Capítulo 1

Combate Automático às *Fake News* nos Meios Digitais

Paulo Márcio Souza Freire, Argus Antônio Barbosa Cavalcante, Ronaldo Ribeiro Goldschmidt

Abstract

Combating Fake News (i.e., false news intentionally spread) is not a recent problem. However, its complexity has increased mainly due to the growth of volume and speed of news dissemination provided by the digital media of news distribution (e.g.: social networks, online newspaper, etc). In this scenario, computational approaches are becoming essential devices to combat this type of news. Thus, this chapter presents a study about the main computational approaches to combat Fake News, besides some comments on related areas and recent research on this theme.

Resumo

O problema de combater Fake News (isto é, notícias falsas veículadas de forma intencional) não é recente. Contudo, sua complexidade vem aumentando em função do crescimento do volume e da velocidade de divulgação de notícias proporcionado pelos meios digitais de divulgação de notícias (por exemplo, redes sociais, jornais on-line, etc.). Diante deste cenário, abordagens computacionais que possam auxiliar no combate deste tipo de notícia estão se tornando cada vez mais necessárias. Assim sendo, o presente capítulo apresenta um estudo sobre as principais abordagens computacionais de combate às Fake News, além de comentar sobre áreas e pesquisas atuais relacionadas a este tema.

1.1. Introdução

O consumo de notícias on-line vem aumentando a cada dia [Vosoughi et al. 2017]. Dentre as razões para este crescimento pode-se destacar o fácil acesso e o baixo custo proporcionado pelos meios digitais de divulgação de notícias (*MDDN*), compostos, basicamente, pelas mídias virtuais (por exemplo, jornais on-line), redes sociais e aplicativos de troca de

mensagens) [Souza Freire et al. 2021]. A divulgação de uma notícia em um meio digital está relacionada com dois eventos (isto é, publicação e propagação) [Shu et al. 2017]. A publicação corresponde ao evento no qual a notícia é disponibilizada pela primeira vez em um meio digital. Por outro lado, toda e qualquer reação ocorrida após a publicação (por exemplo, um comentário ou compartilhamento) corresponde a um evento de engajamento e, consequentemente, de propagação da notícia.

Apesar de seus benefícios, alguns *MDDN*, tais como as redes sociais, permitem que qualquer pessoa, independentemente de sua credibilidade, divulgue notícias com intenso poder de propagação [Shu et al. 2017, Wang et al. 2018a]. Tal permissividade amplificou a disseminação de *Fake News*, um tipo particular de notícia falsa cuja divulgação acontece de forma proposital [Freire and Goldschmidt 2019, Conroy et al. 2015, Zhang et al. 2018]. A proliferação de *Fake News*, geralmente, afeta não apenas a integridade jornalística, mas também perturba as áreas social, política, econômica, cultural, assim como da saúde e segurança [Mejova and Kalimeri 2020, Mustafaraj and Metaxas 2017, Wang 2017].

Como exemplo do poder de influência desse tipo de notícia, pode ser destacado que nos três meses finais das eleições presidenciais americanas, realizadas em 2016, as notícias falsas publicadas na rede social Facebook, que favoreceram qualquer um dos dois candidatos, foram compartilhadas 37 milhões de vezes [Farajtabar et al. 2017]. Ademais, casos relacionados às *Fake News* não se limitam aos EUA. Conforme divulgado pela *BBC News*[1], após notícias falsas terem, supostamente, levado a linchamentos em 2018 na Índia, o *WhatsApp* anunciou um limitador para a quantidade de encaminhamentos de mensagem. Infelizmente, o Brasil também apresenta acontecimentos vinculados às *Fake News*. Segundo matéria divulgada pelo jornal *on-line G1*[2], uma dona de casa, de 33 anos, morreu dois dias após ter sido espancada por dezenas de moradores do município de Guarujá, no litoral de São Paulo. De acordo com essa matéria, ela foi agredida a partir de uma notícia falsa, divulgada por uma rede social, que afirmava que a dona de casa sequestrava crianças para utilizá-las em rituais de magia negra.

Além disso, as *Fake News* também podem trazer riscos à Segurança Nacional, haja vista a possibilidade de que notícias intencionalmente falsas consigam interferir nas atividades tanto da Inteligência quanto da Contrainteligência. Ressalte-se que tais atividades são consideradas fundamentais e indispensáveis à segurança dos Estados, da sociedade e das instituições nacionais[3]. Portanto, há um apelo urgente para desenvolver estratégias efetivas para combater o impacto desse tipo de notícia falsa, não só no âmbito civil como também no militar, especificamente no que tange à Defesa Cibernética.

Como um último e atual exemplo do poder de influência das *Fake News*, pode-se destacar o caso da pandemia de *COVID-19* (*Coronavirus Disease-19*), causada pelo patógeno *SARS-Cov-2* e que já matou milhares de pessoas no ano de 2020 ao redor do mundo[4], onde inúmeras *Fake News* têm sido divulgadas em MDDN [Mejova and Kalimeri 2020]. Estas divulgações têm dificultado de forma significativa o esclarecimento da população

[1]BBC News - https://www.bbc.com

[2]G1 - https://http://g1.globo.com

[3]Agência Brasileira de Inteligência. http://www.abin.gov.br/atividadeinteligencia/inteligenciaecontrainteligencia/

[4]www.who.int/emergencies/diseases/novel-coronavirus-2019

sobre a disseminação da pandemia e sobre as devidas medidas de enfrentamento da doença a serem adotadas a cada instante.

Devido aos impactos das *Fake News* disponibilizadas em MDDN, diferentes segmentos da sociedade têm pesquisado como combatê-las [Zhou et al. 2019, UNESCO 2019, Flintham et al. 2018, Wang et al. 2018a, Campan et al. 2017, Kshetri and Voas 2017]. Como consequência, algumas ações mitigadoras estão sendo potencializadas, onde é possível enfatizar a criação de legislação punitiva[5], os serviços de checagem de fatos (por exemplo, AosFatos[6]), as iniciativas educacionais presentes na alfabetização midiática [UNESCO 2016] (ex: Jogos Educacionais Digitais - JED que buscam capacitar pessoas para avaliar criticamente a veracidade das notícias [Passos et al. 2020, Passos et al. 2021]) e o emprego de abordagens computacionais nos MDDN [Freire and Goldschmidt 2020].

Dentre as ações mitigadoras, o emprego de abordagens computacionais vem se destacando devido à sua maior velocidade de atuação [Ruchansky et al. 2017]. Uma ação rápida se faz necessária, pois o espalhamento das *Fake News* se apresenta como um problema não trivial, tanto pelo volume de publicações quanto pela velocidade das suas respectivas propagações [Shu et al. 2017].

As abordagens computacionais utilizadas no combate automático às *Fake News* nos MDDN, conforme proposto por [Freire and Goldschmidt 2020], podem possuir duas funcionalidades: *Detecção e Intervenção*. Simplificadamente, enquanto a detecção procura identificar se uma notícia é intencionalmente falsa, a intervenção busca mitigar os efeitos da divulgação (publicação/propagação) dessa notícia em um determinado meio digital de divulgação de notícias.

Baseado nesta necessidade computacional, o presente capítulo provê uma introdução ao referido combate através da seguinte estrutura: a Seção 1.2 apresenta diferentes definições para o termo *Fake News*, assim como aborda o comportamento disseminativo deste tipo de notícia nos MDDN e conceituações sobre *Verdade*. Um levantamento sobre os trabalhos relacionados é realizado na Seção 1.3. Por fim, na Seção 1.4, são abordados os problemas em aberto.

1.2. Fundamentos

Apesar da histórica existência de notícias fraudulentas em nossa sociedade, a utilização do termo *Fake News* é relativamente recente. Desta forma, se faz necessária a caracterização das diferentes definições para *Fake News*, assim como o seu comportamento disseminativo e algumas das conceituações sobre *Verdade*.

1.2.1. Caracterização de *Fake News*

Apesar da originalidade da expressão, as *Fake News* não surgiram com o uso dos MDDN. Haja vista que, mesmo com as mídias tradicionais, já existiam pessoas que, por diferentes razões, divulgavam notícias falsas de forma proposital [Golbeck et al. 2018]. Independente do surgimento, devido à contemporaneidade do termo *Fake News*, surgiram

[5]https://www12.senado.leg.br/noticias/materias/2020/06/02/nova-versao-de-lei-contra-fake-news-tera-restricoes-a-contas-anonimas-e-mais-poder-a-denuncias-de-usuarios

[6]https://www.aosfatos.org/

diferentes definições, as quais podem ser organizadas em dois grupos.

O primeiro grupo considera que o aspecto proposital é fundamental, pois define as *Fake News* como publicações intencionais e verificadamente falsas [Conroy et al. 2015, Reis et al. 2019, Zhou et al. 2019, Campan et al. 2017, Wang et al. 2018a, Shu et al. 2017, Zhou and Zafarani 2018, Flintham et al. 2018, Mustafaraj and Metaxas 2017]. Assim, para esse grupo, não basta a notícia ser falsa para ser caracterizada como *Fake News*, é também preciso ter sido divulgada intencionalmente. Para enfatizar a diferença entre uma notícia falsa e uma intencionalmente falsa, pode-se utilizar dois termos denominados *misinformation* e *disinformation* [Golbeck et al. 2018, Campan et al. 2017]. Enquanto *misinformation* corresponde às notícias falsas divulgadas pela falta da informação verdadeira, a *disinformation* diz respeito às notícias falsas divulgadas com algum propósito. Com base nessas correspondências, para o primeiro grupo, é possível caracterizar *Fake News* como sendo uma *disinformation* [Kshetri and Voas 2017]. Cabe ressaltar que, apesar de pertencente ao primeiro grupo, o trabalho [Zhou and Zafarani 2018] é ainda mais específico em sua definição, pois só considera *Fake News* quando a notícia intencionalmente falsa é divulgada por uma agência de notícias.

Ademais, ainda de acordo com esse primeiro grupo, existem outros campos de estudo que, apesar de não se enquadrarem na área de *Fake News*, apresentam uma relação com o combate às notícias intencionalmente falsas. Alguns desses campos se encontram descritos abaixo:

- Classificação de Rumores (*Rumor Classification*) - Rumor é uma informação em circulação cuja veracidade não foi verificada no momento da publicação. Um rumor pode ser classificado como verdadeiro, falso ou ainda não verificado [Ma et al. 2015, Shu et al. 2017, Liu and Xu 2016, Vosoughi et al. 2017]. Portanto, uma *Notícia* não verificada antes da publicação é um *Rumor*, que pode ser caracterizado como *Fake News* a partir do momento que seja identificado como falso e intencional. A tarefa mais relacionada com o combate às *Fake News* é a classificação da veracidade dos rumores;

- Descoberta da Verdade (*Truth Discovery*) - é a descoberta da verdade de fatos conflitantes entre diferentes fontes [Shu et al. 2017, Li et al. 2015]. Assim, uma mesma *Notícia* pode conter afirmações diferentes (isto é, distintas opiniões), onde as intencionalmente falsas podem ser caracterizadas como *Fake News*. Assim, o combate às *Fake News* pode se beneficiar da Descoberta da Verdade para determinar a veracidade das afirmações;

- Detecção de Iscas de Cliques (*Clickbait Detection*) – procura identificar, nas páginas *Web*, as chamadas iscas de cliques que, praticamente, forçam o usuário a selecionar a opção apresentada. Nesse caso, o corpo do texto (*bodytext*) dos artigos é, frequentemente, pobre em relação ao seu cabeçalho (*headline*). Essa discrepância pode ser encontrada não só em *Clickbait*, como também em *Fake News*. Sendo assim, o *Clickbait* pode ser usado como um indicador de *Fake News* [Shu et al. 2017];

- Detecção de Bots (*Bot Detection*) – procura identificar o envio automático de informações nas redes sociais por meio de robôs [Braz and Goldschmidt 2017]. Esses

envios podem potencializar tanto a publicação quanto a respectiva propagação da *Fake News* [Wang et al. 2018a, Nasim et al. 2018, Ferrara et al. 2016];

- Checagem de fatos (*Fact Checking*) - são *Websites* ou *Frameworks* responsáveis pela verificação, normalmente realizada com a ajuda de especialistas, da veracidade de fatos divulgados em MDDN [Ciampaglia et al. 2015, Vo and Lee 2018, Sethi 2017, Ruchansky et al. 2017]. Inclusive, existem abordagens voltadas para a seleção automática de notícias a serem enviadas para a referida checagem de fatos [Kim et al. 2018, Tschiatschek et al. 2018]. A verificação da verdade dos fatos pode ser utilizada na tarefa de detecção de *FakeNews* [Cazalens et al. 2018], assim como na criação de *datasets* [da Silva et al. 2020];

- Sistemas de Reputação (*Reputation System*) - são sistemas que buscam determinar o nível de confiança em MDDN baseados na obtenção de graus de reputação dos usuários [Vavilis et al. 2014, Hendrikx et al. 2015, Seo J. 2013, Deng et al. 2014, Sherchan et al. 2013]. A determinação de graus de reputação pode ser utilizada na tarefa de identificação das *Fake News*.

O segundo grupo, entretanto, tem uma definição mais genérica. Para esse segmento, as *Fake News* são todas as notícias falsas, independente da sua natureza intencional [Sharma et al. 2019, Castelo et al. 2019, Ajao et al. 2019]. Inclusive, consideram-se como *Fake News* outros tipos de notícia, como, por exemplo, Rumor.

Este trabalho adota a definição do primeiro grupo. Consequentemente, considera *Fake News* como sendo uma notícia intencionalmente falsa. A principal razão da escolha é que uma notícia propositalmente divulgada tende a ser mais bem elaborada, podendo causar mais malefícios aos usuários.

1.2.2. Comportamento Disseminativo

A disseminação e, consequente, divulgação de uma notícia se inicia pela sua publicação e provável propagação (Efeito de Câmara de Eco) [Shu et al. 2017]. Dessa forma, é importante destacar o momento no qual uma notícia pode ser caracterizada como *Fake News*. Basicamente, uma notícia intencionalmente falsa pode surgir de três formas. A primeira é quando a *Fake News* é iniciada (publicada) em uma mídia virtual, podendo, posteriormente, ser republicada em uma rede social ou em um aplicativo de troca de mensagens. Na segunda forma, esse tipo de notícia é iniciada (publicada) diretamente na rede social ou em um aplicativo de troca de mensagens. Independente de ter surgido ou não na rede social ou em um aplicativo de troca de mensagens, a partir da sua chegada, essa notícia pode ser potencializada pela sua propagação. A terceira é quando uma notícia não *fake* é publicada, porém se torna *fake*, a partir do seu espalhamento, de acordo com as contribuições intencionalmente falsas feitas durante a sua propagação.

Independente do momento de criação, a recente proliferação de notícias falsas e mal-intencionadas nos MDDN, em especial nas redes sociais, tem sido uma fonte de preocupação generalizada. Essa apreensão se deve pelo seu poder de espalhamento e, consequente, influência na sociedade [Flintham et al. 2018]. As razões que potencializam a divulgação das *Fake News* nos MDDN podem ser divididas em quatro categorias. A

primeira tem relação com poder de influência ocasionado pelos fatores inerentes ao ser humano, dentre eles podemos destacar que as pessoas [Shu et al. 2017]:

- Preferem receber informações que confirmem as suas opiniões sem, necessariamente, verificarem a veracidade da notícia;
- Tendem a aceitar as informações não pela análise da verdade, mas pela relação de ganhos e perdas que a notícia pode trazer para elas;
- Tendem a avaliar as informações sem a busca da veracidade, pois acabam acompanhando a aceitação dos outros.

A segunda categoria é a carência de legislação punitiva, uma das alegações para tal fato é que as referidas leis poderiam cercear a liberdade de expressão. A terceira categoria está vinculada ao potencial ganho financeiro com a divulgação de determinadas notícias nos MDDN [Kshetri and Voas 2017]. Já a quarta categoria advém da facilidade de criação de contas nas redes sociais [Conroy et al. 2015]. Um aspecto importante inerente à essa facilidade é a criação de contas digitais maliciosas por meio de divulgadores de natureza humana e/ou computacional [Shu et al. 2017]. Esses divulgadores subdividem-se em:

- *Bot* - robôs responsáveis por divulgar *Fake News*;
- Humano - pessoas (*trolls*) intencionadas em disseminar *Fake News*;
- *Cyborg* - mecanismos híbridos (*Bot*/Humano) que divulgam *Fake News*.

Ainda se tratando da facilidade de divulgação de notícias intencionalmente falsas nas redes sociais, uma das formas mais simples de criar uma *Fake News* é se infiltrar em uma comunidade de pessoas engajadas em discutir um determinado assunto. Para tanto, segundo [Mustafaraj and Metaxas 2017], devem ser realizados os seguintes passos para divulgação de *Fake News*: Criar um domínio falso (*website*), criar contas anônimas, identificar comunidades e usuários interessados em um determinado assunto, contaminar esses usuários com a notícia falsa e, finalmente, incentivar a discussão para que a *Fake News* seja espalhada.

1.2.3. Conceituações sobre Verdade

Conforme introduzido neste trabalho, os dois grupos que definem *Fake News* convergem ao considerá-las como falsas. Diante disto, torna-se importante procurar conceituar o que é *Verdade*.

Essa conceituação sobre *Verdade* de uma afirmação é ampla, podendo-se destacar a existência de diferentes concepções filosóficas sobre a natureza do conhecimento verdadeiro. Inclusive, tal amplitude pode ser caracterizada pelas concepções ceticista e relativista. A primeira concepção, na sua forma clássica (ou ceticismo pirrônico) defende uma postura suspensiva, haja vista que a multiplicidade de explicações acerca de uma mesma afirmação constitui, por si só, razão suficiente para nada se afirmar de forma absoluta [Empirico 1997]. A segunda concepção, baseada no relativismo absoluto, aceita

diferentes verdades para uma mesma afirmação [Rodrigues 2013]. Inclusive, é possível ressaltar a refutação tanto do ceticismo quanto do relativismo, pois é difícil para alguém declarar-se ceticista ou relativista sem se colocar fora ou acima de tal declaração. Isso acontece porque, se uma pessoa declara que *"não há verdade absoluta"* ou que *"todas as verdades são relativas"*, aparece a dúvida se tal afirmação é ou não, respectivamente, cética ou relativa.

Em uma perspectiva jornalística, uma notícia pode ter a verdade constatada tanto pela verificação dos fatos quanto dos argumentos a respeito do seu conteúdo [Sponholz 2009]. Partindo da premissa de que um argumento consiste na justificativa de uma opinião, a agregação das opiniões (argumentações) pode ser uma forma de avaliar uma notícia como verdadeira ou não.

1.3. Trabalhos Relacionados

Uma revisão bibliográfica foi realizada em busca de sumarizar as evidências existentes acerca do combate automático às *Fake News* nos MDDN, assim como identificar lacunas no estado-da-arte a fim de sugerir áreas a serem mais investigadas (isto é, problemas em aberto). O processo de revisão bibliográfica foi realizado a partir de pesquisas em algumas bases de dados científicas que apresentam publicações atualizadas e relevantes, tais como: *ACM Digital Library*[7], *IEEE Xplore*[8], *Scopus*[9] e *Science Direct*[10]. Nessas pesquisas foram normalmente considerados somente os trabalhos com foco principal no combate às *Fake News* escritos em inglês e português, cujo *qualis*[11] seja superior ou igual a *B*4 e publicados a partir do ano de 2015. O ano de 2015 foi escolhido como período de corte, pois, até onde foi possível observar, os trabalhos de anos anteriores ainda não demonstravam resultados robustos no combate às notícias intencionalmente falsas (por exemplo, experimentos realizados exclusivamente com *datasets* criados para detecção de rumores). Cabe ressaltar que também foram considerados alguns artigos referenciados pelos trabalhos analisados.

Um dos produtos resultantes dessa revisão bibliográfica foi a criação de uma visão geral dos trabalhos vinculados ao combate automático às *Fake News*. Essa visão começa pela apresentação de um modelo comparativo que viabiliza a distinção entre abordagens computacionais voltadas para o referido combate [Freire and Goldschmidt 2020]. Em seguida, os trabalhos, juntamente com os seus respectivos conjuntos de dados (*datasets*), são brevemente descritos e enquadrados no citado modelo.

1.3.1. Modelo Comparativo

O combate automático às *Fake News* nos MDDN possui uma variedade de aspectos que podem ser considerados. Com o objetivo de facilitar a comparação e a consequente classificação das referidas abordagens, tais aspectos são categorizados na Figura 1.1. As próximas subseções detalham cada um desses aspectos.

[7] https://dl.acm.org/
[8] https://ieeexplore.ieee.org/Xplore/home.jsp
[9] https://www.elsevier.com/pt-br/solutions/scopus
[10] https://www.sciencedirect.com/
[11] https://qualis.ic.ufmt.br/

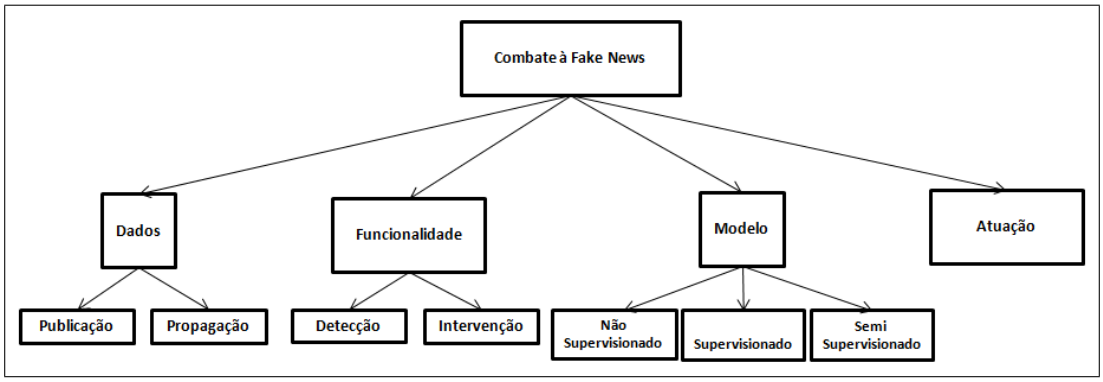

Figura 1.1: Aspectos considerados em Abordagens de Combate Automático às *Fake News*

1.3.1.1. Dados

Aspecto relacionado aos dados que podem ser utilizados pelas abordagens computacionais de combate às *Fake News*. Este aspecto subdivide-se em dados obtidos a partir da *Publicação* da notícia, como também aqueles associados com a sua *Propagação*.

Os dados de *Publicação* representam as informações inerentes ao surgimento da notícia no meio digital. Esses dados podem ser classificados em *Notícia, Usuário, Assunto e Temporalidade*. No que diz respeito à *Notícia*, a abordagem pode ser capaz de analisar dados oriundos da publicação a partir de diferentes tipos de *Mídia* (*Texto, Áudio e Imagem*). Independente da *Mídia*, a análise do *Conteúdo* pode ser realizada de forma *Léxica, Sintática, Semântica e Legibilidade*. Com relação ao *Usuário* publicador, a abordagem pode identificar diferentes *Tipos*, tais como: humano, *bot* ou *cyborg*. Pode-se analisar também dados referentes ao *Perfil* do usuário na rede social, tais como: identificação e idade. Outro aspecto relevante está relacionado à *Reputação* do publicador, que pode estar vinculada à sua capacidade em identificar ou publicar *Fake News*. A abordagem pode também utilizar o *Assunto* abordado no momento da publicação. Assim, é possível tratar Especificidades, tais como: relacionamento entre assuntos, assuntos controversos ou análise de tópicos. Outro aspecto leva em consideração a *Relevância* do assunto publicado, uma vez que assuntos em voga motivam a criação de *Fake News*. A variação das características de uma notícia de acordo com o período de tempo da sua divulgação, torna a *Temporalidade* mais um relevante recurso para a identificação de *Fake News*.

Os dados de *Propagação* representam as informações obtidas após a publicação, consequentemente, aquelas inerentes às contribuições devido ao espalhamento da notícia na rede social (ex: curtida/*like*, comentário/*reply* ou compartilhar/*retweet*). Portanto, esses dados podem ser classificados em *Contribuição, Usuário, Assunto, Temporalidade e Rede*. No que diz respeito à *Contribuição*, *Usuário*, *Assunto* e *Temporalidade* a abordagem pode ser capaz de analisar os dados oriundos da *Propagação*, a partir dos mesmos aspectos anteriormente citados na *Publicação*. Ademais, as informações relacionadas à *Rede* criada, a partir da propagação da notícia, possibilitam não só a identificação de uma *Fake News* como uma possível atuação contra a mesma.

1.3.1.2. Funcionalidade

Além dos dados coletados, as abordagens automáticas de combate às *Fake News* podem, basicamente, possuir duas funcionalidades: *Detecção e Intervenção*.

A *Detecção* automática da *Fake News* pode ser, basicamente, um problema de classificação binária onde, dada uma rede social $\mathscr{G}$ (caso particular de um meio digital de divulgação de notícia), uma notícia n é divulgada (publicada/propagada) por meio das suas respectivas postagens $\mathscr{P}_n$. Sendo que as divulgações pertencentes à $\mathscr{P}_n$ são realizadas por um conjunto de usuários U_n em um instante t. Assim o referido classificador binário $\mathscr{F}$ deve, aprendendo a partir dos dados, prever se n é uma *Fake News* ou não, como formalmente indicado na equação 1. Uma outra forma é a utilização de técnicas mais subjetivas que definam a probabilidade, peso ou pertinência de uma notícia n ser *fake*.

$$\mathscr{F}(n, \mathscr{P}_n, U_n, t) = \begin{cases} 1, & \text{se } n \text{ é uma } \textit{Fake News}; \\ 0, & \text{caso contrário.} \end{cases} \tag{1}$$

Independente da forma, para que uma notícia n possa ser detectada como *Fake News* é necessária a realização de duas subfuncionalidades: *Autenticidade e Intencionalidade* [Janze and Risius 2017, Vosoughi et al. 2017]. A *Autenticidade* analisa se a notícia é verdadeira ou falsa, enquanto que a *Intencionalidade* busca determinar a intenção dos divulgadores em ludibriar os receptores. Essa *Intencionalidade* pode ser mensurada como pontuação, peso ou score e obtida, por exemplo, por intermédio da análise de sentimentos que a notícia disponibiliza, pela associação entre usuários, assim como pelas características de perfil, tipo e reputação (credibilidade/confiança) dos divulgadores.

Já a *Intervenção* automática procura atacar as *Fake News*, nos MDDN, de forma proativa ou reativa [Shu et al. 2017, Farajtabar et al. 2017]. A intervenção reativa busca combater os efeitos da notícia a partir do momento da sua detecção como notícia propositalmente falsa. Por outro lado, a intervenção proativa tenta atuar antes mesmo da referida detecção, agindo então como uma forma de prevenção. Além disso, a tarefa de intervenção pode ser dividida em dois segmentos: *o Bloqueio e a Mitigação*. O *Bloqueio* atua de forma reativa. Na sua forma mais branda, o bloqueio interrompe a propagação da notícia e/ou a atuação do(s) usuário(s) divulgador(es). Uma outra forma mais incisiva seria remover a(s) notícia(s) e/ou o(s) usuário(s) divulgador(es). Já a *Mitigação* pode agir de forma reativa ou proativa buscando enfraquecer as consequências causadas pela *Fake News*. Na reatividade, a mitigação pode, por exemplo, imunizar os usuários provendo notícias verdadeiras [Farajtabar et al. 2017]. Uma forma de proatividade na mitigação é prover alertas, mesmo que a notícia ainda não tenha sido detectada com propositalmente falsa. Esses alertas podem estar relacionados com o nível de reputação da fonte (usuário) ou sobre o assunto estar relacionado com outras *Fake News* já identificadas.

Independente da funcionalidade da abordagem, a coleta dos dados inerentes à divulgação da notícia se faz necessária para subsidiar a detecção e a intervenção das notícias intencionalmente falsas. Assim, tanto a coleta de dados no meio digital quanto as tarefas de detecção e intervenção são fases iterativas, conforme ilustra a Figura 1.2. Cabe ressal-

tar que quanto mais cedo acontecer a detecção e a intervenção da *Fake News*, os impactos negativos dessa notícia tendem a ser menores.

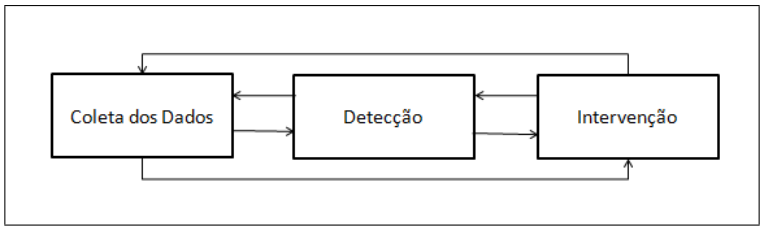

Figura 1.2: Fluxo do processo de combate às Fake News

1.3.1.3. Modelo

Quando a solução é por aprendizado de máquina, pode-se utilizar modelos computacionais para, a partir dos dados coletados, detectar as *Fake News*. Esses modelos são categorizados em *Não Supervisionado, Supervisionado e Semi-Supervisionado*.

No modelo *Não Supervisionado* são enquadradas as técnicas que normalmente levam mais tempo para realizar a identificação, porém, como não necessitam de rótulos, podem utilizar *datasets* mais simples [Shu et al. 2017].

Os modelos *Supervisionados* são lentos na fase do treinamento, entretanto, tendem a ser mais rápidos do que os não supervisionados no momento da sua utilização na identificação das *Fake News*. Devido à necessidade de treinamento, os modelos supervisionados precisam de *datasets* mais completos [Shu et al. 2017].

O modelo *Semi-Supervisionado* procura realizar a tarefa de identificação da *Fake News*, em MDDN, de uma forma mista que busque utilizar, tanto as técnicas supervisionadas quanto não supervisionadas. Essa abordagem pode utilizar *datasets* mais simples do que aqueles manipulados pelos modelos supervisionados, porém mais complexos do que os utilizados pelos não supervisionados [Shu et al. 2017].

1.3.1.4. Atuação

As abordagens computacionais que visam o combate às *Fake News*, independemente dos dados coletados, funcionalidade e modelo utilizados, podem ter diferentes formas de atuação.

Uma das possibilidades de atuação está associada à localização física do combate dentro do meio digital. Uma abordagem *Centralizada* encontra-se, fisicamente, em um único ponto. Portanto, todas as tarefas relacionadas com a detecção/intervenção da *Fake News* são executadas em um mesmo local.

Por outro lado, uma abordagem *Descentralizada* encontra-se, fisicamente, espalhada. Assim, essa forma de atuação possibilita, inclusive, uma execução paralela e/ou ditribuida [Wu and Liu 2018] no combate às *Fake News*.

1.3.2. Revisão dos Trabalhos

Nesta subseção são apresentados alguns trabalhos relacionados ao combate automático às *Fake News* nos MDDN. Para tal, foram realizadas buscas de acordo com a revisão bibliográfica já descrita, onde as principais fontes de consulta foram os artigos [Guo et al. 2020, Zhou and Zafarani 2018, Zhou and Zafarani 2019, Reis et al. 2019, Sharma et al. 2019] [Zhou et al. 2019, Shu et al. 2017, Conroy et al. 2015].

Para um melhor entendimento, os citados trabalhos são identificados e enquadrados no, já apresentado, Modelo Comparativo, conforme mostram as Tabelas 1.1, 1.2 e 1.3. Cabe ressaltar que, nessas três tabelas, as células não preenchidas indicam a não utilização do respectivo aspecto no trabalho correspondente.

Com base no referido enquadramento, uma constatação a ser destacada é a existência de um número significativo de abordagens computacionais que realizam a detecção de *Fake News*, nos MDDN, através da análise do texto existente na notícia. Uma possível razão para essa aparente preferência é que, inicialmente, as notícias intencionalmente falsas poderiam ser criadas sem uma forte preocupação de aparentar uma similaridade de conteúdo com as notícias não $fake$. Entretanto, atualmente, detecção através da análise do texto pode ser complexa, pois a cada dia os divulgadores intencionais de notícias falsas estão cada vez mais preocupados em manipular o conteúdo para que essas notícias pareçam não *fake* [Wu and Liu 2018, Liu and BrookWu 2018]. Sendo assim, as abordagens que não necessitam exclusivamente dos dados relativos ao conteúdo da notícia para a detecção de *Fake News* vêm se destacando. Dentre elas, as soluções baseadas na reputação dos usuários se apresentam como uma alternativa promissora, pois apresentam duas características interessantes. A primeira é a não necessidade de utilização do conteúdo da notícia, devido à, já citada, crescente similaridade entre as notícias *fake* e não *fake* que acaba por dificultar essa forma de detecção. A segunda é a não obrigatoriedade do uso dos dados relativos ao perfil do usuário na rede social, haja vista a dificuldade em se obter tais informações cada vez mais sigilosas [Shu et al. 2019b].

Além disso, os trabalhos são brevemente descritos, podendo seus detalhes serem consultados através das suas respectivas referências:

T1) *A Deep Transfer Learning Approach for Fake News Detection* [Saikh et al. 2020]: Esta pesquisa busca avaliar se o título de uma notícia está de acordo com o seu respectivo conteúdo utilizando *deep transfer learning*, onde o título é considerado como uma hipótese e a notícia é uma premissa. Assim, o objetivo é verificar se o corpo da notícia faz inferência ao seu respectivo título utilizando *stance detection*, tendo as seguintes respostas possíveis: *Agree, Disagree, Discuss e Unrelated*. Foram utilizados *Bi-LSTM e Bi-LSTM com max-pooling*. Os resultados foram comparados com a detecção humana, onde a acurácia máxima obtida foi de 0.90;

T2) *A Linguistic-Based Method that Combines Polarity, Emotion and Grammatical Characteristics to Detect Fake News in Portuguese* [de Souza et al. 2020]: O artigo apresenta o método (FNE) que, além da classificação gramatical e análise de sentimento baseada em polaridade, também usa a análise de emoções ao detectar notícias intencionalmente falsas escritas em português. Foram realizados experimentos com o FNE variando os classificadores *Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBo-*

ost (AB), Gradient Boost (GB) e Naive Bayes (NV). Os resultados foram comparados com os obtidos em [Moraes et al. 2019] (T12). A acurácia máxima obtida foi de 0.92;

T3) *A Topic-Agnostic Approach for Identifying Fake News Pages* [Castelo et al. 2019]: O trabalho propõe um *topic-agnostic* (TAG) classificador que usa dados linquísticos e *Web-Markup* (padrões de layout das páginas) para detectar *Fake News*. Assim, ao invés de usar o *bag of words*, o trabalho explora as *topic-agnostic*, incluindo características morfológicas, psicológicas e de legibilidade que são comuns em *Fake News*. O trabalho propõe que páginas com *Fake News* normalmente têm inclinação sensacionalista, assim como a ocorrência de termos, tais como: *"Just in" e "Read this"*. Foram utilizados 3 classificadores *Support Vector Machine (SVM), K-Nearest Neighbors (KNN) e Random Forest (RF)*. Comparou o TAG com os resultados obtidos em [Pérez-Rosas et al. 2018] (T4), separando-os ano a ano (2013 até 2018). A acurácia máxima obtida foi de 0.86;

T4) *Automatic Detection of Fake News* [Pérez-Rosas et al. 2018]: Este trabalho cria uma ferramenta de detecção de *Fake News* por classificação com *Support Vector Machines* (SVM), combinando informações léxicas, sintáticas, semânticas e de legibilidade. O presente trabalho compara os resultados com a detecção humana. A acurácia máxima obtida foi de 0.91;

T5) *Automatic Detection of Fake News on Social Media Platforms* [Janze and Risius 2017]: Este artigo implementa a detecção com os classificadores binários *Logistic Regression, Support Vector Machines (SVM), Decision Tree, Random Forest e Extreme Gradient Boosting*. O referido trabalho compara os resultados entre os classificadores. A acurácia máxima obtida foi de 0.80;

T6) *Automatically Identifying Fake News in Popular Twitter Threads* [Buntain and Golbeck 2017]: O trabalho apresenta um método para detecção de *Fake News* no *Twitter* que acumula, ao longo do tempo, as características de rede, usuário e conteúdo para gerar uma regressão linear. Assim, a abordagem realiza a sua análise, levando em consideração os aspectos temporais relacionados à notícia a ser detectada. O artigo procura avaliar os resultados nos *datasets PHEME (Twitter* para rumor), *CredBank (Twitter)* e *BuzzFeed News Fact-Checking Dataset (Facebook)* que precisaram ser alinhados com as mesmas características e rótulos. A acurácia máxima obtida foi de 0.70 nos experimentos realizados com o *dataset CredBank* para a detecção automática de *Fake News*;

T7) *BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection* [Zhang et al. 2020]: Compreende três módulos principais, sendo um extrator de recursos multimodais, um classificador de domínio e um detector de notícias falsas. Nesta proposta as características textuais são extraídas pelo modelo *BERT* e as características da imagem são obtidas pelo modelo *VGG-19*. Este método compara os seus resultados com outros trabalhos, como [Wang et al. 2018b] (T14). A acurácia máxima obtida foi de 0.85;

T8) *Beyond News Contents: The Role of Social Context for Fake News Detection* [Shu et al. 2019b]: Este artigo explora as correlações da postura da notícia, o bias e engajamento do usuário. Assim, é apresentado um Tri-Relacionamento (TriFN) onde tanto

informações partidárias quanto níveis de confiança do usuário podem ser utilizados para detecção de *Fake News*. Além disso, os usuários tendem a formar relacionamentos com pessoas afins que podem aumentar o espalhamento das *Fake News*. Essa abordagem compara os seus resultados com outros trabalhos, como [Rubin et al. 2015] (T32). A acurácia máxima obtida foi de 0.87;

T9) *CIMTDetect: A Community Infused Matrix-Tensor Coupled Factorization Based Method for Fake News Detection* [Gupta et al. 2018]: Através da modelagem de Câmara de Ecos, o trabalho representa uma notícia como um *3-mode tensor <News, User, Community>* e propõe um método baseado em *tensor factorization*. Além disso, apresenta uma extensão desse método com a junção de modelos que utilizam o conteúdo da notícia através de um *framework coupled matrix-tensor factorization*. Esse artigo usou o algoritmo de detecção da comunidade *Girvan-Newman* para identificar, na rede social, comunidades representativas de câmaras de eco. Os seus resultados são comparados com métodos que utilizam o classificador SVM, porém com diferentes formas de análise de conteúdo (ex. N-Gram). Os dois métodos propostos *CITDetect (community-infused tensor information) e CIMTDetect (community-infused tensor information + conteúdo da notícia)* utilizam o classificador SVM. O F1-score máximo obtido foi de 0.81;

T10) *Combining Neural, Statistical and External Features for Fake News Stance Identification* [Bhatt et al. 2018]: Neste estudo a ferramenta, desenvolvida para o primeiro desafio (FNC-1)[12], não tem o objetivo final de detectar se a notícia é *Fake News*. Nessa abordagem, as notícias são classificadas de acordo com a relação existente entre a manchete e o corpo do texto. Portanto os possíveis resultados da classificação são *Agree* - o texto do corpo concorda com a manchete, *Disagree* - o texto do corpo discorda da manchete, *Discuss* - o texto do corpo discute a mesma afirmação que o título, mas não toma uma posição ou *Unrelated* - o texto do corpo discute uma alegação que difere do título. A ferramenta combina as abordagens neural e estatística com recursos externos. Para isto, a solução implementa um modelo profundo recorrente (*Neural Embedding*), um modelo ponderado de características estatísticas (*n-gram bag-of-words*) e recursos externos criados à mão com a ajuda de uma heurística de engenharia de atributos. Por fim, usando uma rede neural profunda, todas as referidas abordagens são combinadas. Os resultados foram comparados com as demais ferramentas participantes do referido desafio (FNC-1). A acurácia máxima obtida foi de 0.89;

T11) *CSI: A Hybrid Deep Model for Fake News Detection* [Ruchansky et al. 2017]: O trabalho procura melhorar a acurácia na detecção de *Fake News* por meio de um modelo híbrido de rede neural profunda chamado CSI. Esse modelo utiliza três características: o texto da notícia, a resposta do usuário que recebeu a notícia e o usuário fonte da notícia. O CSI trabalha com o comportamento temporal dos usuários e da notícia. Esse modelo se divide em três partes: *Capture, Score e Integrate*. O primeiro módulo é baseado no texto e na resposta, por meio de uma rede neural recorrente (LSTM) para capturar um padrão temporal de atividades do usuário sobre a notícia e a representação *Doc2Vec*. O segundo usa uma rede neural para aprender as características da fonte, baseado nas interações dos usuários, gerando um score por meio de um grafo. Os dois módulos são integrados com o terceiro para caracterizar ou não a notícia como *Fake News*. O trabalho propõe a sua

[12]http://www.fakenewschallenge.org/

utilização em diferentes domínios, inclusive, em bancos de dados. Os resultados foram comparados com técnicas criadas para detecção de rumores. A acurácia máxima obtida foi de 0.95;

T12) *Data mining applied in fake news classification through textual patterns* [Moraes et al. 2019]: Este estudo realiza o levantamento de características textuais por meio do processamento do texto com um léxico de sentimentos. Assim, o método proposto calcula a polaridade total de cada texto que é adicionada ao conjunto de atributos (por exemplo, classes gramaticais) utilizados. Este método foi testado com os algoritmos de classificação *Support Vector Machines, Naive Bayes e Adaboost*. A acurácia máxima obtida foi de 0.93;

T13) *DistrustRank: Spotting False News Domains* [Woloszyn and Nejdl 2018]: Esta solução propõe uma estratégia de aprendizagem semi-supervisionada para separar automaticamente notícias falsas a partir de fontes não confiáveis de notícias. O trabalho utiliza como fonte *experts* de portais de checagem de fatos para classificar manualmente as notícias. A partir disto, é criado um grafo de pesos com os *ranks* de confiança sobre os *sites* e as arestas representam a similaridade dos mesmos. A pesquisa computa a centralidade, utilizando o *PageRank* em busca de uma similaridade entre os *sites* não confiáveis. O resultado da análise é a classificação em *Trust* ou *Distrust* para a fonte da notícia. O trabalho verificou que a semelhança entre os sites de notícias falsas é estatisticamente superior aos sites de notícias verdadeiras. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. O F1-score máximo obtido foi de 0.78;

T14) *EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection* [Wang et al. 2018b]: O artigo aponta que a maioria das abordagens existentes aprendem a detectar *Fake News* a partir de características específicas do evento, consequentemente, não podem ser transferidas para outros eventos ainda não aplicados. Assim, esse trabalho desenvolveu um *framework*, de ponta a ponta, denominado *EANN*, que pode derivar características invariantes de um evento para outro. Dessa forma, propõe uma detecção de *Fake News* para eventos recém-chegados. Isso consiste de três componentes principais: o extrator de características multimodais para texto e imagem (rede neural Convolucional), o detector de *Fake News (fully connected layer com softmax)* e o discriminador de eventos (rede neural) que é o responsável por remover as características específicas do evento e manter as características compartilháveis entre os eventos para poder rotulá-los. Assim, o *framework* mede as características não similares entre diferentes eventos e remove-os para capturar as características invariantes entre eventos. Para avaliar seus resultados, realizou testes com técnicas de identificação de texto e imagem, porém utilizadas em trabalhos não ligados à detecção de *Fake News*. A acurácia máxima obtida foi de 0.82;

T15) *Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks* [Liu and BrookWu 2018]: O artigo propõe um modelo para detecção precoce de Fake News através da classificação dos caminhos de propagação da notícia. O referido trabalho modela o caminho de propagação de cada notícia como uma série temporal multivariada, na qual cada tupla é um vetor numérico que representa as características do usuário empenhado em espalhar a notícia. Para tal, é construído um classificador de série temporal que incorpora redes recorrente e

convolucional. Essas redes capturam as variações globais e locais das características do usuário, ao longo do caminho de propagação, para detectar *Fake News*. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.92;

T16) *Evaluating Machine Learning Algorithms for Fake News Detection* [Gilda 2017]: Este artigo explora técnicas de linguagem natural para a detecção de *Fake News*. O trabalho aplicou *term frequency-inverse document frequency (TF-IDF) de bi-grams e probabilistic context free grammar (PCFG)* para um conjunto de 11.000 artigos em um *dataset* obtido pela *Signal Media*[13] e uma lista de fontes da *OpenSources.com*[14]. Este *dataset* foi testado com os algoritmos de classificação *Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees e Random Forests*. A acurácia máxima obtida foi de 0.77, com modelos treinados apenas no conjunto de recursos do TF-IDF;

T17) *FActCheck: Keeping Activation of Fake News at Check* [Srivastava et al. 2018]: Esta abordagem de intervenção sobre *Fake News* propõe uma melhoria na abordagem *competing cascades*, onde os *AFC (algoritmos polynomial time greedy) e RAFC (fast graph-pruning)* procuram escolher quais usuários têm maior poder de mitigação. Assim, os usuários com maior capacidade de influência na rede social realizam a mitigação através da divulgação de notícias alternativas *(Real News)*. A máxima qualidade da redução obtida foi de 0.03;

T18) *Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning* [Souza Freire et al. 2021]: Denominada *HCS (Hybrid Crowd Signals)*, a abordagem proposta neste trabalho busca detectar se uma notícia *n*, divulgada em um meio digital, é *fake* ou não utilizando as reputações dos usuários (membros do *Crowd*) para ponderar as suas respectivas opiniões (*Signals*) sobre *n*. No entanto, a principal diferença entre a *HCS* e a abordagem baseada em *Crowd Signals* é a forma como as opiniões dos membros do *Crowd* são obtidas. Enquanto a abordagem baseada em *Crowd Signals* exige a opinião explícita dos usuários sobre os rótulos das notícias, a *HCS* considera obter as opiniões que se encontram *implícitas* nos padrões de comportamento dos usuários ao divulgar (publicar/propagar) essas notícias. Além disso, inspirada em meta-aprendizagem, a *HCS* permite ainda a formação de um *Crowd* híbrido, uma vez que os membros do *Crowd* podem ser tanto usuários divulgadores do meio digital quanto máquinas (modelos de classificação de notícias) disponibilizadas para o uso da *HCS*. O enquadramento da *HCS* no modelo comparativo, proposto neste capítulo, foi realizado somente com o método que utiliza as opiniões implícitas dos usuários (isto é, não utiliza as máquinas). A acurácia máxima obtida foi de 0.99;

T19) *Fake News Detection in Social Networks via Crowd Signals* [Tschiatschek et al. 2018]: A ferramenta desenvolvida trabalha na detecção e consequente intervenção de *Fake News*. Essa solução possui um algoritmo, chamado de *Detective* que usa inferência Bayesiana para detectar *Fake News* a partir de *Crowd Signals*. Esse *Crowd* é formado pela opinião dos usuários sobre a notícia, juntamente com a sua capacidade em opinar corretamente. O objetivo é detectar, de forma antecipada, a *Fake News* e bloqueá-

[13]https://research.signal-ai.com/newsir16/signal-dataset.html
[14]http://www.opensources.co

la. Os resultados foram comparados a partir de variações na própria abordagem, sendo as mesmas denominadas pelo artigo como Opt, Oracle, Fixed-CM e No-Learn. A máxima utilidade média obtida foi de 0.98;

T20) *Fake News Detection Using One-Class Classification* [Faustini and Covões 2019]: O principal objetivo desse trabalho é eliminar a necessidade de contar com notícias rotuladas (como *fake* e não *fake*) na detecção de *Fake News*. Para que isso seja possível, seus autores utilizam o conteito de *One-Class Classification* (OCC). Nessa abordagem de aprendizado de máquina, o modelo de classificação é treinado utilizando apenas dados de uma classe, nesse caso, notícias rotuladas como *fake*, que costumam ser mais facilmente encontradas. Para avaliar essa abordagem no âmbito da detecção de *Fake News*, um algoritmo baseado em técnicas de redução de dimensionalidade foi estendido de modo a incorporar dados de publicação da notícia em formato de texto. O F1-score obtido a partir dos experimentos realizados variou entre 65% e 67%, onde os resultados foram comparados com abordagens similares;

T21) *Fake News Mitigation Via Point Process Based Intervention* [Farajtabar et al. 2017]: Neste artigo, o enfoque está na intervenção de *Fake News*. A proposta é intervir, mitigando a notícia falsa, fornecendo recompensas na forma de notícias verdadeiras para quem recebeu a *Fake News*. A nível de influência da *Fake News* e a respectiva mitigação são quantificadas por contadores. O modelo utilizado foi baseado em *least-squares temporal difference learning* (LSTD). Um dos experimentos foi real, com a criação de cinco contas no *Twitter*. A melhora máxima de 0.2 foi obtida em relação a uma intervenção randômica;

T22) *FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization* [Shu et al. 2019a]: Apresenta o FakeNewsTracker, um sistema para detecção de notícias falsas. O FakeNewsTracker pode coletar, automaticamente, dados para notícias e contexto social. Esse tratalho propõe um *framework end to end* para realizar a coleta de dados, a detecção das *Fake News* e a visualização dos resultados. Essa pesquisa usa *autoencoders* para aprender o conteúdo de notícias e *RNN* para capturar o padrão temporal dos usuários de acordo com o seu engajamento com a notícia. O trabalho compara os seus resultados, internamente, a partir de variações do próprio *FakeNewsTracker*, onde são considerados somente o conteúdo da notícia ou o contexto social. Além disso, os resultados são comparados também com *Support Vector Machine, Logistic Regression and Naive Bayes*. A acurácia máxima obtida foi de 0.74;

T23) *IARNet: An Information Aggregating and Reasoning Network over Heterogeneous Graph for Fake News Detection* [Yu et al. 2020]: Sistema heterogêneo que considera a origem do post, os comentários e os respectivos usuários divulgadores, sendo estas informações armazenadas em um grafo. A detecção de *Fake News* utiliza *GloVe Embeddings, BERT representations, Bidirectional Gated recurrent units (BiGRU)* para o texto e *two-layer fully-connected MLP* para os dados do usuário. Esse trabalho compara os seus resultados com outros classificadores clássicos (por exemplo, *SVM e GRU*) a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.96;

T24) *Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection* [Wang 2017]: Além de propor um *dataset*, cria uma técnica de detecção de *Fake News* híbrida, usando redes neurais convolucionais (CNNs) para analisar, não somente textos,

mas também os dados do usuário. O artigo apresenta os resultados ao ser comparado com outros três detectores implementados com *Logistic Regression Classifier* (LR), *Support Vector Machine Classifier* (SVM) e *bi-directional long short-term memory* (Bi-LSTMs). A acurácia máxima obtida foi de 0.27;

T25) *Neural User Response Generator: Fake News Detection with Collective User Intelligence* [Qian et al. 2018]: O trabalho enfatiza a rápida propagação das *Fake News* nas redes sociais e, portanto, destaca a importância da sua detecção nos estágios iniciais, onde considera que apenas o texto da notícia está disponível. Tal afirmação se baseia no fato de que informações adicionais, como respostas dos usuários e padrões de propagação, podem ser obtidas somente após a notícia se espalhar. Contudo, como as respostas propagadas podem ajudar na tarefa de detecção, os autores propõe um *Two-Level Convolutional Neural Network with User Response Generator (TCNN-URG)* onde o TCNN captura a semântica do texto da notícia e o URG cria um modelo generativo de resposta dos usuários propagadores. O URG, a partir de respostas históricas, é treinado para aprender como os usuários respondem às notícias publicadas, gerando respostas de usuários para ajudar a TCNN na detecção da *Fake News*. Essa abordagem cita e compara os seus resultados com outros trabalhos a partir do mesmo *dataset*. A acurácia máxima obtida foi de 0.89;

T26) *Ranking-based Method for News Stance Detection* [Zhang et al. 2018]: Mais uma pesquisa relacionada ao primeiro desafio (FNC-1). A solução do artigo é criada a partir de uma rede neural *Multi-Layer Perceptron*. Os resultados foram comparados com as demais ferramentas participantes do referido desafio. A acurácia máxima obtida foi de 0.86;

T27) *Real-time Detection of Content Polluters in Partially Observable Twitter Networks* [Nasim et al. 2018]: Esta pesquisa procura encontrar um tipo específico de *bots*, chamados de poluidores de conteúdo, para poder distinguir notícias verdadeiras de *Fake News*. Segundo o artigo, o estado da arte de detecção de *bots*, normalmente, necessita de um histórico completo da rede. Assim, o trabalho propõe uma abordagem baseada em informações parciais onde, ao invés de mapear um grafo com seguidores e seguidos, utiliza um grafo com a (dupla de Usuário) x (Evento). Essa dupla é obtida a partir do momento em que o par tenha *tweetado* no mesmo dia do evento. Dessa forma, os dados são clusterizados para que os usuários possam ser classificados como *bots* pela análise dos respectivos perfis e a frequência dos *tweets*. Os resultados do trabalho foram comparados com os obtidos por uma ferramenta citada pelo artigo, denominada de *Truthy*. A porcentagem de verdadeiros positivos obtida foi próxima de 0.65;

T28) *Sentiment Aware Fake News Detection on Online Social Networks* [Ajao et al. 2019]: O trabalho se aplica tanto a *Fake News* como Rumor. Assim, o artigo propõe a hipótese de que existe uma relação entre mensagens falsas ou rumores com os sentimentos dos textos. Foram utilizados dois modelos para extrair os escores de sentimento (positividade, negatividade ou neutralidade) do texto: *Latent Semantic Analysis (LSA) e Latent Dirichlet Allocation (LDA)*. O objetivo foi desenvolver um classificador que utilize os escores de sentimento. Assim, utilizando classificadores distintos, compara os resultados a partir da abordagem proposta com sentimentos. A acurácia máxima obtida foi de 0.89;

T29) *Sentiment-Based Multimodal Method to Detect Fake News* [Maia et al. 2021]: O artigo propõe o método (FNEXT) que considera o uso da análise de polaridade e da emoção extraídas tanto dos textos quanto das imagens existentes nas notícias para detecção de *Fake News* escritas em língua portuguesa. Foram realizados experimentos com o FNEXT variando os classificadores *Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBoost (AB), Gradient Boost (GB) e Naive Byes (NV)*. Os resultados foram comparados com os obtidos em [de Souza et al. 2020] (T2). A acurácia máxima obtida foi de 0.99;

T30) *Supervised Learning for Fake News Detection* [Reis et al. 2019]: Este artigo sobre *machine learning* explicável realiza uma pesquisa sistemática sobre *Fake News*, identificando os dados existentes, assim como propõe novos dados para detecção. Para implementar e avaliar esses recursos, o trabalho aplica os algoritmos de classificação *Extreme Gradient Boosting, Random Forest e SVM*. Ao aplicar os classificadores, os dados são subdivididos em características de conteúdo da notícia (por exemplo, texto e imagem), características da fonte (por exemplo, credibilidade e viés político), e características extraídas do ambiente (por exemplo, engajamento - likes e perfil do usuário). O trabalho propõe a utilização da combinação dessas características e comparou os seus resultados variando as características utilizadas por cada classificador de forma individual. A AUC máxima obtida foi de 0.88;

T31) *This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News* [Horne and Adali 2017]: Este trabalho usa um classificador SVM para detecção por meio da análise do texto, comparando os seus resultados entre detecção de *Fake News, Real News* e Sátira. Esse estudo determinou que as *Fake News* são mais próximas das Sátiras do que as notícias reais. A acurácia máxima obtida foi de 0.77;

T32) *Towards News Verification: Deception Detection Methods for News Discourse* [Rubin et al. 2015]: O trabalho propõe a ferramenta RST-SVM que analisa a notícia para extrair o estilo por meio da combinação do *Rhetorical Structure Theory (RST) e Vector Space Modeling (VSM)* para Clusterização. A detecção da notícia como enganosa ou real foi feita por meio de um classificador SVM. Os resultados obtidos foram comparados com a detecção humana. A acurácia máxima obtida foi de 0.56;

T33) *Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate* [Wu and Liu 2018]: Este trabalho busca a detecção de *Fake News*, pela modelagem da propagação da notícia através da mineração de grafos em Florestas. Segundo o artigo, classificar notícias pelo seu conteúdo é muito difícil, devido à similaridade entre as divulgações *fake* e não *fake*. Em contra partida, as *Fake News* tendem a ter as mesmas fontes e sequências. O trabalho propõe a ferramenta paralelizável chamada TraceMiner que utiliza *Recurrent Neural Networks* (LSTM-RNNs), para classificar o caminho de propagação das mensagens no *Twitter*. O artigo comparou os seus resultados por meio de técnicas de análise de conteúdo criadas com SVM e XGBoost. O F1-score máximo obtido foi de 0.93;

T34) *Um Método Linguístico que combina Polaridade, Emoção e Aspectos Gramaticais para Detecção de Fake News em Inglês* [Testoni et al. 2021]: O artigo utiliza o já proposto método (FNE), porém propõe um protótipo adaptado que, além da classificação gramatical e análise de sentimento baseada em polaridade, também usa a análise de emoções ao detectar notícias intencionalmente falsas escritas em inglês. Foram realizados experimentos com o FNE variando os classificadores *Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBoost (AB), Gradient Boost (GB) e Naive Byes (NV)*. Os resultados foram comparados com os baselines sem a utilização da emoção. A acurácia máxima obtida foi de 0.99;

T35) *Weakly Supervised Learning for Fake News Detection on Twitter* [Helmstetter and Paulheim 2018]: Neste estudo, como existe uma dificuldade em conseguir um grande volume de dados para análise *(datasets)*, os *tweets* são rotulados, automaticamente, durante a coleta, de acordo com a confiança na sua fonte. Assim é criado um *dataset*, denominado de *Large-scale Training Dataset*, onde cada tweet de uma fonte confiável é rotulado como uma notícia real, assim como, cada tweet de uma fonte não confiável é rotulado como uma *Fake News*. Espera-se que nesse *dataset* a classe de notícias reais contenha apenas uma quantidade negligenciável de ruído, pois fontes confiáveis raramente divulgam *Fake News*. Também é criado um segundo *dataset*, denominado de *Small-scale Evaluation Dataset*, possuindo *tweets* rotulados manualmente, como *fake* e não *fake*, a partir do site PolitiFact[15]. O objetivo principal do trabalho é treinar um classificador, a partir do primeiro *dataset*, para aplicá-lo no segundo *dataset*. Portanto, esse classificador, apesar de ter sido treinado em um *dataset* desenvolvido a partir da confiança, é utilizado para detectar *tweets fakes* e não *fakes* no segundo *dataset*. Portanto, o artigo considera que o classificador foi treinado e avaliado com alvos distintos. Para a referida detecção foram levadas em consideração as características do usuário (ex: engajamento e quantidade de seguidores), do tweet (ex: dia da semana, hora e texto), do tópico (assunto) e do sentimento. Como algoritmos de aprendizado foram usados o *Naive Bayes*, Árvores de Decisão, *SVM* e Redes Neurais. Além disso, foram usados dois *comitês*, o *Random Forest e XGBoost*. Os resultados foram comparados, utilizando diferentes combinações para os classificadores. O F1-score máximo obtido foi de 0.89;

T36) *XFake: Explainable Fake News Detector with Visualizations* [Yang et al. 2019]: O detector *XFake* é composto por 3 *frameworks: MIMIC, ATTN e PERT*. O *MIMIC* é construído para análise de atributos (ex. contexto da notícia e publicador) por meio de uma *deep neural network*. O *ATTN* é para análise semântica através de *pre-trained word embedding*, rede neural convolucional e *self-attention mechanism*. O *PERT* é para análise linguística utilizando um classificador *XGBoost*. A ferramenta, além de realizar as predições, também possui um módulo de interface para prover os usuários de explicações sobre as predições. O *XFake* é implementado em Python e deployed em *FLASK* com *front-end em HTML*. Para comparar seus resultados, o trabalho utilizou mão-de-obra humana realizada pela Amazon Mechanical Turk[16]. A acurácia máxima obtida foi de 0.67.

[15]https://www.politifact.com/

[16]https://www.mturk.com/

Tabela 1.1: Comparação entre abordagens - Dados de Publicação

Id	Dados							
	Publicação							
	Notícia		Usuário			Assunto		Temporalidade
	Mídia (Texto, Áudio e Imagem)	Conteúdo (Léxica, Sintática, Semântica e Legibilidade)	Tipo (Humano, Bot e Cyborg)	Perfil	Reputação	Especificidades	Relevância	
T1	Texto	Léxica e Semântica						
T2	Texto	Léxica						
T3	Texto	Léxica e Semântica						
T4	Texto	Léxica, Sintática, Semântica e Legibilidade						
T5	Texto e Imagem	Léxica						
T6	Texto	Léxica e Semântica		X				X
T7	Texto e Imagem	Léxica e Semântica						
T8	Texto	Léxica e Semântica		X	X			
T9	Texto	Léxica		X				
T10	Texto	Semântica						
T11	Texto	Léxica e Semântica		X	X			X
T12	Texto	Léxica						
T13					X	Assuntos controversos		
T14	Texto e Imagem	Léxica						
T15				X				X
T16	Texto	Léxica e Semântica						
T17								
T18					X			
T19					X			X
T20	Texto	Léxica e Semântica						
T21								
T22	Texto	Léxica e Semântica		X				X
T23	Texto	Léxica e Semântica		X				
T24	Texto	Léxica e Semântica		X		Relaciona Assuntos		
T25	Texto	Semântica						
T26	Texto	Semântica						
T27			Bot	X				X
T28	Texto	Léxica e Semântica						
T29	Texto e Imagem	Léxica						
T30	Texto e Imagem	Léxica		X				X
T31	Texto	Léxica e Sintática						
T32	Texto	Semântica						
T33					X			
T34	Texto	Léxica						
T35	Texto	Léxica e Semântica		X		Análise dos Tópicos		X
T36	Texto	Léxica e Semântica		X		Análise dos Tópicos		

Tabela 1.2: Comparação entre abordagens - Dados de Propagação

Id	Dados								
	Propagação								
	Contribuição		Usuário			Assunto		Temporalidade	Rede
	Mídia (Texto, Áudio e Imagem)	Conteúdo (Léxica, Sintática, Semântica e Legibilidade)	Tipo (Humano, Bot e Cyborg)	Perfil	Reputação	Especificidades	Relevância		
T1									
T2									
T3									
T4									
T5	Texto	Léxica							X
T6	Texto	Léxica e Semântica		X				X	X
T7									
T8				X	X				X
T9				X					X
T10									
T11	Texto	Léxica e Semântica		X	X			X	X
T12									
T13									
T14									
T15				X				X	X
T16									
T17									X
T18					X				
T19					X			X	X
T20									
T21									X
T22				X				X	X
T23	Texto	Léxica e Semântica		X					
T24									
T25	Texto	Semântica							
T26									
T27			Bot	X				X	X
T28	Texto	Léxica e Semântica							
T29									
T30								X	X
T31									
T32									
T33					X				X
T34									
T35	Texto	Léxica e Semântica							
T36									

Tabela 1.3: Comparação entre abordagens - Modelo, Funcionalidade e Atuação

Id	Modelo			Funcionalidade				Atuação (Centralizada ou Descentralizada)
				Detecção		Intervenção		
	Não Supervisionado	Semi Supervisionado	Supervisionado	Autenticidade	Intencionalidade	Bloqueio (Reativa)	Mitigação (Proativa e Reativa)	
T1		X	X	X				Centralizada
T2			X	X	Análise de Sentimentos			Centralizada
T3			X	X				Centralizada
T4			X	X				Centralizada
T5			X	X				Centralizada
T6			X	X	Análise das características dos usuários			Centralizada
T7		X	X	X				Centralizada
T8		X		X	Pontuação de credibilidade para os usuários			Centralizada
T9			X	X				Centralizada
T10			X	X				Centralizada
T11			X	X	Score para os usuários			Centralizada
T12			X	X	Análise de Sentimentos			Centralizada
T13		X		X	Atribui pesos de confiança aos websites			Centralizada
T14			X	X				Centralizada
T15			X	X				Centralizada
T16			X	X				Centralizada
T17							Reativa	Centralizada
T18			X	X	Reputação do usuário			Centralizada
T19			X	X		X		Centralizada
T20			X	X				Centralizada
T21							Reativa	Centralizada
T22		X		X				Centralizada
T23			X	X				Centralizada
T24			X	X	Associação com o usuário			Centralizada
T25			X	X				Centralizada
T26			X	X				Centralizada
T27		X		X	identificação de bots			Centralizada
T28			X	X	Análise de Sentimentos			Centralizada
T29			X	X	Análise de Sentimentos			Centralizada
T30			X	X	Pontuação de credibilidade para as fontes			Centralizada
T31			X	X				Centralizada
T32		X		X				Centralizada
T33			X	X	Relação entre os usuários			Centralizada (pode ser paralelizada)
T34			X	X	Análise de Sentimentos			Centralizada
T35			X	X	Análise de Sentimentos			Centralizada
T36			X	X				Centralizada

Com o objetivo de apresentar alguns *datasets*, a Tabela 1.4 relaciona os trabalhos acima apresentados com os seus respectivos *datasets*. Em seguida, a Tabela 1.5 enquadra esses repositórios de acordo com os dados fornecidos por cada um deles. Esse enquadramento é realizado no apresentado Modelo Comparativo, restrito ao aspecto *Dados*. Cabe ressaltar que, na Tabela 1.5, as células não preenchidas indicam o não fornecimento do respectivo dado no *dataset* correspondente. Além disso, são utilizadas as seguintes formas de disponibilização dos dados:

- No Dataset: a informação está armazenada na própria base de dados;
- Link para o dado: a informação não está armazenada na base de dados, mas o *dataset* disponibiliza um link direto para o dado específico;
- Link para a notícia: Nesta caso, o *dataset* simplesmente disponibiliza o link para a notícia. Assim, se faz necessário o acesso à notícia original para a retirada das informações desejadas.

Com base no referido enquadramento dos *datasets*, apesar da relevância do problema de combate às *Fake News* nos MDDN, foi possível constatar que os repositórios que contêm dados reais ainda estão raramente disponíveis para download. Como consequência, algumas das pesquisas relacionadas ao combate às *Fake News* adaptou *datasets* originalmente criados para investigar outros problemas em redes sociais, como divulgação de

Rumor. Esses datasets adaptados, geralmente, não contêm informações importantes para a detecção de *Fake News*, como rótulos *fake* / não *fake*. Além disso, a maioria desses *datasets*, adaptados ou originalmente criados para detecção de *Fake News*, não descrevem a propagação das notícias nas redes sociais, como uma mesma notícia divulgada por vários usuários e várias notícias divulgadas por um mesmo usuário. Assim, não há um consenso sobre os *datasets* de referência para esse problema [Shu et al. 2017]. Outro fator complicador para a criação de *datasets* é a carência de informação, proveniente das redes sociais, para combate às *Fake News*. Tal carência acontece pois, muitas vezes essas informações são apagadas, impossibilitando a sua análise [Mustafaraj and Metaxas 2017].

Além disso, os referidos *datasets* são brevemente descritos, podendo seus detalhes serem consultados através das respectivas referências:

D1) *BS Detector* [Shu et al. 2017]: Este *dataset* é coletado de uma extensão de *browser* chamada *BS Detector* que foi desenvolvido para checagem da veracidade de notícias. Os rótulos existentes são *Fake news, Satire, Extreme bias, Conspiracy theory, Rumor mill, State news, Junk science, Hate group e Clickbait*;

D2) *BuzzFace* [Santia and Williams 2018]: Este repositório foi criado pela equipe do BuzzFeed. Ele contém 2.282 artigos rotulados como *Mostly true, Mixture of true and false, Mostly false e No factual content*;

D3) *BuzzFeedNews (2016-10-facebookfact-check modificado)*
[Janze and Risius 2017]: Conjunto de dados criado a partir do *BuzzFeedNews (2016-10-facebookfact-check)* (D4), sendo os artigos rotulados como *Fake e Non-Fake*;

D4) *BuzzFeedNews (2016-10-facebookfact-check)* [Shu et al. 2017]: Este *dataset* compreende as notícias, do *Facebook*, oriundas de nove agências para a eleição presidencial americana de 2016. Os eventos e artigos ligados foram checados por jornalistas do *BuzzFeed*. Ele contém 1.627 artigos rotulados como *Mostly true, Mixture of true and false, Mostly false e No factual content*;

D5) *Celebrity* [Pérez-Rosas et al. 2018]: Este *dataset* fornece os dados da notícia para análise de texto. As notícias verdadeiras e falsas foram retiradas da *Web*, sendo relacionadas com assuntos de celebridades;

D6) *CredBank* [Shu et al. 2017]: Conjunto de dados criado a partir do cruzamento de várias fontes, com aproximadamente 60 milhões de *tweets*, que cobrem 96 dias, iniciados em outubro de 2015. Todos os *tweets* são relacionados com mais de 1.000 eventos de notícias. Cada evento foi avaliado por 30 anotadores da *Amazon Mechanical Turk*. Os rótulos existentes são *[-2]Certainly inaccurate, [-1]Probably inaccurate, [0] Uncertain (doubtful), [+1] Probably accurate e [+2] Certainly accurate*;

D7) *DataSet Emergent* [Zhang et al. 2018][Bhatt et al. 2018]: Neste repositório, as notícias são rotuladas como *Agree* (o texto do corpo concorda com a manchete), *Disagree* (o texto do corpo discorda da manchete), *Discuss* (o texto do corpo discute a mesma afirmação que o título, mas não toma uma posição) e *Unrelated* (o texto do corpo discute uma alegação que difere do título). Esta base faz parte do primeiro desafio (FNC-1) e foi criada a partir do *dataset* para detecção de rumor chamado *Emergent*;

D8) *DistrustRank Datasets* [Woloszyn and Nejdl 2018]: Foram desenvolvidos dois *datasets*. O primeiro, gerado com sites confiáveis, por meio do *SimilarWeb*[17], tem 502 domínios e 396.422 *URLs* de notícias. O segundo, obtido com sites não confiáveis, através do *Wikipedia's list of prominent Fake News*[18], possui 47 dominios e 37.320 *URLs* de notícias;

D9) *Facebook para Detective* [Tschiatschek et al. 2018]: Repositório que considera os círculos sociais do *Facebook*, consistindo de 4.039 usuários (nós) e 88.234 arestas;

D10) *Factck.Br* [Moreno and Bressan 2019]: *Dataset* composto por textos de 1.313 notícias no idioma português que foram verificadas individualmente pelas agências de checagem de fatos brasileiras 'Lupa', 'Aos Fatos' e 'Truco'. Possui como principal característica o fato de ser atualizável, ou seja, há uma API que possibilita a atualização do *Dataset* com novas notícias analisadas por agências de checagem. As notícias são rotuladas em *Ainda é cedo para dizer, de olho, discutível, distorcido, exagerado, impossível provar, impreciso, insustentável, verdadeiro, falso e outros*;

D11) *FakeBr* [Monteiro et al. 2018]: Um dos primeiros *datasets* disponibilizados para viabilizar a construção de modelos de detecção de *Fake News* no idioma português. Possui 7.200 notícias, das quais 3.600 são rotuladas *verdadeiras*, obtidas em versões digitais de mídias virtuais tradicionais. As outras 3.600, obtidas em sites considerados propagadores de notícias falsas, são rotuladas como *fake*;

D12) *FakePedia* [Charles and Oliveira 2018]: *Dataset* composto por textos de notícias no idioma português que foram verificadas através de *sites* brasileiros: 'e-farsas' e 'boatos.org'. As notícias são rotuladas em *fake news ou true news*;

D13) *Fake News vs Satire* [Golbeck et al. 2018]: *Dataset* para diferenciar *Fake News* e Sátiras onde as notícias são codificadas manualmente. A base, oriunda de diversas fontes, é composta por 283 relatos rotulados como *Fake News* e 203 como *Satirical*. Estes relatos são compostos pelo título, texto e um link para cada artigo;

D14) *Fakeddit* [Kai Nakamura 2019]: *Dataset* composto por 628.501 notícias rotuladas como *Fake* e 527.049 rotuladas como *True*. Estas notícias contém o texto, imagem e os dados dos comentários;

D15) *FakeNewsAMT* [Pérez-Rosas et al. 2018]: As notícias falsas e legítimas são fornecidas em duas pastas separadas. Cada pasta contém 40 notícias de seis domínios diferentes: tecnologia, educação, negócios, esportes, política e entretenimento;

D16) *FakeNewsData1* [Horne and Adali 2017]: São dois *datasets* onde o primeiro contém notícias rotuladas como *Fake e Real* retiradas a partir do *BuzzFeed*. Já o segundo contém notícias políticas rotuladas como *Real, Fake* e Sátira obtidas, randomicamente, durante as eleições americanas de 2016;

D17) *FakeNewsNet1* [Shu et al. 2017] [Shu et al. 2019b] [Sharma et al. 2019] [Gupta et al. 2018] [Shu et al. 2019a]: Esta base de dados, coletada do Twitter, fornece 211 notícias *Fake* e 211 notícias *Real*, rotuladas a partir do *BuzzFeed* e *PolitiFact*. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

[17]https://www.similarweb.com/top-websites/category/News-and-media

[18]https://en.wikipedia.org/wiki/List-of-fake-News-websites

D18) *FakeNewsNet2* [Shu et al.][Sharma et al. 2019]: Esta base de dados, coletada do Twitter, fornece 6.480 notícias *Fake* e 17.441 notícias *Real*, rotuladas a partir do *GossipCop*[19] e *PolitiFact*. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

D19) *FakeNewsSet* [da Silva et al. 2020]: Repositório, coletado do Twitter, contendo 300 notícias *Fake* e 300 notícias *Not Fake* em português, rotuladas a partir de agências de checagem de fatos nacionais. Este trabalho armazena tanto os dados de publicação quanto de propagação das notícias;

D20) *Kaggle*[20]: Este conjunto de dados contém texto e metadados de 244 sites, totalizando 12.999 postagens. Os dados foram extraídos usando a *API webhose.io*. Cada site foi rotulado de acordo com o *BS Detector*, sendo que as fontes de dados sem rótulo foram categorizadas como *Bs*;

D21) *KV* [Dong et al. 2014]: Nesta base as notícias têm sujeito, predicado e objeto. Cada notícia tem um rótulo que indica a probabilidade da mesma ser verdadeira. A ferramenta, por meio de uma fusão de conhecimentos, cria um grafo relacionando o sujeito com o objeto para medir a quantidade de interações e, assim, gerar automaticamente o *dataset*;

D22) *Large-scale Training Dataset e Small-scale Evaluation Dataset* [Helmstetter and Paulheim 2018]: No *Large-scale Training Dataset* cada tweet de uma fonte confiável é rotulado como notícia real e cada tweet de um uma fonte não confiável é rotulado como uma *Fake News*. As 46 fontes confiávies e 65 não confiáveis foram obtidas através de pesquisas em sites e os *tweets* foram coletados a partir dessas fontes. No total, foram coletados 401.414 exemplos, nos quais 110.787 foram rotulados como *Fake News*, enquanto 290.627 foram rotulados como *Real News*. O *Small-scale Evaluation Dataset* contém 116 *tweets* rotulados manualmente a partir do *PolitiFact*;

D23) *LIAR* [Wang 2017]: Esta base de dados é coletada do *PolitiFact*. Ele inclui 12.836 notícias rotuladas manualmente como *Pants-fire, False, Barely-true, Half-true, Mostly true e True*. Cabe salientar que os dados referentes ao usuário se resumem ao nome do autor da postagem;

D24) *PoliticalNews* [Castelo et al. 2019]: Para criar o dataset foram usados os sites *Politifact, BuzzFeed, OpenSources.co e Alexa's top 500 news*[21]. O resultado foi um *dataset* com 14.240 páginas de notícias sendo 7.136 páginas vindas de 79 sites não confiáveis e 7.104 vindos de 58 sites confiáveis;

D25) *PolitiFact para XFake* [Yang et al. 2019]: Repositório criado a partir do site *PolictFact* com 5.104 notícias contendo os atributos *Subject, Context, Speaker, Targeting e Statement*. As notícias foram rotuladas como *True e False*;

D26) *RST-SVM Dataset* [Rubin et al. 2015]: Esta base de dados foi criada a partir de codificadores, usando notícias do *Bluff the Listener*[22]. Esse repositório consiste de 144

[19]https://https://www.gossipcop.com/
[20]https://www.kaggle.com/datasets
[21]https://www.alexa.com/topsites/category/News
[22]https://www.npr.org/bluff-the-listener

notícias selecionadas, aleatoriamente, de 2010 até 2014;

D27) *Signal Media para Evaluating Machine Learning Algorithms for Fake News Detection* [Gilda 2017]: *Dataset* rotulado com *Fake* ou *Not Fake* criado a partir de uma base de notícias da *Signal Media* e uma lista do repositório de confiança de fontes *OpenSources.co*. Esse *dataset* contém 11.051 artigos, sendo 3.217 rotulados com falsos;

D28) *Soc-LiveJournal* [Srivastava et al. 2018]: Este repositório não rotulado contém uma rede de relacionamentos formada por 4.847.571 nós e 68.475.391 arestas;

D29) *Twitter e Sina Weibo para CSI* [Ruchansky et al. 2017]: *Dataset* criado com 2.811 artigos rotulados como *Fake* e 2.845 como *True*. A citada base de dados foi obtida a partir do repositório, para detecção de rumores, gerado no artigo [Ma et al. 2016];

D30) *Twitter e Sina Weibo para EANN* [Ruchansky et al. 2017]: A base de dados foi criada a partir de dois *datasets* não originários de *Fake News*. O primeiro repositório foi obtido a partir do *Sina Weibo* contendo 4.749 notícias com rótulos adaptados para *Fake* e 4.779 para *Real*, além de 9.528 imagens. O segundo repositório foi obtido a partir do *Twitter* contendo 7.898 notícias com rótulos adaptados para *fake* e 6.026 para real, além de 514 imagens;

D31) *Twitter e Sina Weibo para Early Detection Through Propagation Path* [Liu and BrookWu 2018]: Este repositório foi criado a partir de três *datasets* usados para detecção de rumores. O primeiro, oriundo da rede social *Weibo*, com os rótulos *rumor (fake) e otherwise (true)*. Já os outros dois *datasets*, obtidos do *Twitter*, são rotulados como *fake, true, unverified e non-rumor (debunking of fake)*. As características dos usuários foram obtidas por meio de pesquisas realizadas nas respectivas redes sociais.

D32) *Twitter e Sina Weibo para TCNN-URG* [Qian et al. 2018]: Base de dados que utilizou dois *datasets*. O primeiro *dataset* foi obtido, automaticamente, a partir do *Sina Weibo*. Já o segundo *dataset* foi gerado por um processo manual de coleta de dados. Para tal, foram selecionadas notícias em sites avaliados como confiáveis (*The Guardian*[23]) e notoriamente falsos. Com as URLs de todas as notícias coletadas, pesquisas foram realizadas no *Twitter* para cada uma das notícias rotuladas como falsas ou reais.

D33) *Twitter para Automatically Identifying Fake News* [Buntain and Golbeck 2017]: Base de dados que utilizou os *datasets* PHEME (rumor no *Twitter*), CredBank (credibilidade no *Twitter*) e *BuzzFeed News Fact-Checking Dataset* (Checagem de fatos no *Facebook*). Os três *datasets* precisaram ser alinhados com as mesmas características e rótulos;

D34) *Twitter para Content Polluters* [Nasim et al. 2018]: Repositório de dados criado para detecção de *bots*. Esse *dataset*, obtido a partir do *Twitter*, foi rotulado manualmente como *Bot* ou *Não Bot*;

D35) *Twitter para Mitigation via Point Process* [Farajtabar et al. 2017]: Este trabalho realizou experimentos com contas reais no Twitter e com uma base de dados sintética, onde foi assumido que 20 nós criaram *Fake News* e outros 20 nós divulgaram notícias verdadeiras;

[23] https://www.theguardian.com/

D36) *Twitter para TraceMiner* [Wu and Liu 2018]: Conjunto de dados gerado pela coleta de informações do *Twitter* com rotulação a partir do site de checagem de fatos *Snopes*[24]. Nessa base, os rótulos atribuidos são *Real news ou Fake news*;

D37) *Twitter Trec* [Srivastava et al. 2018]: Conjunto de dados gerado pela coleta de informações do *Twitter*, sem rotulação, contendo uma rede de relacionamentos formada por 3.919.215 nós e 5.399.949 arestas;

D38) *Twitter Using One-Class* [Faustini and Covões 2019]: Conjunto de dados gerado pela coleta de informações do *Twitter* e de sites de checagem de fatos contendo 4.392 notícias "fake"e 4.589 notícias "real";

D39) *WhatssApp Using One-Class* [Faustini and Covões 2019]: Conjunto de dados gerado pela coleta de informações do aplicativo de troca de mensagens *WhatsApp* e de sites de checagem de fatos contendo 165 notícias "fake"e 12 notícias "real".

Tabela 1.4: Trabalhos x Datasets

Id	DataSet
T1	DataSet Emergent (D7) com Benchmark NLI: Stanford Natural Language Inference (SNLI)
T2	FakeBr (D11)
T3	FakeNewsAMT (D15), Celebrity (D5) e PoliticalNews (D24)
T4	FakeNewsAMT (D15) e Celebrity (D5)
T5	BuzzFeedNews (2016-10-facebookfact-check modificado) (D3)
T6	Twitter para Automatically Identifying Fake News (D33)
T7	Twitter e Sina Weibo para EANN (D30)
T8	FakeNewsNet1 (D17)
T9	FakeNewsNet1 (D17)
T10	DataSet Emergent (D7)
T11	Twitter e Sina Weibo para CSI (D29)
T12	FakeBr (D11) e FakePedia (D12)
T13	DistrustRank Datasets (D8)
T14	Twitter e Sina Weibo para EANN (D30)
T15	Twitter e Sina Weibo para Early Detection Through Propagation Path (D31)
T16	Signal Media para Evaluating Machine Learning Algorithms for Fake News Detection (D27)
T17	Soc-LiveJournal (D28) e Twitter Trec (D37)
T18	FakeBr (D11), FakeNewsNet2 (D18) e FakeNewsSet (D19)
T19	Facebook para Detective (D9)
T20	Twitter Using One-Class (D38) e WhatsApp Using One-Class (D39) e FakeBr (D11)
T21	Twitter para Mitigation via Point Process (D35)
T22	FakeNewsNet1 (D17)
T23	Fakeddit (D14)
T24	LIAR (D23)
T25	Twitter e Sina Weibo para TCNN-URG (D32)
T26	DataSet Emergent (D7)
T27	Twitter para Content Polluters (D34)
T28	Pheme (dataset para Rumor)
T29	FakeNewsSet (D19)
T30	BuzzFace (D2)
T31	FakeNewsData1 (D16)
T32	RST-SVM Dataset (D27)
T33	Twitter para TraceMiner (D36)
T34	FakeNewsNet2 (D18)
T35	Large-scale Training Dataset e Small-scale Evaluation Dataset (D22)
T36	PolitiFact para XFake (D25)

1.4. Problemas em Aberto

O combate automático às *Fake News* nos MDDN é uma nova e emergente área de pesquisa que, mesmo com estudos já realizados, ainda carece de maior aprofundamento científico. Desta forma, os seguintes problemas são descritos como áreas ainda férteis para o desenvolvimento de novos trabalhos:

• Carência de *datasets* que forneçam, de forma suficiente, os diferentes dados necessários para combater as *Fake News* em diferentes MDDN;

• Trabalhos que levem em consideração aspectos temporais do ciclo de vida da *Fake News* e que, consequentemente, possam intervir mais rapidamente;

• Estudos que analisem o aspecto intencional, assim não se limitam a verificar a

[24]https://www.snopes.com

Tabela 1.5: Comparação entre Datasets

Id	Dados									URL
	Publicação				Propagação					
	Notícia			Usuário	Contribuição			Usuário	Rede	
	Texto	Áudio	Imagem		Texto	Áudio	Imagem			
D1	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	Link para notícia	Link para notícia	Link para notícia	Link para notícia	https://github.com/thiagovas/bs-detector-dataset
D2	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	https://github.com/gsantia/BuzzFace
D3	No Dataset		Link para imagem	No Dataset	No Dataset			Link para notícia	Link para notícia	
D4	Link para notícia	Link para notícia	Link para notícia	No Dataset	Link para notícia	Link para notícia	Link para notícia	Link para notícia	Link para notícia	https://github.com/BuzzFeedNews/2016-10-facebook-fact-check
D5	No Dataset			No Dataset						http://lit.eecs.umich.edu/downloads.html#undefined
D6	No Dataset			No Dataset				No Dataset	No Dataset	http://compsocial.github.io/CREDBANK-data/
D7	No Dataset			No Dataset						https://github.com/FakeNewsChallenge/fnc-1
D8	Link para notícia	Link para notícia	Link para notícia	No Dataset						
D9				No Dataset				No Dataset		
D10	No Dataset									https://github.com/jghm-f/FACTCK.BR
D11	No Dataset									https://github.com/roneysco/Fake.br-Corpus
D12	No Dataset									http://www.fakepedia.org/
D13	No Dataset	Link para notícia	Link para notícia	No Dataset						https://github.com/jgolbeck/fakenews
D14	No Dataset		No Dataset	No Dataset	No Dataset		No Dataset	No Dataset		https://github.com/entitize/Fakeddit
D15	No Dataset			No Dataset						http://lit.eecs.umich.edu/downloads.html#undefined
D16	No Dataset									https://github.com/BenjaminDHorne/fakenewsdata1/blob/master/Horne2017_FakeNewsData.zip
D17	No Dataset		Link para imagem	No Dataset				No Dataset	No Dataset	https://github.com/KaiDMML/FakeNewsNet
D18	No Dataset		Link para notícia	No Dataset				No Dataset	No Dataset	https://github.com/KaiDMML/FakeNewsNet
D19	No Dataset		Link para notícia	No Dataset	No Dataset		No Dataset	No Dataset	No Dataset	https://github.com/kamplus/FakeNewsSetGen
D20	No Dataset		Link para imagem	No Dataset						https://www.kaggle.com/mrisdal/fake-news/data
D21	No Dataset			No Dataset						
D22	No Dataset			No Dataset						http://dws.informatik.uni-mannheim.de/en/research/twitter-fake-news-detection
D23	No Dataset			No Dataset						https://github.com/nishitpatel01/Fake_News_Detection/tree/master/liar_dataset ou https://www.cs.ucsb.edu/~william/software.html
D24	No Dataset			No Dataset						https://osf.io/ez5q4/
D25	No Dataset			No Dataset						
D26	No Dataset			No Dataset						
D27	No Dataset			No Dataset						
D28									No Dataset	https://snap.stanford.edu/data/soc-LiveJournal1.html
D29	No Dataset			No Dataset	No Dataset			No Dataset	No Dataset	https://github.com/majingCUHK/Rumor_RvNN ou http://alt.qcri.org/~wgao/data/rumdect.zip
D30	No Dataset		No Dataset	No Dataset						
D31				No Dataset				No Dataset	No Dataset	Twitter 15 e 16 (https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0) e Weibo(http://alt.qcri.org/~wgao/data/rumdect.zip)
D32	No Dataset			No Dataset	No Dataset					False (https://drive.google.com/open?id=1WRoRVV9j4CStMFKDwP7DVGAHFJZX4t5a) e True(https://drive.google.com/open?id=1JgbW4suN2yWHx65P4QU8HkrrB3OMHsuo)
D33	No Dataset			No Dataset				No Dataset	No Dataset	
D34				No Dataset				No Dataset	No Dataset	
D35									No Dataset	
D36				No Dataset				No Dataset	No Dataset	
D37	No Dataset			No Dataset				No Dataset	No Dataset	https://trec.nist.gov/data/tweets/
D38	Link para notícia									https://github.com/phfaustini/BRACIS2019_FAKENEWS
D39	No Dataset									https://github.com/phfaustini/BRACIS2019_FAKENEWS

autenticidade (veracidade) das notícias;

• Extração de características a partir de imagem e/ou áudio, portanto não se limitando as análises de texto (multimodal);

• Métodos que abordem características baseadas na rede que representa a propagação da notícia. Neste caso, inclusive, podem ser aplicadas técnicas baseadas em grafos;

• Pesquisas que, ao invés de realizarem uma classificação binária, utilizem probabilidades e/ou pertinências na detecção. Esta linha de trabalho se baseia no fato de que, normalmente, a *Fake News* é uma mistura de afirmações falsas e verdadeiras;

• Utilização de um comitê de classificadores para determinar se uma notícia é *fake*. Desta forma, pode-se agregar diferentes técnicas de classificação na detecção;

• Utilização de modelos não supervisionados ou semi-supervisionados devido à carência de *datasets* rotulados que possuam variedade de dados;

• Estudo sobre o comportamento distinto da *Fake News* em diferentes comunida-

des (escolar, trabalho e etc) e/ou redes sociais *(Weibo, WhatsApp e etc)*. Isto se deve pela possível mudança de comportamento das notícias de acordo com o meio;

• Classificar os usuários de *Fake News* com o objetivo de identificar o seu tipo *(humanos, bots e cyborgs)*. Isto se deve pela possível alteração de comportamento das notícias propositalmente falsas de acordo com o seu tipo de usuário divulgador;

• Trabalhos relacionados à intervenção de *Fake News*, tanto para bloqueio quanto para mitigação. Haja vista que o combate às *Fake News* não se limita à detecção, sendo necessária, também, a intervenção sobre a mesma;

• Abordagens que atuem descentralizadas na rede. Esta atuação se destaca, pois quanto mais rápido e extensivo for o combate, menor serão os efeitos nocivos da notícia;

• Abordagens que utilizem o assunto para a análise da notícia, pois assuntos relevantes, normalmente, motivam a criação de notícias intencionalmente falsas;

• Pesquisas que levem em consideração a reputação dos usuários, pois usuários com baixa reputação tendem a ser potenciais divulgadores de *Fake News*;

• Detecção de Fake News por meio de recursos da Web Semântica;

• Explicabilidade na Detecção de Fake News;

• Aplicação de TranferLearning na Detecção de Fake News.

1.5. Considerações Finais

Com a crescente popularidade dos meios digitais de divulgação de notícias (MDDN), cada vez mais pessoas consomem notícias on-line, em vez dos tradicionais meios de comunicação. No entanto, os MDDN também são usados para divulgar notícias intencionalmente falsas, as chamadas *Fake News*, que podem causar fortes impactos negativos. Nesse capítulo foi explorado o combate automático às Fake News em MDDN. Para tal, a literatura existente foi revisada objetivando, por meio de um levantamento do estado da arte, fornecer subsídios para pesquisas que busquem desenvolver métodos para o combate automático às Fake News em MDDN. Tendo como base essa revisão da literatura, dois aspectos significativos podem ser destacados. O primeiro é a carência de datasets, rotulados com fake e não fake, que disponibilizem não somente os dados da publicação, mas, também, as informações relacionadas à propagação das notícias. O segundo aspecto é que os métodos computacionais, voltadas para a detecção, vêm se adaptando de acordo com as mudanças nas características das *Fake News*. Uma dessas mudanças é a maior similaridade nas características de escrita, presentes no texto, entre as notícias *fake* e não *fake*. Portanto, os métodos que não utilizam somente o conteúdo da notícia na tarefa de detecção de *Fake News* têm apresentado resultados promissores. Nesse grupo particular de métodos, aqueles baseados na reputação dos usuários dos MDDN se apresentam como uma alternativa para a detecção de notícias intencionalmente falsas.

Referências

[Ajao et al. 2019] Ajao, O., Bhowmik, D., and Zargari, S. (2019). Sentiment aware fake news detection on online social networks. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2507–2511.

[Bhatt et al. 2018] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., and Mittal, A. (2018). Combining neural, statistical and external features for fake news stance identification. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1353–1357, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[Braz and Goldschmidt 2017] Braz, P. and Goldschmidt, R. (2017). Um método para detecção de bots sociais baseado em redes neurais convolucionais aplicadas em mensagens textuais. In SBSeg 2017, pages 501–508. 10/11/2017.

[Buntain and Golbeck 2017] Buntain, C. and Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In 2017 IEEE International Con on Smart Cloud (SmartCloud), pages 208–215.

[Campan et al. 2017] Campan, A., Cuzzocrea, A., and Truta, T. M. (2017). Fighting fake news spread in online social networks: Actual trends and future research directions. In 2017 IEEE International Con on Big Data (Big Data), pages 4453–4457.

[Castelo et al. 2019] Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E., and Freire, J. (2019). A topic-agnostic approach for identifying fake news pages. In Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, pages 975–980, New York, NY, USA. ACM.

[Cazalens et al. 2018] Cazalens, S., Lamarre, P., Leblay, J., Manolescu, I., and Tannier, X. (2018). A content management perspective on fact-checking. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 565–574, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[Charles and Oliveira 2018] Charles, A. and Oliveira, J. (2018). Checking fake news on web browsers: an approach using collaborative datasets.

[Ciampaglia et al. 2015] Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. PLOS ONE, 1:1–13.

[Conroy et al. 2015] Conroy, N., Rubin, V., and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Association for Information Science and Technology, 52:1–4.

[da Silva et al. 2020] da Silva, F. R. M., Freire, P. M. S., de Souza, M. P., de A. B. Plenamente, G., and Goldschmidt, R. R. (2020). Fakenewssetgen: A process to build datasets that support comparison among fake news detection methods. WebMedia '20, page 241–248, New York, NY, USA. Association for Computing Machinery.

[de Souza et al. 2020] de Souza, M. P., da Silva, F. R. M., Freire, P. M. S., and Goldschmidt, R. R. (2020). A linguistic-based method that combines polarity, emotion and grammatical characteristics to detect fake news in portuguese. In Proceedings of the Brazilian Symposium on Multimedia and the Web, WebMedia '20, page 217–224, New York, NY, USA. Association for Computing Machinery.

[Deng et al. 2014] Deng, S., Huang, L., and Xu, G. (2014). Social network-based service recommendation with trust enhancement. Expert Systems with Applications, 41(18):8075 – 8084.

[Dong et al. 2014] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In ACM SIGKDD international Con on Knowledge discovery and data mining, pages 601–610.

[Empirico 1997] Empirico, S. (1997). Hipotiposes pirrônicas livro i (tradução de danilo marcondes). O que nos faz pensar, 9(12):115–122.

[Farajtabar et al. 2017] Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., and Zha, H. (2017). Fake news mitigation via point process based intervention. In Proceedings of the 34th International Con on Machine Learning - Volume 70, ICML'17, pages 1097–1106. JMLR.org.

[Faustini and Covões 2019] Faustini, P. and Covões, T. (2019). Fake news detection using one-class classification. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), pages 592–597.

[Ferrara et al. 2016] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. Commun. ACM, 59(7):96–104.

[Flintham et al. 2018] Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., and Moran, S. (2018). Falling for fake news: Investigating the consumption of news via social media. In Proceedings of the 2018 CHI Con on Human Factors in Computing Systems, CHI '18, pages 376:1–376:10, New York, NY, USA. ACM.

[Freire and Goldschmidt 2020] Freire, P. and Goldschmidt, R. (2020). Combatendo fake news nas redes sociais via crowd signals implícitos. In Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, pages 424–435, Porto Alegre, RS, Brasil. SBC.

[Freire and Goldschmidt 2019] Freire, P. M. S. and Goldschmidt, R. R. (2019). Fake news detection on social media via implicit crowd signals. In Proceedings of the 25th Brazilian Symposium on Multimedia and the Web, WebMedia '19, pages 521–524, New York, NY, USA. ACM.

[Gilda 2017] Gilda, S. (2017). Evaluating machine learning algorithms for fake news detection. In 2017 IEEE 15th Student Con on Research and Development (SCOReD), pages 110–115.

[Golbeck et al. 2018] Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J. B., Falak, W., Gieringer, C., Graney, J., Hoffman, K. M., Huth, L., Ma, Z., Jha, M., Khan, M., Kori, V., Lewis, E., Mirano, G., Mohn IV, W. T., Mussenden, S., Nelson, T. M., Mcwillie, S., Pant, A., Shetye, P., Shrestha, R., Steinheimer, A., Subramanian, A., and Visnansky, G. (2018). Fake news vs satire: A dataset and analysis. In Proceedings of

the 10th ACM Con on Web Science, WebSci '18, pages 17–21, New York, NY, USA. ACM.

[Guo et al. 2020] Guo, B., Ding, Y., Yao, L., Liang, Y., and Yu, Z. (2020). The future of false information detection on social media: New perspectives and trends. ACM Comput. Surv., 53(4).

[Gupta et al. 2018] Gupta, S., Thirukovalluru, R., Sinha, M., and Mannarswamy, S. (2018). Cimtdetect: A community infused matrix-tensor coupled factorization based method for fake news detection. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 278–281.

[Helmstetter and Paulheim 2018] Helmstetter, S. and Paulheim, H. (2018). Weakly supervised learning for fake news detection on twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 274–277.

[Hendrikx et al. 2015] Hendrikx, F., Bubendorfer, K., and Chard, R. (2015). Reputation systems: A survey and taxonomy. Journal of Parallel and Distributed Computing, pages 184–197.

[Horne and Adali 2017] Horne, B. D. and Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Association for the Advancement of Artificial Intelligence.

[Janze and Risius 2017] Janze, C. and Risius, M. (2017). Automatic detection of fake news on social media platforms. In PACIS 2017.

[Kai Nakamura 2019] Kai Nakamura, Sharon Levy, W. Y. W. (2019). r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. ArXiv.

[Kim et al. 2018] Kim, J., Tabibian, B., Oh, A., Schölkopf, B., and Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In Proceedings of the Eleventh ACM International Con on Web Search and Data Mining, WSDM '18, pages 324–332, New York, NY, USA. ACM.

[Kshetri and Voas 2017] Kshetri, N. and Voas, J. (2017). The economics of fake news. IT Professional, 19(06):8–12.

[Li et al. 2015] Li, Y., Gao, J., Meng, C., Li, Q., Su, L., Zhao, B., Fan, W., and Han, J. (2015). A survey on truth discovery. ACM SIGKDD Explorations Newsletter, 17:1–16.

[Liu and BrookWu 2018] Liu, Y. and BrookWu, Y. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In AAAI Con on Artificial Intelligence, pages 354–361.

[Liu and Xu 2016] Liu, Y. and Xu, S. (2016). Detecting rumors through modeling information propagation networks in a social media environment. IEEE Transactions on Computational Social Systems, 3(2):46–62.

[Ma et al. 2016] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K., and Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In International Joint Con on Artificial Intelligence.

[Ma et al. 2015] Ma, J., Gao, W., Wei, Z., Lu, Y., and Wong, K.-F. (2015). Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Con on Information and Knowledge Management, CIKM '15, pages 1751–1754, New York, NY, USA. ACM.

[Maia et al. 2021] Maia, I. M. L., de Souza, M. P., da Silva, F. R. M., Freire, P. M. S., and Goldschmidt, R. R. (2021). A sentiment-based multimodal method to detect fake news. In Proceedings of the Brazilian Symposium on Multimedia and the Web, WebMedia '21, New York, NY, USA. Association for Computing Machinery.

[Mejova and Kalimeri 2020] Mejova, Y. and Kalimeri, K. (2020). Advertisers jump on coronavirus bandwagon: Politics, news, and business. ArXiv, abs/2003.00923.

[Monteiro et al. 2018] Monteiro, R. A. et al. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In Computational Processing of the Portuguese Language, pages 324–334, Cham. Springer International.

[Moraes et al. 2019] Moraes, M. P., de Oliveira Sampaio, J., and Charles, A. C. (2019). Data mining applied in fake news classification through textual patterns. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia '19, page 321–324, New York, NY, USA. Association for Computing Machinery.

[Moreno and Bressan 2019] Moreno, J. a. and Bressan, G. (2019). Factck.br: A new dataset to study fake news. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia '19, page 525–527, New York, NY, USA. ACM.

[Mustafaraj and Metaxas 2017] Mustafaraj, E. and Metaxas, P. T. (2017). The fake news spreading plague:was it preventable? In Web Science Con, pages 236–239.

[Nasim et al. 2018] Nasim, M., Nguyen, A., Lothian, N., Cope, R., and Mitchell, L. (2018). Real-time detection of content polluters in partially observable twitter networks. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1331–1339, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[Passos et al. 2020] Passos, C., Silva, F., Souza, I., Freire, P., and Goldschmidt, R. (2020). Jogos educacionais digitais como ferramentas de apoio à capacitação discente na identificação de fake news escritas em língua portuguesa: Um estudo de caso. In Anais do XXXI Simpósio Brasileiro de Informática na Educação, pages 401–410, Porto Alegre, RS, Brasil. SBC.

[Passos et al. 2021] Passos, C., Silva, F., Souza, I., Freire, P., and Goldschmidt, R. (2021). Jedi – um jogo educacional digital para apoiar a capacitação discente na identificação de fake news escritas em língua portuguesa: Estudos de caso nos ensinos médio e superior. Revista Brasileira de Informática na Educação, 29(0):634–661.

[Pérez-Rosas et al. 2018] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2018). Automatic detection of fake news. In International Conference on Computational Linguistics, pages 3391–3401, Santa Fe, New Mexico, USA.

[Qian et al. 2018] Qian, F., Gong, C., Sharma, K., and Liu, Y. (2018). Neural user response generator: Fake news detection with collective user intelligence. In International Joint Con on Artificial Intelligence, pages 3834–3840.

[Reis et al. 2019] Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Explainable machine learning for fake news detection. In Proceedings of the 10th ACM Conference on Web Science, WebSci '19, pages 17–26, New York, NY, USA. ACM.

[Reis et al. 2019] Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Supervised learning for fake news detection. IEEE Intelligent Systems, 34(2):76–81.

[Rodrigues 2013] Rodrigues, J. G. (2013). O relativismo acerca da verdade refuta-se a si mesmo? Revista Portuguesa de Filosofia, 69(3/4):777–806.

[Rubin et al. 2015] Rubin, V. L., Conroy, N. J., and Chen, Y. (2015). Towards news verification: Deception detection methods for news discourse.

[Ruchansky et al. 2017] Ruchansky, N., Seo, S., and Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Con on Information and Knowledge Management, CIKM '17, pages 797–806, New York, NY, USA. ACM.

[Saikh et al. 2020] Saikh, T., Haripriya, B., Ekbal, A., and Bhattacharyya, P. (2020). A deep transfer learning approach for fake news detection. In International Joint Conference on Neural Networks (IJCNN), pages 1–8.

[Santia and Williams 2018] Santia, G. C. and Williams, J. R. (2018). Buzzface: A news veracity dataset with facebook user commentary and egos. In AAAI Con on Web and Social Media, pages 531–540.

[Seo J. 2013] Seo J., Choi S., H. S. (2013). The method of trust and reputation systems based on link prediction and clustering. In IFIP International Con on Trust Management, pages 223–230.

[Sethi 2017] Sethi, R. J. (2017). Crowdsourcing the verification of fake news and alternative facts. In Proceedings of the 28th ACM Con on Hypertext and Social Media, HT '17, pages 315–316, New York, NY, USA. ACM.

[Sharma et al. 2019] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Trans. Intell. Syst. Technol., 10(3):21:1–21:42.

[Sherchan et al. 2013] Sherchan, W., Nepal, S., and Paris, C. (2013). A survey of trust in social networks. ACM Comput. Surv., 45(4):47:1–47:33.

[Shu et al. 2019a] Shu, K., Mahudeswaran, D., and Liu, H. (2019a). Fakenewstracker: A tool for fake news collection, detection, and visualization. Comput. Math. Organ. Theory, 25(1):60–71.

[Shu et al.] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 8(3):171–188.

[Shu et al. 2017] Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. SIGKDD Explor. Newsl., 19(1):22–36.

[Shu et al. 2019b] Shu, K., Wang, S., and Liu, H. (2019b). Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM '19, pages 312–320, New York, NY, USA. ACM.

[Souza Freire et al. 2021] Souza Freire, P. M., Matias da Silva, F. R., and Goldschmidt, R. R. (2021). Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning. Expert Systems with Applications, 183:115414.

[Sponholz 2009] Sponholz, L. (2009). O que é mesmo um fato? conceitos e suas conseqüências para o jornalismo. In Revista Galáxia, pages 56–69. PUC-SP.

[Srivastava et al. 2018] Srivastava, A., Kannan, R., Chelmis, C., and Prasanna, V. K. (2018). Factcheck: Keeping activation of fake news at check. In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18, pages 2079–2081, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

[Testoni et al. 2021] Testoni, G., Souza, M., Freire, P. M., and Goldschimidt, R. (2021). Um método linguístico que combina polaridade, emoção e aspectos gramaticais para detecção de fake news em inglês. In Anais do X Brazilian Workshop on Social Network Analysis and Mining, pages 151–162, Porto Alegre, RS, Brasil. SBC.

[Tschiatschek et al. 2018] Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., and Krause, A. (2018). Fake news detection in social networks via crowd signals. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 517–524, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[UNESCO 2016] UNESCO (2016). Marco de Avaliação Global da Alfabetização Midiática e Informacional (AMI): disposição e competências do país. Cetic.br.

[UNESCO 2019] UNESCO (2019). Manual para Educação e Treinamento em Jornalismo. UNESCO.

[Vavilis et al. 2014] Vavilis, S., PetkoviÄ‡, M., and Zannone, N. (2014). A reference model for reputation systems. Decision Support Systems, 61:147 – 154.

[Vo and Lee 2018] Vo, N. and Lee, K. (2018). The rise of guardians: Fact-checking url recommendation to combat fake news. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18, pages 275–284, New York, NY, USA. ACM.

[Vosoughi et al. 2017] Vosoughi, S., Mohsenvand, M. N., and Roy, D. (2017). Rumor gauge: Predicting the veracity of rumors on twitter. ACM Trans. Knowl. Discov. Data, 11(4):50:1–50:36.

[Wang et al. 2018a] Wang, P., Angarita, R., and Renna, I. (2018a). Is this the era of misinformation yet: Combining social bots and fake news to deceive the masses. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 1557–1561, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[Wang 2017] Wang, W. Y. (2017). "liar, liar pants on fire": A new benchmark dataset for fake news detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

[Wang et al. 2018b] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., and Gao, J. (2018b). Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '18, pages 849–857, New York, NY, USA. ACM.

[Woloszyn and Nejdl 2018] Woloszyn, V. and Nejdl, W. (2018). Distrustrank: Spotting false news domains. In Proceedings of the 10th ACM Con on Web Science, WebSci '18, pages 221–228, New York, NY, USA. ACM.

[Wu and Liu 2018] Wu, L. and Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the Eleventh ACM International Con on Web Search and Data Mining, WSDM '18, pages 637–645, New York, NY, USA. ACM.

[Yang et al. 2019] Yang, F., Pentyala, S. K., Mohseni, S., Du, M., Yuan, H., Linder, R., Ragan, E. D., Ji, S., and Hu, X. B. (2019). Xfake: Explainable fake news detector with visualizations. In The World Wide Web Conference, WWW '19, pages 3600–3604, New York, NY, USA. ACM.

[Yu et al. 2020] Yu, J., Huang, Q., Zhou, X., and Sha, Y. (2020). Iarnet: An information aggregating and reasoning network over heterogeneous graph for fake news detection. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–9.

[Zhang et al. 2018] Zhang, Q., Yilmaz, E., and Liang, S. (2018). Ranking-based method for news stance detection. In Companion Proceedings of the The Web Con 2018, WWW '18, pages 41–42, Republic and Canton of Geneva, Switzerland. International World Wide Web Con Steering Committee.

[Zhang et al. 2020] Zhang, T., Wang, D., Chen, H., Zeng, Z., Guo, W., Miao, C., and Cui, L. (2020). Bdann: Bert-based domain adaptation neural network for multi-modal fake news detection. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–8.

[Zhou and Zafarani 2018] Zhou, X. and Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315.

[Zhou and Zafarani 2019] Zhou, X. and Zafarani, R. (2019). Fake news detection: An interdisciplinary research. In Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, pages 1292–1292, New York, NY, USA. ACM.

[Zhou et al. 2019] Zhou, X., Zafarani, R., Shu, K., and Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. In Proceedings of the Twelfth ACM International Con on Web Search and Data Mining, WSDM '19, pages 836–837, New York, NY, USA. ACM.