



Trabalho de Conclusão de Curso

**Uma abordagem para extrair relatos de agressões
contra mulheres no Twitter e enquadrar na Lei
Maria da Penha**

Júlia Albuquerque Aguiar
jaa@ic.ufal.br

Orientador:
Prof. Dr. André Lage Freitas

Maceió, 22 de setembro de 2021.

Júlia Albuquerque Aguiar

**Uma abordagem para extrair relatos de agressões
contra mulheres no Twitter e enquadrar na Lei
Maria da Penha**

Monografia apresentada como requisito parcial para
obtenção do grau de Bacharel em Ciência da Com-
putação do Instituto de Computação da Universidade
Federal de Alagoas.

Orientador:

Prof. Dr. André Lage Freitas

Maceió, 22 de setembro de 2021.

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

A282a Aguiar, Júlia Albuquerque.

Uma abordagem para extrair relatos de agressões contra mulheres no Twitter e enquadrar na Lei Maria da Penha / Júlia Albuquerque Aguiar. – 2021.
25 f. : il.

Orientador: André Lage Freitas.

Monografia (Trabalho de conclusão de curso em Ciência da Computação) – Universidade Aberta do Brasil. Universidade Federal de Alagoas, Instituto de Computação. Maceió.

Bibliografia: f. 23-25.

1. Twitter (Rede social on-line). 2. Violência contra as mulheres. 3. Violência doméstica. 4. Brasil. Lei n. 11.340, de 7 de agosto de 2006. 5. Processamento de linguagem natural (Computação). 6. Direito. 7. Aprendizado do computador. 8. Inteligência artificial. I. Título.

CDU: 004.81:159.953..5

FOLHA DE APROVAÇÃO

JÚLIA ALBUQUERQUE AGUIAR

INSTITUTO DE COMPUTAÇÃO

Trabalho Final de Graduação apresentado ao Curso de Ciência da Computação do Instituto de Computação da Universidade Federal de Alagoas, como requisito parcial para obtenção do grau de bacharel em Ciência da Computação.

BANCA EXAMINADORA

Prof. Dr. André Lage Freitas (Orientador)

Profa. Dra. Mônica de Sá Dantas Paz

Prof. Dr. Pablo Bravo Hurtado

Agradecimentos

Agradeço ao meu pai Jadson Lima de Aguiar (in memoriam), que foi o meu maior incentivador em vida. Sua lembrança me inspira e seus ensinamentos permanecem comigo.

À minha mãe Vânia Albuquerque dos Santos Aguiar, por ter investido na minha educação com todo o seu coração e força.

Ao Prof. André Lage Freitas, por ter me orientado neste trabalho com tanta paciência, dedicação e amizade.

Aos membros da banca examinadora, por sua avaliação e feedback que serão essenciais no meu desenvolvimento acadêmico e profissional.

À minha companheira Priscylla Karla Calado Cavalcanti, por estar sempre ao meu lado me incentivando e me fortalecendo.

Aos amigos que foram minha família sempre que mais precisei.

E a todos que direta ou indiretamente fizeram parte da minha formação.

—

Resumo

A violência de gênero é uma violência de repetição que causa não só danos físicos, mas também psicológicos a uma quantidade razoável de mulheres ao redor do mundo. No Brasil, a Lei Maria da Penha, com seus 15 anos de existência, continua sendo um símbolo nacional de luta contra a violência de gênero e acumula milhões de processos na justiça. Com a pandemia da COVID-19, aumentou-se consideravelmente a quantidade de casos de violência devido ao isolamento social e à carência de projetos de apoio a este público.

Como um canal alternativo às mídias tradicionais de conteúdo, as redes sociais deram voz às mulheres nesta busca por justiça e criaram frentes de discussões sobre uma possível modernização da aplicação da legislação brasileira. Por exemplo, a literatura da área de Ciências Sociais fala sobre conceitos importantes como *tecnopolítica*, *ciberfeminismo* e *ciberespaço* e também traz um debate essencial sobre teoria do afeto nas redes sociais. Já outros trabalhos analisam manifestações online como o movimento *#MeToo*, “Não mereço ser estuprada” e dezenas de outras *hashtags* que viralizaram nas mídias sociais, e para além desta análise, estudam formas de classificações destes relatos.

O objetivo do nosso trabalho é investigar se os relatos de violência doméstica postadas em redes sociais podem ser enquadrados na Lei Maria da Penha assim como analisar a frequência nesses relatos dos tipos de violência descritos na Lei Maria da Penha. Assim, desenvolvemos uma metodologia para a coleta de relatos de redes sociais e classificação dos relatos de acordo com os cinco tipos de violência doméstica descritos na Lei Maria da Penha: Moral, Sexual, Psicológica, Física e Patrimonial. Essa metodologia foi validada através da implementação de um protótipo de pesquisa que automatiza toda a abordagem proposta. Esse protótipo, que foi implementado na linguagem de programação Python, utiliza técnicas de Aprendizado de Máquina e Processamento de Linguagem Natural.

Utilizamos o protótipo desenvolvido para coletar e analisar dados do Twitter. Percebemos que o tipo de violência contra mulher que apareceu com mais frequência nesses relatos é a Violência Moral, seguida da Violência Sexual. Assim, validamos parcialmente nossa hipótese de pesquisa que afirmava que a maioria dos relatos seriam de Violência Sexual. Uma consequência importante dessa conclusão é sobre a atenção que deve ser dada também às violências não físicas contra mulher.

Por fim, a abordagem proposta neste trabalho oferece um arcabouço para o auxílio à elaboração, fiscalização e fomento de políticas públicas para o combate à violência de gênero. Ou

seja, instituições privadas ou o próprio Poder Público – seja através do Executivo, Legislativo, Judiciário, Ministério Público ou Defensoria Pública – podem utilizar o conceito ou a tecnologia propostos neste trabalho para combater as agressões contra mulheres.

Palavras-chave: Violência contra mulher. Violência doméstica. Lei Maria da Penha. Processamento de Linguagem Natural. Twitter. Redes Sociais. Direito. Aprendizado de Máquina. Inteligência Artificial.

Abstract

Gender violence is a repeated violence that causes both physical and psychological damages in a fair amount of women around the world. In Brazil the Maria da Penha law through its 15 years of existence remains a national symbol in the fight against gender violence and accumulates millions of lawsuits at court. The COVID-19 pandemic increased considerably the number of violent cases in social isolation and there is a lack of support projects for this group. As an alternative channel to traditional media, the social networks gave women voice in this search for justice and created various research groups who are approaching possible law enforcement from innovative perspectives.

For instance, literature in this field highlights key concepts as *technopolitics*, *cyberfeminism*, *cyberspace* and also brings into discussion the theory about the affect of social networks. Other projects analyze online movements such as *#MeToo* movement, “Não mereço ser estuprada” and dozens of other *hashtags* that went viral on social media and study ways about how to classify these reports.

The goal of our project is to investigate if the reporting of aggression posted on social networks can be covered by Maria da Penha law and to analyze the frequency which type of violence described in the Maria da Penha law. So we developed a methodology to collect posts from social networks and classify them according to the five types of domestic violence covered by Maria da Penha law, i.e., Moral, Sexual, Psychological, Physical, and Property. We validated this methodology by implementing a working prototype that automates the overall proposed approach and takes advantage of Artificial Intelligence and Natural Language Processing techniques. This prototype is implemented in Python programming language and uses the social network Twitter.

Moreover, we used the prototype to collect posts from Twitter and concluded that moral violence is the most common type of violence against women followed by sexual violence. So, our research hypothesis was partially validated as we had assumed that most posts would refer to sexual violence. An important consequence of this conclusion is that we must also pay attention to non-physical violence against women.

In conclusion, the approach proposed here provides a framework for elaboration, supervision, and promotion of public policy for the fight against gender violence. In other words, private institutions or Public Power - whether through the Executive, Legislative, Judicial, Public Ministry or Public Defenders - may use the concept or the technology proposed in this work

to fight against gender violence.

Key-words: Violence against women. Domestic violence. Maria da Penha law. Natural language processing. Twitter. Social Network. Law. Legal. Machine Learning. Artificial Intelligence. Machine Learning.

Lista de Figuras

2.1	Número de processos judiciais abertos com acusações de Assédio Sexual junto à <i>EEOC</i> (<i>Equal Employment Opportunity Commission</i>) e à <i>FEPA</i> (<i>Fair Employment Practices Agency</i>), no período de 1992-2018. Fonte: U.S. Equal Employment Opportunity Commission, Charges Alleging Sex-Based Harassment (Charges filed with EEOC) FY 2010 - FY 2018, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment_new.cfm ; U.S. Equal Employment Opportunity Commission, Sexual Harassment Charges EEOC FEPAs Combined: FY 1997 - FY 2011, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment.cfm ; U.S. Equal Employment Opportunity Commission, Sexual Harassment Charges EEOC FEPAs Combined: FY 1992 - FY 1996, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment-a.cfm	5
2.2	Power and Control Wheel: Ferramenta com os padrões de comportamentos abusivos de um agressor com sua parceira. Fonte: https://connect.springerpub.com/content/book/978-0-8261-7991-3/back-matter/app3	9
3.1	Metodologia - Etapas da abordagem de solução proposta. Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	11
4.1	Classificação de Violência de Gênero de todos os dados coletados do Twitter. A violência moral é tipo de violência identificado com maior presença nas postagens. Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	17
4.2	Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a Violência Física . Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	18
4.3	Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a Violência Moral . Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	18
4.4	Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a Violência Patrimonial . Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	19
4.5	Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a Violência Psicológica . Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	19
4.6	Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a Violência Sexual . Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.	20

Sumário

Lista de Figuras	v
1 INTRODUÇÃO	1
1.1 Contexto	1
1.2 Problema de Pesquisa	2
2 REVISÃO DE LITERATURA	4
3 METODOLOGIA	10
3.1 Abordagem Proposta	10
3.1.1 Etapa 1: Elaboração do Dicionário de Sinônimos de Violência de Gênero	10
3.1.2 Etapa 2: Coleta de Relatos de Redes Sociais e Representação Numérica	12
3.1.3 Etapa 3: Enquadramento dos Relatos na Lei Maria da Penha	12
3.2 Detalhes de Implementação	13
4 RESULTADOS	16
5 CONCLUSÃO	21
Referências bibliográficas	23

1

INTRODUÇÃO

O nosso trabalho se inicia contextualizando, na Seção 1.1, o tema da pesquisa que é “*Deteção de casos de violências no Brasil contra a mulher utilizando redes sociais*”, apresentando os pontos focais que nortearam o estudo. Em seguida, estão destacados a pergunta científica, hipótese e objetivos da nossa pesquisa na Seção 1.2.

1.1 Contexto

A violência de gênero é qualquer violência motivada com base no gênero de um indivíduo e pode incluir qualquer ato de força verbal ou física, coerção ou privação com risco de vida, dirigido a uma mulher ou menina que cause dano físico ou psicológico, humilhação ou privação arbitrária de liberdade e que perpetue a subordinação feminina Heise et al. (1994). De acordo com a Organização Mundial da Saúde (OMS) [World Health Organization \(2021\)](#), 1 em cada 3 (30%) das mulheres em todo o mundo sofreram violência física e/ou sexual por pessoa íntima ou não íntima; como também, 27% das mulheres entre 15 e 49 anos sofreram violência dentro de seus relacionamentos.

A agressão contra mulher é uma violência de repetição que passa por ciclos entre momentos de carinho, agressão psicológica/moral e agressão física/sexual; o que torna a identificação dos abusos ainda mais difícil. Apesar da liberdade de expressão e o direito constitucional protegido, a vítima pode ter seu relato invalidado ou, até mesmo, se tornar ré em um processo por injúria, calúnia ou difamação movida por seu próprio agressor.

Como mostra um relatório da Organização das Nações Unidas [Headquarters \(2020\)](#), a pandemia da COVID-19 agravou a violência de gênero no momento em que o isolamento social forçou a convivência das mulheres com seus abusadores, ao mesmo tempo que os projetos de apoio social voltados a este público foram interrompidos. No ano atual em que a Lei Maria da Penha completa 15 anos de existência e os desafios se acentuaram com a pandemia, os números do Fórum Brasileiro de Segurança Pública mostram que uma mulher foi assassinada a cada

6 horas e meia em 2020 e registraram uma agressão a cada 2 minutos [Redação, O Estado de S.Paulo \(2021\)](#).

No Brasil, a Lei Maria da Penha nº. 11.340 (2006) se tornou um símbolo nacional de luta contra a violência de gênero. Em 1983, o marido de Maria da Penha Maia Fernandes tentou assassiná-la duas vezes: na primeira, com um tiro, quando ela ficou paraplégica; e na segunda, por eletrocussão e afogamento. Após 19 anos de luta por direitos contra a opressão e violência contra mulheres, Maria da Penha teve seu nome sancionado como lei em 7 de agosto de 2006 pelo atual presidente da época, Luiz Inácio Lula da Silva. A Lei Maria da Penha tem um papel muito expressivo no combate à violência contra a mulher. Por exemplo, segundo números do CNJ [Conselho Nacional de Justiça \(2020\)](#), mais de 1 milhão de processos relacionados à lei Maria da Penha correram na Justiça até março de 2020 e quase 400 mil medidas protetivas tiveram de ser aplicadas em 2018.

As redes sociais vieram como uma explosão de informações e um canal alternativo às mídias tradicionais de conteúdos e notícias e, em paralelo a este movimento, as redes deram voz e empoderamento às mulheres no que parece ser uma busca por justiça. Uma pesquisa do Instituto Avon em parceria com a Folks Netnográfica [Avon and Netnográfica \(2018\)](#) revelou que foram feitas 14.043.912 menções relacionadas aos assuntos “assédio” e “violência contra a mulher” e termos variados entre Janeiro de 2015 a Dezembro de 2017 no Facebook, Instagram e Twitter. Ao mesmo tempo, o poder das redes sociais acaba por expor as vítimas a violências diversas dentro do meio virtual, cumprindo também um papel de violador dos direitos humanos. Neste contraste, são inúmeras as recomendações e discussões de como este movimento por reivindicação dentro das redes sociais pode, de maneira completa, contribuir para a erradicação da violência de gênero.

Tendo em vista os inúmeros relatos de violência no meio virtual, surgiram várias frentes de discussões sobre uma possível modernização de aplicação da lei para garantir a luta contra a violência de gênero, como também uma reflexão sobre a importância de um canal alternativo que amplia a voz de grupos minoritários.

1.2 Problema de Pesquisa

Apesar de significativas contribuições dos trabalhos citados no Capítulo 2, estes não fornecem uma metodologia ou arcabouço teórico e/ou tecnológico para responder perguntas científicas mais específicas. Devido à grande relevância e importância do tema introduzido na Seção 1.1, propomos a seguinte pergunta científica: **os relatos de agressões contra mulheres nas redes sociais podem ser enquadrados na Lei Maria da Penha utilizando o modelo LSI de classificação de texto?**

A Hipótese para esta pergunta é que *ao menos 60% dos relatos de agressões contra mulheres nas redes sociais se enquadram nos parágrafos 1^o e 3^o do Artigo 7 da Lei Maria da Penha*

utilizando o modelo LSI de classificação de texto. Os parágrafos 1 e 3 do Artigo 7 da Lei Maria da Penha (Lei Nº11.340) estão transcritos a seguir:

I - a violência física, entendida como qualquer conduta que ofenda sua integridade ou saúde corporal;

III - a violência sexual, entendida como qualquer conduta que a constranja a presenciar, a manter ou a participar de relação sexual não desejada, mediante intimidação, ameaça, coação ou uso da força; que a induza a comercializar ou a utilizar, de qualquer modo, a sua sexualidade, que a impeça de usar qualquer método contraceptivo ou que a force ao matrimônio, à gravidez, ao aborto ou à prostituição, mediante coação, chantagem, suborno ou manipulação; ou que limite ou anule o exercício de seus direitos sexuais e reprodutivos;

A motivação para a construção dessa hipótese baseia-se no fato de que a violência física e sexual podem aparecer com mais frequência nos relatos, uma vez que estas violências são destacadas em notícias nos jornais ou na TV (ver Seção 1.1).

O **principal objetivo** deste trabalho é responder à pergunta de pesquisa introduzida anteriormente assim como testar a sua hipótese. Para facilitar a execução dessa tarefa, esse objetivo foi dividido nos seguintes objetivos específicos:

1. Identificar as descrições de violência doméstica e familiar contra mulher e suas formas segundo a Lei Maria da Penha;
2. Selecionar as abordagens de Processamento de Linguagem Natural e Aprendizado de Máquina mais adequadas para serem utilizadas;
3. Desenvolver um protótipo de pesquisa que colete os dados de relatos de redes sociais, implemente as técnicas selecionadas e gere resultados que possam ser utilizados para testar a hipótese;
4. Validar um protótipo de pesquisa através da sua execução em ambiente real (não simulado);
5. Analisar se a hipótese é verdadeira ou falsa.

2

REVISÃO DE LITERATURA

Quando tratamos da possibilidade de denúncias postadas em redes sociais serem enquadradas em alguma lei vigente, [Alexander \(2019\)](#) apresenta em sua pesquisa uma abordagem dos efeitos que o movimento *#MeToo* pode ter causado no sistema jurídico. O movimento viralizou nas mídias sociais através de uma hashtag com o mesmo nome em que várias vítimas de assédio denunciaram suas agressões depois que a atriz Alyssa Milano denunciou o produtor Harvey Weinstein, acusado de abuso e violência sexual. De acordo com um levantamento do *The New York Times* [Carlsen et al. \(2021\)](#), pelo menos 200 homens influentes perderam seus empregos após alegações públicas de assédio sexual através do *#MeToo*. O trabalho da Profa. Charlotte Alexander [Alexander \(2019\)](#) cita o *#MeToo* como um canal de denúncia não tradicional para estudar a possibilidade de que os relatos sejam tratados em tribunais.

É importante ressaltar as estatísticas do *Equal Employment Opportunity Commission (EEOC)* mostradas no estudo de Charlotte [Alexander \(2019\)](#) e também apresentadas na [Figura 2.1](#) onde as denúncias de assédio sexual aumentam consideravelmente em 2017-2018 devido ao movimento *#MeToo*. Os números das estatísticas mostram também que as acusações sofreram um aumento em mais de 12%, o que pode gerar impacto na conscientização e progresso no sistema jurídico.

A pesquisa de [Alexander \(2019\)](#) observa as possibilidades de que os movimentos de denúncias em redes sociais se tornem frentes fora do sistema legal jurídico, como também que possam ser interpretadas dentro da lei futuramente. Apesar de que não conseguimos ter uma previsão para que estas hipóteses ocorram, é um fato que o movimento pode dar voz a estas mulheres e consequentemente aumentar a quantidade de relatos que chegam aos tribunais.

Seguindo com a leitura sobre os impactos do movimento *#MeToo*, [Tippett \(2018\)](#) que estudou a interseção da lei com a tecnologia na Faculdade de Direito da Universidade de Oregon cita o movimento como uma *era*, de uma hashtag que se tornou viral a uma potencialização jurídica de denúncias de assédio. Tomando o contexto de interpretação judicial, o estudo de Eliza-

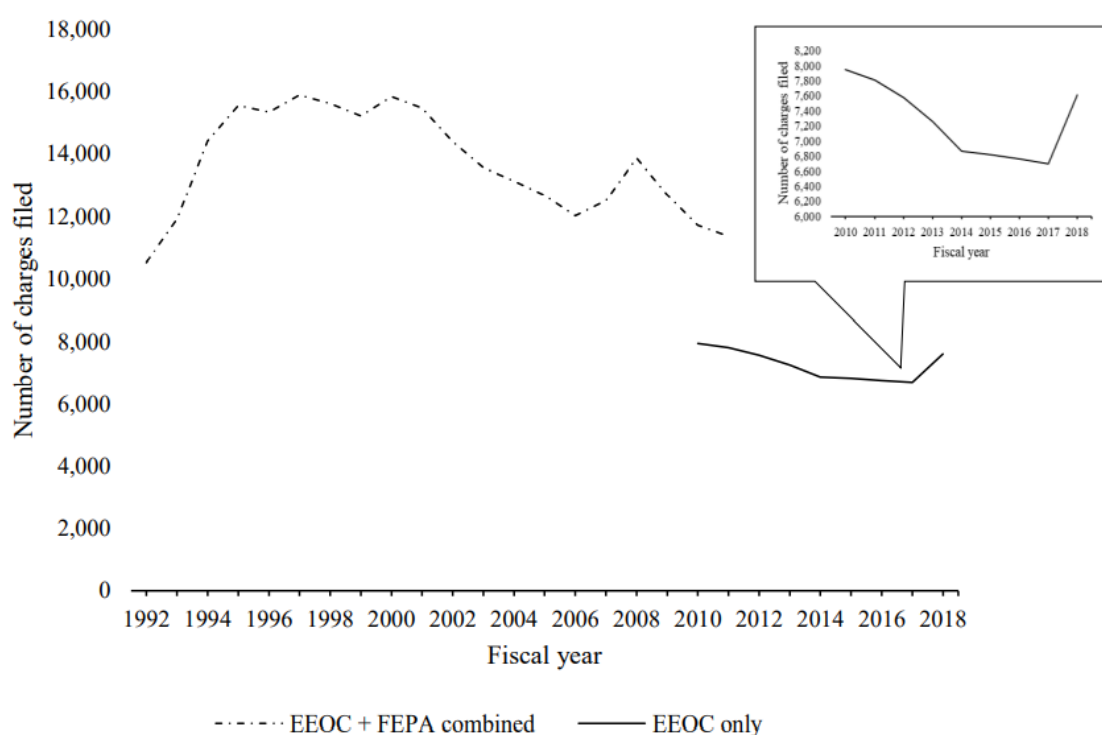


Figura 2.1: Número de processos judiciais abertos com acusações de Assédio Sexual junto à *EEOC* (*Equal Employment Opportunity Commission*) e à *FEPA* (*Fair Employment Practices Agency*), no período de 1992-2018. **Fonte:** U.S. Equal Employment Opportunity Commission, Charges Alleging Sex-Based Harassment (Charges filed with EEOC) FY 2010 - FY 2018, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment_new.cfm; U.S. Equal Employment Opportunity Commission, Sexual Harassment Charges EEOC FEPA's Combined: FY 1997 - FY 2011, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment.cfm; U.S. Equal Employment Opportunity Commission, Sexual Harassment Charges EEOC FEPA's Combined: FY 1992 - FY 1996, disponível em https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment-a.cfm.

beth Tippett (2018) mostra projeções positivas dos impactos que a digitalização de denúncias podem alcançar, incluindo a limitação do poder daqueles que antes reprimiam as reivindicações.

Na era *#MeToo* e das mudanças tecnológicas, como descreve Tippett (2018), as percepções de “razoabilidade” das tomadas de decisões dos empregadores que tinham funcionários com denúncias de assédio se tornaram menos abstratas, pois a forma como buscamos informações e geramos estatísticas mudou com o advento da tecnologia. Após o movimento *#MeToo* e outros movimentos digitais surgirem, estudos apontaram que as empresas mudaram a postura de desligar funcionários de alto nível profissional. Em contrapartida, a funcionária que decide por denunciar seu assediador em mídias sociais ainda não goza de uma proteção adequada e pode sofrer retaliações, ressalta Tippett (2018).

Na sua pesquisa, Elizabeth Tippett (2018) conclui que a indústria de tecnologia está propensa a desenvolver novas abordagens de análises em tempo real. Além disso, seu trabalho também prospecta mudanças na atuação das empresas diante a casos de assédio e da forma como as denúncias em meios digitais podem mover adaptações nas ações legais.

Em outro estudo de Tippett (2017) é construída uma análise detalhada sobre treinamento de

assédio em empresas cujo objetivo seria educar os funcionários sobre violência sexual. Algumas das críticas abordadas em sua análise é a simplificação das possibilidades de assédios tendo em vista que a violência sexual é apenas um dos abusos descritos em lei; como também a falta de atualização no conteúdo dos treinamentos e conseqüentemente sua adaptação a tecnologias e informações atuais.

Analisando esta pesquisa, é importante ressaltar que [Tippett \(2017\)](#) propõe o uso de tecnologias para melhoria dos programas de treinamento. Baseado na análise do seu conteúdo desatualizado, entendemos a importância de um projeto futuro que sirva como ferramenta de automação de treinamento como os programas analisados por [Tippett \(2017\)](#).

Existem muitas variações de pesquisa em que o trabalho prático é a busca de textos em redes sociais para fazer classificações. [Gil Vera et al. \(2018\)](#) criou *word clouds* para analisar quais eram os sentimentos de colombianos e mexicanos sobre o feminicídio. Utilizando palavras chaves na busca por postagens e fazendo o tratamento dos textos com o software R 3.4.3, a pesquisa conseguiu classificar os tweets em uma escala que varia entre “muito negativo” e “muito positivo”, chegando a conclusão de que os colombianos e mexicanos têm opiniões e percepções negativas sobre o feminicídio.

Há um debate em torno da teoria do afeto através de mídias sociais abordado por [Keller et al. \(2016\)](#), que estão particularmente interessadas em entender como ocorre a “solidariedade afetiva” com as conexões digitais e como estas conexões moldam as práticas ativistas offline. Para explorar as formas de conectividade e compartilhamento, [Jessalynn Keller et al. \(2016\)](#) utilizou análise textual discursiva e teorizações de afeto focando em três estudos de caso: (1) postagens online no site Hollaback sobre a luta contra a violência de rua; (2) relatos de uso da hashtag *#BeenRapedNeverReported* no Twitter e (3) o uso de plataformas de mídia social por adolescentes feministas para desafiar a cultura do estupro dentro e ao redor das escolas.

Utilizando a abordagem ‘in situ’ para compreender a experiência na perspectiva dos usuários, foram selecionadas 159 postagens e executadas entrevistas pessoais como uma parte do material de pesquisa. Foram classificadas até três tipos de assédio com as postagens, e ‘catcalling’¹ no geral foram as mais frequentes e registradas em 47% das postagens. Todos os relatos estudados na pesquisa demonstraram como é desafiador fazer denúncia às autoridades, trazendo à tona também os diversos motivos para as denúncias formais não ocorrerem.

[Navarro et al. \(2018\)](#) cita três conceitos muito importantes: tecnopolítica, ciberfeminismo e ciberespaço. [Toret \(2013\)](#) destaca a tecnopolítica como a “capacidade organizacional massiva mediada pela rede, capaz de produzir estados de espírito empoderados e um padrão de auto-organização política na sociedade em rede”². Dado o contexto ao termo ciberfeminismo como uma série de posicionamentos políticos e ideológicos da comunidade femista dentro do ciberespaço, o artigo une os três conceitos para analisar se o meio digital quebra a dicotomia de

¹Termo em inglês utilizado para se referir ao assédio a mulher na rua.

²“capacidad organizativa masiva mediada por la red y capaz de producir estados de ánimos empoderados y un patrón de autoorganización política en la sociedad red.”

gênero ou é mais um meio que reproduz as estruturas heteropatriarcais. Para além das barreiras de gênero, Navarro et al. (2018) também pontua a necessidade de pensarmos em uma atuação no meio digital que inclua todas as mulheres, especialmente idosas e mulheres com recursos escassos.

Navarro et al. (2018) conta em sua narrativa que para tentar quebrar a desigualdade de conteúdo existente também em meios digitais, os movimentos feministas começaram a criar sua própria mídia especializada, embora em sua maioria, alcancem um público pequeno. Fazendo uma análise crítica do alcance destes movimentos, há uma suspeita de que os algoritmos das redes sociais filtram os conteúdos de tal forma a despriorizar postagens com assuntos desta natureza, o que nos faz questionar se é possível fazer um uso efetivo das redes para este objetivo. Com base nisto, a pesquisa buscou páginas no Facebook cujo tema principal das postagens é a violência de gênero no México e entrevistou três grupos principais de organizações sobre sua atuação e experiência na rede: órgãos públicos, entidades civis e agências de notícias.

Com as entrevistas, o trabalho de Navarro et al. (2018) destacou que todas as entidades usam a plataforma como possibilidade de interação com usuários e outros grupos. Outro elemento fortemente destacado são as denúncias que as páginas recebem de mulheres que sofreram violência, o que pode nos mostrar a importância que as redes sociais possuem de facilitar a procura por ajuda e aumentar as possibilidades de encontrar suporte tendo em vista a ampliação de rede de contatos que a internet trouxe. Por fim, as páginas que recebem as denúncias contaram em suas entrevistas que constroem relatórios com dados estatísticos sintetizados e que servem como expansão das bases de dados oficiais do governo, destacando novamente a importância de pesquisas com foco no tema.

A área de Segurança e Cidadania da Diretoria de Análise de Políticas Públicas da Fundação Getúlio Vargas (FVG) coletou os relatos do programa Disque-Denúncia do estado do Rio de Janeiro com dados de crimes contra mulher Sanches et al. (2018), que correspondem a 34,7% do total de denúncias do estado. Dentre várias análises feitas, foi registrado no relatório um mapa baseado no cálculo do *Índice de Moran* que busca medir a autocorrelação espacial entre a proximidade dos bairros e a tipificação de violência. O objetivo final do trabalho destacou as áreas com maior número de denúncias e que, portanto, precisam intensificar as ações de políticas públicas.

Para análise do movimento “Não mereço ser estuprada” presente em uma página no Facebook, o estudo BORTOLON et al. (2015) faz uso do conceito sociológico de repertório de ação coletiva que trata de movimentos sociais, trazendo a noção de ciberespaço como um termo futuramente legado por considerar a rede online uma parte da vivência real e não oposta a ela. Para extrair dados da página o projeto utilizou a aplicação *Netvizz* e o método *Top Modeling* para classificá-los e por fim fazer a auditoria dos dados. O resultado do principal contingente de mensagens postadas foram relatos de violação que, de acordo com a análise, motivou outras mulheres a compartilhar suas experiências.

O trabalho Recuero et al. (2015) analisou as manifestações online da grande onda de pro-

testos ocorridos em 2013, em uma tentativa de entender como as mídias sociais podem contribuir na organização e fortalecimento dos protestos, além das ações online. Utilizando métodos quantitativos e qualitativos de análise de conteúdo, foi possível construir classificações que nos inspiram com a aplicação para outros movimentos em um trabalho correlacionado a nossa pesquisa. O estudo concluiu, com o uso das funções de linguagem de *Jakobson*, que a maior ocorrência de postagens foram com hashtags conativas e referenciais.

Contribuindo ao debate sobre a importância dos dados que podem ser coletados em rede social, [Purohit et al. \(2015\)](#) faz um paralelo entre as dificuldades existentes nos métodos tradicionais de pesquisa e análise de violência de gênero com as contribuições de tecnologias modernas que ampliam a cobertura do estudo. Utilizando o termo ‘*gender-based violence (GBV)*’ para se referir a violência exercida contra alguém com base em seu gênero, o estudo reflete sobre o uso do Big Data para levar a pesquisa a uma escala global em termos de dados coletando relatos com o Twitter inicialmente. Finalmente, temos uma reflexão sobre a abordagem de suporte à pesquisa tradicional, como também críticas relacionadas ao controle de amostragem, disponibilidade de dados de determinadas localizações, a linguagem utilizada no meio digital, dentre outros.

[McCauley et al. \(2018\)](#) fez um trabalho interessante ao analisar relatos de *IPV (intimate partner violence)* dentro das ferramentas “*Duluth Power and Control Wheel*” e “*Women’s Experience with Battering (WEB)*”. *Power and Control Wheel* é um guia para identificar as diferentes maneiras que um parceiro pode ser abusivo em um relacionamento utilizando poder e controle, uma ferramenta bastante utilizada por assistentes sociais no suporte às vítimas de violência e está descrita na Figura 2.2. As consequências dos abusos foram identificadas dentro do framework *WEB* citado neste parágrafo, que por definição é uma escala que busca medir a experiência de mulheres com práticas abusivas.

[Heather McCauley et al. \(2018\)](#) utilizou a hashtag *#MaybeHeDoesntHitYou* na rede social Twitter para filtrar relatos de *IPV* com o software *NCapture* durante uma semana. A hashtag viralizou após *Zahira Kelly* começar um movimento na rede em resistência a minimização de violências não físicas contra mulheres. A equipe responsável codificou amostras dos tweets inicialmente para obter consistência e depois codificar todos os dados recuperados.

Ao final da pesquisa, houve relatos classificados em todas as dimensões do *Power and Control Wheel* e do *Women’s Experience with Battering (WEB)*, como também houveram experiências de *IPV* que não estavam representadas dentro da ferramenta *Power and Control Wheel* como a *coerção reprodutiva*. O estudo compreende, com todas as observações, que a rede social pode ser um lugar para criar conexão através da identificação com as histórias e, por fim, o desenvolvimento da empatia em torno do tema. Porém, também ressalta a necessidade de alguns usuários em repensar o uso de gênero com a hashtag, tendo em vista que outros gêneros podem sofrer violência.

O trabalho brasileiro intitulado “Feminismo por hashtags: as potencialidades e riscos tecidos pela rede” [Silva Reis \(2017\)](#) debate as dinâmicas sociotécnicas por trás das dezenove



Figura 2.2: Power and Control Wheel: Ferramenta com os padrões de comportamentos abusivos de um agressor com sua parceira.

Fonte: <https://connect.springerpub.com/content/book/978-0-8261-7991-3/back-matter/app3>

principais campanhas movimentadas por hashtags em mídias sociais entre o ano de 2015 e o primeiro semestre de 2016; como também leva a discussão para o lado do poder de criação dessas tecnologias que estão nas mãos de uma maioria masculina de profissionais. Como resultado das observações classificatórias, as conclusões do trabalho trouxe que a violência sexual é um dos principais temas dos relatos. Para além das análises quantitativas, o artigo também trata o *slackativismo* (ativismo preguiçoso) sobre o qual é utilizado para definir o ativismo digital em caráter pejorativo e se questiona sobre os tipos de sujeitos que as campanhas têm atendido, quais as necessidades de incremento social, entre outras reflexões.

3

METODOLOGIA

A Metodologia proposta inclui três etapas principais como ilustrado na Figura 3.1. As seções a seguir apresentam os detalhes de cada Etapa da Metodologia assim como detalhes de implementação do protótipo de pesquisa desenvolvido para validar a abordagem proposta.

3.1 Abordagem Proposta

3.1.1 Etapa 1: Elaboração do Dicionário de Sinônimos de Violência de Gênero

O primeiro passo da metodologia é a criação de um Dicionário de Sinônimos de Violência de Gênero. Para isso, é necessário utilizar uma base de relatos similares entre si, discriminados por tipos de violência de gênero. Por exemplo, no caso da língua inglesa, essa já existe disponível [Karystianis et al. \(2018\)](#). No momento em que essa abordagem foi proposta, não foi possível encontrar uma base de relatos similares para a língua portuguesa, então, foi necessário construir tal base. Assim, a primeira contribuição desse trabalho é um Dicionário de Sinônimos de Violência de Gênero.

A metodologia utilizada para a sua construção do Dicionário utilizou uma base de relatos de violência contra a mulher categorizadas por tipo de violência. O primeiro passo da metodologia proposta consiste na Etapa **1.1 Busca de Relatos** (ver Figura 3.1) e compreende reunir relatos oriundos de diferentes fontes. Nessa busca, foram utilizados livros, matérias jornalísticas e relatórios de prefeituras que continham descrições de relatos de violência contra a mulher. Então, criamos uma lista de relatos, realizamos a **1.2 Identificação dos Tipos de Violência** (ver Figura 3.1) e esses relatos foram categorizados dentro dos cinco tipos de violência doméstica identificados na Lei Maria da Penha (vide Seção 1.2) discriminados a seguir: Moral, Sexual, Psicológica, Física e Patrimonial.

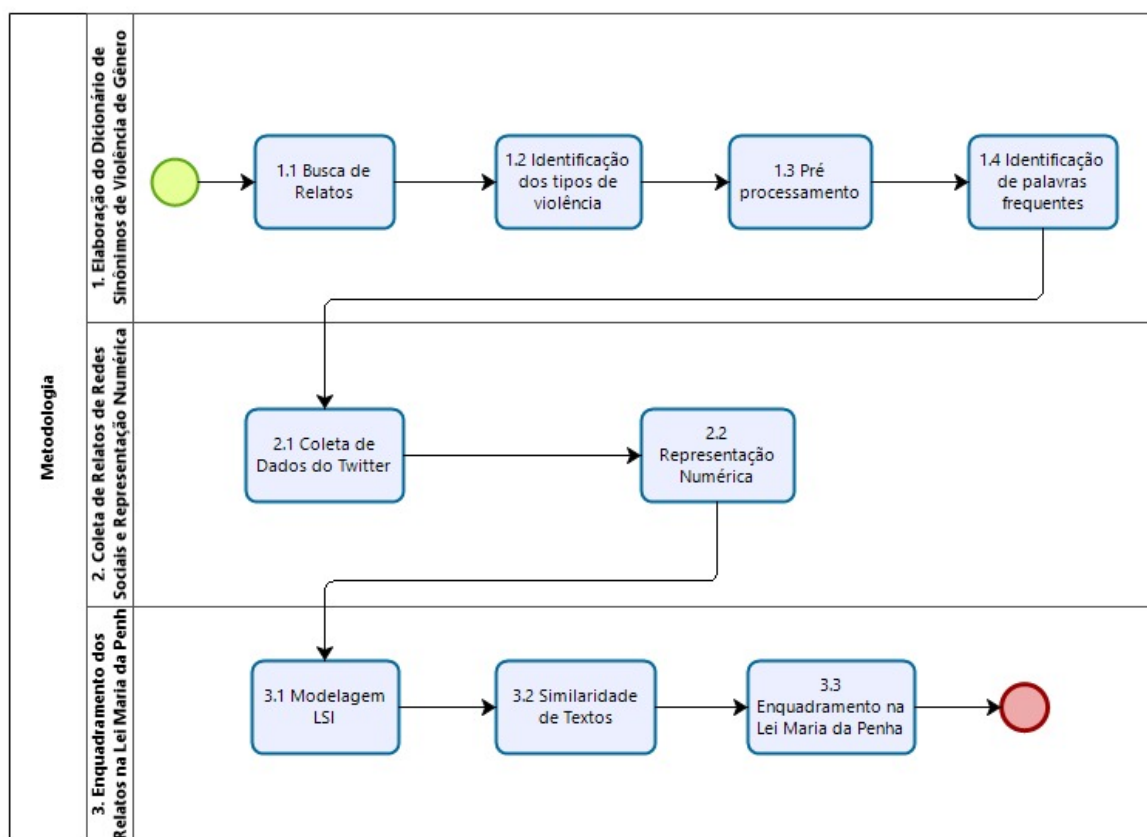


Figura 3.1: Metodologia - Etapas da abordagem de solução proposta.

Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

Após, fizemos um **1.3 Pré-processamento** (ver Figura 3.1) no Dicionário de Sinônimos retirando as *stopwords* – ou “palavras de parada” em tradução livre –, que remove palavras com menos expressões semânticas tais como preposições e conjunções “e, com, de, não” por exemplo. Com as palavras (ou, *tokens*) que restaram da lista, fizemos a lematização para deixá-las em suas formas básicas, sem sufixos flexionais. Por exemplo, “bateu, bate, batia” têm como lema comum a palavra “bater”. Por fim, retiramos também palavras que aparecem nos relatos apenas uma vez com o intuito de termos somente palavras com mais significância ao problema estudado.

Além disso, executamos um algoritmo que faz a **1.4 Identificação de Palavras Frequentes** (ver Figura 3.1) presentes na lista e, dessa maneira, obter criamos um **Dicionário de Sinônimos de Violência de Gênero** (ver Figura 3.1). A seguir, mostramos um exemplo didático de Dicionário, armazenado em forma de vetores e que contém um conjunto de palavras associadas a cada tipo de violência.

- Violência Sexual = [estuprar, abusar, assediar, agarrar]

- Violência Física = [bater, machucar, espancar, empurrar, agarrar]
- Violência Psicológica = [ameaçar, humilhar, xingar, ofender]
- Violência Moral = [acusar, chantagear, ofender, assediar]
- Violência Patrimonial = [quebrar, destruir, pegar, roubar]

3.1.2 Etapa 2: Coleta de Relatos de Redes Sociais e Representação Numérica

Após a Etapa de criação de um Dicionário de Sinônimos devidamente mapeados nos tipos de violência que a Lei Maria da Penha descreve, realizamos a próxima Etapa, resumida em coletar relatos de redes sociais e os representá-los numericamente. A metodologia proposta neste trabalho baseia-se em relatos do Twitter e a **2.1 Coleta de Dados do Twitter** (ver Figura 3.1) foi dividida em coleta por *hashtag* e coleta por palavras-chave. As *hashtag* utilizadas foram *#EuViviUmRelacionamentoAbusivo*, *#maselenuncamebateu*, *#MeuExAbusivo*, *#relacionamentoabusivoquando* e *#elenãotebate* pois foram movimentos virtuais de denúncias de violência que geraram grande repercussão. Já as palavras-chave utilizadas na busca foram "violência contra mulher", "violência doméstica" e "violência de gênero" pois são termos utilizados com frequência para se referir ao assunto que buscávamos.

Após a coleta dos relatos do Twitter baseado nos critérios descritos acima, utilizamos os mesmos passos de pré-processamento de textos descritos na Seção 3.1.1. Após esse tratamento dos relatos, criamos uma **2.2 Representação Numérica** (ver Figura 3.1) para cada relato utilizando o Dicionário de Sinônimos como referência. A técnica utilizada para representar numericamente os relatos é a *bag-of-words*. Essa técnica é apelidada de “saco de palavras”, pois ela registra a quantidade que cada palavra de um dicionário pré-definido aparece em um determinado texto. Por exemplo, temos o dicionário D , com suas respectivas tuplas “índice:palavra”: $D = 0 : bater, 1 : machucar, 2 : espancar, 3 : empurrar$, o texto “ele me empurrou e me bateu, depois me empurrou de novo e eu caí”, modelado através da *bag-of-words*, resulta no vetor $[1, 0, 0, 2]$, onde cada índice do vetor representa a frequência que cada palavra do dicionário aparece no texto.

3.1.3 Etapa 3: Enquadramento dos Relatos na Lei Maria da Penha

Uma vez que os relatos foram representados numericamente (ver Seção 3.1.2, passamos para a próxima etapa que enquadra os relatos de violência contra a mulher na Lei Maria da Penha. Para isso, utilizamos o modelo *Latent Semantic Indexing* (LSI) [Deerwester \(1988\)](#), que cria um espaço vetorial para a representação numérica que foi gerada para cada relato; ver **3.1 Modelagem LSI** na Figura 3.1. Esse espaço vetorial é mais adequado para modelar textos e extrair informações mais detalhadas da similaridades entre eles.

Em seguida, realizamos a operação **3.2 Similaridade de Textos** (ver Figura 3.1), que compara o quão similar é o relato modelado em LSI é com em relação ao Dicionário de Sinônimos. Como resultado, temos uma **Vetor de Similaridades** que indica o quão similar o relato é para cada tipo de violência da Lei Maria da Penha. Esse vetor é ordenado por ordem de decrescente de similaridade. Por exemplo, para o relato “Mulher recebe indenização por assédio de patrão em Araxá A violência contra a trabalhadora foi provada pelas conversas trocadas no WhatsApp”, o vetor resultante é [*'Violência Moral'*, *'Violência Sexual'*, *'Violência Psicológica'*, *'Violência Patrimonial'*, *'Violência Física'*], que indica que esse relato há mais similaridade com violência moral do que as demais.

Dessa maneira, utilizamos o Vetor de Similaridades para realizar a classificação de cada relato à Lei Maria da Penha; ver **3.3 Enquadramento na Lei Maria da Penha** na Figura 3.1. Ou seja, no exemplo do relato acima, o enquadramento seria no Art. 7º, Inciso V: “Art. 7º São formas de violência doméstica e familiar contra a mulher, entre outras: [...] V - a violência moral, entendida como qualquer conduta que configure calúnia, difamação ou injúria”. Assim, nos termos deste trabalho, enquadrar significa classificar relatos de acordo com suas semelhanças em relação ao Dicionário de Sinônimos.

3.2 Detalhes de Implementação

A fim de validar a abordagem proposta na Metodologia, implementamos um protótipo de pesquisa cujo código-fonte está disponível publicamente em [Aguiar \(2021\)](#). Para construção do protótipo [Aguiar \(2021\)](#), utilizamos a linguagem de programação Python e precisamos desenvolver partes do projeto fora do algoritmo com o objetivo de construir uma base de relatos para a construção do Dicionário de Sinônimos. Os próximos passos do desenvolvimento do algoritmo foram criar funcionalidades e implementar bibliotecas existentes como detalharemos a seguir.

Os primeiros passos da construção do protótipo ocorrem antes da implementação do algoritmo de classificação dos relatos, onde criamos o Dicionário de Sinônimos. O processo de criação passa por uma busca manual de relatos com uma tipificação de violência pré-definida, armazenadas construção de uma estrutura de dados chamada *Documents*, e a identificação das palavras mais frequentes nos *Documents* inicial. O Dicionário de Sinônimos é o resultado final desta fase, que é utilizada no algoritmo para realizar classificações dos relatos.

Os *Documents* criados são importados para processamento que passam por três tratamentos iniciais em que utilizamos a biblioteca *Spacy* para Processamento de Linguagem Natural com um modelo pré-treinado em português. Com um arquivo de lista de *stopwords*, inicialmente processamos a base para a remoção destas palavras para continuar o tratamento de lematização com os tokens restantes. Para o processo de lematização tratamos os *Documents* como um objeto *Doc* do *Spacy* no qual cada objeto Token que o forma possui o atributo *Token* que precisávamos. Com as palavras em suas formas básicas sem sufixos flexionais, o algoritmo

remove os tokens com frequência menor a dois. A seguir, temos a função responsável pelo pré processamento da base:

```
def preprocessing(typeViolence):
    documents = createListFromFile("corpus/documents"+typeViolence+".txt")

    documentsLem = []

    nlp = spacy.load('pt_core_news_sm')
    for document in documents:
        doc = nlp(document)

        docLem = []
        for t in doc:
            if t.text.lower() not in stoplist:
                docLem.append(t.lemma_)

        documentsLem.append(docLem)

    texts = [
        [word for word in document]
        for document in documentsLem
    ]

    frequency = defaultdict(int)
    for text in texts:
        for token in text:
            frequency[token] += 1

    texts = [[token for token in text if frequency[token] > 1] for text in texts]

    return texts
```

Além do *Spacy*, utilizamos a biblioteca *Gensim* de código aberto que também disponibiliza suporte para o processamento de linguagem natural e aprendizagem de máquina. Usamos três classes do *Gensim* para construção do algoritmo: *Dictionary* do corpora, *LSIModel* de *models* e *MatrixSimilarity* de *similarities*. A *MatrixSimilarity* calcula a similaridade de textos utilizando a distância do cosseno dos ângulos dos vetores que representam os relatos e o Dicionário de Sinônimos.

Inicialmente a classe *Dictionary* cria um dicionário que será usado para montar uma representação de *bag-of-words* através do método *doc2bow*. Esta representação é formada por uma

lista de vetores. A classe *LSIModel* que implementa o modelo *LSI* (*Latent Semantic Analysis*) utiliza o corpus, formado por uma lista de vetores, e o nosso dicionário, retornando assim, uma modelagem da transformação entre espaço vetoriais. E, por fim, implementando a classe *MatrixSimilarity* para calcular a similaridade e armazenar em uma matriz de índices.

Com os dicionários, modelos LSI e indexação, que criamos dos tipos de violência, conseguimos calcular a similaridade dos textos das postagens do Twitter com os tipos de violência que consta no relato. Por fim, ordenamos os pesos e obtivemos as violências mais recorrentes nos relatos de violência contra mulher, gravando em uma tabela (*DataFrame*) e posteriormente em um arquivo CSV, que guarda dados no formato texto puro separando-os por vírgula.



RESULTADOS

O conjunto de dados para gerar os resultados foram coletados durante uma semana, pois é o período limite da funcionalidade gratuita da interface de programação (*Application Programming Interface – API*) do Twitter. Foi 399 o número total de postagens coletadas do Twitter em agosto de 2020. Para fins de validação da abordagem proposta neste trabalho, consideramos que essas postagens são relatos de vítimas de violência de gênero. Já os relatos coletados e classificados manualmente, que compõem o Dicionário de Sinônimos, totalizaram 122 relatos com cobertura máxima de todos os cinco tipos de violência que trabalhamos.

Lembramos que o protótipo desenvolvido automatiza a coleta e a classificação de relatos de violência doméstica contra mulher em cinco tipos distintos de acordo com a Lei Maria da Penha: violência física, psicológica, sexual, patrimonial e moral. Assim, o resultado da tipificação gera uma variável categórica onde, para cada tipo, existe uma quantidade associada; sendo assim, escolhemos plotar os resultados em um gráfico *lollipop plot* que é um híbrido entre um gráfico de dispersão e um gráfico de barras em que mostra a relação de uma variável numérica e uma variável categórica.

A Figura 4.1 foi gerada com os resultados do tipo mais frequente em cada relato analisado. Foi identificada com mais frequente a violência moral e violência sexual ficou em segunda lugar. Esse resultado não era esperado, pois a hipótese de pesquisa era que as violências predominantes seriam a física e sexual. Sobre os detalhes a respeito da construção da Figura 4.1, consultamos a tabela gerada pelo algoritmo de classificação para buscar o primeiro tipo mais frequente para cada relato e fizemos incrementos de valor em um *defaultdict* de *Python* até conseguirmos os valores totais para plotagem. Utilizando a biblioteca *Pandas* criamos um *DataFrame* com as informações geradas e com a coleção *Pyplot* de funções da biblioteca *Matplotlib* usamos o *DataFrame* para plotar o gráfico de classificação geral da pesquisa.

Já para as figuras com o segundo tipo de violência mais frequente nos relatos, fizemos um *defaultdict* para cada tipo no qual também fizemos incrementos de valor até conseguirmos os

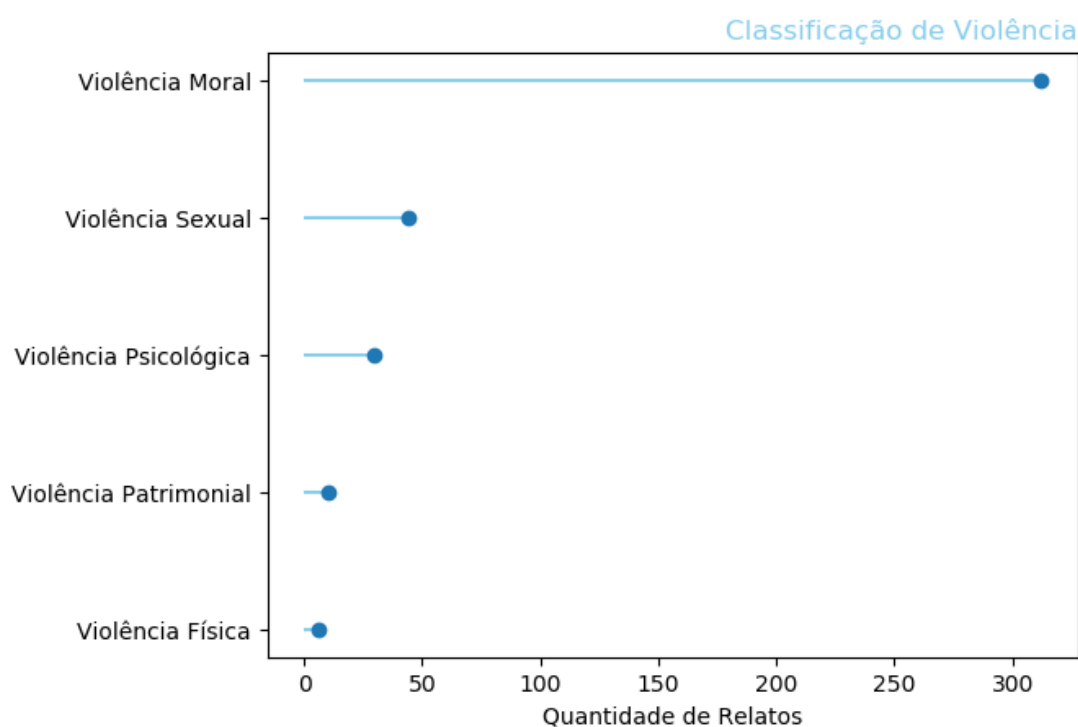


Figura 4.1: **Classificação de Violência de Gênero de todos os dados coletados do Twitter.** A violência moral é tipo de violência identificado com maior presença nas postagens.

Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

valores totais para plotagem dos próximos cinco gráficos. Também utilizamos *dataframes* e o *Matplotlib* para coleta e geração de gráficos para o segundo nível de classificação.

A Figura 4.2 relaciona os valores em quantidade do segundo tipo mais frequente nos relatos quando o primeiro tipo de violência é a física. Foram poucos relatos, então os resultados não são tão significantes.

As Figuras 4.3 e 4.4 relacionam o segundo tipo de violência mais frequente nos relatos quando o primeiro tipo de violência é moral e patrimonial, respectivamente. Podemos destacar que, para a violência moral, temos que o segundo tipo mais frequente é violência psicológica, o que já era esperado pois são violências com definições diretamente correlacionadas na lei. Quando o primeiro tipo de violência é patrimonial, também há poucos relatos, logo o resultado não é significativo.

Seguindo com as análises, temos os últimos gráficos 4.5 e 4.6 que relacionam o segundo tipo mais frequente nos relatos quando o primeiro é psicológica e sexual, respectivamente. Como citado anteriormente, existe uma correlação direta das definições de violência psicológica e moral, o que explica a violência moral ser o segundo tipo mais frequente quando o primeiro do relato é violência psicológica. Com relação à Figura 4.6, esperava-se que o tipo de violência predominante fosse física, mas a violência moral foi mais marcante quando o primeiro tipo de violência é a sexual.

Para verificar a acurácia do algoritmo, nós pegamos uma amostra dos relatos coletados pelo

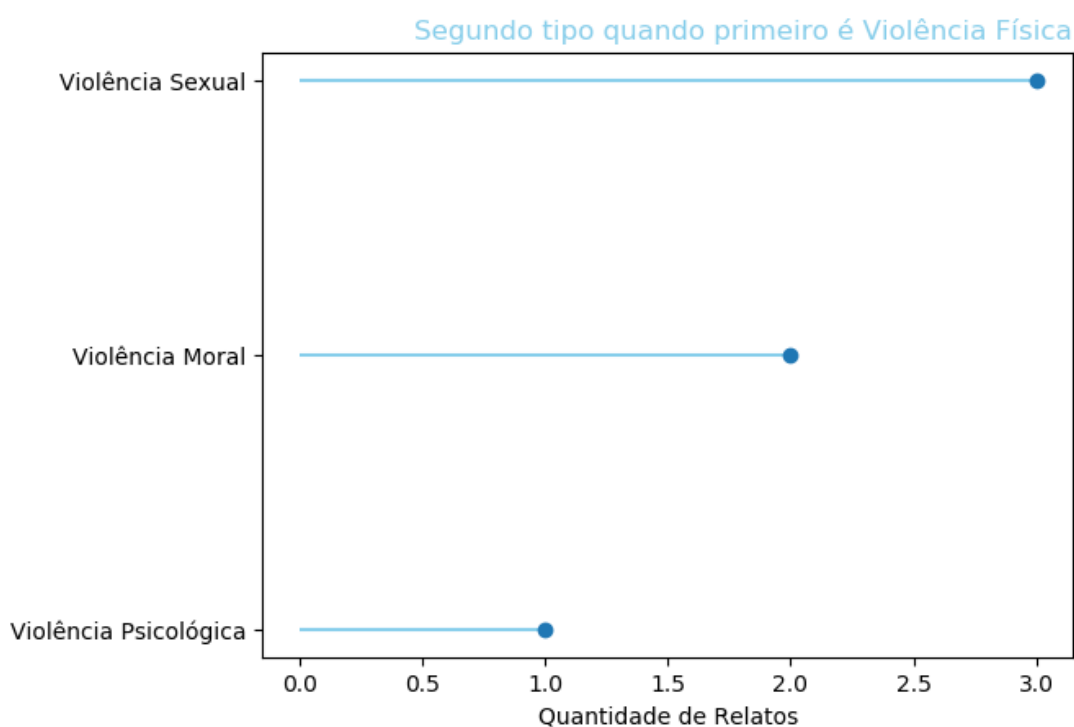


Figura 4.2: Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a **Violência Física**.
Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

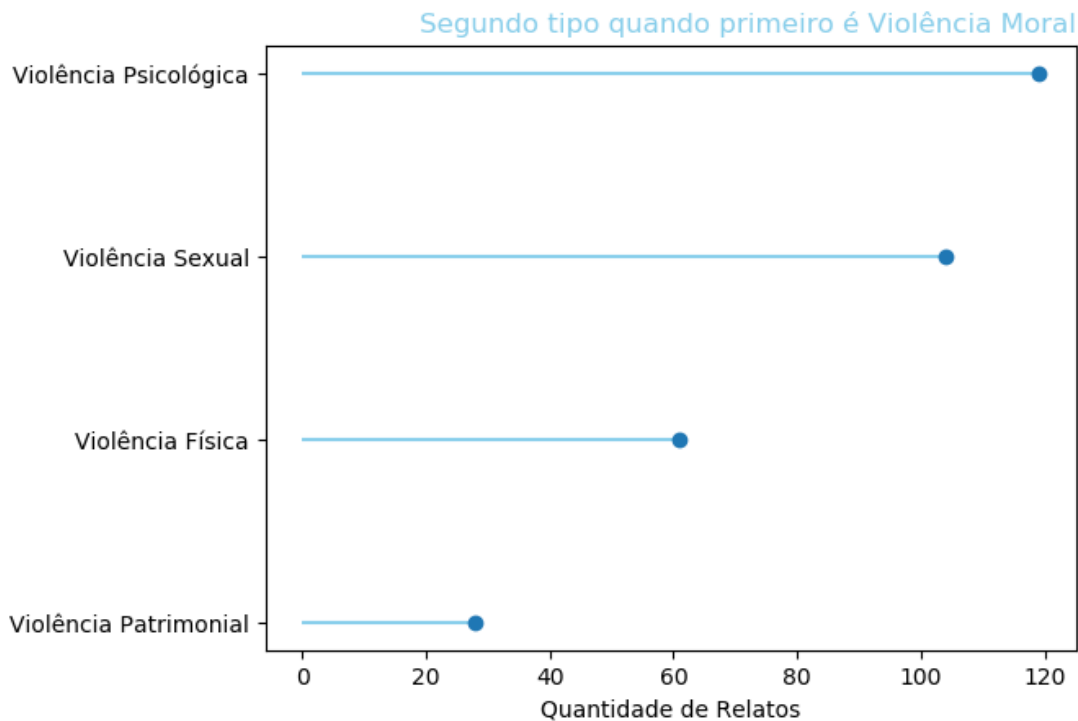


Figura 4.3: Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a **Violência Moral**.
Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

software para fazer uma classificação manual. Em 39,1% dos casos, não foi possível identificar um tipo específico de violência pois não há um detalhamento do que ocorreu. Em 85,25% des-

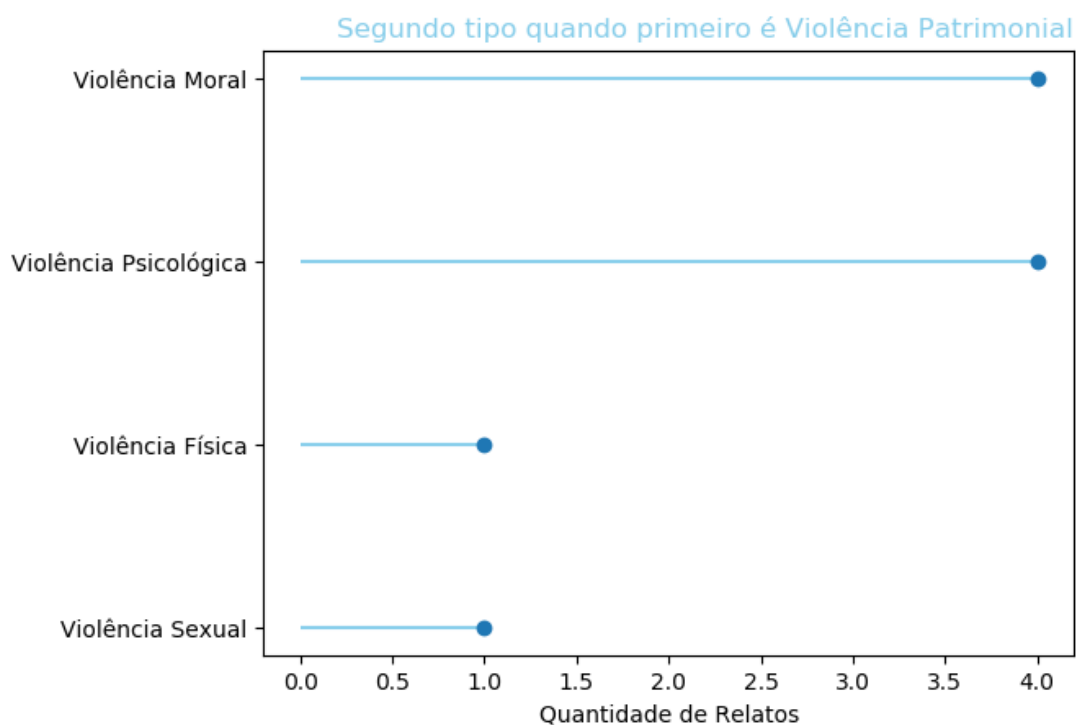


Figura 4.4: Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a **Violência Patrimonial**.
Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

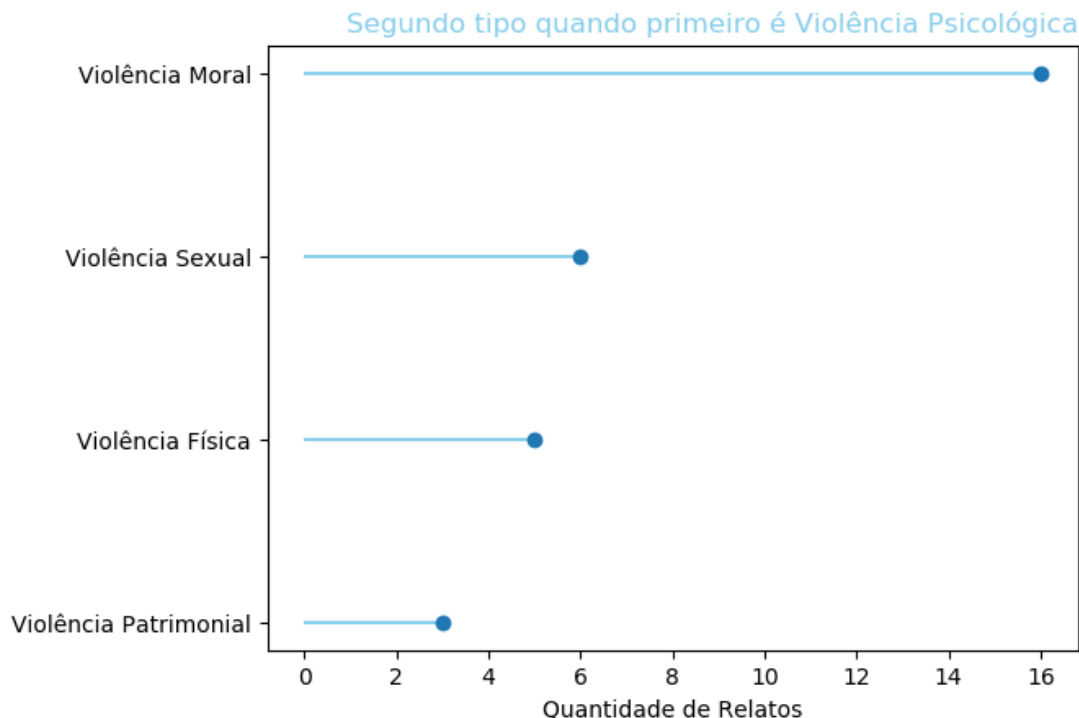


Figura 4.5: Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a **Violência Psicológica**.
Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

tes casos, verificamos que o algoritmo classificou o relato com a violência moral em primeiro lugar como o tipo mais provável. No entanto, para os 60,9% dos casos que conseguimos fa-

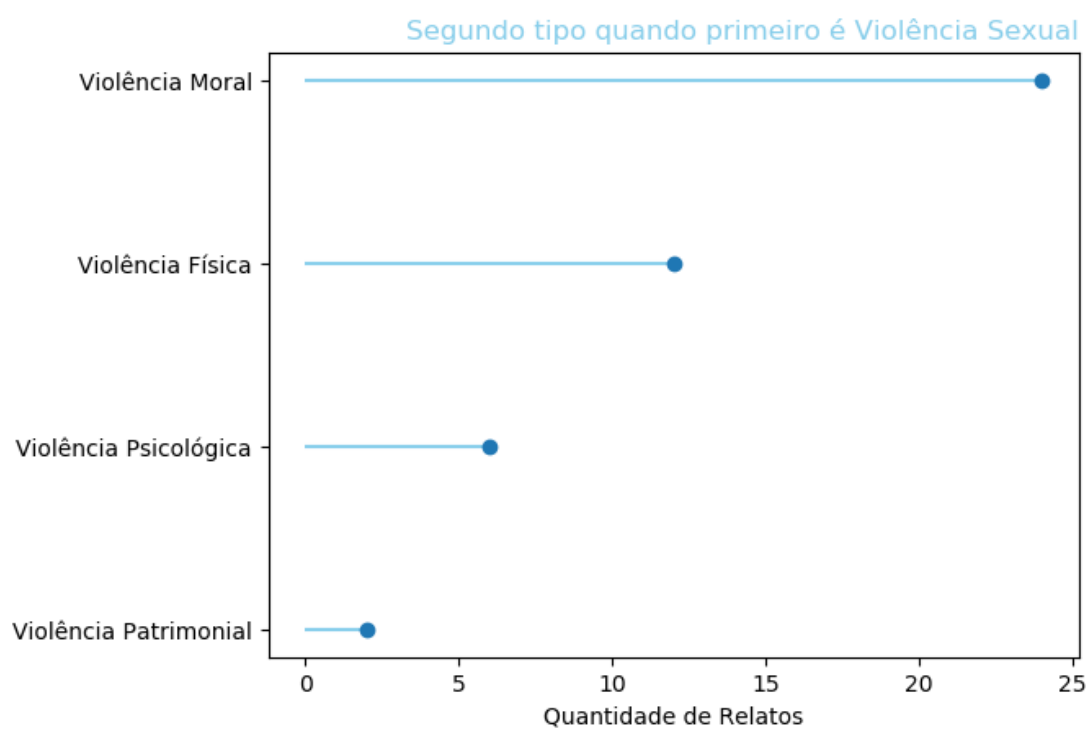


Figura 4.6: Segundo tipo de Violência de Gênero quando o primeiro tipo de violência é a **Violência Sexual**.

Fonte: Júlia Albuquerque Aguiar. Acervo pessoal.

zer a identificação manual, 35,79% destes foram enquadrados em violência sexual como mais provável e 33,68% como violência física.

5

CONCLUSÃO

Este trabalho propõe uma abordagem para coletar relatos de violência contra mulheres em redes sociais e enquadrá-los nos Artigos da Lei Maria da Penha. Como estudo de caso, foram relatos coletados da rede social Twitter e foram utilizadas técnicas de processamento de linguagem natural para realizar o enquadramento através da classificação dos relatos.

A classificação dessa violência se torna possível pois a vítima constrói sua narrativa baseada em uma mesma experiência de violência doméstica, o que muda, neste caso, é o meio utilizado para expor o relato. Com o contato frequente às notícias na televisão, jornais e histórias contadas por pessoas no âmbito social, construímos nossa hipótese de que a maioria dos casos coletados se enquadrariam no Artigo 7 §1 e §3 da Lei Maria da Penha, que discorrem sobre violência física e sexual respectivamente. Porém, uma parte da nossa hipótese não foi validada com a classificação feita pelo algoritmo, já que o tipo de violência identificado com mais frequência foi a moral, seguida da violência sexual. Assim, o nosso protótipo valida parcialmente a hipótese de pesquisa já que a maioria dos casos relatados puderam ser classificados como Violência Sexual.

Em contrapartida, ao fazer a análise manual dos relatos no processo de verificar a acurácia do algoritmo, nós identificamos que boa parte das postagens sem detalhamento da violência que ocorreu, foram classificadas como violência moral pelo protótipo. Sendo assim, se desconsiderarmos os relatos que não são possíveis tipificar pela falta de detalhamento, a classificação manual aponta que a violência mais frequente é a sexual, seguida pela violência física totalizando 69,47% dos casos, o que estaria de acordo com a nossa hipótese.

Acreditamos que este trabalho empodera o debate sobre os ciclos de violência e como os abusos morais e psicológicos estão diretamente envolvidos com as agressões físicas e sexuais; trazendo à tona também a importância de atentarmos a todas as fases deste ciclo de violência.

Ademais, esperamos que este trabalho sirva de base para outros trabalhos de pesquisa interdisciplinares. Por exemplo, na área das ciências sociais, o estudo de ações afirmativas que diminuam os impactos negativos sob as mulheres que denunciam casos de violência e/ou o

desenvolvimento de relatórios estatísticos com estimativas mais próximas do real sobre a subnotificação de violência contra mulher.

Já na área da Computação, uma possibilidade de trabalho futuro seria a implementação de Inteligências Artificiais que reconheçam mensagens com teor abusivo ou violento e/ou o desenvolvimento de chatbots em páginas de redes sociais. Assim, poderemos orientar mulheres em como denunciar casos de violência e como a lei pode protegê-la considerando seu caso específico. Além disso, a pergunta científica proposta neste trabalho pode ser testada em um conjunto de dados maior, com mais relatos coletados do Twitter durante um período mais significativo, por exemplo, um ano.

Além disso, considerando a análise manual feita dos resultados do protótipo, podemos futuramente colocar uma porcentagem mínima de similaridade para gerar casos de não identificado e enviar para uma “mesa” de análise manual. E não apenas isso, mas incluir um filtro de hashtags considerando o contexto do momento atual da pesquisa para evitar um viés no resultado e encontrar uma forma de considerar agressões de outros parentes próximos, da mulher em um relacionamento homoafetivo, por exemplo, e não apenas de homens.

Ainda na área da Computação, o protótipo de pesquisa desenvolvido é extensível e pode ser adaptado para outras redes sociais como o Facebook, por exemplo. Outro trabalho futuro poderá utilizar melhor modelagens como BERT embeddings [Devlin et al. \(2018\)](#) para discriminar melhor contextos e poder discernir, por exemplo, entre “assediar” no sentido moral e no sentido sexual ou para saber se “quebrar” refere-se a quebrar um objeto ou fraturar um osso do corpo. Por último, uma melhoria significativa seria criar um algoritmo que classificasse as postagens em redes sociais entre relatos de vítimas de violência de gênero e discussões sobre esse tema. Dessa maneira, teríamos mais significância em relação aos dados analisados, ou seja, trabalhando apenas com postagens que referem-se à relatos das vítimas.

Em relação à área do Direito, a metodologia proposta neste trabalho pode ser utilizada para classificar documentos jurídicos para outras leis por exemplo. Pode-se também utilizar a técnica de similaridade de documentos utilizada na metodologia para comparar jurisprudência, agrupando precedentes judiciais similares por exemplo.

Referências bibliográficas

- Júlia Albuquerque Aguiar. Protótipo de pesquisa do trabalho de conclusão de curso. <https://github.com/juliaguilar/bachelor-thesis>, 2021.
- Charlotte Alexander. #MeToo and the Litigation Funnel. *Employee Rights and Employment Policy Journal, Forthcoming, Georgia State University College of Law, Legal Studies Research Paper No. 2019-04*, 2019. Available at SSRN: <https://ssrn.com/abstract=3330795>.
- Instituto Avon and Folks Netnográfica. A voz das redes: o que elas podem fazer pelo enfrentamento das violências contra as mulheres. <https://dossies.agenciapatriciagalvao.org.br/dados-e-fontes/pesquisa/voz-das-redes/>, 3 2018.
- Bianca BORTOLON, Marianne MALINI, and Fábio MALINI. Gênero e ativismo online: um estudo de caso da campanha Não Mereço Ser Estuprada no Facebook. *XXXVIII Congresso Brasileiro de Ciências da Comunicação*, 09 2015.
- Audrey Carlsen, Maya Salam, Claire Cain Miller, Denise Lu, ASH NGU, Jugal K. Patel, and Zach Wichter. #MeToo brought down 201 powerful men. nearly half of their replacements are women., 2021. URL <https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html>.
- Conselho Nacional de Justiça. Violência doméstica: Brasil ganha formulário unificado de avaliação de risco, 2020. URL <https://www.cnj.jus.br/violencia-domestica-brasil-ganha-formulario-unificado-de-avaliacao-de-risco/>.
- Scott Deerwester. Improving information retrieval with latent semantic indexing. 1988.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Technical report, oct 2018. URL <https://github.com/tensorflow/tensor2tensor><http://arxiv.org/abs/1810.04805>.

- Victor Gil Vera, Isabel Puerta, Juan Arias, and Lina Montoya Suarez. The femicide in colombia and mexico: A text mining analysis. *The Turkish Online Journal of Design, Art and Communication*, pages 170–177, 03 2018. **DOI** [10.7456/1080MSE/121](https://doi.org/10.7456/1080MSE/121).
- UN Women Headquarters. Policy brief: The impact of covid-19 on women. Technical report, United Nations, 2020.
- Lori L. Heise, Jacqueline Pitanguy, and Adrienne Germain. Violence against women : the hidden health burden (english). *World Bank discussion papers ; no. WDP 255 Washington, D.C. : World Bank Group*, 1994.
- George Karystianis, Armita Adily, Peter Schofield, David Greenberg, Louisa Jorm, Goran Nenadic, and Tony Butler. Automated analysis of domestic violence police reports to explore abuse types and victim injuries (preprint). *Journal of Medical Internet Research*, 21, 12 2018. **DOI** [10.2196/13067](https://doi.org/10.2196/13067).
- Jessalynn Keller, Kaitlynn Mendes, and Jessica Ringrose. Speaking 'unspeakable things:' documenting digital feminist responses to rape culture. *Indian Journal of Gender Studies*, 27, 08 2016. **DOI** [10.1080/09589236.2016.1211511](https://doi.org/10.1080/09589236.2016.1211511).
- Heather McCauley, Amy Bonomi, Megan Maas, Katherine Bogen, and Teagen O'Malley. #MaybeHeDoesntHitYou: Social Media Underscore the Realities of Intimate Partner Violence. *Journal of Women's Health*, 27, 03 2018. **DOI** [10.1089/jwh.2017.6560](https://doi.org/10.1089/jwh.2017.6560).
- Celina Navarro, Luisa Martínez García, and Simón Vialás Fernández. Usos y valoración de Facebook como herramienta para la lucha contra la violencia de género en México. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 15(2):285–299, sep. 2018. **DOI** [10.5209/TEKN.59546](https://doi.org/10.5209/TEKN.59546). URL <https://revistas.ucm.es/index.php/TEKN/article/view/59546>.
- Lei nº. 11.340. Lei maria da penha. Cria mecanismos para coibir a violência doméstica e familiar contra a mulher, ago. 2006. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/111340.htm. Acesso em: 5 ago. 21.
- Hemant Purohit, Tanvi Banerjee, Andrew Hampton, Valerie Shalin, Nayanesh Bhandutia, and Amit Sheth. Gender-Based Violence in 140 Characters or Fewer: A #BigData Case Study of Twitter. *First Monday*, 03 2015. **DOI** [10.5210/fm.v21i1.6148](https://doi.org/10.5210/fm.v21i1.6148).
- Raquel Recuero, Gabriela Zago, Marco T. Bastos, and Ricardo Araújo. Hashtags Functions in the Protests Across brazil. *SAGE Open*, 5(2):2158244015586000, 2015. **DOI** [10.1177/2158244015586000](https://doi.org/10.1177/2158244015586000). URL <https://doi.org/10.1177/2158244015586000>.

- Redação, O Estado de S.Paulo. Maria da penha: os 15 anos da lei que respeita a mulher, 2021. URL <https://brasil.estadao.com.br/noticias/geral,maria-da-penha-os-15-anos-da-lei-que-respeita-a-mulher,70003803092>.
- Danielle Sanches, Andressa Contarato, and Renan Barbosa. Um outro olhar sobre a violência a mulher: o que dizem as denúncias no estado do Rio de Janeiro. Technical report, FGV DAPP, Rio de Janeiro, 2018.
- Josemira Silva Reis. Feminismo por hashtags: As potencialidades e riscos tecidos pela rede. *Seminário Internacional Fazendo Gênero 11 13th Women's Worlds Congress (Anais Eletrônicos)*, 2017.
- Elizabeth Chika Tippett. Harassment Trainings: A Content Analysis. *Berkeley Journal of Employment and Labor Law*, 12 2017. Available at SSRN: <https://ssrn.com/abstract=2994571>.
- Elizabeth Chika Tippett. The Legal Implications of the MeToo Movement. *Minnesota Law Review, Forthcoming*, 7 2018. Available at SSRN: <https://ssrn.com/abstract=3170764>.
- Javier Toret. Tecnopolítica: la potencia de las multitudes conectadas. El sistema red 15m, un nuevo paradigma de la política distribuida. *IN3 Working Paper Series*, 06 2013. DOI [10.7238/in3wps.v0i0.1878](https://doi.org/10.7238/in3wps.v0i0.1878). tradução do autor.
- World Health Organization. Violence against women, 2021. URL <https://www.who.int/news-room/fact-sheets/detail/violence-against-women>.