artigos analisados, aqueles que utilizaram essa técnica foram [Al-Ramahi et al. \(2021\)](#), [Wahbeh et al. \(2020\)](#), [Domalewska \(2021\)](#), [Romeiro Paulino e Pires Ventura \(2021\)](#), [Jang et al. \(2021\)](#), [Liao et al. \(2020\)](#), [Huang et al. \(2020\)](#), [Gerts et al. \(2021\)](#), [Kwok, Vadde e Wang \(2021\)](#), [Glowacki, Wilcox e Glowacki \(2021\)](#), [Ruffer, Knitza e Krusche \(2020\)](#), [Pastor \(2020\)](#), [Gonçalves et al. \(2021\)](#), [Leão et al. \(2020\)](#), [Santos, Carmargo e Menossi \(2020\)](#). Alguns dos *softwares/ferramentas* mais utilizados foram *WEKA*, *SenticNet*, *NodeXL*, *TextBlob Tool*, *QDA Miner*, *WordStat 8*, *IBM Analyze*, *Syuzhet Tool*. A lista com os artigos referentes é apresentada a seguir:

1. *WEKA*: [Yigitcanlar et al. \(2020\)](#), [Permana et al. \(2021\)](#);
2. *SenticNet*: [\(Valle Cruz et al., 2021\)](#); [Maia; Oliveira; Gallegos, 2021\)](#);
3. *NodeXL*: [Chong e Park \(2021\)](#), [Park, Park e Chong \(2020\)](#);
4. *TextBlob Tool*: [Singh, Jakhar e Pandey \(2021\)](#), [Gottipati et al. \(2021\)](#), [Kaur, Kaul e Zadeh \(2020\)](#), [Gupta et al. \(2020\)](#), [Hussain et al. \(2021\)](#), [Liu, Xia e Lang \(2021\)](#), [Pinto et al. \(2020\)](#);

5. *QDA Miner WordStat 8*: Al-Rawi et al. (2021);
6. *IBM Analyze* Kaur, Kaul e Zadeh (2020), Kaur, Kaul e Zadeh (2020), Nurbakova, Ermakova e Ovchinnikova (2020);
7. *Syuzhet Tool*: Goel e Sharma (2021), Tao, Miao e Ng (2020).

Também foi possível observar modelos recentes ou pouco utilizados e que contribuíram para a Análise de Dados no contexto da pandemia da COVID-19, um deles é o COVID *TwitterBERTmodel* e outro é o modelo SIR - *Susceptible Infected Recovered*. O COVID *TwitterBERTmodel* foi importante pois é um modelo que foi treinado a partir desse contexto da pandemia, ele foi utilizado por exemplo para prever possíveis casos de COVID-19 (Klein et al., 2021) ou fornecer subsídios para a Análise de Sentimentos através de uma excelente precisão em relação a polaridade e aos sentimentos (Viviani et al., 2021). O modelo SIR não é um modelo criado recentemente, mas a forma com que foi usado no contexto atual é um grande diferencial. Ele consiste de um modelo que é capaz de calcular a taxa de difusão, nesse contexto foi possível ver a taxa de difusão de mitos sobre a COVID-19 (Yang et al., 2020), por exemplo.

5.6 Apresentação dos Resultados

Após realizar o Processamento dos Dados e também aplicar alguma Técnica de Análise, pode ser feita a Validação do Modelo e também a Apresentação dos Resultados obtidos. Nos artigos analisados foi possível identificar uma forma principal de Validação do Modelo, sendo ela a Validação Cruzada.

A Validação Cruzada é uma técnica onde os pesquisadores particionam o conjunto de dados em subconjuntos de tamanho praticamente iguais. Esses subconjuntos são utilizados para fazer o treinamento de um preditor, o qual será utilizado para testar os outros subconjuntos. Esse processo será repetido várias vezes e o desempenho será sempre anotado, de forma que ao final dessas execuções se obtenha a média dos desempenhos em cada subconjunto. Dessa forma, será possível observar qual é o modelo mais adequado para realizar o processamento dos dados e análise (Faceli et al., 2011). Alguns dos trabalhos aonde foram feitas Validações Cruzadas são Hadj Ameer e Aliane (2021), Cotfas et al. (2021), Hasni e Faiz (2021), Gupta et al. (2020), Amanatidis et al. (2021), Chang, Monselise e Yang (2021), Yang, Zhou e Zafarani (2021), Agarwal et al. (2020), Shan e Li (2020), Nguyen e Chaturvedi (2020), Naseem et al. (2021), Sari e Ruldeviyani (2020), Maia, Oliveira e Gallegos (2021), Li, Chaudhary e Zhang (2020), Chandrasekaran et al. (2020), Wang et al. (2021), Alzamzami e Saddik (2021), Pran et al. (2020), Duong et al. (2020), Long, Alharthi e Saddik (2020), Caldera, Meedin e Perera (2020), Chen et al. (2021).

Os dados foram apresentados de diversas formas, onde foi possível identificar três formas principais sendo elas tabelas, gráficos e *Word Cloud* (Elgazzar; Spurlock; Bogart, 2021; Valle Cruz et al., 2021; Goel; Sharma, 2021; Singh; Jakhar; Pandey, 2021; Romeiro Paulino; Pires Ventura, 2021; Raheja; Asthana, 2021; Balasubramaniam; Nayak; Bashar, 2020; Cotfas et al., 2021; Bahja; Safdar, 2020; Malagoli et al., 2021; Lee et al., 2020). Foram apresentados gráficos contendo comparações de sentimentos antes e depois da pandemia, tabelas com a contagem de *tweets* por categoria de tópico, ferramentas para visualização (ex.: *NetworkX Library*, *MatplotLib*, *WEKA Visualization*, *Gephi*) (Elgazzar; Spurlock; Bogart, 2021; Yigitcanlar et al., 2020; Amara; Taieb; Aouicha, 2021; Ceron et al., 2021; Romeiro Paulino; Pires Ventura, 2021; Liao et al., 2020; Massaad; Cherfan, 2020; Permana et al., 2021; Huang et al., 2020; Zhao et al., 2020; Mourad et al., 2020; Mohamed Ridhwan; Hargreaves, 2021) e principalmente *Word Cloud*, que é um agrupamento visual de palavras onde as palavras mais importantes e relevantes são as maiores. A Figura 28 exemplifica o uso do *Word Cloud*, no qual os títulos dos trabalhos analisados foram utilizados para gerá-lo.

Também foram utilizados algoritmos que permitiram a visualização de resultados relacionados a agrupamentos. Alguns dos algoritmos são *Harel-Koren Fast Multiscale Layout Algorithm* (Chong; Park, 2021; Park; Park; Chong, 2020) e *UMAP Algorithm* (Lee et al., 2021).

Figura 28 – *Word Cloud* utilizando o título dos artigos analisados.



Fonte: Gráfico elaborado pela autora através do *Word Cloud Generator* (Chrome, 2020).

6 Conclusão

Neste trabalho, as pesquisas que realizaram a Análise de Dados de Redes Sociais no contexto da pandemia da COVID-19 foram reunidos e analisados. O objetivo foi realizar uma revisão sistemática da literatura para contribuir na identificação das abordagens existentes no contexto da pandemia da COVID-19 para apoio à Análise de Dados de Redes Sociais. Para alcançar esse objetivo foi estabelecido um protocolo para a execução da revisão. O protocolo consistiu em:

1. Estudo primário;
2. Coletar artigos relacionados ao tema;
3. Aplicar os critérios de inclusão/exclusão;
4. Realizar a Análise Quantitativa;
5. Realizar a Análise Qualitativa;
6. Redação.

Na etapa de Análise Quantitativa foi feito um estudo quantitativo em relação às questões de pesquisa definidas na Seção 4.1 do Capítulo 4, foram apresentados gráficos que permitiram analisar quais foram as categorias de métodos de análise mais exploradas, como a distribuição de artigos ao curso da pandemia se parece, quais países mais contribuíram, entre outros. Na etapa de Análise Qualitativa foram abordados os métodos mais utilizados nas pesquisas selecionadas a partir do processo de análise de dados apresentado na Seção 5.1 do Capítulo 5, sendo ele: determinar a rede social, extração de dados, processamento dos dados, aplicar técnicas de análise e apresentação dos resultados. Dessa forma foi possível realizar uma síntese dos artigos selecionados.

6.1 Trabalhos Futuros

Neste trabalho foi realizada uma revisão sistemática da literatura com pesquisas relacionadas a Análise de Dados de Redes Sociais no contexto da pandemia da COVID-19. Foram incluídos trabalhos estritamente relacionados à análise de redes sociais que permitem um alto poder de conexão e interação pública entre usuários do mundo todo, por exemplo o *Twitter*, *Facebook*, *Reddit*, entre outras.

Estudos futuros poderão incluir mais tipos de redes sociais, principalmente mídias sociais que são plataformas voltadas para promoção de conteúdo. Alguns exemplos de

mídias sociais são o *Pinterest*, *Youtube*, etc. Além de outras redes sociais, poderão ser incluídas mais bases de dados a serem utilizadas para a coleta de artigos.

Referências

- Abd-Alrazaq, A. et al. Top concerns of tweeters during the covid-19 pandemic: Infveillance study. *J Med Internet Res*, v. 22, n. 4, p. e19016, Apr 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/4/e19016/>>. Citado 3 vezes nas páginas 49, 53 e 57.
- Abdelminaam, D. S. et al. Coaid-deep: An optimized intelligent framework for automated detecting covid-19 misleading information on twitter. *IEEE Access*, v. 9, p. 27840–27867, 2021. ISSN 2169-3536. Citado 3 vezes nas páginas 53, 54 e 55.
- ACM. *ACM Digital Library Resource Center*. 2020. Acessado em 10 de Outubro de 2021. Disponível em: <<https://libraries.acm.org/>>. Citado na página 26.
- Adikari, A. et al. Emotions of covid-19: Content analysis of self-reported information using artificial intelligence. *J Med Internet Res*, v. 23, n. 4, p. e27341, Apr 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/4/e27341>>. Citado 6 vezes nas páginas 49, 53, 57, 58, 59 e 60.
- Agarwal, A. et al. Leveraging natural language processing to mine issues on twitter during the covid-19 pandemic. p. 886–891, Dec 2020. Citado 4 vezes nas páginas 53, 54, 57 e 62.
- Ahmed, W. et al. Covid-19 and the 5g conspiracy theory: Social network analysis of twitter data. *J Med Internet Res*, v. 22, n. 5, p. e19458, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e19458/>>. Citado 3 vezes nas páginas 52, 57 e 59.
- Al-Ramahi, M. et al. Public discourse against masks in the covid-19 era: Infodemiology study of twitter data. *JMIR Public Health Surveill*, v. 7, n. 4, p. e26780, Apr 2021. ISSN 2369-2960. Disponível em: <<https://publichealth.jmir.org/2021/4/e26780>>. Citado 4 vezes nas páginas 53, 57, 58 e 61.
- Al-Rawi, A. et al. Investigating public discourses around gender and covid-19: a social media analysis of twitter data. *Journal of Healthcare Informatics Research*, v. 5, n. 3, p. 249–269, Sep 2021. ISSN 2509-498X. Disponível em: <<https://doi.org/10.1007/s41666-021-00102-x>>. Citado 7 vezes nas páginas 49, 50, 52, 54, 58, 59 e 62.
- Al-Shargabi, A. A.; Selmi, A. Social network analysis and visualization of arabic tweets during the covid-19 pandemic. *IEEE Access*, v. 9, p. 90616–90630, 2021. ISSN 2169-3536. Citado 3 vezes nas páginas 49, 52 e 60.
- Alanazi, E. et al. Identifying and ranking common covid-19 symptoms from tweets in arabic: Content analysis. *J Med Internet Res*, v. 22, n. 11, p. e21329, Nov 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/11/e21329/>>. Citado 3 vezes nas páginas 49, 52 e 57.
- Almossa, S. Y. University students’ perspectives toward learning and assessment during covid-19. *Education and Information Technologies*, Apr 2021. ISSN 1573-7608. Disponível em: <<https://doi.org/10.1007/s10639-021-10554-8>>. Citado 4 vezes nas páginas 49, 52, 57 e 58.

Alshehabi, S. et al. Analysis of twitter data using evolutionary clustering during the covid-19 pandemic. *Computers, Materials & Continua*, v. 65, 07 2020. Citado 3 vezes nas páginas 49, 57 e 59.

Alzamzami, F.; Saddik, A. E. Monitoring cyber sentihate social behavior during covid-19 pandemic in north america. *IEEE Access*, v. 9, p. 91184–91208, 2021. ISSN 2169-3536. Citado 7 vezes nas páginas 52, 54, 55, 57, 59, 60 e 62.

Amanatidis, D. et al. Mining textual and imagery instagram data during the covid-19 pandemic. *Applied Sciences*, v. 11, n. 9, 2021. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/11/9/4281>>. Citado 5 vezes nas páginas 49, 50, 54, 59 e 62.

Amara, A.; Taieb, M. A. H.; Aouicha, M. B. Multilingual topic modeling for tracking covid-19 trends based on facebook data analysis. *Applied Intelligence*, v. 51, n. 5, p. 3052–3073, May 2021. ISSN 1573-7497. Disponível em: <<https://doi.org/10.1007/s10489-020-02033-3>>. Citado 3 vezes nas páginas 52, 57 e 63.

Bahja, M.; Safdar, G. A. Unlink the link between covid-19 and 5g networks: An nlp and sna based approach. *IEEE Access*, v. 8, p. 209127–209137, 2020. ISSN 2169-3536. Citado 5 vezes nas páginas 49, 57, 58, 59 e 63.

Balasubramaniam, T.; Nayak, R.; Bashar, M. A. Understanding the spatio-temporal topic dynamics of covid-19 using nonnegative tensor factorization: A case study. p. 1218–1225, Dec 2020. Citado 5 vezes nas páginas 49, 52, 57, 58 e 63.

BBC. *CoronaVac: os quatro países além do Brasil que planejam usar a vacina contra covid-19*. BBC, 2020. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.bbc.com/portuguese/internacional-55224934>>. Citado na página 33.

Bhatnagar, S.; Choubey, N. Making sense of tweets using sentiment analysis on closely related topics. *Social Network Analysis and Mining*, v. 11, n. 1, p. 44, May 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00752-0>>. Citado 7 vezes nas páginas 49, 53, 54, 55, 57, 58 e 60.

Blei, D. M. *Probabilistic Topic Models*. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 55, n. 4, p. 77–84, abr. 2012. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/2133806.2133826>>. Citado na página 21.

Boon-Itt, S.; Skunkan, Y. Public perception of the covid-19 pandemic on twitter: Sentiment analysis and topic modeling study. *JMIR Public Health Surveill*, v. 6, n. 4, p. e21978, Nov 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/4/e21978/>>. Citado 6 vezes nas páginas 49, 52, 53, 54, 57 e 58.

Brum, P. V. et al. A characterization of portuguese tweets regarding the covid-19 pandemic. SBC, Porto Alegre, RS, Brasil, p. 177–184, 2020. ISSN 2763-8944. Disponível em: <<https://sol.sbc.org.br/index.php/kdmile/article/view/11974>>. Citado 3 vezes nas páginas 49, 52 e 58.

Böttger, T.; Ibrahim, G.; Vallis, B. How the internet reacted to covid-19: A perspective from facebook's edge network. Association for Computing Machinery, New York, NY, USA, p. 34–41, 2020. Disponível em: <<https://doi.org/10.1145/3419394.3423621>>. Citado 3 vezes nas páginas 49, 52 e 61.

- Caldera, H. M. M.; Meedin, G. S. N.; Perera, I. Time series based trend analysis for hate speech in twitter during covid 19 pandemic. p. 1–2, Nov 2020. ISSN 2472-7598. Citado 4 vezes nas páginas 52, 60, 61 e 62.
- Canva. *What will you design?* 2012. Acessado em 18 de Outubro de 2021. Disponível em: <<https://www.canva.com/>>. Citado 7 vezes nas páginas 16, 17, 21, 22, 25, 45 e 50.
- Ceron, W. et al. Covid-19 fake news diffusion across latin america. *Social Network Analysis and Mining*, v. 11, n. 1, p. 47, May 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00753-z>>. Citado 3 vezes nas páginas 52, 59 e 63.
- Chandrasekaran, R. et al. Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. *J Med Internet Res*, v. 22, n. 10, p. e22624, Oct 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/10/e22624/>>. Citado 6 vezes nas páginas 53, 55, 57, 58, 59 e 62.
- Chang, C.-H.; Monselise, M.; Yang, C. C. What are people concerned about during the pandemic? detecting evolving topics about covid-19 from twitter. *Journal of Healthcare Informatics Research*, v. 5, n. 1, p. 70–97, Mar 2021. ISSN 2509-498X. Disponível em: <<https://doi.org/10.1007/s41666-020-00083-3>>. Citado 5 vezes nas páginas 48, 57, 58, 61 e 62.
- Chen, E.; Lerman, K.; Ferrara, E. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health Surveill*, v. 6, n. 2, p. e19273, May 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/2/e19273/>>. Citado 3 vezes nas páginas 49, 53 e 57.
- Chen, H. et al. Country image in covid-19 pandemic: A case study of china. *IEEE Transactions on Big Data*, v. 7, n. 1, p. 81–92, March 2021. ISSN 2332-7790. Citado 6 vezes nas páginas 49, 52, 54, 58, 59 e 62.
- Chen, P.-L.; Cheng, Y.; Chen, K. *Analysis of Social Media Data: An Introduction to the Characteristics and Chronological Process*. p. 297–321, 01 2018. Citado na página 45.
- Chen, S. et al. A novel machine learning framework for comparison of viral covid-19-related sina weibo and twitter posts: Workflow development and content analysis. *J Med Internet Res*, v. 23, n. 1, p. e24889, Jan 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/1/e24889>>. Citado 4 vezes nas páginas 52, 54, 57 e 59.
- Chong, M.; Park, H. W. Covid-19 in the twitterverse, from epidemic to pandemic: information-sharing behavior and twitter as an information carrier. *Scientometrics*, v. 126, n. 8, p. 6479–6503, Aug 2021. ISSN 1588-2861. Disponível em: <<https://doi.org/10.1007/s11192-021-04054-2>>. Citado 5 vezes nas páginas 49, 52, 60, 61 e 63.
- Chrome. *Word Cloud Generator™*. Chrome, 2020. Acessado em 13 de Outubro de 2021. Disponível em: <<https://chrome.google.com/webstore/detail/word-cloud-generator/abeaddpndnoipjlbkchghbpoilcmkdoj?hl=pt-BR>>. Citado na página 63.
- CNPq; CAPES; FINEP. *Nova Tabela Das Áreas Do Conhecimento*. 2005. Disponível em: <https://www.ppg.uema.br/uploads/files/cee-areas_do_conhecimento.pdf>. Citado na página 38.

- Cotfas, L.-A. et al. The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access*, v. 9, p. 33203–33223, 2021. ISSN 2169-3536. Citado 4 vezes nas páginas 52, 54, 62 e 63.
- Cousineau, D.; Chartier, S. Outliers detection and treatment: a review. *International Journal of Psychological Research*, v. 3, n. 1, p. 58–67, 2010. Disponível em: <<https://revistas.usb.edu.co/index.php/IJPR/article/view/844>>. Citado na página 23.
- Cresswell, K. et al. Understanding public perceptions of covid-19 contact tracing apps: Artificial intelligence-enabled social media analysis. *J Med Internet Res*, v. 23, n. 5, p. e26618, May 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/5/e26618>>. Citado 7 vezes nas páginas 48, 49, 52, 54, 57, 58 e 59.
- Creswell, J. W.; Clark, V. L. P. *Designing and conducting mixed methods research*. SAGE Publications, 2017. 188-206 p. Disponível em: <<https://us.sagepub.com/en-us/sam/designing-and-conducting-mixed-methods-research/book241842>>. Acesso em: 16 maio 2021. Citado na página 16.
- Danfá, L. *Alteridade, racismo e representações sociais: o caso do ebola no Brasil*. AT-TENA: Repositório Digita da UFPE, 2015. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/17791>>. Acesso em: 02 maio 2021. Citado na página 12.
- Das, S.; Dutta, A. Characterizing public emotions and sentiments in covid-19 environment: A case study of india. *Journal of Human Behavior in the Social Environment*, Routledge, v. 31, n. 1-4, p. 154–167, 2021. Disponível em: <<https://doi.org/10.1080/10911359.2020.1781015>>. Citado 3 vezes nas páginas 52, 57 e 58.
- Daughton, A. R. et al. Mining and validating social media data for covid-19–related human behaviors between january and july 2020: Infodemiology study. *J Med Internet Res*, v. 23, n. 5, p. e27059, May 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/5/e27059>>. Citado 4 vezes nas páginas 53, 54, 57 e 60.
- Domalewska, D. An analysis of covid-19 economic measures and attitudes: evidence from social media mining. *Journal of Big Data*, v. 8, n. 1, p. 42, Mar 2021. Disponível em: <<https://doi.org/10.1186/s40537-021-00431-z>>. Citado 5 vezes nas páginas 49, 52, 53, 58 e 61.
- DOOGAN, C. et al. Public perceptions and attitudes toward covid-19 nonpharmaceutical interventions across six countries: A topic modeling analysis of twitter data. *J Med Internet Res*, v. 22, n. 9, p. e21419, Sep 2020. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2020/9/e21419>>. Citado 3 vezes nas páginas 49, 53 e 57.
- Duong, V. et al. The ivory tower lost: How college students respond differently than the general public to the covid-19 pandemic. p. 126–130, Dec 2020. ISSN 2473-991X. Citado 6 vezes nas páginas 52, 54, 55, 57, 59 e 62.
- Elgazzar, H.; Spurlock, K.; Bogart, T. Evolutionary clustering and community detection algorithms for social media health surveillance. *Machine Learning with Applications*, v. 6, p. 100084, 2021. ISSN 2666-8270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2666827021000426>>. Citado 5 vezes nas páginas 49, 53, 59, 60 e 63.
- Facebook. *It's quick and easy*. 2004. Acessado em 13 de Outubro de 2021. Disponível em: <<https://www.facebook.com/>>. Citado na página 12.

- Faceli, K. et al. *Inteligência artificial: uma abordagem de aprendizado de máquina*. [S.l.]: LTC, 2011. Citado na página 62.
- Fazenda, I. et al. *O que é Interdisciplinaridade?* Cortez Editora, v. 1, 2008. Citado na página 38.
- Forte, E. C. N.; Pires, D. Nursing appeals on social media in times of coronavirus. *Revista Brasileira de Enfermagem [online]*., v. 9, n. 2021, p. 2020–0225, October 2020. Disponível em: <<https://doi.org/10.1590/0034-7167-2020-0225>>. Citado 5 vezes nas páginas 49, 54, 55, 57 e 58.
- G1. *OMS declara pandemia de coronavírus*. Globo, 2020. Acessado em 10 de Outubro de 2021. Disponível em: <<https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de-coronavirus.ghtml>>. Citado 2 vezes nas páginas 33 e 43.
- Galvão, M. C. B.; Pluye, P.; Ricarte, I. L. M. *Métodos de pesquisa mistos e revisões de literatura mistas: conceitos, construção e critérios de avaliação*. INCID: Revista de ciência da informação e documentação, p. 188–206, 2017. Disponível em: <<https://doi.org/10.11606/issn.2178-2075.v8i2p4-24>>. Citado 2 vezes nas páginas 16 e 17.
- Galvão, M. C. B.; Ricarte, I. L. M. *Revisão Sistemática da Literatura: Conceituação, produção e publicação*. Logeion: Filosofia da Informação, 2019. Disponível em: <<http://revista.ibict.br/fiinf/article/view/4835>>. Acesso em: 02 maio 2021. Citado 2 vezes nas páginas 13 e 15.
- Gao, Y.; Xie, Z.; Li, D. Electronic cigarette users' perspective on the covid-19 pandemic: Observational study using twitter data. *JMIR Public Health Surveill*, v. 7, n. 1, p. e24859, Jan 2021. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2021/1/e24859/>>. Citado 4 vezes nas páginas 52, 57, 58 e 59.
- Gerts, D. et al. "thought i'd share first" and other conspiracy theory tweets from the covid-19 infodemic: Exploratory study. *JMIR Public Health Surveill*, v. 7, n. 4, p. e26527, Apr 2021. ISSN 2369-2960. Disponível em: <<https://publichealth.jmir.org/2021/4/e26527>>. Citado 5 vezes nas páginas 49, 53, 55, 58 e 61.
- Glowacki, E. M.; Wilcox, G. B.; Glowacki, J. B. Identifying #addiction concerns on twitter during the covid-19 pandemic: A text mining analysis. *Substance Abuse*, Taylor & Francis, v. 42, n. 1, p. 39–46, 2021. PMID: 32970973. Disponível em: <<https://doi.org/10.1080/08897077.2020.1822489>>. Citado 2 vezes nas páginas 59 e 61.
- Goel, R.; Sharma, R. Studying leaders & their concerns using online social media during the times of crisis - a covid case study. *Social Network Analysis and Mining*, v. 11, n. 1, p. 46, May 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00756-w>>. Citado 4 vezes nas páginas 57, 60, 62 e 63.
- Gonçalves, M. I. A. et al. Tempos de pandemia: educação em saúde via redes sociais. *Revista de Extensão da UPE*, v. 6, n. 1, p. 38–45, jun. 2021. Disponível em: <<https://www.revistaextensao.upe.br/index.php/reupe/article/view/145>>. Citado 2 vezes nas páginas 50 e 61.
- Google. *Google Acadêmico*. 2004. Acessado em 10 de Outubro de 2021. Disponível em: <<https://scholar.google.com.br/?hl=pt>>. Citado na página 27.

- Gottipati, S. et al. Analyzing tweets on new norm: Work from home during covid-19 outbreak. p. 0500–0507, Jan 2021. Citado 5 vezes nas páginas 53, 57, 58, 59 e 61.
- Graziosi, M. E. S.; Liebano, R. E.; Nahas, F. X. *Pesquisa em Bases de Dados*. Acervo de Recursos Educacionais em Saúde, 2012. Disponível em: <<http://ares.unasus.gov.br/acervo/handle/ARES/232>>. Citado na página 26.
- Guo, J.-W. et al. Mining twitter to explore the emergence of covid-19 symptoms. *Public Health Nursing*, v. 37, n. 6, p. 934–940, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/phn.12809>>. Citado 3 vezes nas páginas 49, 53 e 58.
- Gupta, P. et al. Sentiment analysis of lockdown in india during covid-19: A case study on twitter. *IEEE Transactions on Computational Social Systems*, v. 8, n. 4, p. 992–1002, Aug 2020. ISSN 2329-924X. Citado 6 vezes nas páginas 53, 54, 58, 59, 61 e 62.
- Hadj Ameer, M. S.; Aliane, H. Aracovid19-mfh: Arabic covid-19 multi-label fake news & hate speech detection dataset. *Procedia Computer Science*, v. 189, p. 232–241, 2021. ISSN 1877-0509. AI in Computational Linguistics. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050921012059>>. Citado 6 vezes nas páginas 47, 48, 52, 53, 57 e 62.
- Hasni, S.; Faiz, S. Word embeddings and deep learning for location prediction: tracking coronavirus from british and american tweets. *Social Network Analysis and Mining*, v. 11, n. 1, p. 66, Jul 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00777-5>>. Citado 7 vezes nas páginas 49, 52, 53, 54, 55, 60 e 62.
- Hou, Q.; Han, M.; Cai, Z. *Survey on data analysis in social media: A practical application aspect*. *Big Data Mining and Analytics*, v. 3, n. 4, p. 259–279, 2020. Acesso em: 21 maio 2021. Citado 5 vezes nas páginas 18, 20, 21, 23 e 24.
- Huang, C. et al. Mining the characteristics of covid-19 patients in china: Analysis of social media posts. *J Med Internet Res*, v. 22, n. 5, p. e19087, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e19087/>>. Citado 4 vezes nas páginas 49, 52, 61 e 63.
- Hussain, A. et al. Artificial intelligence-enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study. *J Med Internet Res*, v. 23, n. 4, p. e26627, Apr 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/4/e26627>>. Citado 5 vezes nas páginas 49, 53, 58, 59 e 61.
- IEEE. *The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity*. 2000. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.ieee.org/>>. Citado na página 26.
- Instagram. *Capture and share the world's moments*. 2010. Acessado em 13 de Outubro de 2021. Disponível em: <<https://www.instagram.com/>>. Citado na página 12.
- Ito, H.; Chakraborty, B. Social media mining with dynamic clustering: A case study by covid-19 tweets. p. 1–6, Dec 2020. ISSN 2325-5994. Citado 6 vezes nas páginas 48, 49, 53, 54, 57 e 59.

Jang, H. et al. Tracking covid-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis. *J Med Internet Res*, v. 23, n. 2, p. e25431, Feb 2021. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2021/2/e25431/>>. Citado 5 vezes nas páginas 49, 50, 53, 57 e 61.

Jelodar, H. et al. Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics*, v. 24, n. 10, p. 2733–2742, Oct 2020. ISSN 2168-2208. Citado 5 vezes nas páginas 53, 54, 55, 57 e 59.

Jiang, T. et al. Impact of the covid-19 pandemic on patient preferences and decision making for symptomatic urolithiasis. *Journal of Endourology*, v. 35, n. 8, p. 1250–1256, 2021. PMID: 33478351. Disponível em: <<https://doi.org/10.1089/end.2020.1141>>. Citado 5 vezes nas páginas 49, 52, 57, 58 e 60.

JMIR. *JMIR Journal of Medical Internet Research*. 1999. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.jmir.org/>>. Citado na página 26.

Kamiński, M.; Muth, A.; Bogdański, P. Smoking, vaping, and tobacco industry during covid-19 pandemic: Twitter data analysis. *Cyberpsychology, Behavior, and Social Networking*, v. 23, n. 12, p. 811–817, 2020. PMID: 32757951. Disponível em: <<https://doi.org/10.1089/cyber.2020.0384>>. Citado 3 vezes nas páginas 52, 58 e 59.

KARAMI, A.; ANDERSON, M. Social media and covid-19: Characterizing anti-quarantine comments on twitter. *Proceedings of the Association for Information Science and Technology*, v. 57, n. 1, p. e349, 2020. Disponível em: <<https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/prai.2020.0349>>. Citado 2 vezes nas páginas 52 e 57.

Kaur, S.; Kaul, P.; Zadeh, P. M. Monitoring the dynamics of emotions during covid-19 using twitter data. *Procedia Computer Science*, v. 177, p. 423–430, 2020. ISSN 1877-0509. The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) / The 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020) / Affiliated Workshops. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877050920323243>>. Citado 3 vezes nas páginas 49, 52 e 62.

Kaur, S.; Kaul, P.; Zadeh, P. M. Study the impact of covid-19 on twitter users with respect to social isolation. p. 1–6, Dec 2020. Citado 5 vezes nas páginas 48, 49, 52, 61 e 62.

Kirabo, L.; Namara, M.; Mcneese, N. The power of the blue tick (): Ugandans' experiences and engagement on twitter at the onset of the covid-19 pandemic. *Association for Computing Machinery*, New York, NY, USA, p. 84–93, 2021. Disponível em: <<https://doi.org/10.1145/3448696.3448746>>. Citado 3 vezes nas páginas 57, 58 e 59.

Klein, A. Z. et al. Toward using twitter for tracking covid-19: A natural language processing pipeline and exploratory data set. *J Med Internet Res*, v. 23, n. 1, p. e25314, Jan 2021. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2021/1/e25314/>>. Citado 3 vezes nas páginas 49, 53 e 62.

Kwok, S. W. H.; Vadde, S. K.; Wang, G. Tweet topics and sentiments relating to covid-19 vaccination among australian twitter users: Machine learning analysis. *J Med Internet Res*, v. 23, n. 5, p. e26953, May 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/5/e26953>>. Citado 4 vezes nas páginas 49, 53, 57 e 61.

- Lee, C. et al. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In: _____. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021. ISBN 9781450380966. Disponível em: <<https://doi.org/10.1145/3411764.3445211>>. Citado 7 vezes nas páginas 48, 49, 50, 52, 57, 59 e 63.
- Lee, H. et al. Determining public opinion of the covid-19 pandemic in south korea and japan: Social network mining on twitter. *hir*, v. 26, n. 4, p. 335–343, 2020. Disponível em: <<https://synapse.koreamed.org/articles/1145365?viewtype=pubreader>>. Citado 4 vezes nas páginas 52, 57, 60 e 63.
- Leskovec, J.; Adamic, L. A.; Huberman, B. A. *The dynamics of viral marketing*. ACM Transactions on the Web (TWEB), 2007. Disponível em: <<https://cs.stanford.edu/~jure/pubs/viral-tweb.pdf>>. Acesso em: 02 maio 2021. Citado na página 12.
- Leão, C. F. et al. Covid-19 em revistas no instagram: uma análise de conteúdo relacionada às orientações do unicef e da oms. *Paradoxos*, v. 5, n. 1, p. 20–35, jul. 2020. Disponível em: <<http://www.seer.ufu.br/index.php/paradoxos/article/view/55104>>. Citado 3 vezes nas páginas 49, 52 e 61.
- Li, D.; Chaudhary, H.; Zhang, Z. Modeling spatiotemporal pattern of depressive symptoms caused by covid-19 using social media data mining. *International Journal of Environmental Research and Public Health*, v. 17, n. 14, 2020. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/17/14/4988>>. Citado 5 vezes nas páginas 53, 54, 57, 60 e 62.
- Li, J. et al. Data mining and content analysis of the chinese social media platform weibo during the early covid-19 outbreak: Retrospective observational infoveillance study. *JMIR Public Health Surveill*, v. 6, n. 2, p. e18700, Apr 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/2/e18700/>>. Citado 4 vezes nas páginas 49, 52, 60 e 61.
- Liao, Q. et al. Public engagement and government responsiveness in the communications about covid-19 during the early epidemic stage in china: Infodemiology study on social media data. *J Med Internet Res*, v. 22, n. 5, p. e18796, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e18796/>>. Citado 7 vezes nas páginas 49, 52, 57, 59, 60, 61 e 63.
- Lima, P. R. S. et al. Redes sociais como ferramentas de transparência em tempos de covid-19: uma análise das publicações dos boletins epidemiológicos do estado de alagoas. *Logeion: Filosofia da Informação*, v. 7, n. 2, p. 88–107, mar. 2021. Disponível em: <<http://revista.ibict.br/fiinf/article/view/5650>>. Citado 4 vezes nas páginas 49, 50, 52 e 59.
- LinkedIn. *Relationships Matter*. 2003. Acessado em 13 de Outubro de 2021. Disponível em: <<https://www.linkedin.com/>>. Citado na página 12.
- Liu, C.; Xia, S.; Lang, C. Clothing consumption during the covid-19 pandemic: Evidence from mining tweets. *Clothing and Textiles Research Journal*, v. 39, n. 4, p. 314–330, 2021. Disponível em: <<https://doi.org/10.1177/0887302X211014973>>. Citado 3 vezes nas páginas 53, 58 e 61.

Liu, Y. et al. Monitoring covid-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Information Science and Systems*, v. 9, n. 1, p. 25, Jun 2021. ISSN 2047-2501. Disponível em: <<https://doi.org/10.1007/s13755-021-00158-4>>. Citado 5 vezes nas páginas 49, 52, 53, 54 e 57.

Long, Z.; Alharthi, R.; Saddik, A. E. Needfull – a tweet analysis platform to study human needs during the covid-19 pandemic in new york state. *IEEE Access*, v. 8, p. 136046–136055, 2020. ISSN 2169-3536. Citado 3 vezes nas páginas 53, 54 e 62.

Lyu, J. C.; Luli, G. K. Understanding the public discussion about the centers for disease control and prevention during the covid-19 pandemic using twitter data: Text mining analysis study. *Journal of medical Internet research*, v. 23, n. 2, p. e25108, February 2021. ISSN 1439-4456. Disponível em: <<https://europepmc.org/articles/PMC7879718>>. Citado 3 vezes nas páginas 49, 52 e 57.

Maakoul, O. et al. Towards evaluating the covid'19 related fake news problem: Case of morocco. p. 1–6, Dec 2020. Citado 3 vezes nas páginas 50, 53 e 54.

Mackey, T. et al. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with covid-19 on twitter: Retrospective big data infoveillance study. *JMIR Public Health Surveill*, v. 6, n. 2, p. e19509, Jun 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/2/e19509/>>. Citado 5 vezes nas páginas 49, 53, 58, 59 e 60.

Mackey, T. K. et al. Big data, natural language processing, and deep learning to detect and characterize illicit covid-19 product sales: Infoveillance study on twitter and instagram. *JMIR Public Health Surveill*, v. 6, n. 3, p. e20794, Aug 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/3/e20794/>>. Citado 4 vezes nas páginas 49, 53, 55 e 58.

Maia, M.; Oliveira, E.; Gallegos, L. Covid-19 e tweets no brasil: coleta, tratamento e análise de textos com evidências de estados afetivos alterados em momentos impactantes. SBC, Porto Alegre, RS, Brasil, p. 79–90, 2021. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/16127>>. Citado 8 vezes nas páginas 49, 53, 57, 58, 59, 60, 61 e 62.

Malagoli, L. et al. Caracterização do debate no twitter sobre a vacinação contra a covid-19 no brasil. SBC, Porto Alegre, RS, Brasil, p. 55–66, 2021. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/16125>>. Citado 4 vezes nas páginas 48, 49, 50 e 63.

Massaad, E.; Cherfan, P. Social media data analytics on telehealth during the covid-19 pandemic. *Cureus*, Cureus, v. 12, n. 4, p. e7838–e7838, Apr 2020. ISSN 2168-8184. 32467813[pmid]. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/32467813>>. Citado 5 vezes nas páginas 49, 53, 57, 59 e 63.

Melo, T. Análise exploratória das dúvidas sobre a covid-19 publicadas no twitter. SBC, Porto Alegre, RS, Brasil, p. 175–180, 2021. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/16136>>. Citado 3 vezes nas páginas 52, 57 e 60.

- Miller, G. A. *WordNet: A Lexical Database for English*. *Communications of the ACM*, v. 38, n. 11, p. 39–41, 1995. Disponível em: <<https://dl.acm.org/doi/10.1145/219717.219748>>. Citado na página 22.
- Miranda-Escalada, A. et al. The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora. Association for Computational Linguistics, Mexico City, Mexico, p. 13–20, jun. 2021. Disponível em: <<https://aclanthology.org/2021.smm4h-1.3>>. Citado 4 vezes nas páginas 49, 52, 54 e 57.
- Mohamed Ridhwan, K.; Hargreaves, C. A. Leveraging twitter data to understand public sentiment for the covid-19 outbreak in singapore. *International Journal of Information Management Data Insights*, v. 1, n. 2, p. 100021, 2021. ISSN 2667-0968. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2667096821000148>>. Citado 7 vezes nas páginas 48, 49, 52, 53, 55, 57 e 63.
- Morshed, S. A. et al. Impact of covid-19 pandemic on ride-hailing services based on large-scale twitter data analysis. *Journal of Urban Management*, v. 10, n. 2, p. 155–165, 2021. ISSN 2226-5856. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2226585621000200>>. Citado 4 vezes nas páginas 49, 52, 58 e 60.
- Mourad, A. et al. Critical impact of social networks infodemic on defeating coronavirus covid-19 pandemic: Twitter-based study and research directions. *IEEE Transactions on Network and Service Management*, v. 17, n. 4, p. 2145–2155, Dec 2020. ISSN 1932-4537. Citado 6 vezes nas páginas 49, 50, 53, 57, 59 e 63.
- Mutanga, M. B.; Abayomi, A. Tweeting on covid-19 pandemic in south africa: Lda-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development*, Routledge, v. 0, n. 0, p. 1–10, 2020. Disponível em: <<https://doi.org/10.1080/20421338.2020.1817262>>. Citado 2 vezes nas páginas 53 e 57.
- Naseem, U. et al. Covidsent: A large-scale benchmark twitter data set for covid-19 sentiment analysis. *IEEE Transactions on Computational Social Systems*, v. 8, n. 4, p. 1003–1015, Aug 2021. ISSN 2329-924X. Citado 3 vezes nas páginas 54, 57 e 62.
- NGUYEN, J.; CHATURVEDI, R. Quarantine quibbles: A sentiment analysis of covid-19 tweets. p. 0346–0350, Nov 2020. ISSN 2644-3163. Citado 4 vezes nas páginas 52, 53, 55 e 62.
- NLTK. 2021. Acessado em 12 de Outubro de 2021. Disponível em: <<https://www.nltk.org/>>. Citado na página 48.
- NURBAKOVA, D.; ERMAKOVA, L.; OVCHINNIKOVA, I. Understanding the personality of contributors to information cascades in social media in response to the covid-19 pandemic. p. 45–52, Nov 2020. ISSN 2375-9259. Citado 2 vezes nas páginas 52 e 62.
- Nurlaila, I. et al. Provoking tweets by indonesia media twitter in the initial month of coronavirus disease hit. p. 409–414, Aug 2020. Citado 3 vezes nas páginas 49, 52 e 59.
- NYT. *Coronavirus World Map: Tracking the Global Outbreak*. The New York Times, 2020. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.nytimes.com/interactive/2021/world/covid-cases.html>>. Citado na página 33.

Osakwe, Z. T. et al. Identifying public concerns and reactions during the covid-19 pandemic on twitter: A text-mining analysis. *Public Health Nursing*, v. 38, n. 2, p. 145–151, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/phn.12843>>. Citado na página 52.

Park, H. W.; Park, S.; Chong, M. Conversations and medical news frames on twitter: Infodemiological study on covid-19 in south korea. *J Med Internet Res*, v. 22, n. 5, p. e18897, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e18897/>>. Citado 5 vezes nas páginas 49, 52, 59, 61 e 63.

Pastor, C. K. Sentiment analysis of filipinos and effects of extreme community quarantine due to coronavirus (covid-19) pandemic (april 13, 2020). *Social Science Research Network*, 2020. Disponível em: <<https://ssrn.com/abstract=3574385>>. Citado 2 vezes nas páginas 52 e 61.

Paula, F. R. de; Mello, M. G. d. S. Análise de redes sociais: a formação de grupos do facebook frente à epidemia da covid-19 no brasil. *VITTALLE - Revista de Ciências da Saúde*, v. 32, n. 1, p. 32–42, jul. 2020. Disponível em: <<https://periodicos.furg.br/vittalle/article/view/11406>>. Citado 4 vezes nas páginas 49, 50, 52 e 60.

Permana, F. C. et al. Perception analysis of the indonesian society on twitter social media on the increase in BPJS kesehatan contribution in the covid 19 pandemic era. *Journal of Physics: Conference Series*, IOP Publishing, v. 1722, p. 012022, jan 2021. Disponível em: <<https://doi.org/10.1088/1742-6596/1722/1/012022>>. Citado 6 vezes nas páginas 49, 53, 55, 59, 61 e 63.

Petersen, K.; Vakkalanka, S.; Kuzniarz, L. *Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology*, v. 64, p. 1–18, 2015. ISSN 0950-5849. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584915000646>>. Citado na página 18.

Pinto, M. A. S. et al. Relacionando modelagem de tópicos e classificação de sentimentos para análise de mensagens do twitter durante a pandemia da covid-19. SBC, Porto Alegre, RS, Brasil, p. 61–64, 2020. ISSN 2596-1683. Disponível em: <https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/13064>. Citado 5 vezes nas páginas 52, 53, 57, 59 e 61.

Pran, M. S. A. et al. Analysis of bangladeshi people's emotion during covid-19 in social media using deep learning. p. 1–6, July 2020. Citado 5 vezes nas páginas 49, 54, 55, 58 e 62.

PUCRS. *Qualis Ciência da Computação*. 2021. Acessado em 13 de Outubro de 2021. Disponível em: <<https://ppgcc.github.io/discentesPPGCC/pt-BR/qualis/>>. Citado na página 35.

Queiroz, F. C. B. P. et al. Previsão e séries temporais para tomada de decisão empresarial em uma indústria moveleira da região de criciúma-sc. *Revista Catarinense da Ciência Contábil*, v. 11, n. 32, p. p. 26–42, 2012. Disponível em: <<https://revista.crcsc.org.br/index.php/CRCSC/article/view/1316>>. Citado na página 23.

Rafi, A. M. et al. Understanding global reaction to the recent outbreaks of covid-19: Insights from instagram data analysis. p. 3413–3420, Oct 2020. ISSN 2577-1655. Citado 4 vezes nas páginas 49, 50, 52 e 55.

Raheja, S.; Asthana, A. Sentimental analysis of twitter comments on covid-19. p. 704–708, Jan 2021. Citado 3 vezes nas páginas 52, 58 e 63.

Recuero, R.; Soares, F. O discurso desinformativo sobre a cura do covid-19 no twitter: Estudo de caso. *E-Compós*, v. 24, jul. 2021. Disponível em: <<https://www.e-compos.org.br/e-compos/article/view/2127>>. Citado 4 vezes nas páginas 49, 52, 57 e 59.

Recuero, R.; Soares, F.; Zago, G. Polarização, hiperpartidarismo e câmaras de eco: Como circula a desinformação sobre covid-19 no twitter. *Revista Contracampo*, v. 1, n. 40, p. 2021, 2021. Disponível em: <<https://periodicos.uff.br/contracampo/article/view/45611>>. Citado 3 vezes nas páginas 49, 52 e 59.

Rodrigues, F. S. Métodos de agrupamento na análise de dados de expressão gênica. *Universidade Federal de São Carlos*, 2009. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/4537>>. Citado na página 23.

Romeiro Paulino, R. d. C.; Pires Ventura, M. O engajamento no Twitter: Métodos de análise para #Somos70porcento. *Cuadernos.info*, scielocl, p. 51 – 71, 00 2021. ISSN 0719-367X. Disponível em: <http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0719-367X2021000200051&nrm=iso>. Citado 4 vezes nas páginas 49, 52, 61 e 63.

Rosenberg, J. M. How k-12 school districts communicated during the covid-19 pandemic: A study using facebook data. *International Conference on Learning Analytics & Knowledge*, 2020. Disponível em: <<https://joshuamrosenberg.com/post-prints/LAK21>>. Citado 3 vezes nas páginas 52, 57 e 60.

Ruffer, N.; Knitza, J.; Krusche, M. #covid4rheum: an analytical twitter study in the time of the covid-19 pandemic. *Rheumatology International*, v. 40, n. 12, p. 2031–2037, Dec 2020. ISSN 1437-160X. Disponível em: <<https://doi.org/10.1007/s00296-020-04710-5>>. Citado 7 vezes nas páginas 49, 50, 52, 53, 54, 58 e 61.

Rusho, M. A.; Ahmed, M. A.; Sadri, A. M. Social media response and crisis communications in active shootings during covid-19 pandemic. *Transportation Research Interdisciplinary Perspectives*, v. 11, p. 100420, 2020. ISSN 2590-1982. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2590198221001263>>. Citado 3 vezes nas páginas 52, 53 e 57.

Saha, K. et al. Psychosocial effects of the covid-19 pandemic: Large-scale quasi-experimental study on social media. *J Med Internet Res*, v. 22, n. 11, p. e22600, Nov 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/11/e22600/>>. Citado 4 vezes nas páginas 49, 52, 58 e 61.

Saire, J. E. C.; Cruz, J. F. O. Study of coronavirus impact on parisian population from april to june using twitter and text mining approach. p. 242–246, Dec 2020. Citado 2 vezes nas páginas 53 e 57.

Santis, E. D.; Martino, A.; Rizzi, A. An infoveillance system for detecting and tracking relevant topics from italian tweets during the covid-19 event. *IEEE Access*, v. 8, p. 132527–132538, 2020. ISSN 2169-3536. Citado 4 vezes nas páginas 49, 53, 57 e 60.

Santos, G. S. H. d.; Carmargo, C. C. d.; Menossi, B. R. d. S. Projeto de extensão universitário no combate a obesidade infantil através das mídias sociais em face de pandemia

- por covid-19: Um estudo transversal. *Brazilian Journal of Development*, v. 6, n. 9, p. 69886–69900, 2020. Citado 3 vezes nas páginas 49, 52 e 61.
- Sari, I. C.; Ruldeviyani, Y. Sentiment analysis of the covid-19 virus infection in indonesian public transportation on twitter data: A case study of commuter line passengers. p. 23–28, Oct 2020. Citado 3 vezes nas páginas 49, 54 e 62.
- SBC. *Biblioteca Digital da Sociedade Brasileira de Computação*. 2019. Acessado em 10 de Outubro de 2021. Disponível em: <<https://sol.sbc.org.br/>>. Citado na página 26.
- Scannavino, K. R. F. et al. *Revisão Sistemática da Literatura em Engenharia de Software: teoria e prática*. Elsevier, 2017. 144 p. Disponível em: <<https://repositorio.usp.br/item/002944453>>. Acesso em: 25 maio 2021. Citado na página 15.
- Schück, S. et al. Concerns discussed on chinese and french social media during the covid-19 lockdown: Comparative infodemiology study based on topic modeling. *JMIR Form Res*, v. 5, n. 4, p. e23593, Apr 2021. ISSN 2561-326X. Disponível em: <<https://formative.jmir.org/2021/4/e23593>>. Citado 2 vezes nas páginas 53 e 58.
- SciELO. *Scientific Electronic Library Online*. 2002. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.scielo.br/>>. Citado na página 26.
- ScienceDirect. *Science, health and medical journals, full text articles and books*. Science Direct, 1997. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.sciencedirect.com/>>. Citado na página 26.
- Shan, S.; Li, Y. A warning system for the return of covid-19 using social media data. p. 1–5, Oct 2020. Citado 4 vezes nas páginas 52, 54, 61 e 62.
- Shen, C. et al. Using reports of symptoms and diagnoses on social media to predict covid-19 case counts in mainland china: Observational infoveillance study. *J Med Internet Res*, v. 22, n. 5, p. e19421, May 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/5/e19421/>>. Citado 5 vezes nas páginas 48, 49, 52, 54 e 61.
- Shen, L. et al. Emotional attitudes of chinese citizens on social distancing during the covid-19 outbreak: Analysis of social media data. *JMIR Med Inform*, v. 9, n. 3, p. e27079, Mar 2021. Disponível em: <<https://medinform.jmir.org/2021/3/e27079>>. Citado 5 vezes nas páginas 49, 52, 54, 57 e 58.
- Shim, J.-G. et al. Text mining approaches to analyze public sentiment changes regarding covid-19 vaccines on social media in korea. *International Journal of Environmental Research and Public Health*, v. 18, n. 12, 2021. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/18/12/6549>>. Citado 4 vezes nas páginas 52, 54, 55 e 57.
- Siddaway, A. P.; Wood, A. M.; Hedges, L. V. *How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses*. *Annual Review of Psychology*, v. 70, n. 1, p. 747–770, 2019. PMID: 30089228. Disponível em: <<https://doi.org/10.1146/annurev-psych-010418-102803>>. Acesso em: 16 maio 2021. Citado na página 16.
- Silva, A. O. da. A sua revista tem qualis? *Mediações - Revista de Ciências Sociais*, v. 14, n. 1, p. 117–124, 2009. Citado na página 36.

- Silva, T.; Machado, C.; Maciel, A. *Identificação de Possíveis Influenciadores Digitais em Fóruns de Discussão em Cursos a Distância*. *Revista de Engenharia e Pesquisa Aplicada*, v. 5, n. 1, p. 117–123, 2020. Citado na página 24.
- Silva, V. et al. *A Further Analysis of the Importance of Systematic Reviews to Computer Science*. *International Journal of Social Science and Humanity*, v. 5, 2015. Disponível em: <<https://www.redalyc.org/articulo.oa?id=465660194008>>. Citado na página 15.
- Singh, M.; Jakhar, A. K.; Pandey, S. Sentiment analysis on the impact of coronavirus in social life using the bert model. *Social Network Analysis and Mining*, v. 11, n. 1, p. 33, Mar 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00737-z>>. Citado 5 vezes nas páginas 49, 54, 59, 61 e 63.
- Slavik, C. E. et al. Examining tweet content and engagement of canadian public health agencies and decision makers during covid-19: Mixed methods analysis. *J Med Internet Res*, v. 23, n. 3, p. e24883, Mar 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/3/e24883>>. Citado 5 vezes nas páginas 48, 49, 50, 52 e 59.
- Springer. *Springer: Our business is publishing*. Springer Nature, 1996. Acessado em 10 de Outubro de 2021. Disponível em: <<https://www.springer.com/>>. Citado na página 26.
- Statista. *Most used social media 2021*. Statista, 2021. Acessado em 12 de Outubro de 2021. Disponível em: <<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>>. Citado na página 46.
- Tao, G.; Miao, Y.; Ng, S. Covid-19 topic modeling and visualization. p. 734–739, Sep. 2020. ISSN 2375-0138. Citado 3 vezes nas páginas 49, 57 e 62.
- Tao, Z.-Y. et al. Nature and diffusion of covid-19–related oral health information on chinese social media: Analysis of tweets on weibo. *J Med Internet Res*, v. 22, n. 6, p. e19981, Jun 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/6/e19981/>>. Citado 5 vezes nas páginas 49, 53, 57, 59 e 60.
- Trust, T. et al. #remoteteaching & #remoteteaching: Educator tweeting during the covid-19 pandemic. *Journal of Technology and Teacher Education*, Society for Information Technology & Teacher Education, Waynesville, NC USA, v. 28, n. 2, p. 151–159, 2020. ISSN 1059-7069. Disponível em: <<https://www.learntechlib.org/p/216094>>. Citado 2 vezes nas páginas 49 e 59.
- Twitter. *What's happening?* 2006. Acessado em 13 de Outubro de 2021. Disponível em: <<https://twitter.com/>>. Citado na página 12.
- Twitter. *Data dictionary: Standard v1.1*. 2021. Acessado em 10 de outubro de 2021. Disponível em: <<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>>. Citado na página 18.
- Twitter. *TweetDeck*. 2021. <<https://tweetdeck.twitter.com/>>. Acessado em 10 de Outubro de 2021. Citado na página 23.
- Valdez, D. et al. Social media insights into us mental health during the covid-19 pandemic: Longitudinal analysis of twitter data. *J Med Internet Res*, v. 22, n. 12, p. e21418, Dec 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/12/e21418/>>. Citado 5 vezes nas páginas 52, 57, 58, 59 e 60.

- Valle Cruz, D. et al. Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive Computation*, Jan 2021. ISSN 1866-9964. Disponível em: <<https://doi.org/10.1007/s12559-021-09819-8>>. Citado 3 vezes nas páginas 53, 61 e 63.
- Viviani, M. et al. Assessing vulnerability to psychological distress during the covid-19 pandemic through the analysis of microblogging content. *Future Generation Computer Systems*, v. 125, p. 446–459, 2021. ISSN 0167-739X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167739X21002399>>. Citado 5 vezes nas páginas 52, 57, 58, 59 e 62.
- Wahbeh, A. et al. Mining physicians' opinions on social media to obtain insights into covid-19: Mixed methods analysis. *JMIR Public Health Surveill*, v. 6, n. 2, p. e19276, Jun 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/2/e19276/>>. Citado 4 vezes nas páginas 49, 52, 60 e 61.
- Wang, D.; Qian, Y. Echo chamber effect in rumor rebuttal discussions about covid-19 in china: Social media content and network analysis study. *J Med Internet Res*, v. 23, n. 3, p. e27009, Mar 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/3/e27009>>. Citado 3 vezes nas páginas 50, 52 e 58.
- Wang, H. et al. Using tweets to understand how covid-19 related health beliefs are affected in the age of social media: Twitter data analysis study. *J Med Internet Res*, v. 23, n. 2, p. e26302, Feb 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/2/e26302>>. Citado 4 vezes nas páginas 53, 54, 60 e 62.
- Wang, J. et al. Concerns expressed by chinese social media users during the covid-19 pandemic: Content analysis of sina weibo microblogging data. *J Med Internet Res*, v. 22, n. 11, p. e22152, Nov 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/11/e22152/>>. Citado 4 vezes nas páginas 49, 57, 58 e 59.
- We Are Social, Hootsuite. *The global state of digital in 2021*. In: . We Are Social, 2021. Disponível em: <<https://datareportal.com/reports/digital-2021-global-overview-report>>. Acesso em: 02 maio 2021. Citado 2 vezes nas páginas 12 e 46.
- Wibowo, N. S.; Mahardika, R.; Kusrini, K. Twitter data analysis using machine learning to evaluate community compliance in preventing the spread of covid-19. p. 1–4, Oct 2020. Citado 2 vezes nas páginas 52 e 54.
- WICKE, P.; BOLOGNESI, M. M. Framing covid-19: How we conceptualize and discuss the pandemic on twitter. *PLOS ONE*, Public Library of Science, v. 15, n. 9, p. 1–24, 09 2020. Disponível em: <<https://doi.org/10.1371/journal.pone.0240010>>. Citado 2 vezes nas páginas 53 e 57.
- Wrycza, S.; Maślankowski, J. Social media users' opinions on remote work during the covid-19 pandemic. thematic and sentiment analysis. *Information Systems Management*, Taylor & Francis, v. 37, n. 4, p. 288–297, 2020. Disponível em: <<https://doi.org/10.1080/10580530.2020.1820631>>. Citado 5 vezes nas páginas 49, 53, 57, 58 e 59.
- Xavier, F. et al. Análise de redes sociais como estratégia de apoio à vigilância em saúde durante a covid-19. *Estudos Avançados*, v. 10, n. 1590, p. 103–4014, Julho 2020. Disponível em: <<https://doi.org/10.1590/s0103-4014.2020.3499.016>>. Citado 2 vezes nas páginas 52 e 54.

- Xu, P.; Dredze, M.; Broniatowski, D. A. The twitter social mobility index: Measuring social distancing practices with geolocated tweets. *J Med Internet Res*, v. 22, n. 12, p. e21499, Dec 2020. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2020/12/e21499>>. Citado 3 vezes nas páginas 49, 52 e 59.
- Xu, Q. et al. Characterizing weibo social media posts from wuhan, china during the early stages of the covid-19 pandemic: Qualitative content analysis. *JMIR Public Health Surveill*, v. 6, n. 4, p. e24125, Dec 2020. ISSN 2369-2960. Disponível em: <<http://publichealth.jmir.org/2020/4/e24125/>>. Citado 4 vezes nas páginas 49, 52, 58 e 60.
- Xue, J. et al. The hidden pandemic of family violence during covid-19: Unsupervised learning of tweets. *J Med Internet Res*, v. 22, n. 11, p. e24361, Nov 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/11/e24361/>>. Citado 4 vezes nas páginas 52, 53, 54 e 57.
- Xue, J. et al. Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PLOS ONE*, Public Library of Science, v. 15, n. 9, p. 1–12, 09 2020. Disponível em: <<https://doi.org/10.1371/journal.pone.0239441>>. Citado 5 vezes nas páginas 49, 50, 52, 57 e 58.
- Xue, J. et al. Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach. *J Med Internet Res*, v. 22, n. 11, p. e20550, Nov 2020. ISSN 1438-8871. Disponível em: <<http://www.jmir.org/2020/11/e20550/>>. Citado 4 vezes nas páginas 53, 57, 58 e 59.
- Yang, C.; Zhou, X.; Zafarani, R. Checked: Chinese covid-19 fake news dataset. *Social Network Analysis and Mining*, v. 11, n. 1, p. 58, Jun 2021. ISSN 1869-5469. Disponível em: <<https://doi.org/10.1007/s13278-021-00766-8>>. Citado 5 vezes nas páginas 48, 49, 55, 58 e 62.
- Yang, M.; Han, C. Revealing industry challenge and business response to covid-19: a text mining approach. *International Journal of Contemporary Hospitality Management*, Emerald Publishing Limited, v. 33, n. 4, p. 1230–1248, Jan 2021. ISSN 0959-6119. Disponível em: <<https://doi.org/10.1108/IJCHM-08-2020-0920>>. Citado 4 vezes nas páginas 50, 53, 57 e 60.
- Yang, S. et al. Analysis and insights for myths circulating on twitter during the covid-19 pandemic. *IEEE Open Journal of the Computer Society*, v. 1, p. 209–219, 2020. ISSN 2644-1268. Citado 6 vezes nas páginas 50, 52, 53, 55, 58 e 62.
- Yigitcanlar, T. et al. How can social media analytics assist authorities in pandemic-related policy decisions? insights from australian states and territories. *Health Information Science and Systems*, v. 8, n. 1, p. 37, Oct 2020. ISSN 2047-2501. Disponível em: <<https://doi.org/10.1007/s13755-020-00121-9>>. Citado 4 vezes nas páginas 52, 55, 61 e 63.
- Zhang, C. et al. Understanding concerns, sentiments, and disparities among population groups during the covid-19 pandemic via twitter data mining: Large-scale cross-sectional study. *J Med Internet Res*, v. 23, n. 3, p. e26482, Mar 2021. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2021/3/e26482>>. Citado 6 vezes nas páginas 49, 52, 54, 57, 58 e 59.

Zhao, Y. et al. Chinese public's attention to the covid-19 epidemic on social media: Observational descriptive study. *J Med Internet Res*, v. 22, n. 5, p. e18825, May 2020. ISSN 1438-8871. Disponível em: <<https://www.jmir.org/2020/5/e18825>>. Citado na página 63.

ZHOU, J. et al. Examination of community sentiment dynamics due to covid-19 pandemic: A case study from a state in australia. *SN Computer Science*, v. 2, n. 3, p. 201, Apr 2021. ISSN 2661-8907. Disponível em: <<https://doi.org/10.1007/s42979-021-00596-7>>. Citado 3 vezes nas páginas 49, 58 e 59.