Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Bacharelado em Ciência da Computação

Detecção automática de *fake news* em redes sociais: um mapeamento e revisão sistemática da literatura

Fernando Marques de Souza Filho

JUIZ DE FORA
DEZEMBRO, 2018

Detecção automática de *fake news* em redes sociais: um mapeamento e revisão sistemática da literatura

Fernando Marques de Souza Filho

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza
Coorientadora: Alessandreia Marta de Oliveira Julio

JUIZ DE FORA
DEZEMBRO, 2018

Detecção automática de *fake news* em redes sociais: um mapeamento e revisão sistemática da literatura

Fernando Marques de Souza Filho

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Jairo Francisco de Souza
Doutor em Informática - PUC-Rio

Alessandreia Marta de Oliveira Julio
Doutora em Computação - UFF

Victor Ströele de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação - COPPE/UFRJ

Saulo Moraes Villela
Doutor em Engenharia de Sistemas e Computação - COPPE/UFRJ

JUIZ DE FORA
03 DE DEZEMBRO, 2018

Resumo

Contexto: A proliferação de notícias falaciosas vem aumentando com a evolução dos meios de comunicação e, principalmente, pelo consumo de notícias através de redes sociais. Com isto, se torna inviável realizar uma curadoria manual de todo o conteúdo publicado nestes meios. Graças a isto, muitos trabalhos em diversas áreas tem sido realizados para tentar conter os males causados pela disseminação das *fake news*. **Objetivo:** O objetivo deste estudo é apresentar uma revisão e mapeamento sistemático capaz de apresentar o estado da arte em detecção de textos falaciosos aplicados no contexto de redes sociais, de forma a identificar os métodos mais frequentes e mais relevantes que tem sido utilizados. **Método:** Para conduzir a revisão e mapeamento sistemático foram utilizadas diretrizes propostas na literatura sobre engenharia de software baseada em evidências. **Resultado:** Este estudo foi capaz de identificar 63 trabalhos que reportam abordagens capazes de detectar *fake news* em redes sociais. A partir destes foram descobertas as principais estratégias de suporte utilizadas, bem como os atributos mais utilizados pelos trabalhos da literatura. **Conclusão:** Este campo de pesquisa está cada vez atraindo mais interesse, por isto uma estruturação para que os mais diversos trabalhos possam ser comparados é essencial para o progresso da área. Sendo isto possível com a criação e utilização de *benchmarks* e bases de dados com curadoria realizada por especialistas.

Palavras-chave: *fake news*, notícias falaciosas, redes sociais, revisão sistemática

Abstract

Context: The proliferation of fallacious news has been increasing with the evolution of the media and the consumption of news through social networks, with this it becomes impracticable to conduct a manual fact-checking of all the content published in these media. Causing a growing search for methods to diminish this problem. Thanks to this many works in many areas have been realized focusing on attempts to contain the troubles caused by the dissemination of fake news. **Objective:** The aim of this study is to present a systematic review and mapping of the literature in order to present the state of the art in the detection of fallacious texts applied in the context of social networks, with the intention of identify the most frequent and relevant methods that have been used. **Method:** In order to conduct systematic review and mapping, we have used guidelines proposed in the evidence-based software engineering literature. **Results:** This study was able to identify 63 papers that report some approach capable of detecting fake news in social networks, from these were discovered the main support strategies used, as well as the attributes most used by the works of the literature. **Conclusion:** This research field is increasingly attracting more interest, therefore a structure so that the most diverse works can be compared is essential for the progress of the area. This is possible with the creation and use of benchmarks and databases curated by specialists.

Keywords: fake news, deception detection, social network, systematic review

"The best books... are those that tell you what you know already."

George Orwell, 1984

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

API - *Application Program Interface*

EC - *Exclusion Criteria*

MQ - *Mapping Question*

RQ - *Resource Question*

1 Introdução

Com a proliferação de conteúdos gerados por usuários através do uso de ferramentas como blogs, Twitter, Facebook, dentre outros, o modo como as notícias são consumidas e publicadas passou por grandes mudanças durante os últimos anos (CONROY; RUBIN; CHEN, 2015). A nova forma de publicação e consumo de notícias possibilitou um maior alcance desses textos, bem como um maior volume de publicações e aceleração no tempo de disseminação destas notícias.

Apesar das grandes vantagens proporcionadas por estes mecanismos, a disseminação de notícias falsas se apresenta como um problema atual, devido ao fato de que, ao possibilitar a disseminação de notícias de forma rápida e em larga escala, impossibilita uma curadoria e checagem da veracidade de forma manual na mesma proporção (CIAMPAGLIA et al., 2015).

Devido à questionável natureza das notícias publicadas por meio de redes sociais, associada à inabilidade humana em julgá-las de acordo com sua genuinidade (JR; DEPAULO, 2006), surge a necessidade de ferramentas que sejam capazes de auxiliar ou determinar a autenticidade destas informações de forma automática.

Com isto, o estudo de métodos que visam facilitar a detecção de textos falaciosos tem aumentado rapidamente nos últimos tempos (ÖZGÖBEK; GULLA, 2017). Porém, a diversidade de técnicas aplicadas para a resolução desse problema, bem como a finalidade de cada uma destas propostas, faz com que novos pesquisadores tenham dificuldade ao começar o estudo do tema.

Graças à recente proliferação das *fake news*, termo utilizado para englobar textos falaciosos que são publicados de forma a assemelhar-se a notícias verdadeiras (JR; LIM; LING, 2018), a busca por um combate a este tipo de publicação tomou a atenção dos pesquisadores. Apesar da existência de alguns trabalhos recentes que abordam algumas destas técnicas na literatura (RUBIN; CHEN; CONROY, 2015; CONROY; RUBIN; CHEN, 2015; ZUBIAGA et al., 2018), não foi encontrado nenhum que aborda especificamente o problema das *fake news* e os métodos utilizados para a detecção das mesmas.

Além disto, a metodologia utilizada nesses trabalhos de revisão da literatura não foram feitas de forma sistemática, o que não garante a abrangência dos trabalhos realizados na área (KITCHENHAM, 2004). Assim, é necessário um estudo mais completo da literatura quanto aos métodos para a classificação automática das *fake news*, de forma a analisar os métodos mais comuns, sua aplicabilidade, além de apontar os principais desafios na área.

1.1 Objetivos

Este trabalho tem como objetivo apresentar o estado da arte na detecção de *fake news* no contexto de redes sociais. Embora grande parte das técnicas utilizadas em outros contextos possam ser aplicadas em redes sociais, abordagens específicas são encontradas na literatura, pois além de ser um meio mais propenso à disseminação desse tipo de texto, este âmbito também possui como especificidade a propagação das informação em rede (ALLCOTT; GENTZKOW, 2017). Assim, espera-se identificar os métodos mais frequentes e mais atuais na identificação de textos falaciosos em redes sociais.

1.2 Metodologia

Para atingir os objetivos desse trabalho, foi realizada uma revisão sistemática e um mapeamento sistemaático da literatura, de maneira a seguir as orientações propostas por Kitchenham (2004). Para auxiliar a tarefa foi utilizada a ferramenta *Parsifal*[1], que apoiou a organização do trabalho e proveu suporte para o armazenamento dos dados encontrados. No mapeamento foram selecionados trabalhos em inglês que apresentam técnicas para identificação de *fake news* no contexto de redes sociais. Os trabalhos foram extraídos de algumas das principais bases de artigos científicos e analisados de acordo com os seguintes aspectos: técnica utilizada pelo trabalho, os atributos utilizados por estas técnicas, contexto do trabalho, forma de avaliação da proposta, o domínio em que este trabalho é apresentado e bases de dados utilizadas.

[1] ⟨http://www.parsif.al⟩

1.3 Organização do Trabalho

Este trabalho está organizado em quatro capítulos, além desta introdução. O Capítulo 2 apresenta uma definição formal do que é uma *fake news*, além de apresentar alguns dos conceitos que serão utilizados neste trabalho e apresentar alguns dos trabalhos relacionados presentes na literatura. O Capítulo 3 descreve o processo utilizado para realizar o mapeamento e a revisão da literatura de forma sistemática. O Capítulo 4 apresenta os resultados encontrados por este processo e realiza a análise dos resultados. Por fim, o Capítulo 5 discute as conclusões deste trabalho e apresenta os trabalhos futuros.

2 Fundamentação Teórica

O problema das *fake news* pode ser caracterizado de algumas formas, assim como possuir diferentes abordagens para tratar este tema. Este capítulo traz uma revisão dos principais conceitos encontrados na literatura para caracterização de uma *fake news*.

Além da caracterização, também são apresentadas definições das técnicas que têm sido utilizadas na área de computação para abordar este assunto e, por fim, também são apresentadas classificações dos atributos que podem ser utilizados para distinguir notícias falsas de verídicas.

O capítulo está dividido como segue: na Seção 2.1 são demonstrados os conceitos básicos que definem o que é uma *fake news*, assim como a etimologia do termo. Ao longo da Seção 2.2 as características dos principais métodos para a identificação de uma *fake news* são discutidas. A Seção 2.3 apresenta os termos que foram utilizados por este trabalho como base para classificação das atributos para identificação das *fake news* e na Seção 2.4 são discutidos outros trabalhos que visam compilar as descobertas na área e suas contribuições.

2.1 Definição do Termo *Fake News*

O termo *fake news* foi cunhado em meados de 2016 por uma uma reportagem realizada pela *Buzz Feed* para caracterizar uma série de notícias absurdas produzidas e vinculadas através de diversos sites e páginas no Facebook (WENDLING, 2018). Desde então o termo foi amplamente utilizado pela mídia, sendo inclusive escolhida como a "palavra do ano"em 2017 pela American Dialect Society (2018). Este termo porém já havia sido utilizado na literatura por alguns autores com a mesma definição. Na área da computação, a primeira formalização do termo está presente em (RUBIN; CHEN; CONROY, 2015).

Desde então o termo vem aumentando sua popularidade e o problema ligado a ele se tornando cada vez mais presente no cotidiano das pessoas. Graças à vasta popularidade do termo e a sua ampla utilização, torna-se difícil entender o que seria uma *fake news*, por

isto alguns trabalhos da literatura buscaram caracterizar *fake news* em alguns subtipos. Rubin, Chen e Conroy (2015) dividem em três tipos: fabricação séria, farsas de larga escala e textos humorísticos, enquanto em (JR; LIM; LING, 2018) são demonstradas seis definições para *fake news* que têm sido utilizadas em todas as áreas de pesquisa, sendo essas: sátira, paródia, fabricação séria, propaganda política, anúncios e manipulação.

Como neste trabalho não são feitas distinções perante ao tipo de *fake news* que a abordagem intenciona resolver, será utilizada a definição que está presente em (JR; LIM; LING, 2018) e engloba o termo para qualquer informação falaciosa que é apresentada de forma a parecer verídica.

Outra diferenciação importante de ser caracterizada é entre os termos rumor e *fake news*. Apesar de alguns trabalhos utilizam equivocadamente o termo rumor para indicar uma informação falsa (ZUBIAGA et al., 2018), os termos possuem significados distintos, sendo rumor uma notícia cuja veracidade ainda não possui confirmação, podendo esta notícia vir ou não se caracterizar como *fake news*.

2.2 Métodos Computacionais para Detecção de *Fake News*

Um conceito importante para classificação das abordagens é o método computacional utilizado por ela. Em (CONROY; RUBIN; CHEN, 2015), o autor propõe uma topologia capaz de dividir as abordagens em duas principais categorias, sendo uma para definir as abordagens que utilizam das características linguísticas presentes e outra nas categorias de rede.

2.2.1 Abordagens Linguísticas

Uma vez que muitas das pessoas que tentam criar uma notícia falsa ou falar uma mentira utilizam de um vocabulário de forma estratégica para que não sejam identificadas, alguns dos aspectos presentes em textos com conteúdos enganosos passam desapercebidos pelo autor. Sendo assim, o trabalho das abordagens linguísticas é identificar estas dicas presentes no texto de forma a classificar a veracidade do texto (ZHOU et al., 2004; CON-

ROY; RUBIN; CHEN, 2015). Como muitas técnicas podem ser utilizadas na abordagem linguística, os autores dividiram em cinco subcategorias, que são explicitadas a seguir:

- Representação de dados: a ideia deste método é utilizar formas de representação dos textos como *bag-of-words*, n-grams ou palavras individuais. A frequência de utilização de palavras pode ser utilizada como possível distinção do que é verídico para o que é falacioso nesta abordagem. Apesar de bem simples, algumas das abordagens da literatura conseguiram utilizar bem este método em combinação com outras análises complementares;

- Análise de sintaxe: a análise das palavras frequentemente não é suficiente para prever textos enganosos, por isto análises mais complexas através da sintaxe podem ser utilizadas. Através dessa abordagem é possível criar gramáticas probabilísticas que utilizam a estrutura sintática;

- Análise semântica: a ideia principal desta abordagem é utilizar o contexto e algumas nuances do texto, para que possam ser comparadas com outros textos que abordam o mesmo assunto. Esta abordagem porém apresenta algumas limitações, tais como a necessidade de associar corretamente as referências entre os textos, além da necessidade de uma grande quantidade de conteúdo do mesmo assunto para que seja eficiente;

- Estrutura retórica e análise de discurso: com abordagens deste tipo, visa-se utilizar características presentes nas relações entre os elementos linguísticos. A estratégia principal desta abordagem é utilizar a relação entre os elementos linguísticos, baseando-se na diferença em termos de coerência e estrutura. Uma das maneiras de fazer isto é utilizando um modelo espacial de vetor que dispõe cada parte de um tópico no vetor e utiliza a distância até uma *ground truth* como forma a indicar a possibilidade de uma mensagem ser falsa;

- Classificadores: as abordagens por meio de classificadores utilizam de técnicas de aprendizado de máquina, tais como *Support Vector Machines* (SVM) e Naïve Bayes, para criar um modelo matemático que associa alguns atributos com a chance de

veracidade. Com este tipo de abordagem é necessário alguns exemplos prévios para treinamento do modelo e, a partir disto, estes métodos conseguem ter sucesso em sua classificação.

As abordagens linguísticas normalmente são de execução bem rápida, o que faz com que elas possam ser utilizadas em conjunto. Mais de um tipo de abordagem linguística pode ser utilizado, bem como abordagens linguísticas podem ser utilizadas para auxiliar em abordagens de redes.

2.2.2 Abordagens de Redes

Além das abordagens que utilizam métodos linguísticos também existem os subgrupos das que utilizam métodos baseados em redes como estratégia. Este tipo pode ser cada vez mais utilizadas graças ao uso das estratégias em ambientes como redes sociais. As subcategorias deste grupo são explicitadas a seguir:

- Dados ligados: os dados ligados são estruturas capazes de fornecer significado semântico para certas afirmações. Desta forma esse tipo de abordagem consegue ser bastante útil na checagem de fatos. Métodos que utilizam desta estratégia realizam buscas em bases de conhecimento e ligam partes da afirmação tentando estabelecer relações entre os conceitos representados por ela. Uma abordagem possível é encontrar o menor caminho entre as afirmações, sendo que caminhos mais curtos entre os nós possuem uma probabilidade maior de veracidade. Um exemplo desta abordagem pode ser visto na Figura 2.1. Neste exemplo é buscado o caminho mais curto em um grafo de conhecimento para verificar a veracidade da afirmação de que Obama é muçulmano;

- Comportamento em redes sociais: outra estratégia possível para detectar uma falácia em redes sociais é a utilização do comportamento dos usuários em uma rede social, este tipo de abordagem utiliza características como a propagação das mensagens, a interação dos usuários com uma determinada postagem.

Figura 2.1: Exemplo de aplicação de uma técnica usando dados ligados (CONROY; RUBIN; CHEN, 2015).

2.3 Atributos Utilizados Para Caracterização das *Fake News*

A grande maioria dos trabalhos que visa classificar uma notícia ou rumor como *fake news* utiliza de uma série de características presentes na mensagem ou na publicação como forma de identificar a veracidade. Com isto, a partir dos atributos utilizados pelos trabalhos, pode-se separar em algumas categorias. O primeiro trabalho a classificar as diferentes categorias foi Castillo, Mendoza e Poblete (2011), sendo as características definidas neste trabalho listadas a seguir:

- Atributos em nível de mensagem: as técnicas que utilizam destes atributos consideram as características da mensagem para determinar a sua veracidade, estas podendo ser baseadas na rede social abordada, como no caso do Twitter a utilização de *hashtags*, ou independentes da rede social em que é enquadrado o número de palavras, o conteúdo das mensagens, tamanho do textos, dentre outras características;

- Atributos em nível de usuário: estas características relacionam-se com o autor do texto que está sendo analisado, observando idade, sexo, número de seguidores (para redes sociais como Twitter), e até uma classificação baseada na relevância deste autor para a comunidade;

- Atributos em nível de tópico: este atributo visa classificar abordagens que aproveitam de caracterizações presentes em algum contexto. Um exemplo disto é a utilização de dados anteriores de sentimento e *hashtags* presentes no contexto de um determinado assunto para verificar a veracidade da notícia como um todo;

- Atributos em nível de propagação: A propagação é uma propriedade que pode ser utilizada principalmente para análise de redes sociais. Esta classificação serve para denominar técnicas que fazem o emprego de características como número de interações com uma mensagem, bem como a estrutura de rede inerente de estruturas sociais.

Cada um destes atributos pode auxiliar em alguma das partes da análise de credibilidade de uma notícia, por isto é comum a utilização de abordagens híbridas para melhores resultados. Esta estratégia é inclusive utilizada no artigo que introduz a separação das características utilizando estes termos (CASTILLO; MENDOZA; POBLETE, 2011).

2.4 Trabalhos Relacionados

Nesta seção são destacados alguns dos trabalhos que estudam temas relacionados a detecção de *fake news*, bem como suas contribuições para a literatura. Para isto, foram selecionados alguns *surveys* contendo informações sobre o assunto.

Uma das principais divergências entre os trabalhos é perante a definição do escopo estudado, enquanto (CONROY; RUBIN; CHEN, 2015; SHU et al., 2017; PARIKH; ATREY, 2018) estudam métodos para detecção automática de *fake news*, outros autores, como Zubiaga et al. (2018), abordam todo o aspecto de rumor e tratando o problema de detecção *fake news* como uma resolução do rumor, ou seja classificar sua veracidade. Também existem abordagens de cunho semelhantes que foram aplicadas antes do termo ser definido, como em (SHARIFF; ZHANG, 2014), em que os autores buscam caracterizar os principais modos que o consumidor de uma notícia pode ser enganado por meio de redes sociais, incluindo *spams* e críticas. Porém, mesmo com estas divergências, grande parte do que é abordado nestes trabalhos se assemelha com os problemas de detecção de *fake news*, por isto as abordagens que podem ser utilizadas são semelhantes.

Pode ser notado nos artigos que fazem algum tipo de levantamento literário um consenso de que existe uma grande necessidade de bases de dados melhores e mais expressivas do que as existentes para que seja possível um maior desenvolvimento em detecção de falácias, sendo este problema relatado em (PARIKH; ATREY, 2018; CONROY; RUBIN; CHEN, 2015; VIVIANI; PASI, 2017; ZUBIAGA et al., 2018; SHU et al., 2017). Em (PARIKH; ATREY, 2018; CONROY; RUBIN; CHEN, 2015; ZUBIAGA et al., 2018) também é sugerida a utilização de abordagens que não dependam apenas de um tipo de característica de uma determinada postagem, assim indicando a necessidade de abordagens híbridas nesta área.

Alguns destes trabalhos apresentam proposições de *frameworks* que podem ser utilizados para abordar o problema. Desta forma, utilizam artigos que propõem soluções para cada etapa deste *framework*, não contemplando assim todas as abordagens da área no mesmo trabalho. Em (SHU et al., 2017), são abordados apenas artigos que utilizam mineração de dados para solucionar o problema, já em (SHARIFF; ZHANG, 2014; VIVIANI; PASI, 2017) um foco maior é dado em relação a credibilidade do autor da notícia. Em (ZUBIAGA et al., 2018) o foco principal do trabalho é analisar a detecção de rumor, com isto o texto não faz uma ampla abordagem em detecção de *fake news* e (CONROY; RUBIN; CHEN, 2015) como é um dos trabalhos precursores da área, não realiza uma análise dos trabalhos para encontrar *fake news*, mas sim demonstra quais as técnicas que

podem ser utilizadas para isto. Desta forma, este trabalho se diferencia dos trabalhos prévios realizando exclusivamente uma análise dos métodos propostos para a detecção de *fake news* e fazendo isto utilizando uma abordagem sistemática, como o objetivo de tentar encontrar os trabalhos de forma mais imparcial. Este trabalho também faz um levantamento quantitativo de uma série de características dos trabalhos encontrados, possibilitando uma visão geral do que tem sido feito na área e os locais em que estes trabalhos são apresentados, de forma a facilitar um primeiro contato com este tema, além de ajudar pesquisadores em relação a escolha de meios para publicar sobre a área.

3 Revisão e Mapeamento Sistemáticos

O mapeamento e a revisão da literatura possibilitam que seja feita um levantamento dos trabalhos existentes de um determinado assunto de forma mais imparcial. Com isto, os resultados podem servir como ponto de partida para interessados na área, bem como diminuir esforços na busca de algumas informações por pesquisadores (KITCHENHAM; BUDGEN; BRERETON, 2011).

No contexto dessa pesquisa, tem-se como intenção encontrar os principais estudos realizados na área de detecção automática de *fake news* em redes sociais, bem como avaliar e interpretar os estudos disponíveis na literatura de acordo com as questões de interesse levantadas. As etapas de mapeamento e revisão sistemática descritas neste trabalho são baseadas nas instruções descritas em (NEIVA; SILVA, 2016; BUDGEN; BRERETON, 2006).

3.1 Planejamento da Revisão e Mapeamento Sistemático

Durante o planejamento da atividade descrita nesta monografia foram definidos os objetivos e um protocolo para execução do método para revisão, de forma a diminuir a parcialidade do procedimento. Com isto, todos os resultados encontrados devem ser reproduzíveis e são descritos nesta seção do trabalho.

3.1.1 Questões de Pesquisa

O mapeamento sistemático proposto visa responder as questões descritas a seguir:

- MQ1: Quantos estudos foram publicados ao longo dos anos?
- MQ2: Quais são os autores mais ativos na área?
- MQ3: Em quais veículos de publicação são publicados os trabalhos da área?
- MQ4: Em quais domínios as pesquisas são utilizadas? (e.g. pós-desastre, política, propósito geral)

- MQ5: Quais países são os mais ativos na área?
- MQ6: Em quais redes sociais as pesquisas são aplicadas? (e.g. Twitter, Facebook, Sina Weibo)
- MQ7: Quais são os atributos mais utilizados na detecção de *fake news*?
- MQ8: Quais abordagens são utilizadas na detecção de *fake news*?
- MQ9: Quais são os métodos utilizados para suporte da decisão (e.g. Decision Tree, SVM, Random Forest)?
- MQ10: Quais são as bases de dados mais utilizadas para validar as pesquisas?

Os objetivos da revisão sistemática são responder as questões a seguir:

- RQ1: Quais métodos têm sido utilizados para detectar *fake news* em redes sociais?
- RQ2: Quais práticas podem ser utilizadas para melhorar a detecção de *fake news* em redes socais?

Baseadas nas questões de pesquisa e mapeamento, foi utilizando o método PICOC proposto por Petticrew e Roberts (2006), que possui como propósito definir um escopo para uma questão de pesquisa. Para o escopo deste trabalho o PICOC desenvolvido se encontra na Tabela 3.1.

	Descrição
Population (P)	Soluções que identificam informações falsas
Intervention (I)	Comportamentos analisados pela solução
Comparison(C)	Não se aplica
Outcome (O)	Soluções
Context (C)	Redes Sociais

Tabela 3.1: Definição do PICOC para o escopo desta monografia.

3.1.2 Critérios de Exclusão

Alguns critérios de exclusão tiveram que ser levados em conta na execução desta pesquisa. Os mesmos são listados a seguir:

- EC1: O artigo não apresenta um método ou solução automática para detecção de *fake news*;
- EC2: O artigo não apresenta um método autoral para solução do problema (e.g. survey ou reviews);
- EC3: O artigo não está escrito em inglês;
- EC4: O artigo não apresenta uma solução aplicável em redes sociais.

Após a definição das questões de pesquisa e dos critérios de exclusão relacionados a mesma, foram tomados os seguintes passos para definições das bases em que a pesquisa seria realizada, baseado em (COSTA; MURTA, 2013). Os requerimentos para a pesquisa foram:

- As bases são capazes de executar pesquisas utilizando expressões lógicas ou mecanismos similares;
- Elas permitem que as pesquisas sejam feitas de modo a contemplar todo o texto ou apenas campos específicos (e.g. título, *abstract*);
- Elas devem estar disponíveis na instituição do pesquisador;
- Elas devem cobrir a área de pesquisa de interesse, no caso Ciência da Computação.

Além destes critérios, foi levado em conta que a base não oferecesse uma limitação no número de palavras da *string* de busca inferior a utilizada. Com isto as bases escolhidas para realizar a busca são as citadas a seguir:

Bases de artigos científicos	Link de acesso
Scopus	http://www.scopus.com
EI Compendex	http://www.engineeringvillage.com
ISI Web of Science	http://www.isiknowledge.com

Tabela 3.2: Bases utilizadas para executar a busca.

3.1.3 *String* de Busca

Para a criação da *string* de busca, foram levadas em considerações os termos mais relevantes relacionados ao PICOC levantado, bem como seus sinônimos. Para tal, foram consultados alguns *surveys* e uma lista de artigos encontrados a partir de uma busca comum.

A partir dos termos encontrados, uma busca lógica foi definida utilizando os termos definidos para cada parte do PICOC separados por "AND"e entre si por "OR", uma série de estudos que deveriam ser retornados pela *string* de busca também foi definida, para que sua validação pudesse ser feita, como é sugerido em (ZHANG; BABAR, 2010). Após a realização das pesquisas e inclusão de novas palavras chaves relacionadas ao assunto, a *string* de Busca final obtida foi:

(deception OR "fake news"OR "fake information"OR rumor OR rumour OR misinformation OR disinformation OR hoax) AND (detection OR detect OR spread OR analysis OR classify OR predict OR prediction) AND (method OR benchmark OR technique OR algorithm OR approach OR application OR system OR automatic) AND ("social network"OR news OR "social media"OR "online communication"OR tweet)

Para englobar somente o contexto de interesse, também foram utilizadas as configurações das bases de artigos científicos para possibilitar a busca apenas de artigos da área de Ciência da Computação.

Alguns artigos previamente encontrados foram definidos como artigos de interesse (CONROY; RUBIN; CHEN, 2015; TACCHINI et al., 2017; ZUBIAGA et al., 2018). Sendo estes utilizados para controlar se a *string* de busca estava encontrando os estudos relevantes da área.

3.1.4 Condução da Revisão e Mapeamento Sistemático

Para a condução do trabalho foi executada a *string* de busca nas bases de artigos científicos selecionadas. Após isto o estudo dos resultados foi conduzido da seguinte forma:

1. Os resultados retornados foram inseridos na ferramenta Parsifal, na qual os artigos duplicados foram detectados e removidos;

2. Foi realizada uma primeira etapa de exclusão baseada no título e *abstract* dos artigos, em que, considerando os critérios de exclusão, foram removidos os artigos que não possuíam relevância para este mapeamento. Os artigos nos quais existia dúvida da relevância ou foram considerados relevantes passaram para um novo filtro em que foi conduzida uma análise baseada na introdução e conclusão. A partir do resultado deste segundo filtro, as pesquisas restantes foram analisadas;

3. Os artigos foram lidos integralmente e analisados perante as questões de mapeamento e pesquisa. Nesta etapa além dos critérios de exclusão também foi levada em conta a qualidade do artigo perante as questões levantadas, de forma a excluir os artigos que não possuíam respostas para as questões de mapeamento e revisão.

A partir das etapas definidas, o procedimento foi realizado e os resultados obtidos podem ser visualizados na Figura 3.1. Para obter uma maior idoneidade dos resultados da pesquisa, as etapas foram realizadas em par. Com as divergências encontradas entre os resultados discutidas em encontros e com apoio dos orientadores do trabalho.

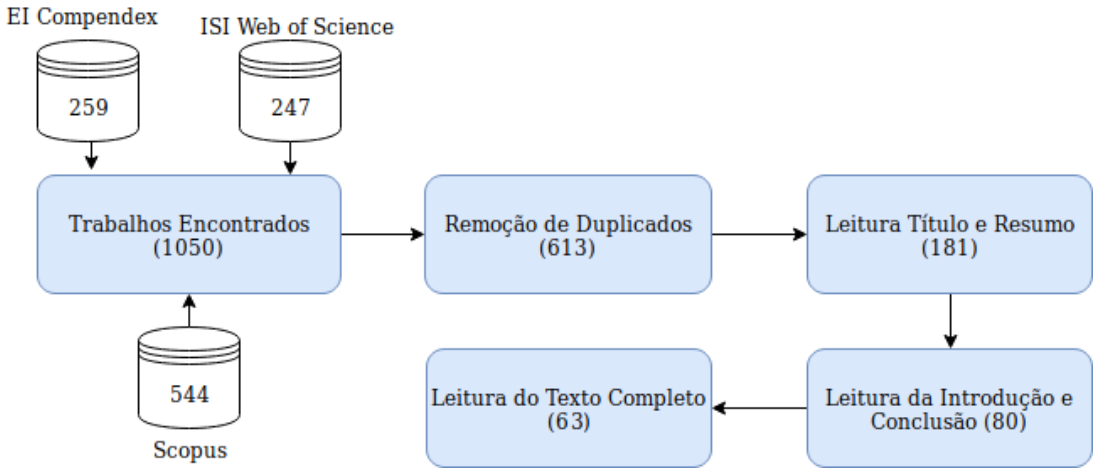

Figura 3.1: Processo de seleção de trabalhos.

Na primeira etapa foram obtidos 1050 artigos através do conjunto das três bases de artigos científicos, com a distribuição demonstrada na Figura 3.2, sendo este procedimento realizado no dia 09/08/2018, sendo que deste conjunto 473 (45,04%) foram removidos por serem duplicados. Os 577 artigos restantes foram analisados com uma leitura de título e resumo, em que 396 (68,6%) foram excluídos e 181 (31,4%) foram mantidos para análise posterior. Dos 181 artigos selecionados na segunda fase foram lidas a

introdução e conclusão. Ao final desta etapa, 80 (44,2%) artigos restaram baseados nos critérios de exclusão. Estes artigos selecionados correspondem a 7,6% dos artigos inicialmente selecionados. O baixo número de artigos selecionados perante o resultado total obtido pela *string* de busca, ocorreu devido a definição do escopo da pesquisa para uma abordagem com foco restrito na detecção de *fake news*. Sendo assim, apesar da utilização do termo *rumor* na *string* de busca, os artigos que não distinguiam os rumores verídicos de falaciosos não se enquadram no propósito deste trabalho e por isto foram excluídos.

Após realizar a leitura do texto completo, já utilizando as questões de mapeamento e revisão como critérios de qualidade, foram encontrados 63 artigos (6% dos artigos iniciais) que respondiam todas ou quase todas as questões levantadas.

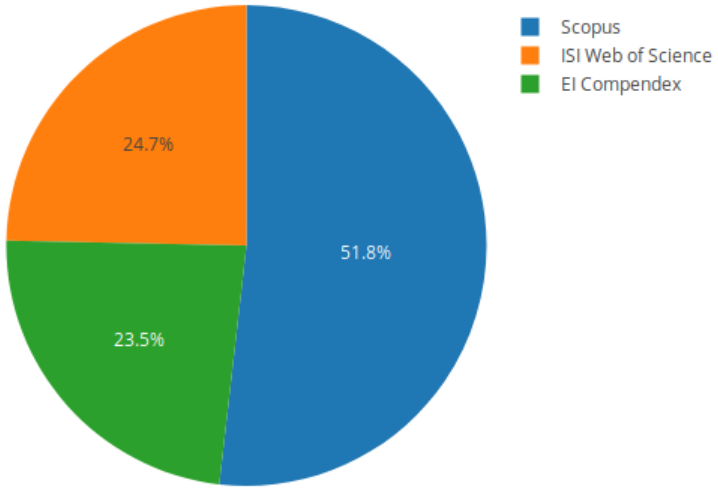

Figura 3.2: Distribuição do conjunto de artigos por fonte de origem.

4 Resultados da Revisão e Mapeamento Sistemáticos

Dos artigos inicialmente encontrados, foram selecionados 63 artigos, estes foram analisados e as suas principais características foram extraídas de acordo com as questões levantadas. Neste capítulo as questões serão divididas em seções em que cada uma delas será discutida separadamente, bem como as considerações sobre os resultados e ameaça a validade.

4.1 Respostas às questões do mapeamento

Nesta seção são demonstrados os dados quantitativos relativos às questões de mapeamentos levantadas na Subseção 3.1.1. Além de uma análise nos resultados encontrados a partir deste mapeamento.

4.1.1 MQ1: Quantos estudos foram publicados ao longo dos anos?

A primeira publicação relevante que aborda detecção de notícias falaciosas em redes sociais é de 2011, data que precede a utilização do termo *fake news*. Desde então é possível perceber um aumento significativo de pesquisas no assunto, como pode ser notado na Figura 4.1. A queda do número de artigos em 2018, possivelmente, não indica um desinteresse na área em questão, uma vez que nem todos os artigos publicados no decorrer deste ano puderam ser indexados para a pesquisa, que foi realizada no dia 9 de agosto de 2018 e os resultados podem ser conferidos através do seguinte link ⟨https://github.com/Fernandoms/fake_news_literature_review⟩.

4.1.2 MQ2: Quais são os autores mais ativos na área?

Poucos autores possuem mais de uma publicação que enquadra nos requisitos desta pesquisa. Dentre os 220 autores com publicações nos resultados, apenas 11 foram responsáveis

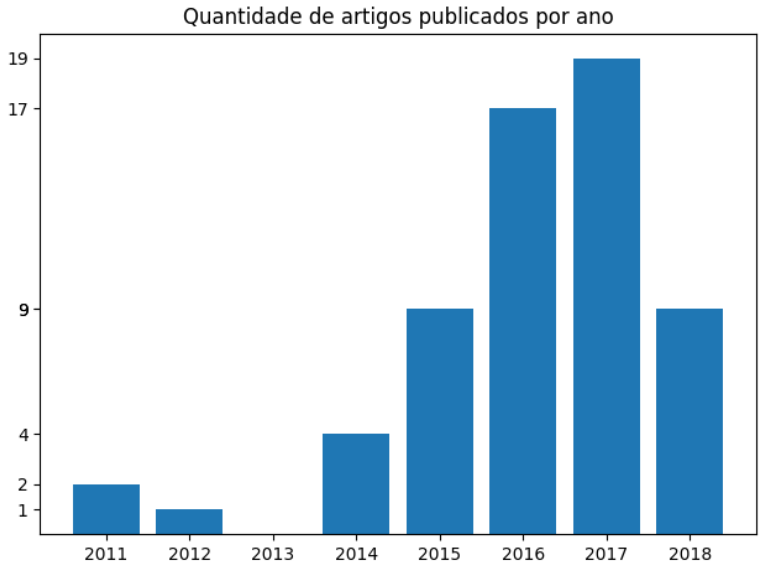

Figura 4.1: Quantidade de artigos publicados por ano.

pela publicação de mais de um artigo, sendo "Zhang, Y"o único a publicar 3 artigos (ZHANG et al., 2016; CHANG et al., 2016; JIN et al., 2017). Um grafo para demonstrar como são distribuídas as co-autorias foi gerado utilizando o algoritmo de Fruchterman Reingol (FRUCHTERMAN; REINGOLD, 1991), o mesmo pode ser visualizado na Figura 4.2, em que as linhas mais escuras representam relações entre autores que publicaram mais de um trabalho em conjunto.

Uma característica que pode ser notada é a falta de interação entre os grupos de pesquisadores da área, o que poderia ser benéfico para este campo de pesquisa, como pode ser visualizado no grafo de coautoria, a maioria dos autores realizou trabalhos isolados, com raras exceções em que autores participaram de trabalhos com diferentes grupos de pesquisa.

4.1.3 MQ3: Em quais veículos de publicação são publicados os trabalhos da área?

Os trabalhos na área são em sua maioria publicados em conferências, como pode ser visto na Figura 4.3, sendo que a maioria destas conferências aconteceram nos Estados Unidos, sendo a distribuição apontada na Figura 4.4.

Quanto ao assunto das conferências, foi notado que devido a amplitude de modos de se trabalhar com detecção de *fake news*, os artigos são publicados em diversas áreas

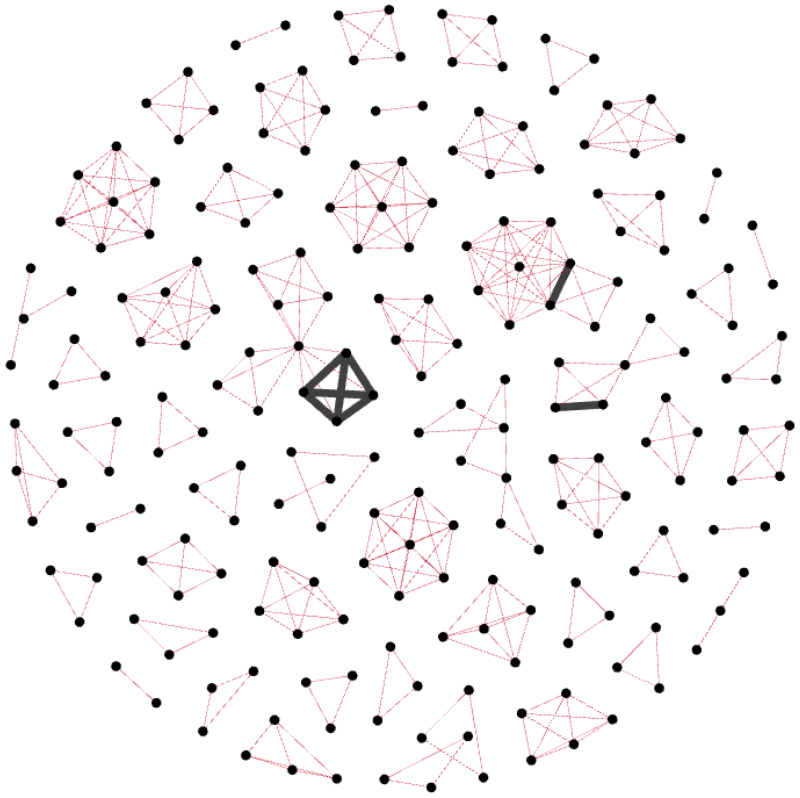

Figura 4.2: Grafo de coautoria.

da computação. As conferências com maior número de publicações foram: International Conference On Big Data e International Conference On Social Informatics, ambas com 3 artigos publicados, sendo que os artigos foram publicados em um total de 40 conferências diferentes. As conferências que possuem mais de um artigo publicado podem ser vistas na Tabela 4.1.

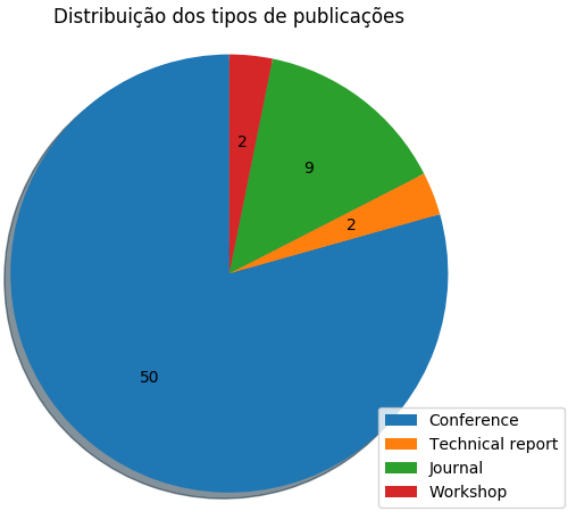

Figura 4.3: Meio de publicação dos artigos.

Nome da Conferência	**Publicações**
International Conference On Big Data	3
International Conference On Social Informatics	3
Lecture Notes In Computer Science	2
International Conference On Natural Computation, Fuzzy Systems And Knowledge Discovery	2
International Conference On Information And Knowledge Management	2
International Conference On World Wide Web	2
International Conference On Data Mining	2
International Conference On Advances In Social Networks Analysis And Mining	2

Tabela 4.1: Relação das conferências com o número de publicações.

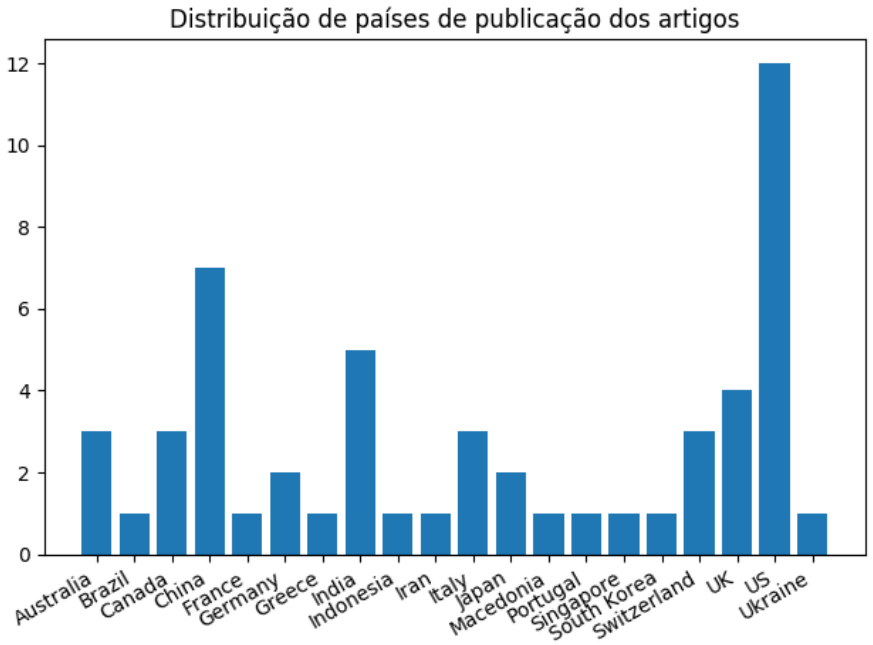

Figura 4.4: Países em que as conferências foram realizadas.

4.1.4 MQ4: Em quais domínios as pesquisas são utilizadas?

Foi notado que a detecção de *fake news* pode ser utilizada de modo generalizado, ou seja, podendo classificar qualquer informação veiculada em redes sociais, ou aplicações para

um domínio específico. Os domínios, bem como o número de trabalhos associados podem ser vistos na Figura 4.5.

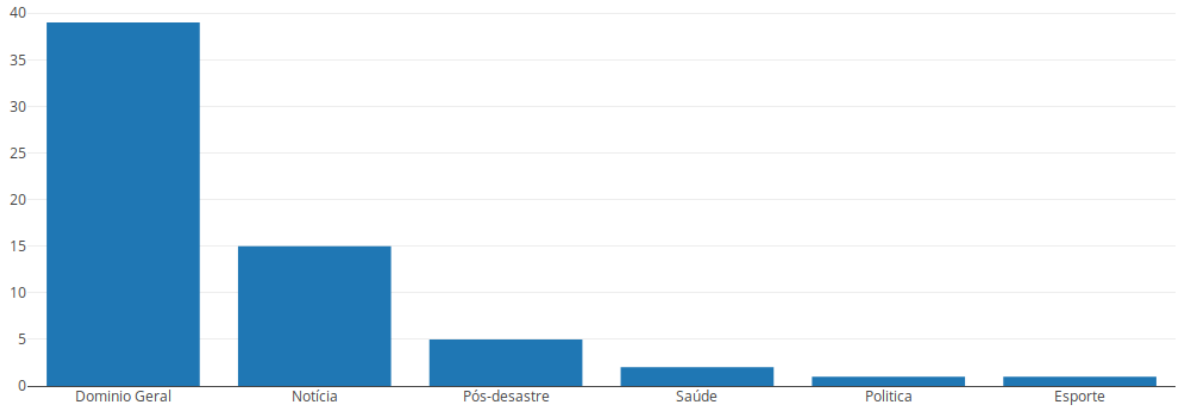

Figura 4.5: Domínio em que as pesquisas são aplicadas.

Nos artigos selecionados, foram encontrados oito diferentes domínios em que as pesquisas são utilizadas. Os domínios são:

- Domínio Geral: nesta categoria enquadram as pesquisas que não realizam nenhuma distinção do conjunto de dados, podendo, segundo os autores, ser aplicada em qualquer publicação realizada através de redes sociais;

- Notícias: os trabalhos classificados neste grupo possuem sua abordagem delimitada ao conjunto de notícias jornalisticas. Este tipo de abordagem pode levar em consideração a fonte do dado como um fator relevante, bem como informações veiculadas em sites tidos como confiáveis, por conta disto, esse tipo de abordagem não pode ser utilizado em todos os contextos;

- Cenários Pós-desastre: estas abordagens propõe soluções que consigam verificar a veracidade de uma informação de maneira rápida e de modo que qualquer usuário da rede social pode ser um provedor de informação. A partir dos trabalhos observados, foi notado que os autores com o foco neste tipo de cenário utilizam informações de geolocalização e sentimento em sua abordagem;

- Saúde: dois dos trabalhos (GUO et al., 2015; SICILIA et al., 2018) fizeram abordagens para o domínio específico de *posts* sobre saúde. Nestes trabalhos foi notado a utilização de informações disponíveis em bases médicas como *ground truth*;

- Político: apenas um dos artigos (CHANG et al., 2016) se enquadrou nesta categoria. A especificidade do domínio para o contexto apenas politico fez com que este artigo utilizasse de características como a polarização de uma eleição para determinar a possível veracidade da publicação. Um dos pontos diferenciais abordados por este trabalho é a identificação de usuários extremistas e utilização disto como um dos fatores;

- Esportes: um artigo (JANSSEN; HABIB; KEULEN, 2017) realizou a pesquisa no domínio específico de esportes, mais precisamente no contexto da Copa do Mundo de 2014. A abordagem deste trabalho se assemelha a utilizada no contexto de Saúde, porém com a *ground truth* advinda de uma base de dados especifica para futebol e o site da FIFA.

Apesar da maioria das pesquisas descreverem abordagens sendo utilizadas em um domínio generalizado, algumas diferenciações podem ser notadas em pesquisas de domínios específicos. Um exemplo disto são as pesquisas com foco em pós-desastre que possuem uma abordagem com ênfase maior na análise de sentimento e posicionamento geográfico do usuário. Também pode ser notado que é possível delimitar um contexto específico para que sejam utilizados dados disponíveis para melhores resultados.

4.1.5 MQ5: Quais países são os mais ativos na área?

Para responder esta questão, foi levado em conta quais os países das instituições dos autores de cada trabalho. Caso o trabalho possuísse autores de mais de um país, todos eles eram incluídos. Com isto foi possível mapear quais os países em que as instituições possuem mais trabalhos voltados para o tema.

A partir desta abordagem, foi possível notar que os Estados Unidos (20 trabalhos publicados) e China (17 trabalhos publicados) lideram os esforços em pesquisas relacionadas ao assunto. A lista completa dos países bem como o número de trabalhos publicados pode ser vista na Tabela 4.2.

País	Publicações
Estados Unidos	20
China	17
Índia	5
Alemanha, Cingapura	4
Catar, Itália, Japão, Reino Unido	3
Austrália, Chile, Grécia, Holanda, Suíça	2
Arábia Saudita, Bulgária, Egito, Espanha, Hungria, Irã, Paquistão, Portugal, Turquia, Ucrânia	1

Tabela 4.2: Relação de publicações por país dos autores.

Também foi observada a relação do país de publicação com país de autoria do trabalho, para tal levantamento, foi considerado quando ao menos um dos autores trabalha em alguma instituição do mesmo país que o trabalho foi publicado, a partir disto chegamos nos valores apresentados na Tabela 4.3. A partir deste levantamento é possível que todos os trabalhos publicados nos Estados Unidos tem pelo menos um autor atuante no país. O mesmo ocorre em para Alemanha, Irã e Ucrânia, porém estes países possuem apenas um trabalho publicado.

Países	Publicações
Estados Unidos	12
China	5
Índia	4
Alemanha, Austrália, Irã, Japão, Reino Unido, Ucrânia	1

Tabela 4.3: Número de trabalhos por autores do país que o trabalho foi publicado.

4.1.6 MQ6: Em quais redes sociais as pesquisas são aplicadas?

A grande maioria dos trabalhos escolheu o Twitter[2] para o desenvolvimento dos trabalhos, seguido pelo Sina Weibo[3] que é um microblog chinês. Os trabalhos e utilizações podem

[2]⟨https://twitter.com/⟩
[3]⟨https://www.weibo.com/⟩

ser vistas na Tabela 4.4.

Grande parte dos trabalhos que utilizaram o Twitter e o Sina Weibo, justificaram a escolha devido a presença de APIs que facilitam a utilização dos dados presentes nessas redes. Foi notado o baixo número de trabalhos relacionados as principais redes sociais utilizadas e até a ausência de uma das redes sociais mais utilizadas no Brasil e no mundo o WhatsApp[4] (Statista Inc., 2018), apesar de alguns trabalhos terem sido desenvolvidos para o monitoramento de grupos na rede social[5], nenhum trabalho relacionado à detecção automática de *fake news* nesta plataforma foi encontrada.

Rede Social	Número de artigos
Twitter	47
Sina Weibo	14
Facebook	3
BuzzFeed	2
Google Plus	1
YouTube	1

Tabela 4.4: Relação das redes sociais com o número de trabalhos.

4.1.7 MQ7: Quais são os atributos mais utilizados na detecção de *fake news*?

Como descrito na Seção 2.3 alguns atributos são comumente utilizados para a detecção de *fake news* pelos trabalhos. Com isto, esta parte do mapeamento visa encontrar quais destes atributos eram os mais utilizados, e como os mesmos são utilizados em conjunto por mais de uma abordagem.

A maioria dos trabalhos utilizou de informações presentes na mensagem, sendo este atributo considerado por 52 (82,5%) trabalhos, seguido pela utilização de informações do usuário com 38 (60,3%), propagação com 28 (44,4%) e tópico com 9 (14,3%).

Destes trabalhos, apenas 3 (4.8%) utilizaram todos os atributos (CASTILLO; MENDOZA; POBLETE, 2011; CASTILLO; MENDOZA; POBLETE, 2013; MA et al., 2015), 15 (23,8%) utilizaram 3 atributos, 25 (39,7%) utilizaram 2 e 20 (31,7%) apenas 1. A relação completa dos trabalhos e atributos utilizados pode ser vista na Tabela A.1

[4]⟨https://www.whatsapp.com/⟩
[5]⟨http://www.monitor-de-whatsapp.dcc.ufmg.br/⟩

(Apêndice A).

Como pode ser notado, há um número bastante limitado de trabalhos que se baseiam no tópico da postagem como critério para abordagem. Isto pode ser explicado devido à falta de bases de dados que seriam capazes de fazer esta uma abordagem mais precisa.

Outra análise interessante que pôde ser realizada a partir das informações deste mapeamento é em relação a maior utilização de características presentes no texto das postagens em detrimento de informações presentes nas redes sociais, pois apesar do escopo deste trabalho ter englobado apenas métodos que atuam na detecção em redes sociais, doze dos artigos desconsideraram estas informações na efetuação das análises, ou seja 60% das abordagens que utilizaram apenas um atributo.

4.1.8 MQ8: Quais abordagens são utilizadas na detecção de *fake news*?

Nesta questão de mapeamento, a intenção é de responder baseado nas abordagens descritas pela Seção 2.2, quais delas são mais utilizadas pelos trabalhos identificados.

Primeiro é importante ressaltar que alguns dos trabalhos utilizaram abordagens híbridas, ou seja estão presentes mais de um dos métodos descritos para encontrar o resultado. A utilização de abordagens combinadas foi notada em 20 (31.7%) trabalhos, sendo que em apenas 3 trabalhos foram utilizadas mais de 2 das abordagens descritas (SABBEH; BAATWAH, 2018; ZHANG et al., 2015; GUO et al., 2015).

A utilização de cada abordagem pode ser visualizada na Tabela 4.5, na qual é possível notar uma predominância dos métodos linguísticos em relação à abordagens de redes.

Abordagem	Nº de Publicações
Aprendizagem supervisionada	43
Comportamento em Redes Sociais	16
Análise de sintaxe	7
Estrutura Retórica e Análise de Discurso	7
Representação de Dados	5
Dados Ligados	4
Análise Semântica	3

Tabela 4.5: Abordagens utilizadas e número de trabalhos.

4.1.9 MQ9: Quais são os métodos utilizados para suporte da decisão?

Para esta etapa foram considerados o emprego de técnicas de que apoiam as classificações e decisões realizadas pela abordagem do autor de forma automática ou semi-automática.

A maioria dos trabalhos (68,2%) apoiou em uma ou mais técnica de aprendizado de máquina. Na Tabela 4.6 são demonstrados os métodos que foram utilizados por mais de um artigo, bem como o número de artigos que utilizaram.

Abordagem Utilizada	Número de Trabalhos
Support Vector Machine	19
Random Forest	11
Naive Bayes	10
Neural Network	10
Decision Tree	8
Logic Regression	7

Tabela 4.6: Relação das abordagens utilizadas com o número de trabalhos a utiliza-las.

Uma característica em comum pode ser notada destes principais métodos para suporte de decisão, pois apesar de estratégias diferentes, todos pertencem à categoria de métodos de aprendizado supervisionado, ou seja necessitam de uma base de treinamento para que sejam capazes de generalizar as situações encontradas e aplicar nos demais casos.

4.1.10 MQ10: Quais são as bases de dados mais utilizadas para validar as pesquisas?

Esta pesquisa demonstrou uma baixa utilização de bases de dados existentes na literatura, a publicação de bases de dados mais específicas para este problema é um fator que deve ser levado em consideração, outro fator é a diversidade de informações que podem ser utilizadas para as abordagens e com isto para que sua abordagem seja testada os autores acabam tendo que realizar uma curadoria manual, com isto não utilizando bases de dados existentes na literatura.

A maioria dos pesquisadores optou por realizar a coleta e processamento de dados de forma autônoma. Outro ponto que foi possível notar é que a maioria dos trabalhos que realiza esta coleta não disponibiliza estes dados de maneira fácil. A maioria dos trabalhos relacionados, como pode ser visto na Seção 2.4, descrevem esta ausência de bases únicas como um dos problemas que impedem o melhor desenvolvimento da área de pesquisa como um todo.

As bases de dados presentes nos trabalhos encontrados que foram disponibilizadas de forma fácil podem ser vistas na Tabela 4.7, bem como os artigos que as utilizaram.

Base Utilizada	Trabalho
(CASTILLO; MENDOZA; POBLETE, 2011)	(YU et al., 2017) (LIM; LEE; HSU, 2016) (MA et al., 2015) (GUPTA; ZHAO; HAN, 2012) (CASTILLO; MENDOZA; POBLETE, 2011)
PHEME [6]	(TOLOSI; TAGAREV; GEORGIEV, 2016) (LENDVAI; REICHEL; DECLERCK, 2016) (BUNTAIN; GOLBECK, 2017) (ZUBIAGA; LIAKATA; PROCTER, 2017)
VMU @ Media Eval [7]	(BOIDIDOU et al., 2018) (AGRAWAL; GUPTA; NARAYANAN, 2017)
(ZUBIAGA; JI, 2014)	(ANTONIADIS; LITOU; KALOGERAKI, 2015)
(KWON et al., 2013)	(YU et al., 2017)
(MA et al., 2016)	(YU et al., 2017)
(WANG, 2017)	(BHATTACHARJEE; TALUKDER; BALANTRAPU, 2017)
SemEval [8]	(YAVARY; SAJEDI, 2018)

Tabela 4.7: Relação das bases de dados utilizadas e trabalhos que as utilizaram.

4.2 Respostas às questões de revisão

Nesta seção são realizadas as discussões das informações encontradas nos artigos encontrados em relação às questões de revisão levantadas na Subseção 3.1.1.

[6] ⟨https://www.pheme.eu/⟩
[7] ⟨http://www.multimediaeval.org/⟩
[8] ⟨http://alt.qcri.org/semeval2017/task8/index.php?id=data-and-tools⟩

4.2.1 RQ1: Quais métodos têm sido utilizados para detectar *fake news* em redes sociais?

A partir do levantamento desta pesquisa, foi notado que os mais diversos métodos que vem sendo utilizados para detectar *fake news* em redes sociais, porém a maioria dos métodos utiliza de alguma forma de classificação para encontrar este resultado. Trabalhos como (ANTONIADIS; LITOU; KALOGERAKI, 2015; LIANG; YANG; XU, 2016; BUNTAIN; GOLBECK, 2017; CASTILLO; MENDOZA; POBLETE, 2013; PAL; CHUA, 2018; MA et al., 2015; RAJDEV; LEE, 2015; CASTILLO; MENDOZA; POBLETE, 2011) baseiam-se em uma série de fatores presentes nas postagens para realizar a classificação da mesma, chegando em até 45 características distintas. Este tipo de abordagem baseia-se em (CASTILLO; MENDOZA; POBLETE, 2013) com algumas alterações como adição de atributos temporais (PAL; CHUA, 2018), informações mais especificas de rede (BUNTAIN; GOLBECK, 2017), dentre outros. A vantagem desta abordagem é a generalidade da implementação, possibilidade de evolução dos resultados com a utilização, além da rapidez para classificar uma instância uma vez que o sistema esteja treinado.

Outro tipo de abordagem que pode ser notada para a detecção de *fake news* é a utilização de informações intrínsecas de redes sociais. Nesta abordagem as postagens e interações entre os usuários são modeladas em forma de grafo, a partir desta modelagem é possível analisar a disseminação do rumor, e com isto atribuir credibilidade para as notícias. Esta abordagem pode ser vista nos seguintes trabalhos (ZHANG et al., 2016; CAI; BI; LIU, 2017; TOLOSI; TAGAREV; GEORGIEV, 2016; HASHISH et al., 2017; ZHANG et al., 2015; KUMAR; GEETHAKUMARI, 2014; BAETH; AKTAS, 2017; LIU et al., 2017; GUPTA; ZHAO; HAN, 2012; XIE et al., 2016; WU et al., 2016; JIN et al., 2014; ZHOU et al., 2015; YAVARY; SAJEDI, 2018; WU; LIU, 2018).

A utilização de dados ligados também foi realizada por alguns trabalhos, para casos específicos como em (GUO et al., 2015) em que a intenção é a detecção de *fake news* na área médica, (JANSSEN; HABIB; KEULEN, 2017) que visa o contexto específico da Copa do Mundo, porém esta abordagem também é utilizada em (ULICNY; KOKAR, 2011) cuja intenção é mais generalizada, sendo que nesta abordagem os *tweets* são transformados em um conjunto de dados ligados para consulta posterior.

Outras abordagens também são utilizadas por uma quantidade menor de artigos como (JIN et al., 2017) que utiliza de imagens presentes nas mensagens para classificar a veracidade e (SABBEH; BAATWAH, 2018; ZHANG et al., 2015; MA; GAO; WONG, 2017; LENDVAI; REICHEL; DECLERCK, 2016) que utilizam a estrutura retórica entre as postagens.

4.2.2 RQ2: Quais práticas podem ser utilizadas para melhorar a detecção de *fake news* em redes socais?

A principal questão levantada pelos artigos encontrados foi a impossibilidade de utilizar uma base de dados pronta, na maioria das vezes isto se dá devido ao modo que as bases de dados presentes na literatura são organizadas. Muitas das abordagens utilizam de diversas das características que podem ser encontradas em postagem da rede social, porém as principais bases de dados disponíveis possuem apenas um conjunto limitado de características, por isto sendo ignorada pela maioria dos pesquisadores.

A maioria dos artigos que realiza a coleta dos dados experimentais também não disponibiliza os mesmos de maneira fácil para que outras abordagens possam ser testadas comparativamente de maneira a ambas possuírem as mesmas condições.

Uma vez que esta área tem se tornado cada vez mais foco de pesquisas, a criação de um *benchmark* em que as mais diversas abordagens possam ser testadas e comparadas é uma prática que pode ajudar consideravelmente este campo de pesquisa. Uma vez que os trabalhos da área não apresentam informações comparativas com outras abordagens, o que facilitaria uma análise de quais abordagens são as melhores para cada contexto.

Outra prática que tem grande capacidade de ajudar a área da detecção de *fake news* em redes sociais é a criação e manutenção de bases de dados que contenham informações atualizadas com *ground truth* que possam ser consultadas para verificação de veracidade. Esta constatação baseia-se no fato de apenas abordagens mais específicas como o caso de saúde, esportes e algumas abordagens de âmbito geral com contexto controlado que utilizaram de bases de dados com anotações como forma de verificação de informação.

4.3 Ameaças à validade

Esta revisão e mapeamento sistemático da literatura é intencionada para identificar, categorizar e analisar os trabalhos que apresentam alguma proposta ou solução para a detecção de *fake news* em redes sociais. Porém como em qualquer método, estão presentes algumas ameaças à validade e limitações a este estudo. Uma das limitações que deve ser levada em conta é em relação ao número de bibliotecas digitais que foram consideradas para a pesquisa. Como apenas 3 permitiam que a pesquisa fosse realizada de maneira a incluir as principais palavras chaves e seus sinônimos, algumas bases de dados foram desconsideradas, o que pode acarretar na exclusão de alguns trabalhos importantes. Apesar de ser argumentável que os principais artigos na área foram contemplados, uma vez que a Scopus é capaz de indexar artigos da IEEE, ACM e Elsevier, como argumentado em (KITCHENHAM, 2010).

Outro ponto a ser considerado é a parcialidade do processo, uma vez que, apesar do procedimento ter sido conduzido em par, é possível que algum erro de parcialidade possa ter sido inserido na avaliação dos artigos, ou mesmo no processo de construção da pesquisa. Sendo assim, algum erro pode ter sido acidentalmente inserido no protocolo de pesquisa, causando a ausência de artigos importantes.

5 Considerações Finais

Neste trabalho foram apresentados uma revisão e mapeamento sistemático da literatura para identificar, classificar e analisar as soluções computacionais existentes para a detecção automática de *fake news* no âmbito das redes sociais, de uma forma imparcial e com uma boa cobertura da literatura. Com o procedimento executado, foi possível encontrar e analisar abordagens, técnicas e deficiências na área. Durante este processo, 1050 artigos foram selecionados e após critérios de exclusão definidos, resultaram num conjunto de 63 artigos selecionados.

A partir dos trabalhos selecionados, foi possível identificar que a partir de 2011 foram registradas publicações para área de pesquisa em questão, bem como identificar a variedade de autores que publicam sobre o tema. A partir deste levantamento foi possível perceber que apenas um autor apareceu mais de duas vezes, com três publicações.

Analisando o conteúdo dos artigos também foi possível notar os diferentes focos em que a busca por veracidades pode ser utilizada nas redes sociais, tendo artigos com métodos aplicáveis em contextos mais generalizados, bem como contextos específicos. Além disto foi possível constatar quais os domínios (redes sociais) possuem maior esforço por parte dos pesquisadores e a partir disto notar uma falta de pesquisa em redes sociais muito utilizadas, como é o caso do WhatsApp.

Durante o desenvolvimento da pesquisa, também foi possível notar uma falta de normalização dos dados encontrados entre as pesquisas, uma vez que poucas bases de dados são compartilhadas entre os artigos, o que faz com que os resultados obtidos por diferentes abordagens sejam difíceis de ser comparados. Uma vez que diversas características podem ser utilizadas para a detecção de *fake news*, um *benchmark* construído através de bases de dados que possam ser utilizados por diferentes tipos de abordagens ajudaria em uma avaliação mais precisa da eficiência e eficácia das abordagens.

Outro ponto importante é a utilização de técnicas de aprendizado de máquina e inteligência artificial, que pode ser constatado que é de amplo uso na área, estando presente em 43 dos trabalhos, ou seja 68,2% das pesquisas selecionadas. Como estes

métodos dependem de informações, a criação de sites com curadoria de conteúdo realizada por especialistas também é importante para a evolução da área.

Como trabalhos futuros propõe-se a realização de algumas análises mais completas em relação à eficiência e eficácia dos métodos para que possa ser descoberto não só quais as abordagens são mais utilizadas, mas também de forma a ser possível entender os contextos em que cada uma delas tem o melhor desempenho.

Bibliografia

AGRAWAL, T.; GUPTA, R.; NARAYANAN, S. Multimodal detection of fake social media use through a fusion of classification and pairwise ranking systems. In: IEEE. *25th European Signal Processing Conference (EUSIPCO)*. [S.l.], 2017. p. 1045–1049.

ALLCOTT, H.; GENTZKOW, M. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, v. 31, n. 2, p. 211–36, 2017.

American Dialect Society. *"Fake news" is 2017 American Dialect Society word of the year*. 2018. Disponível em: ⟨https://www.americandialect.org/fake-news-is-2017-american-dialect-society-word-of-the-year⟩.

ANTONIADIS, S.; LITOU, I.; KALOGERAKI, V. A model for identifying misinformation in online social networks. In: SPRINGER. *OTM Confederated International Conferences"On the Move to Meaningful Internet Systems"*. [S.l.], 2015. p. 473–482.

BAETH, M. J.; AKTAS, M. S. Detecting misinformation in social networks using provenance data. In: IEEE. *13th International Conference on Semantics, Knowledge and Grids (SKG)*. [S.l.], 2017. p. 85–89.

BHATTACHARJEE, S. D.; TALUKDER, A.; BALANTRAPU, B. V. Active learning based news veracity detection with feature weighting and deep-shallow fusion. In: IEEE. *International Conference on Big Data (Big Data)*. [S.l.], 2017. p. 556–565.

BODNAR, T. et al. Increasing the veracity of event detection on social media networks through user trust modeling. In: IEEE. *International Conference on Big Data (Big Data)*. [S.l.], 2014. p. 636–643.

BOIDIDOU, C. et al. Detection and visualization of misleading content on twitter. *International Journal of Multimedia Information Retrieval*, Springer, v. 7, n. 1, p. 71–86, 2018.

BUDGEN, D.; BRERETON, P. Performing systematic literature reviews in software engineering. In: ACM. *Proceedings of the 28th international conference on Software engineering*. [S.l.], 2006. p. 1051–1052.

BUNTAIN, C.; GOLBECK, J. Automatically identifying fake news in popular twitter threads. In: IEEE. *International Conference on Smart Cloud (SmartCloud)*. [S.l.], 2017. p. 208–215.

CAI, G.; BI, M.; LIU, J. A novel rumor detection method based on labeled cascade propagation tree. In: IEEE. *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. [S.l.], 2017. p. 2185–2194.

CAI, G.; WU, H.; LV, R. Rumors detection in chinese via crowd responses. In: IEEE PRESS. *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. [S.l.], 2014. p. 912–917.

CASTILLO, C.; MENDOZA, M.; POBLETE, B. Information credibility on twitter. In: ACM. *Proceedings of the 20th international conference on World wide web*. [S.l.], 2011. p. 675–684.

CASTILLO, C.; MENDOZA, M.; POBLETE, B. Predicting information credibility in time-sensitive social media. *Internet Research*, Emerald Group Publishing Limited, v. 23, n. 5, p. 560–588, 2013.

CHANG, C. et al. Extreme user and political rumor detection on twitter. In: SPRINGER. *International Conference on Advanced Data Mining and Applications*. [S.l.], 2016. p. 751–763.

CHATTERJEE, R.; AGARWAL, S. Twitter truths: Authenticating analysis of information credibility. In: IEEE. *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.], 2016. p. 2352–2357.

CHEN, W. et al. Behavior deviation: An anomaly detection view of rumor preemption. In: IEEE. *7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. [S.l.], 2016. p. 1–7.

CIAMPAGLIA, G. L. et al. Computational fact checking from knowledge networks. *PloS one*, Public Library of Science, v. 10, n. 6, p. e0128193, 2015.

CONROY, N. J.; RUBIN, V. L.; CHEN, Y. Automatic deception detection: Methods for finding fake news. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. [S.l.], 2015. p. 82.

COSTA, C.; MURTA, L. Version control in distributed software development: A systematic mapping study. In: IEEE. *8th International Conference on Global Software Engineering (ICGSE)*. [S.l.], 2013. p. 90–99.

FIGUEIRA, Á.; SANDIM, M.; FORTUNA, P. An approach to relevancy detection: contributions to the automatic detection of relevance in social networks. In: *New Advances in Information Systems and Technologies*. [S.l.]: Springer, 2016. p. 89–99.

FRUCHTERMAN, T. M.; REINGOLD, E. M. Graph drawing by force-directed placement. *Software: Practice and experience*, Wiley Online Library, v. 21, n. 11, p. 1129–1164, 1991.

GARG, A. et al. Mining credible and relevant news from social networks. In: SPRINGER. *International Conference on Big Data Analytics*. [S.l.], 2017. p. 90–102.

GIASEMIDIS, G. et al. Determining the veracity of rumours on twitter. In: SPRINGER. *International Conference on Social Informatics*. [S.l.], 2016. p. 185–205.

GRANIK, M.; MESYURA, V. Fake news detection using naive bayes classifier. In: IEEE. *First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. [S.l.], 2017. p. 900–903.

GUO, Q. et al. Information credibility: A probabilistic graphical model for identifying credible influenza posts on social media. In: SPRINGER. *International Conference on Smart Health*. [S.l.], 2015. p. 131–142.

GUPTA, M.; ZHAO, P.; HAN, J. Evaluating event credibility on twitter. In: SIAM. *Proceedings of the 2012 SIAM International Conference on Data Mining*. [S.l.], 2012. p. 153–164.

HASHISH, I. A. et al. An analysis of social data credibility for services systems in smart cities–credibility assessment and classification of tweets. In: *Cloud Infrastructures, Services, and IoT Systems for Smart Cities*. [S.l.]: Springer, 2017. p. 119–130.

JAIN, S.; SHARMA, V.; KAUSHAL, R. Towards automated real-time detection of misinformation on twitter. In: IEEE. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.], 2016. p. 2015–2020.

JANSSEN, B.; HABIB, M.; KEULEN, M. V. Truth assessment of objective facts extracted from tweets: A case study on world cup 2014 game facts. In: SCITEPRESS. *13th International Conference on Web Information Systems and Technologies, WEBIST 2017*. [S.l.], 2017.

JIN, Z. et al. News credibility evaluation on microblog with a hierarchical propagation model. In: IEEE. *International Conference on Data Mining (ICDM)*. [S.l.], 2014. p. 230–239.

JIN, Z. et al. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia*, IEEE, v. 19, n. 3, p. 598–608, 2017.

JR, C. F. B.; DEPAULO, B. M. Accuracy of deception judgments. *Personality and social psychology Review*, Sage Publications Sage CA: Los Angeles, CA, v. 10, n. 3, p. 214–234, 2006.

JR, E. C. T.; LIM, Z. W.; LING, R. Defining "fake news" a typology of scholarly definitions. *Digital Journalism*, Taylor & Francis, v. 6, n. 2, p. 137–153, 2018.

KAWABE, T. et al. A part-of-speech based sentiment classification method considering subject-predicate relation. In: IEEE. *International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2015. p. 999–1004.

KAWABE, T. et al. Tweet credibility analysis evaluation by improving sentiment dictionary. In: IEEE. *Congress on Evolutionary Computation (CEC)*. [S.l.], 2015. p. 2354–2361.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004.

KITCHENHAM, B. What's up with software metrics?–a preliminary mapping study. *Journal of systems and software*, Elsevier, v. 83, n. 1, p. 37–51, 2010.

KITCHENHAM, B. A.; BUDGEN, D.; BRERETON, O. P. Using mapping studies as the basis for further research–a participant-observer case study. *Information and Software Technology*, Elsevier, v. 53, n. 6, p. 638–651, 2011.

KUMAR, K. K.; GEETHAKUMARI, G. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, SpringerOpen, v. 4, n. 1, p. 14, 2014.

KWON, S. et al. Prominent features of rumor propagation in online social media. In: IEEE. *13th International Conference on Data Mining*. [S.l.], 2013. p. 1103–1108.

LENDVAI, P.; REICHEL, U. D.; DECLERCK, T. Factuality drift assessment by lexical markers in resolved rumors. ACM Press, 2016.

LIANG, G.; YANG, J.; XU, C. Automatic rumors identification on sina weibo. In: IEEE. *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. [S.l.], 2016. p. 1523–1531.

LIM, W.-Y.; LEE, M.-L.; HSU, W. Claimfinder: A framework for identifying claims in microblogs. In: *# Microposts*. [S.l.: s.n.], 2016. p. 13–20.

LIU, Y. et al. Do rumors diffuse differently from non-rumors? a systematically empirical analysis in sina weibo for rumor identification. In: SPRINGER. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. [S.l.], 2017. p. 407–420.

MA, B.; LIN, D.; CAO, D. Content representation for microblog rumor detection. In: *Advances in Computational Intelligence Systems*. [S.l.]: Springer, 2017. p. 245–251.

MA, J. et al. Detecting rumors from microblogs with recurrent neural networks. In: *IJCAI*. [S.l.: s.n.], 2016. p. 3818–3824.

MA, J. et al. Detect rumors using time series of social context information on microblogging websites. In: ACM. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. [S.l.], 2015. p. 1751–1754.

MA, J.; GAO, W.; WONG, K.-F. Detect rumors in microblog posts using propagation structure via kernel learning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. [S.l.: s.n.], 2017. v. 1, p. 708–717.

MOIN, R. et al. Framework for rumors detection in social media. *Framework*, v. 9, n. 5, 2018.

MONDAL, T. et al. Analysis and early detection of rumors in a post disaster scenario. *Information Systems Frontiers*, Springer, p. 1–19, 2018.

NEIVA, F. W.; SILVA, R. L. d. S. da. Revisão sistemática da literatura em ciência da computação um guia prático. 2016.

NGUYEN, T. N.; LI, C.; NIEDERÉE, C. On early-stage debunking rumors on twitter: Leveraging the wisdom of weak learners. In: SPRINGER. *International Conference on Social Informatics*. [S.l.], 2017. p. 141–158.

ÖZGÖBEK, Ö.; GULLA, J. A. Towards an understanding of fake news. 2017.

PAL, A.; CHUA, A. Y. Classification of rumors and counter-rumors. In: IEEE. *4th International Conference on Information Management (ICIM)*. [S.l.], 2018. p. 81–85.

PARIKH, S. B.; ATREY, P. K. Media-rich fake news detection: A survey. In: IEEE. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. [S.l.], 2018. p. 436–441.

PETTICREW, M.; ROBERTS, H. Systematic reviews in the social sciences: a practical guide. 2006. *Malden USA: Blackwell Publishing CrossRef Google Scholar*, 2006.

RAJDEV, M.; LEE, K. Fake and spam messages: Detecting misinformation during natural disasters on social media. In: IEEE. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. [S.l.], 2015. v. 1, p. 17–20.

RUBIN, V. L.; CHEN, Y.; CONROY, N. J. Deception detection for news: three types of fakes. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. [S.l.], 2015. p. 83.

SABBEH, S. F.; BAATWAH, S. Y. Arabic news credibility on twitter: An enhanced model using hybrid features. *Journal of Theoretical & Applied Information Technology*, v. 96, n. 8, 2018.

SAHANA, V. et al. Automatic detection of rumoured tweets and finding its origin. In: IEEE. *International Conference on Computing and Network Communications (CoCoNet)*. [S.l.], 2015. p. 607–612.

SAMPSON, J. et al. Leveraging the implicit structure within social media for emergent rumor detection. In: ACM. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. [S.l.], 2016. p. 2377–2382.

SHARIFF, S. M.; ZHANG, X. A survey on deceptions in online social networks. In: IEEE. *Computer and Information Sciences (ICCOINS), 2014 International Conference on*. [S.l.], 2014. p. 1–6.

SHU, K. et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, ACM, v. 19, n. 1, p. 22–36, 2017.

SICILIA, R. et al. Twitter rumour detection in the health domain. *Expert Systems with Applications*, Elsevier, 2018.

Statista Inc. *Most popular social networks worldwide as of October 2018, ranked by number of active users (in millions)*. 2018. Disponível em: ⟨https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/⟩.

TACCHINI, E. et al. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*, 2017.

TOLOSI, L.; TAGAREV, A.; GEORGIEV, G. An analysis of event-agnostic features for rumour classification in twitter. In: *Tenth International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2016.

ULICNY, B.; KOKAR, M. M. Toward formal reasoning with epistemic policies about information quality in the twittersphere. In: IEEE. *Proceedings of the 14th International Conference on Information Fusion (FUSION)*. [S.l.], 2011. p. 1–8.

VIVIANI, M.; PASI, G. Credibility in social media: opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 7, n. 5, p. e1209, 2017.

VOSOUGHI, S.; MOHSENVAND, M.; ROY, D. Rumor gauge: Predicting the veracity of rumors on twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, ACM, v. 11, n. 4, p. 50, 2017.

WANG, S. et al. Early signals of trending rumor event in streaming social media. In: IEEE. *41st Annual Computer Software and Applications Conference (COMPSAC)*. [S.l.], 2017. v. 2, p. 654–659.

WANG, W. Y. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

WENDLING, M. *The (almost) complete history of 'fake news'*. 2018. Disponível em: ⟨https://www.bbc.com/news/blogs-trending-42724320⟩.

WU, L. et al. Gleaning wisdom from the past: Early detection of emerging rumors in social media. In: SIAM. *Proceedings of the 2017 SIAM International Conference on Data Mining*. [S.l.], 2017. p. 99–107.

WU, L.; LIU, H. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: ACM. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. [S.l.], 2018. p. 637–645.

WU, S. et al. Information credibility evaluation on social media. In: *AAAI*. [S.l.: s.n.], 2016. p. 4403–4404.

XIE, B. et al. Gatekeeping behavior analysis for information credibility assessment on weibo. In: SPRINGER. *International Conference on Network and System Security*. [S.l.], 2016. p. 483–496.

YAVARY, A.; SAJEDI, H. Rumor detection on twitter using extracted patterns from conversational tree. In: IEEE. *2018 4th International Conference on Web Research (ICWR)*. [S.l.], 2018. p. 78–85.

YU, F. et al. A convolutional approach for misinformation identification. In: AAAI PRESS. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. [S.l.], 2017. p. 3901–3907.

ZHANG, D. Y. et al. On robust truth discovery in sparse social media sensing. In: IEEE. *International Conference on Big Data (Big Data)*. [S.l.], 2016. p. 1076–1081.

ZHANG, H.; BABAR, M. A. On searching relevant studies in software engineering. British Informatics Society Ltd., 2010.

ZHANG, H.; LI, J.; XIAO, Y. Hadoop cellular automata for identifying rumor in social networks. In: IEEE. *International Conference on Information Science and Cloud Computing Companion (ISCC-C)*. [S.l.], 2013. p. 37–42.

ZHANG, Q. et al. Automatic detection of rumor on social network. In: *Natural Language Processing and Chinese Computing*. [S.l.]: Springer, 2015. p. 113–122.

ZHANG, Y. et al. A distance-based outlier detection method for rumor detection exploiting user behaviorial differences. In: IEEE. *International Conference on Data and Software Engineering (ICoDSE)*. [S.l.], 2016. p. 1–6.

ZHAO, Z.; RESNICK, P.; MEI, Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. In: INTERNATIONAL WORLD WIDE WEB CONFERENCES STEERING COMMITTEE. *Proceedings of the 24th International Conference on World Wide Web*. [S.l.], 2015. p. 1395–1405.

ZHOU, L. et al. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation*, Springer, v. 13, n. 1, p. 81–106, 2004.

ZHOU, X. et al. Real-time news cer tification system on sina weibo. In: ACM. *Proceedings of the 24th International Conference on World Wide Web*. [S.l.], 2015. p. 983–988.

ZOU, J.; FEKRI, F.; MCLAUGHLIN, S. W. Mining streaming tweets for real-time event credibility prediction in twitter. In: ACM. *International Conference on Advances in Social Networks Analysis and Mining*. [S.l.], 2015. p. 1586–1589.

ZUBIAGA, A. et al. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, ACM, v. 51, n. 2, p. 32, 2018.

ZUBIAGA, A.; JI, H. Tweet, but verify: epistemic study of information verification on twitter. *Social Network Analysis and Mining*, Springer, v. 4, n. 1, p. 163, 2014.

ZUBIAGA, A.; LIAKATA, M.; PROCTER, R. Exploiting context for rumour detection in social media. In: SPRINGER. *International Conference on Social Informatics*. [S.l.], 2017. p. 109–123.

A - Listagem dos trabalhos em relação aos atributos coletados

Neste apêndice é possível visualizar a relação de todos os trabalhos encontrados na revisão, bem como quais os atributos foram utilizados por cada um deles para realizar a classificação.

Trabalho	**Atributos utilizados**
(CASTILLO; MENDOZA; POBLETE, 2013)	U, P, M, T
(MA et al., 2015)	M, U, P, T
(CASTILLO; MENDOZA; POBLETE, 2011)	M, U, T, P
(KAWABE et al., 2015a)	T, U, M
(ZHANG et al., 2015)	U, M, T
(LIANG; YANG; XU, 2016)	U, M, P
(MA; GAO; WONG, 2017)	U, P, M
(BUNTAIN; GOLBECK, 2017)	U, P, M
(KUMAR; GEETHAKUMARI, 2014)	M, U, P
(BOIDIDOU et al., 2018)	M, U, P
(GIASEMIDIS et al., 2016)	M, U, P
(LIU et al., 2017)	M, U, P
(GUPTA; ZHAO; HAN, 2012)	M, U, P
(CHANG et al., 2016)	U, P, M
(AGRAWAL; GUPTA; NARAYANAN, 2017)	M, M, U
(ZHOU et al., 2015)	U, P, M
(VOSOUGHI; MOHSENVAND; ROY, 2017)	U, M, P
(CHATTERJEE; AGARWAL, 2016)	T, U, P
(ZHANG et al., 2016)	U, M
(ANTONIADIS; LITOU; KALOGERAKI, 2015)	U, M
(CAI; BI; LIU, 2017)	U, P

Trabalho	**Atributos utilizados**
(TOLOSI; TAGAREV; GEORGIEV, 2016)	U, M
(HASHISH et al., 2017)	U, M
(SABBEH; BAATWAH, 2018)	U, T
(SAHANA et al., 2015)	U, M
(CHEN et al., 2016)	U, M
(PAL; CHUA, 2018)	M, U
(WANG et al., 2017)	M, P
(ZHAO; RESNICK; MEI, 2015)	M, P
(ZUBIAGA; LIAKATA; PROCTER, 2017)	M, U
(RAJDEV; LEE, 2015)	M, U
(XIE et al., 2016)	M, U
(WU et al., 2017)	M, U
(BODNAR et al., 2014)	M, U
(GUO et al., 2015)	M, U
(SAMPSON et al., 2016)	M, P
(GARG et al., 2017)	M, U
(JIN et al., 2014)	M, P
(JIN et al., 2017)	M, P
(NGUYEN; LI; NIEDERÉE, 2017)	M, U
(ULICNY; KOKAR, 2011)	M, U
(JAIN; SHARMA; KAUSHAL, 2016)	T, M
(SICILIA et al., 2018)	U, P
(YU et al., 2017)	M
(BHATTACHARJEE; TALUKDER; BALANTRAPU, 2017)	M
(FIGUEIRA; SANDIM; FORTUNA, 2016)	M
(MONDAL et al., 2018)	M
(LIM; LEE; HSU, 2016)	M
(MA; LIN; CAO, 2017)	M
(BAETH; AKTAS, 2017)	U

Trabalho	**Atributos utilizados**
(LENDVAI; REICHEL; DECLERCK, 2016)	M
(GRANIK; MESYURA, 2017)	M
(MOIN et al., 2018)	M
(ZHANG; LI; XIAO, 2013)	P
(WU et al., 2016)	P
(ZOU; FEKRI; MCLAUGHLIN, 2015)	M
(ZHANG et al., 2016)	P
(YAVARY; SAJEDI, 2018)	P
(CAI; WU; LV, 2014)	M
(TACCHINI et al., 2017)	P
(WU; LIU, 2018)	P
(JANSSEN; HABIB; KEULEN, 2017)	M
(KAWABE et al., 2015b)	T

Tabela A.1: Relação dos trabalhos com atributos utilizados, sendo "M"para mensagem, "U"para usuário, "P"para propagação e "T"para tópico.