arXiv:1812.00315v1 [cs.CL] 2 Dec 2018

Fake News: A Survey of Research, Detection Methods, and Opportunities

XINYI ZHOU, Syracuse University, USA
REZA ZAFARANI, Syracuse University, USA

The explosive growth in fake news and its erosion to democracy, justice, and public trust has increased the demand for fake news analysis, detection and intervention. This survey comprehensively and systematically reviews fake news research. The survey identifies and specifies fundamental theories across various disciplines, e.g., psychology and social science, to facilitate and enhance the interdisciplinary research of fake news. Current fake news research is reviewed, summarized and evaluated. These studies focus on fake news from four perspective: (1) the false *knowledge* it carries, (2) its writing *style*, (3) its *propagation* patterns, and (4) the *credibility* of its creators and spreaders. We characterize each perspective with various analyzable and utilizable information provided by news and its spreaders, various strategies and frameworks that are adaptable, and techniques that are applicable. By reviewing the characteristics of fake news and open issues in fake news studies, we highlight some potential research tasks at the end of this survey.

CCS Concepts: • **Human-centered computing** → *Collaborative and social computing theory, concepts and paradigms*; *Empirical studies in collaborative and social computing*; • **Social and professional topics** → *Computer crime*; • **Applied computing** → *Computer forensics*;

Additional Key Words and Phrases: Fake News

ACM Reference Format:
Xinyi Zhou and Reza Zafarani. 2018. Fake News: A Survey of Research, Detection Methods, and Opportunities. *ACM Comput. Surv.* 1, 1 (December 2018), 40 pages.

1 INTRODUCTION

Fake news is now viewed as one of the greatest threats to democracy, journalism, and freedom of expression. It has weakened public trust in governments and its potential impact on the contentious "Brexit" referendum and the equally divisive 2016 U.S. presidential election – which it might have affected [Pogue 2017] – is yet to be realized. The reach of fake news was best highlighted during the critical months of the 2016 U.S. presidential election campaign, where the top twenty frequently-discussed false election stories generated 8,711,000 shares, reactions, and comments on Facebook, ironically, larger than the total of 7,367,000 for the top twenty most-discussed election stories posted by 19 major news websites [Silverman 2016]. Our economies are not immune to the spread of fake news either, with fake news being connected to stock market fluctuations and massive trades. For example, fake news claiming that Barack Obama was injured in an explosion wiped out $130 billion in stock value [Rapoza 2017]. These events and losses have motivated fake news research and sparked the discussion around fake news, as observed by skyrocketing usage of terms such as "post-truth" – selected as the international word of the year by Oxford Dictionaries in 2016 [Wang 2016].

Authors' addresses: Xinyi Zhou, Data Lab, EECS Department, Syracuse University, Syracuse, NY, 13244, USA, zhouxinyi@data.syr.edu; Reza Zafarani, Data Lab, EECS Department, Syracuse University, Syracuse, NY, 13244, USA, reza@data.syr.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Association for Computing Machinery.
Manuscript submitted to ACM

While fake news is not a new phenomenon [Allcott and Gentzkow 2017], questions such as why has it emerged as a world topic and why is it attracting increasingly more public attention are particularly relevant at this time. The leading cause is that fake news can be created and published online faster and cheaper when compared to traditional news media such as newspapers and television. The rise of social media and its popularity also plays an important role in this surge of interest. As of August 2017, around two third (67%) of Americans get their news from social media.[1] With the existence of an *echo chamber effect* on social media, biased information is often amplified and reinforced [Jamieson and Cappella 2008]. Furthermore, as an ideal platform to accelerate fake news dissemination, social media breaks the physical distance barrier among individuals, provides rich platforms to share, forward, vote, and review, and encourages users to participate and discuss online news [Zhou et al. 2019]. This surge of activity around online news can lead to grave repercussions, but also substantial potential political and economic benefits. Such generous benefits encourage malicious entities to create, publish and spread fake news.

Take the dozens of "well-known" teenagers in the Macedonian town of Veles as an example of users who produced fake news for millions on social media and became wealthy by penny-per-click advertising during the U.S. presidential election. As reported by the NBC, each individual "has earned at least $60,000 in the past six months – far outstripping their parents' income and transforming his prospects in a town where the average annual wage is $4,800." [Smith and Banic 2016]. The tendency of individuals to overestimate the benefits rather than costs, as *valence effect* [Jones and McGillis 1976] indicates, further widens the gap between benefits and costs attracting individuals to engage in fake news activities. Clearly when governments, parties and business tycoons are standing behind fake news generation, seeking its tempting power and profits, there is a greater motivation and capability to make fake news more persuasive and indistinguishable from truth to the public. But, how can fake news gain public trust?

Social and psychological factors play an important role in fake news gaining public trust and further facilitate the spread of fake news. For instance, humans have been proven to be irrational and vulnerable when differentiating between truth and falsehood while overloaded with deceptive information. Studies in social psychology and communications have demonstrated that human ability to detect deception is only slightly better than chance: typical accuracy rates are in the 55%-58% range, with a mean accuracy of 54% over 1,000 participants in over 100 experiments [Rubin 2010]. The situation is more critical for fake news compared to other types of information, as for news, a representative of authenticity and objectivity, is relatively easier to gain public trust. In addition, individuals tend to trust fake news after repeated exposures (i.e., *validity effect* [Boehm 1994]), or if it confirms their pre-existing knowledge (i.e., *confirmation bias* [Nickerson 1998]. *Peer pressure* can also at times "control" our perception and behavior (i.e., *bandwagon effect* [Leibenstein 1950]).

Many perspectives on who creates fake news, how and why it is created, how it propagates, and how it can be detected motivate the need for an in-depth analysis. This survey aims to develop a systematic framework for the comprehensive *study of fake news*. As fake news is not clearly defined and current studies of fake news are limited, we extend our study to related fields that can serve as a foundation for fake news research. We hope this survey can facilitate fake news studies by inspiring researchers vertically, to extend current fake news studies in-depth, and horizontally, to enrich and improve fake news studies by interdisciplinary research. Before we provide a summary of this work in Section 1.3, we define *fake news* (Section 1.1) and summarize its fundamental theories (Section 1.2).

[1] http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/

1.1 What is Fake News?

There has been no universal definition for fake news, even in journalism. A clear and accurate definition helps lay a solid foundation for fake news analysis and evaluating related studies. Here we (I) theoretically distinguish between several concepts that frequently co-occur or have overlaps with fake news, (II) present a broad and a narrow definition for the term *fake news*, providing a justification for each definition, and (III) further highlight the potential research problems raised by such definitions.

I. Related Concepts. Existing studies often connect fake news to terms and concepts such as *maliciously false news* [Allcott and Gentzkow 2017; Shu et al. 2017a; Waldrop 2017], *false news* [Vosoughi et al. 2018], *satire news* [Berkowitz and Schwartz 2016], *disinformation* (i.e., *deception*) [Kshetri and Voas 2017], *misinformation* [Kucharski 2016], and *rumor* [Buntain and Golbeck 2017]. Based on the these terms and concepts are defined, we can distinguish one from the others based on three characteristics: (i) authenticity (false or not), (ii) intention (bad or not), and (iii) whether the information is news or not. Table 1 has the details.

Table 1. A Comparison between Concepts related to Fake News

	Authenticity	**Intention**	**News?**
Maliciously false news	False	Bad	Yes
False news	False	Unknown	Yes
Satire news	Unknown	Not bad	Yes
Disinformation	False	Bad	Unknown
Misinformation	False	Unknown	Unknown
Rumor	Unknown	Unknown	Unknown

For example, disinformation is false information [news or non-news] with a bad intention aiming to mislead the public.

II. Defining Fake News. We first broadly define fake news as follows:

Definition 1 (Broad definition of fake news). *Fake news is false news,*

where news[2] broadly includes claims, statements, speeches, posts, among other types of information related to public figures and organizations. The broad definition aims to impose minimum constraints in accord with the current resources: it emphasizes information authenticity, purposefully adopts a broad definition for the term news [Vosoughi et al. 2018] and weakens the requirement for information intentions. This definition supports most existing fake-news-related studies, and datasets, as provided by the existing fact-checking websites (Section 2.1 has a detailed introduction). Current fake news datasets often provide ground truth for the authenticity of claims, statements, speeches, or posts related to public figures and organizations, while no information is provided regarding intentions.

We provide a more narrow definition of fake news which satisfies the overall requirements for fake news as follows.

Definition 2 (Narrow definition of fake news). *Fake news is intentionally and verifiably false news published by a news outlet.*

This narrow definition addresses the public's perception of fake news, especially following the 2016 U.S. presidential election. Note that deceptive news (i.e., maliciously false news) is more harmful and less distinguishable than incautiously false news, as the former pretends to be truth to better mislead the public. The narrow definition emphasizes both news

[2]Definition of "news" in Oxford Dictionaries: newly received or noteworthy information, especially about recent events.

authenticity and intentions; it also ensures the posted information is news by investigating its publisher (a news outlet or not). Often news outlets publish news in the form of articles with fixed components: a headline, author(s), a body text which includes the claims and statements made by public figures and organizations. This definition supports recent advancements in fake news studies [Allcott and Gentzkow 2017; Shu et al. 2017a; Waldrop 2017].

III. Open Issues. We have theoretically differentiated between fake news and fake-news-related terms such as rumors, but empirical comparative studies are limited leaving many questions unanswered, e.g., how similar (or specific) are writing style or propagation patterns of fake news compared to that of related concepts (e.g. disinformation and rumors)? Does having different characteristics lead to different detection strategies? Can we automatically distinguish these concepts from fake news? We have also provided two definition for fake news, with the narrow definition being the most accurate; however, ground-truth datasets for fake news supporting the narrow definition are rarely seen. Systematically analyzing, identifying, and blocking fake news still has many unexplored arenas, with detailed discussions on these open issues can be seen in Sections 2 to 6.

1.2 Fundamental Theories

Fundamental human cognition and behavior theories developed across various discipline such as psychology, philosophy, social science, and economics provide invaluable insights for fake news analysis. Firstly, these theories introduce new opportunities for qualitative and quantitative studies of big fake news data, which to date, has been rarely available. Secondly, they facilitate building well-justified and explainable models for fake news detection and intervention, as well as introducing means to develop datasets that provide "ground truth" for fake news studies. We have conducted a comprehensive literature survey across various disciplines and have identified twenty well-known theories that can be potentially used to study fake news. These theories are provided in Table 2 along with short descriptions. These theories can be used to study fake news from three different perspectives: (I) *style*: how fake news is written, (II) *propagation*: how fake news spreads, and (III) *users*: how users engage with fake news and the role users play (or can play) in fake news creation, propagation, and intervention. In the following, we detail how each perspective and its corresponding theories facilitate fake news analysis.

I. Style-based Fake News Analysis. As we will further detail in Section 3, these fundamental theories address how fake news content and writing style can be different from true news. For instance, *reality monitoring* indicates that actual events can be expressed by higher levels of sensory-perceptual information.

II. Propagation-based Fake News Analysis. As we will review in Section 4, *epidemic models*, which can mathematically model the progression of an infectious disease, can be used or extended to model fake news propagation. However, selecting or developing proper epidemic models relies on making reasonable assumptions. Some real-world phenomena can help simplify these assumptions and in turn, simply such epidemic models. Examples includes *backfire effect*, *conservatism bias* and *Semmelweis reflex*, which indicate that "fake news is incorrect but hard to correct" [Roets et al. 2017], i.e., it propagates with minimum resistance.

III. User-based Fake News Analysis. These theories investigate fake news from a user's perspective, considering how users engage with fake news and what roles users play in fake news creation, propagation and intervention, as we will detail later in Section 5. In sum, users that participate in fake news activities can be grouped into (i) malicious users, who intentionally create and/or propagate fake news motivated by some benefits and (ii) normal users, some of whom spread fake news along with malicious users. These normal users are often called *naïve users* as their engagement is unintentional and driven by self-influence or social influence, e.g., naïve users can participate in fake news spreading

Table 2. Fundamental Theories in Psychology, Philosophy, Social Sciences, and Economics

		Term	Phenomenon
Style-based		*Undeutsch hypothesis* [Undeutsch 1967]	A statement based on a factual experience differs in content and quality from that of fantasy.
		Reality monitoring [Johnson and Raye 1981]	Actual events are characterized by higher levels of sensory- perceptual information.
		Four-factor theory [Zuckerman et al. 1981]	Lies are expressed differently in terms of arousal, behavior control, emotion, and thinking from truth.
Propagation-based		*Backfire effect* [Nyhan and Reifler 2010]	Given evidence against their beliefs, individuals can reject it even more strongly.
		Conservatism bias [Basu 1997]	The tendency to revise one's belief insufficiently when presented with new evidence.
		Semmelweis reflex [Bálint and Bálint 2009]	Individuals tend to reject new evidence because it contradicts with established norms and beliefs.
User-based (User's Engagement and Role)	**Social influence**	*Attentional bias* [MacLeod et al. 1986]	An individual's perception is affected by his or her recurring thoughts at the time.
		Validity effect [Boehm 1994]	Individuals tend to believe information is correct after repeated exposures.
		Bandwagon effect [Leibenstein 1950]	Individuals do something primarily because others are doing it.
		Echo chamber effect [Jamieson and Cappella 2008]	Beliefs are amplified or reinforced by communication and repetition within a closed system.
		Normative influence theory [Deutsch and Gerard 1955]	The influence of others leading us to conform to be liked and accepted by them.
		Social identity theory [Ashforth and Mael 1989]	An individual's self-concept derives from perceived membership in a relevant social group.
		Availability cascade [Kuran and Sunstein 1999]	Individuals tend to adopt insights expressed by others when such insights are gaining more popularity within their social circles
	Self-influence	*Confirmation bias* [Nickerson 1998]	Individuals tend to trust information that confirms their preexisting beliefs or hypotheses.
		Illusion of asymmetric insight [Pronin et al. 2001]	Individuals perceive their knowledge to surpass that of others.
		Naïve realism [Ward et al. 1997]	The senses provide us with direct awareness of objects as they really are.
		Overconfidence effect [Dunning et al. 1990]	A person's subjective confidence in his judgments is reliably greater than the objective ones.
	Benefit influence	*Prospect theory* [Kahneman and Tversky 2013]	People make decisions based on the value of losses and gains rather than the outcome.
		Valence effect [Frijda 1986]	People tend to overestimate the likelihood of good things happening rather than bad things.
		Contrast effect [Hovland et al. 1957]	The enhancement or diminishment of cognition due to successive or simultaneous exposure to a stimulus of lesser or greater value in the same dimension.

due their preexisting knowledge (as explained by *confirmation bias*) or peer-pressure (as indicated by *bandwagon effect*). These theorems can be help improve fake news detection efficiency and reduce the expense of fake news intervention.

1.3 An Overview of this Survey

This survey aims to present a comprehensive framework to study fake news by introducing means to qualitatively and quantitatively analyze fake news as well as detection and intervention techniques. We review and summarize the existing resources, e.g., theories, patterns, mathematical models, and empirical approaches, and further detail the role they can play in fake news studies. We also point out specific open issues that are critical but have not been (systematically) investigated or addressed in fake news studies. Information utilized to study fake news can be *news-related* (e.g., headline, body text, creator, publisher) or *social-related* (e.g., comments, propagation network, and spreaders), covering the whole life cycle of fake news, from the time it is created to when it is published or spreading. Fake news can

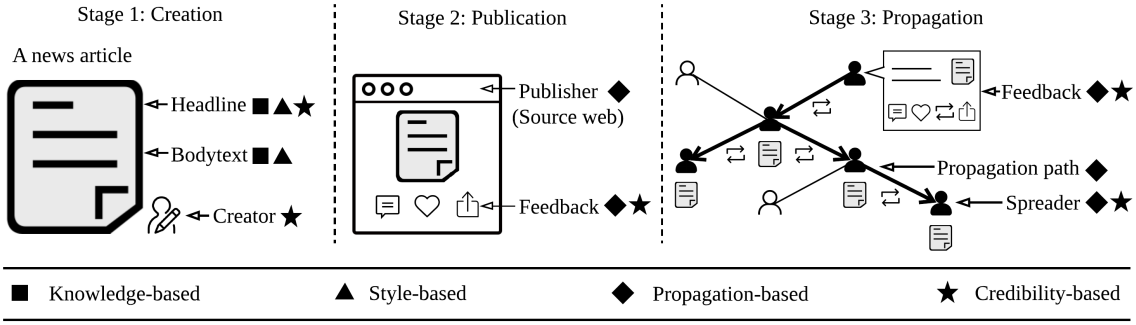


Fig. 1. Fake news Life Cycle and Connections to the Four Fake News Perspectives Presented in this Survey

be studied with respect to four perspectives: (i) *knowledge-based*, focusing on the false knowledge in fake news; (ii) *style-based*, concerned with how fake news is written; (iii) *propagation-based*, focused on how fake news spreads, and; (iii) *credibility-based*, investigating the credibility of its creators and spreaders. Each perspective targets some of fake news characteristics (i.e., its authenticity, intentions, or whether it is news) by using some types of information (news-related or social-related) using different techniques (see discussions in Section 6 and a comparative study in Table 10). As shown using the fake news life cycle in Figure 1, a knowledge-based or style-based study of fake news focuses on news content and thus can be conducted once fake news is created; a propagation-based or credibility-based study of fake news aims to exploit social-related information which appears after fake news is published. In addition to reviewing, summarizing and evaluating the limited number of current fake news studies, we extend our survey to a broader horizon, by presenting theories from disciplines such as psychology or social sciences, mathematical models from fields such as epidemiology and economics, and related topics such as the study of deception, rumors, click-baits, and review spam to facilitate fake news studies. Overall, the specific contributions of this survey are as follows:

(1) We systematically compare several fake-news-related terms and concepts, which can be distinguished based on three characteristics: authenticity, intentions, and whether the information is news or not. We also cautiously provide a clear broad and narrow definition for fake news in view of the current available resources and public concerns, respectively giving the minimum and overall requirements for some information to be fake news.
(2) To our best knowledge, this survey provides the most comprehensive list of fundamental theories that can be utilized when studying fake news. Initially developed in psychology, philosophy, social sciences and economics, these fundamental theories are invaluable for fake news studies, as we will detail in this survey.
(3) This survey comprehensively and extensively studies fake news, presenting (i) methods to qualitatively and quantitatively analyze, detect or intervene with fake news, (ii) four perspectives to study fake news based on knowledge, style, propagation, and credibility, (iii) news-related (e.g., headline, body text, creator, publisher) and social-related information (e.g., comments, propagation path and spreaders) used in fake news studies, (iv) techniques (e.g., feature-based or relational-based) used in fake news research, along with (v) our review, classification, comparison and evaluation of current fake news and fake-news-related studies.

We present the four perspectives to study fake news: knowledge-based, style-based, propagation-based and credibility-based in Sections 2 to 5, respectively. Section 6 is a summary and supplements Sections 2-5, where we further compare fake news studies based on various perspectives and highlight several tasks that can facilitate further development in fake news research. We conclude the survey in Section 7.

2 KNOWLEDGE-BASED STUDY OF FAKE NEWS

When studying fake news from a knowledge-based perspective, one aims to analyze and/or detect fake news, using a process known as *fact-checking*. Fact-checking, initially developed in journalism, aims to assess news authenticity by comparing the knowledge extracted from to-be-verified news content (e.g., its claims or statements) with known facts (i.e., true knowledge). In this section, we will discuss the traditional fact-checking (also known as *manual fact-checking*) and how it can be incorporated into automatic means to analyze and detect fake news (i.e., *automatic fact-checking*).

2.1 Manual Fact-checking

Broadly speaking, manual fact-checking can be divided into (I) expert-based and (II) crowd-sourced fact-checking.

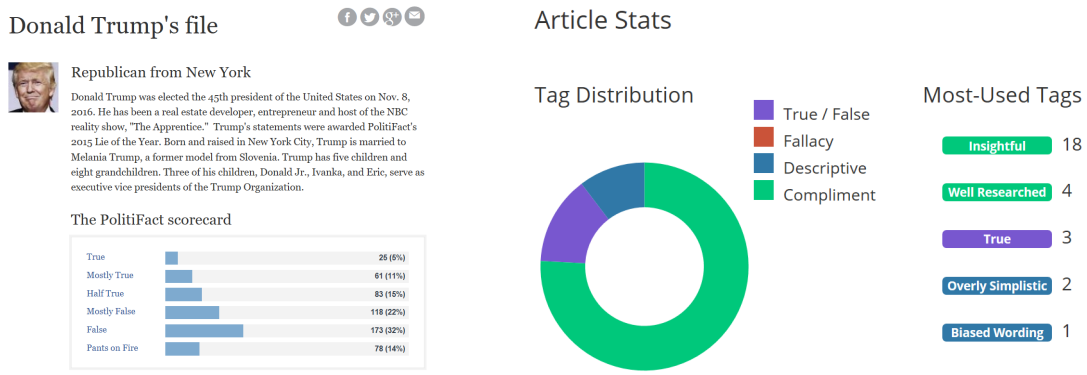
I. Expert-based Manual Fact-checking. Expert-based fact-checking relies on domain-experts (i.e., *fact-checkers*) to verify the given news contents. Expert-based fact-checking is often conducted by a small group of highly credible fact-checkers, is easy to manage, and leads to highly accurate results, but is costly and poorly scales with the increase in the volume of to-be-checked news contents.

Table 3. A Comparison among Expert-based Fact-checking Websites

	Topics Covered	Content Analyzed	Assessment Labels
PolitiFact[3]	American politics	Statements	True; Mostly true; Half true; Mostly false; False; Pants on fire
The Washington Post Fact Checker[4]	American politics	Statements and claims	One pinocchio; Two pinocchio; Three pinocchio; Four pinocchio; The Geppetto checkmark; An upside-down Pinocchio; Verdict pending
FactCheck[5]	American politics	TV ads, debates, speeches, interviews and news	True; No evidence; False
Snopes[6]	Politics and other social and topical issues	News articles and videos	True; Mostly true; Mixture; Mostly false; False; Unproven; Outdated; Miscaptioned; Correct attribution; Misattributed; Scam; Legend
TruthOrFiction[7]	Politics, religion, nature, aviation, food, medical, etc.	Email rumors	Truth; Fiction; etc.
FullFact[8]	Economy, health, education, crime, immigration, law	Articles	Ambiguity (no clear labels)
HoaxSlayer[9]	Ambiguity	Articles and messages	Hoaxes, scams, malware, bogus warning, fake news, misleading, true, humour, spams, etc.

► *Expert-based Fact-checking Websites.* Recently, many websites have emerged to allow expert-based fact-checking better serve the public. We list and provide details on the well-known websites in Table 3. Some websites provide further information, for instance, *PolitiFact* provides "the PolitiFact scorecard", which presents statistics on the authenticity distribution of all the statements related to a specific topic (see an example on Donald Trump, 45th President of the United States, in Figure 2(a)). This information can help identify check-worthy topics (see Section 6 for details) that require further scrutiny for verification. Another example is *HoaxSlayer*, which is different from most fact-checking

[2] http://www.politifact.com/
[3] https://www.factcheck.org/
[4] https://www.washingtonpost.com/news/fact-checker
[5] https://www.snopes.com/
[6] https://www.truthorfiction.com/
[7] https://fullfact.org/
[8] http://hoax-slayer.com/



(a) (Expert-based) PolitiFact: the PolitiFact scorecard

(b) (Crowd-sourced) Fiskkit: the tag distribution

Fig. 2. Illustration of Manual Fact-checking Websites

websites that focus on information authenticity because it further classifies the articles and messages into e.g., hoaxes, spams and fake news. Though the website does not provide clear definitions for these categories, its information can be potentially exploited as ground-truth for comparative studies of fake news. In addition to the list provided here, a comprehensive list of fact-checking websites is provided by Reporters Lab at Duke University[10], where over two hundred fact-checking websites across countries and languages are listed. Generally, these expert-based fact-checking websites can provide ground-truth for the detection of fake news, in particular, under the broad definition (Definition 1). The detailed expert-based analysis these websites provide for checked contents (e.g., what is false and why is it false) carries invaluable insights for various aspects of fake news analysis (e.g., identifying check-worthy content [Hassan et al. 2017, 2015]); however, to date, such insights have not been well utilized (see a discussion in Section 6).

II. Crowd-sourced Manual Fact-checking. Crowd-sourced fact-checking relies on a large population of regular individuals acting as fact-checkers (i.e., the collective intelligence). Compared to expert-based fact-checking, crowd-sourced fact-checking is relatively difficult to manage, less credible and accurate due to the political bias of fact-checkers and their conflicting annotations, and has better (though insufficient) scalability. Hence, in crowd-sourced fact-checking one often needs to (i) filter non-credible users and (ii) resolve conflicting fact-checking results, both requirements becoming more critical as the number of fact-checkers grow. Nevertheless, crowd-sourcing platforms often allow fact-checkers to provide more detailed feedback (e.g., their sentiments or stances), which can be further explored in fake news studies.

► *Crowd-sourced Fact-checking Websites.* Unlike expert-based fact-checking, crowd-sourced fact-checking websites are still in early development. An example is *Fiskkit*[11], where users can upload articles, provide ratings for sentences within articles and choose tags that best describe it. The given sources of articles help (i) distinguish the types of content (e.g., news vs. non-news) and (ii) determine its credibility (Section 5 has details). The tags categorized into multiple dimensions allows one to study the patterns across fake and non-fake news articles (see Figure 2(b) for an example). While crowd-sourced fact-checking websites are not many, we believe more crowd-sourced platforms or tools will

[10] https://reporterslab.org/fact-checking/
[11] http://fiskkit.com/

arise as major Web and social media websites have realized the importance of identifying fake news (e.g., Google[12], Facebook[13], Twitter[14], and Sina Weibo[15].

2.2 Automatic Fact-checking

Manual [expert-based or crowd-sourced] fact-checking does not scale with the volume of newly created information, especially on social media. To address scalability, automatic fact-checking techniques have been developed, heavily relying on Information Retrieval (IR) and Natural Language Processing (NLP) techniques, as well as on network/graph theory [Cohen et al. 2011]. To review these techniques, we first provide a uniform standard representation of knowledge that can be automatically processed by machines and has been widely adopted in related studies [Nickel et al. 2016]:

Definition 3 (Knowledge). *A set of* (`Subject, Predicate, Object`) ***(SPO)*** *triples extracted from the given information that well-represent the given information.*

For instance, the knowledge within sentence "Donald Trump is the president of the U.S." can be (`DonaldTrump`, `Profession`, `President`). Based on the following knowledge representation, we will provide an automatic fact-checking framework to assess news authenticity. We will detail the automatic fact-checking process, the possible tasks within the process, the current standard methods for each task, as well as some open issues in automatic fact-checking. Note that as a systematic framework for automatic fact-checking is lacking, here we give more priority to organizing the related studies to present a clear automatic fact-checking process than to presenting each study in details.

The overall automatic fact-checking process is displayed in Figure 3. It can be divided into two stages: (I) fact extraction (also known as *knowledge-base construction*) and (II) fact-checking (also known as *knowledge comparison*). In fact extraction, knowledge is extracted often from open Web, which provides massive unstructured information in the form of online documents. The extracted knowledge is used to construct a *Knowledge Base* (KB) or a *Knowledge Graph*, each containing a set of facts (i.e., true knowledge) after proper data-cleanup. In fact-checking, the authenticity of the to-be-verified news contents is determined by comparing the knowledge extracted from the news contents to the facts stored in the constructed knowledge base or knowledge graph.

I. Fact Extraction. To collect facts, (i) knowledge is often extracted from the open Web as "raw facts", knowledge that is redundant, outdated, conflicting, unreliable or incomplete. These raw facts are further processed and cleaned up by (ii) knowledge processing tasks to (iii) build a knowledge-base or a knowledge graph.

i. Open Web and Knowledge Extraction. Knowledge extraction, also known as *relation extraction* [Pawar et al. 2017], aims to collect raw facts from the open Web. Broadly speaking, there are four (but not limited to) types of Web content: text, tabular data, structured pages and human annotations that contain relational information and can be utilized for knowledge extraction by different extractors [Dong et al. 2014; Grishman 2015]. Knowledge extraction can be further classified into *single-source* or *open-source* knowledge extraction. Single-source knowledge extraction, which relies on one comparatively reliable source (e.g., Wikepedia) to extract knowledge, is relatively efficient but often leads to incomplete knowledge (see related studies, e.g., in [Auer et al. 2007; Bollacker et al. 2008; Suchanek et al. 2007]). On the other hand, open-source knowledge extraction aims to fuse knowledge from distinct sources, which leads to less

[12] https://blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/
[13] https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/
[14] https://blog.twitter.com/2010/trust-and-safety
[15] http://service.account.weibo.com/ (sign in required)

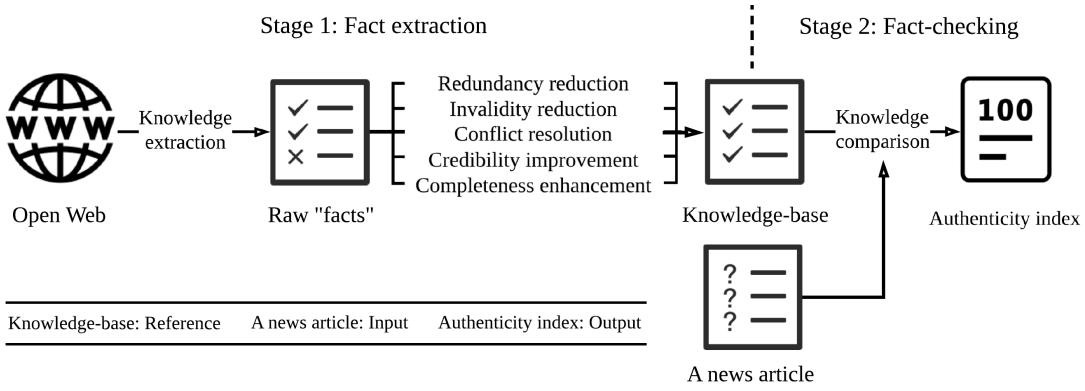


Fig. 3. Illustration of Automatic Fact-checking Process

efficiency, but more complete knowledge (see related studies, e.g., in [Carlson et al. 2010; Dong et al. 2014; Magdy and Wanas 2010; Nakashole et al. 2011, 2012; Niu et al. 2012]).

ii. Knowledge Processing Tasks. Knowledge extraction (i.e., relation extraction) from open Web leads to raw facts. The knowledge in raw facts can be (1) redundant, for example, `(DonaldJohnTrump, profession, President)` is redundant when having `(DonaldTrump, profession, President)` as `DonaldTrump` and `DonaldJohnTrump` refer to the same entity; (2) invalid, for example, `(Britain, joinIn, EuropeanUnion)` has been outdated and should be removed or updated; (3) conflicting, for example, `(DonaldTrump, bornIn, NewYorkCity)` and `(DonaldTrump, bornIn, LosAngeles)` are a pair with conflicting knowledge; (4) unreliabile, for example, the knowledge extracted from The Onion[16], a news satire organization, should be regarded as unreliable knowledge, and (5) incomplete. To address these issues and clean-up raw facts, one needs to conduct the following five tasks:

Task 1: ***Entity Resolution*** *to reduce redundancy.* Entity resolution, also known as *deduplication* [Steorts et al. 2016] or *record linkage* [Christen 2008], is the task of identifying all *mentions* that refer to the same real-world entity within a knowledge-base or across multiple knowledge-bases [Culotta and McCallum 2005]. Most related to fact-checking is *relational entity resolution* [Altowim et al. 2014], where current techniques are either distance-based (e.g., [Kouki et al. 2016]) or dependence-based (e.g., [Bhattacharya and Getoor 2007]). Entity resolution often requires pairwise similarity computations, which is computationally expensive. Blocking (or indexing) techniques are often used to address this computational complexity [Whang and Garcia-Molina 2012].

Task 2: ***Time Recording*** *to remove outdated knowledge.* The correctness of some facts depends on a specific time interval. One way to address this issue is through the Compound Value Type (CVT) construct, allowing facts to have beginning and end date annotations [Bollacker et al. 2008]; or one can *reify* current facts by adding extra assertions to them [Hoffart et al. 2013]. Nevertheless, fact-checking studies that have considered *timeliness* are to date limited, given the importance of timeliness in news[17] and fake news early detection, where facts can get rapidly updated (see Section 6 for further discussion).

Task 3: ***Knowledge Fusion*** *to handle conflicting knowledge.* Conflicting knowledge is common when using open-source knowledge extraction while it less frequent when adopting single-source knowledge extraction. To resolve

[16] https://www.theonion.com/
[17] http://www.axiapr.com/blog/elements-of-news

knowledge conflicts often support values are defined for facts [Magdy and Wanas 2010]. As an alternative, ensemble methods from machine learning [Dietterich 2000] can be utilized to combine multiple truth discovery algorithms, effectively discovering true values from conflicting ones. Credibility of websites from which the knowledge is extracted is often an important information utilized in the knowledge fusion process.

Task 4: ***Credibility Evaluation*** *to improve knowledge credibility.* Knowledge credibility evaluation focuses on analyzing the source website(s) of extracted knowledge, where approaches based on supervised learning [Esteves et al. 2018] and statistical inference [Dong et al. 2015] have been developed to analyze web content and/or links with other websites. We review website credibility evaluation studies in Section 5.

Task 5: ***Link Prediction*** *to infer new facts.* Raw facts extracted from online resources, particularly, using a single source, are far from complete. Hence, reliably inferring new facts based on existing facts is necessary to improve the knowledge-bases being built. Current methods that can help predict new facts can be classified into three groups based on their assumptions: (1) *latent feature models*, that assume the existence of knowledge-base triples is conditionally independent given latent features and parameters (e.g., RESCAL [Nickel et al. 2012]), (2) *graph feature models*, that assume the existence of triples is conditionally independent given observed graph features and parameters (e.g., Path Ranking Algorithm (PRA) [Lao and Cohen 2010]), and (3) *Markov Random Field (MRF) models*, that assume that existing triples have local interactions. A comprehensive review is in [Nickel et al. 2016].

iii. Knowledge-base. A Knowledge-Base is formed by "cleaned up" knowledge (i.e., a set of SPO triples). A graph structure, known as the *knowledge graph*, can be used to represent the SPO triples in a knowledge-base, where the entities (i.e., subjects or objects in SPO triples) are represented as nodes and relationships (i.e., predicates in SPO triples) are represented as edges. Knowledge-bases or knowledge graphs are suitable candidates for providing *ground truth* to fake news studies, i.e., we can reasonably assume the existing triples in a knowledge-base or knowledge graph represent true facts. However, for non-existing triples, there are three common assumptions:

- *Closed-world Assumption*: non-existing triples indicate false knowledge. While this assumption simplifies automatic fact-checking, it is rather dangerous as knowledge-bases are often sparsely populated or incomplete.
- *Open-world Assumption*: non-existing triples indicate unknown knowledge that can be either true or false. This assumption can lead to more accurate fact-checking results, but the results depend on the way the authenticity for non-existing triples is inferred from existing ones.
- *Local Closed-world Assumption [Dong et al. 2014]*: the authenticity of non-existing triple is based on the following rule: suppose $T(s, p)$ is the set of existing triples for a given subject s and predicate p. For any $(s, p, o) \notin T(s, p)$, if $|T(s, p)| > 0$, we say the triple is incorrect; if $|T(s, p)| = 0$, the authenticity of triple (s, p, o) is unknown.

Instead of building a knowledge-base from open Web for each fact-checking task, in recent years, several large-scale knowledge graphs have been constructed, e.g., YAGO [Hoffart et al. 2013; Suchanek et al. 2007], Freebase[18] [Bollacker et al. 2008], NELL [Carlson et al. 2010], PATTY [Nakashole et al. 2012], DBpedia [Auer et al. 2007], Elementary/DeepDive [Niu et al. 2012], and Knowledge Vault [Dong et al. 2014]. However, we should point out that the existing knowledge graphs are insufficient as a source of ground-truth especially for news, and can be improved from various perspectives. For example, due to the timeliness of news, news articles are often not around "common knowledge", but mostly about recent events; hence, a *dynamic* knowledge-base might greatly improve the accuracy of news fact-checking. In addition, it has been verified that fake news spreads faster than true news [Vosoughi et al. 2018], which attaches great importance to fast news fact-checking. Current research on building knowledge-bases has focused on constructing knowledge-bases

[18]Freebase was closed in 2016

with as many facts as possible. However, fast news fact-checking requires not only identifying parts of the to-be-verified news that is check-worthy (see Section 6 for a discussion), but also a knowledge-base that only stores as many "valuable" facts as possible (i.e., a knowledge-base simplification process). The speed of fact checking is also highly related to the strategies and methods used for news fact-checking, which we will discuss in the following.

II. Fact-checking. To evaluate the authenticity of news articles, we need to further compare the knowledge extracted from to-be-verified news contents (i.e., SPO triples) with the facts stored in the constructed or existing knowledge-base(s) or knowledge graph(s), i.e., true knowledge. Generally, the fact-checking strategy for a SPO triple (`Subject`, `Predicate`, `Object`) is to evaluate the possibility that the edge labeled `Predicate` exists from the node labelled `Subject` to the node representing `Object` in a knowledge graph. Specifically,

Step 1: *Entity locating*. In this step, `Subject` (`Object`) is matched with a node in the knowledge graph that represents the same entity as the `Subject` (`Object`). Note that representing the same entity is not equivalent to representing the same string, e.g., `Donald J. Trump` and `Donald John Trump` both represent the same entity; hence, *entity resolution* techniques [Getoor and Machanavajjhala 2012] can be used to identify proper matchings.

Step 2: *Relation verification*. Triple (`Subject`, `Predicate`, `Object`) is considered truth if an edge labeled `Predicate` from the node representing `Subject` to the one representing `Object` exists in the knowledge graph. Otherwise, its authenticity is (1) false based on closed-world assumption, or (2) determined after knowledge inference.

Step 3: *Knowledge inference*. When the triple (`Subject`, `Predicate`, `Object`) does not exist in the knowledge-graph, the probability for the edge labeled `Predicate` to exist from the node representing `Subject` to the one representing `Object` can be computed, e.g., using link prediction methods such as semantic proximity [Ciampaglia et al. 2015], discriminative predicate path [Shi and Weninger 2016], or LinkNBed [Trivedi et al. 2018].

We conclude this section by providing a formal definition for news fact-checking (i.e., news authenticity evaluation) summarizing our discussion:

Problem 1 (News authenticity evaluation). *Assume a to-be-verified news article is represented as a set of knowledge statements (i.e., SPO triples)* $(s_i, p_i, o_i), i = 1, 2, \ldots, n$. *Let* $\boldsymbol{G}_{KB}$ *refer to a knowledge graph containing a set of facts (i.e., true knowledge) denoted as* $(s_{t_j}, p_{t_j}, o_{t_j}), j = 1, 2, \ldots, m$. *The task to evaluate the authenticity of each triple* (s_i, p_i, o_i) *is to identify a function* $\mathcal{A}$ *that assigns an authenticity value* $A_i \in [0, 1]$ *to the corresponding* (s_i, p_i, o_i) *by comparing it with every* $(s_{t_j}, p_{t_j}, o_{t_j})$ *in the knowledge-graph, where* $A_i = 1$ *indicates the triple is true and* $A_i = 0$ *indicates it is false. The final authenticity index* $A \in [0, 1]$ *of the to-be-verified news article is obtained by aggregating all* A_i*'s. To summarize,*

$$\begin{aligned} \mathcal{A} : (s_i, p_i, o_i) &\overset{\boldsymbol{G}_{KB}}{\rightarrow} A_i, \\ A = \mathcal{G}(A_1, A_2, \cdots, A_n), & \end{aligned} \tag{1}$$

where $\mathcal{G}$ *is an aggregation function of choice. The to-be-verified news article is true if* $A = 1$*, and is irrefutable false if* $A = 0$*. Specifically, function* $\mathcal{A}$ *can be formulated as*

$$\mathcal{A}((s_i, p_i, o_i), \boldsymbol{G}_{KB}) = P(\text{edge labeled } p_i \text{ linking } s_i' \text{ to } o_i' \text{ in } \boldsymbol{G}_{KB}), \tag{2}$$

where $P(\cdot)$ *denotes the probability, and* s_i' *and* o_i' *are the matched entities to* s_i *and* o_i *in* $\boldsymbol{G}_{KB}$*, respectively:*

$$s_i' = \arg\min_{s_{t_j}} ||s_i - s_{t_j}||, \tag{3}$$

$$o_i' = \arg\min_{o_{t_j}} ||o_i - o_{t_j}||. \tag{4}$$

3 STYLE-BASED STUDY OF FAKE NEWS

Similar to when fake news is studied from a knowledge-based perspective (Section 2), studying fake news from a style-based perspective also emphasizes on investigating the news content. However, knowledge-based studies aim to evaluate the authenticity of the given news, while style-based studies aim to assess news intention, i.e., is there an intention to mislead the public or not? Formally, fake news style can be defined as

Definition 4 (Fake news style). *A set of quantifiable characteristics (e.g., machine learning features) that can well represent fake news and differentiate fake news from truth.*

While the development of style-based fake news studies is still in its early stages with only a limited number of such studies [Bond et al. 2017; Pisarevskaya 2015; Potthast et al. 2017; Volkova et al. 2017], *deception analysis and detection* has long been an active area of research and has focused on the general style of deceptive (i.e., intentionally false) content across various types of information. We will review studies of deception in Section 3.1, which covers various types of information, e.g., online communication, reviews (which we will review further in Section 5) as well as news articles. These deception studies in general do not consider the inherent characteristics of the style of news articles. We will discuss these characteristics in Section 3.2 with some potential research opportunities. We hope this approach will help the reader better understand the universal style of deception within various types of information and facilitate further research on fake news style based on specific characteristics within news articles.

3.1 Deception Analysis and Detection

Deception analysis aims to investigate style of deceptive content across various types of information, e.g., online communications [Hancock et al. 2007; Pak and Zhou 2015; Rubin 2010; Zhou et al. 2004b], reviews [Li et al. 2014; Mukherjee et al. 2013b; Ott et al. 2011; Popoola 2018; Shojaee et al. 2013; Zhang et al. 2016], statements [Fuller et al. 2009; Humpherys et al. 2011], essays and short text [Afroz et al. 2012; Braud and Søgaard 2017; Pérez-Rosas and Mihalcea 2014, 2015], images and videos [Abouelenien et al. 2017; Gogate et al. 2017; Pérez-Rosas et al. 2015], as well as fake news [Bond et al. 2017; Pisarevskaya 2015; Potthast et al. 2017; Volkova et al. 2017].

Deception studies are mainly concerned with (I) *deception style theories*, i.e., why content style can help investigate deception, (II) *style-based features and patterns* that can (well) represent and capture deception, and (III) *deception detection strategies*: how style can be utilized to detect fake news and other types of deceptive information.

I. Deception Style Theories. Intuitively, the content style of deceptive information (e.g., fake news) that aims to deceive readers (e.g., with exaggerated expressions and strong emotions) should be somewhat different from that of the truth. Indeed, forensic psychological studies (e.g., *Undeutsch hypothesis*, *reality monitoring*, *interpersonal deception theory*, and *four-factor theory* [Siering et al. 2016]) have shown that statements derived from factual experiences differ in content and quality from those that are based on fantasy. These intuitions and fundamental theories have motivated and made possible style-based deception studies, whether for statements, online communications, online reviews, or news articles. The performance of deception detection using content style, discussed later in this section, has further confirmed the validity of these theories, with deception detection accuracy rates varying between 60% to 90% in experiments.

II. Style-based Features and Patterns. As provided in Definition 4, the content style is commonly represented by a set of quantifiable chateristics, often machine learning features. Generally, these features can be grouped into *attribute-based language features* or *structure-based language features*.

Table 4. Attribute-based Language Features

Attribute Type	Feature	[Zhou et al. 2004b]	[Fuller et al. 2009]	[Afroz et al. 2012]	[Shojaee et al. 2013]	[Hauch et al. 2015]	[Pak and Zhou 2015]	[Siering et al. 2016]	[Zhang et al. 2016]	[Braud and Søgaard 2017]	[Bond et al. 2017]	[Potthast et al. 2017]	[Volkova et al. 2017]
Quantity	Character count			✓	✓								
	Word count	✓	✓	✓	✓	✓	✓	✓	✓				
	Noun count								✓				
	Verb count	✓	✓			✓		✓	✓				
	Number of noun phrases	✓											
	Sentence count	✓	✓	✓		✓		✓					
	Paragraph count											✓	
	Number of modifiers (e.g., adjectives and adverbs)	✓	✓	✓				✓	✓				
Complexity	Average number of clauses per sentence	✓						✓					
	Average number of words per sentence	✓	✓	✓	✓	✓		✓	✓			✓	
	Average number of characters per word	✓	✓	✓	✓	✓		✓					
	Average number of punctuations per sentence	✓	✓	✓				✓					
Uncertainty	Percentage of modal verbs	✓	✓	✓		✓		✓					
	Percentage of centainty terms	✓	✓	✓		✓		✓					✓
	Percentage of generalizing terms		✓			✓							
	Percentage of tentative terms		✓	✓		✓							✓
	Percentage of numbers and quantifiers			✓	✓	✓							✓
	Number of question marks			✓	✓								
Subjectivity	Percentage of subjective verbs	✓							✓	✓			
	Percentage of report verbs												✓
	Percentage of factive verbs												✓
	Percentage of imperative commands												✓
Non-immediacy	Percentage of passive voice	✓	✓			✓							
	Percentage of rhetorical questions												✓
	Self reference: 1st person singular pronouns	✓	✓	✓		✓		✓	✓	✓			✓
	Group reference: 1st person plural pronouns	✓	✓	✓		✓		✓	✓	✓			✓
	Other reference: 2nd and 3rd person pronouns	✓	✓	✓		✓		✓	✓	✓			✓
	Number of quotations			✓	✓							✓	✓
Sentiment	Percentage of positive words	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓
	Percentage of negative words	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓
	Number of exclamation marks			✓	✓								
	Activation: the dynamics of emotional state	✓	✓										
Diversity	Lexical diversity: unique words or terms (%)	✓	✓	✓	✓	✓		✓	✓				
	Content word diversity: unique content words (%)	✓	✓			✓			✓				
	Redundancy: unique function words (%)	✓	✓	✓	✓	✓			✓				
Informality	Typographical error ratio: misspelled words (%)	✓				✓		✓	✓				
Specificity	Temporal ratio	✓	✓			✓					✓		
	Spatial ratio	✓	✓			✓					✓		
	Sensory ratio	✓	✓			✓	✓	✓			✓		
	Causation terms		✓			✓				✓			
	Exclusive terms		✓			✓							
Readablity (e.g., Flesch-Kincaid and Gunning-Fog index)				✓								✓	

The studies labeled with gray background color investigate news articles.

Table 5. Patterns of Deceptive Content Attributes

Attribute Type	[Newman et al. 2003]	[Fuller et al. 2009]	[Matsumoto and Hwang 2015]	[Derrick et al. 2013]	[Zhou et al. 2004b]	[Hancock et al. 2007]	[Anderson and Simester 2014]	[Braun and Van Swol 2016]	[Bond and Lee 2005]	[Zhou and Zenebe 2008]	[Ali and Levine 2008]	[Humpherys et al. 2011]
Quantity		+	+	−	+	+	+	−		+	+	+
Complexity					−							+
Uncertainty			−		+	+		+			−	−
Non-immediacy	+	+	+		+	+	+	+	+	+		+
Sentiment	−	+	−			−		+	−		+	+
Diversity		−		−	−		−			−	−	−
Informality					+					+		
Specificity	−	−	+		−					−		−

+: The attribute is positively related to the existence of deception;
−: The attribute is negatively related to the existence of deception.

► *Attribute-based Language Features.* Attribute-based language features, also known as *theory-oriented language features*, are mostly inspired by or directly derived from related aforementioned deception theories. For example, feature "sensory ratio" captures the phenomenon explained by *reality monitoring* that fake events are expressed by lower levels of sensory information compared to true events. Based on our investigation, attribute-based language features that describe the content style can be grouped along ten parallel dimensions: *quantity*, *complexity*, *uncertainty*, *subjectivity*, *non-immediacy*, *sentiment*, *diversity*, *informality*, *specificity*, and *readability* (see Table 4). While attribute-based language features can be highly pertinent, explainable and predictable, they are often poor (or less flexible) in quantifying deception content style compared to structure-based features. Specifically, attributed-based features often demand some additional levels of quantification or computing, which can be time-consuming and attaches greater importance to proper feature evaluation and filtering for deception detection. To investigate which categories of attribute-based features are most relevant for deception detection, we have analyzed studies utilizing such features across different categories in Table 5. As shown in Table 5, deceptive information exhibits higher levels of quantity, non-immediacy, informality, while lower levels of diversity and specificity. We believe there is a strong need for more systematic research on identifying most informative and theory-oriented features that can best capture deception in content.

► *Structure-based Language Features.* Structure-based language features describe content style from (at least) four language levels: (i) *lexicon*, (ii) *syntax*, (iii) *semantic* and (iv) *discourse*. Structure-based features are also known as *technique-oriented features*, as their quantification mostly relies on mature Natural Language Processing (NLP) techniques; hence, are independent of research topics or areas. Specifically, the main task at lexicon level is to assess the frequency statistics of letter(s), word(s), etc., which can be conducted appropriately using n-gram models. At the syntax level, shallow syntax tasks are performed by Part-Of-Speech (POS)-taggers that facilitate POS tagging and analyzing. Deep syntax level tasks are performed by Probabilistic Context-Free Grammars (PCFG) with parse trees that enable Context-Free Grammars (CFG) analysis. At the semantic level, Linguistic Inquiry and Word Count (LIWC) is often used to

Table 6. Performance of Structure-based Language Features for Deception Detection

	Level(s)	Feature(s)	[Ott et al. 2011]	[Feng et al. 2012a]	[Shojaee et al. 2013]	[Mukherjee et al. 2013b]	[Li et al. 2014]	[Pérez-Rosas and Mihalcea 2014]	[Pérez-Rosas et al. 2015]	[Pérez-Rosas and Mihalcea 2015]	[Li et al. 2017b]	[Ott et al. 2011]	[Shojaee et al. 2013]	[Li et al. 2014]	[Pérez-Rosas et al. 2015]	[Abouelenien et al. 2017]	[Braud and Søgaard 2017]	[Pérez-Rosas et al. 2015]
Within Levels	**Lexicon**	UG	.884	.729		.663	**.668**	**.691**	.609	**.695**	.825	.884		.645	**.763**	.585	.717	**.678**
		BG	.896	.708		.661					.804	**.889**					.696	
		UG+BG		.738							.637							
		Others			.810								.700					
	Syntax	POS	.730			.564	.638		**.695**					.690		.513	.717	
		CFG		.742					.654							.513		
		Others	.768		.760					.525			.690		.627			.534
	Semantic	LIWC					.633	***.691***	.602	.534				**.695**	.500	.504		.661
	Discourse	RR															.553	
Across Levels	**Lexicon + Syntax**	UG+POS		.733							**.831**							
		UG+CFG		**.769**														
		BG+POS				**.664**					.808							
		BG+CFG				.659												
		UG+BG+POS															**.760**	
		Others+Others			**.840**								**.740**					
	Lexicon + Semantic	UG+LIWC							.622							**.594**		
		BG+LIWC	**.898**			.661												
	Lexicon + Syntax + Semantic	UG+POS+ LIWC								.653					.636			.576

UG: Unigram BG: Bigram POS: Part-of-Speech tags CFG: Context-Free Grammar (particularly refers to lexicalized production rules)
LIWC: Linguistic Inquiry and Word Count RR: Rhetorical Relations

provide around eighty semantic classes for semantic features. Finally, Rhetorical Structure Theory (RST) and rhetorical parsing tools capture rhetorical relations as features at the discourse level [Pisarevskaya 2015]. While in deception studies, these features are less pertinent, explainable and predictable, computing them is relatively easy compared to attribute-based features. To assess the relative necessity and importance of features at various language levels, we further analyze the performance (accuracy) of a series of deception detection studies that involve features at more than one language level, which is shown in Table 6. For each study, we highlight the feature(s) within or across language levels that achieve optimal performance in bold face, and underline the highest performance that features at a single language level can achieve, leading to the following conclusions:

- Within a single language level, lexicon-level features are almost always performing the best compared to syntax-, semantic-, or discourse-level features, i.e., for studies that involve features within language levels, 11/14 of them achieve better performance at the lexicon level.
- Combining features across language levels almost always performs better than using features within a single language level, i.e., out of the twelve studies containing features within and across language levels, eight of them perform better using cross-language-level features.

While such analyses provides some insights, clearly, more thorough experiments are necessary to systematically assess the relative importance of structure-based features at various language levels.

Table 7. Performance Comparison of Categories of Classification Techniques

Performance \ Algorithms		Decision Trees (e.g., C4.5)	Neural networks (e.g., Deep Nets)	Bayesian (e.g., NB)	Instance-based (e.g., k-NN)	Kernel-based (e.g., SVM)	Rule-based (e.g., RIPPER)
Complexity	Speed of learning	Average	Low	High	High	Low	Average
	Speed of classification	High	High	High	Low	High	High
	Model Complexity (e.g., number of parameters)	Average	High	Low	Average	High	Average
Robustness	Tolerance to irrelevant features	Average	Low	Average	Average	High	Average
	Tolerance to redundant features	Average	Average	Low	Average	High	Average
	Tolerance to dependent features	Average	High	Low	Low	High	Average
	Tolerance to noise	Average	Average	High	Low	Average	Low
	Ability to handle overfitting	Average	Low	High	High	Average	Average
Extendability	Incremental learning ability	Average	Average	High	High	Average	Low
Interpretability	Explainability of classifications	High	Low	High	Average	Low	Good

NB: Naïve Bayes k-NN: k-Nearest Neighbors SVM: Singular Vector Machine RIPPER: Repeated Incremental Pruning to Produce Error Reduction

III. Deception Detection Strategies. A common strategy for style-based deception detection is to utilize a feature vector representing the content style of the given information within a machine learning framework to predict whether the information is deceptive (i.e., a classification problem) or how deceptive it is (a regression problem). So far, most related studies use supervised learning techniques, where labeled training data (i.e., a set of feature vectors with their corresponding labels: `deceptive` vs. `normal`) is necessary. As (i) most of these classifiers are not evaluated on similar deception datasets and (ii) classifiers perform best for machine learning settings they were initially designed for (i.e., *no free lunch theorem*), it is illogical to determine algorithms that perform best for deception detection. Hence, we have only compared several commonly used supervised learning frameworks for deception detection from the following four perspectives: *complexity*, *robustness*, *extendability* and *interpretability* in Table 7. More detailed comparisons for supervised learners can be found in [Kotsiantis et al. 2007], a comprehensive review of several classifiers, and [Fernández-Delgado et al. 2014], an in-depth study of the performance of 179 classifiers from 17 families on 121 datasets.

3.2 Deception in News

The process presented thus far for deception studies does not distinguish between news and other types of information, e.g., statements, online messages, and reviews, and uniformly regards them as deception (i.e., disinformation), false information with an intention to mislead the public. Here, we will focus on how analyzing content style of fake news varies from studying deception in other types of information. We will also discuss how unique characteristics of fake news introduce variations or potential tasks in style-based fake news studies, which so far have not been studied.

Analyzing Fake News Content Style. Analyzing content style of fake news can be fundamentally different from analyzing deception in other types of information. Firstly, fundamental theories that have inspired attribute-based language features listed in Table 4, e.g., Undeutsch hypothesis, are mostly developed in forensic psychology. While these theories are suitable candidates for the analysis of deceptive statements, they do not directly address fake news due to clear differences. For example, while uncertainty and informality can vary from statement to statement, both are rare within news articles. News articles that exhibit high uncertainty or have many typographical errors are considered dubious and often do not even qualify as check-worthy news content (Section 6 provides a detailed discussion). Hence, to detect deception in news articles based on style, more subtle cues and *patterns* should be sought, supported by closely-related *theories*, especially in journalism. Secondly, news articles involve various domains, e.g., politics, economics, education, and health, as well as various languages and topics. Hence, a content-based fake news analysis demands a *cross-domain*, *cross-language*, or *cross-topic* analysis, all of which have been less explored in the current literature. Thirdly, bursts of fake news are often initiated by important events (e.g., a presidential election) [Vosoughi et al. 2018],

with fake-news creators that often financially benefit from such outbreaks. These financial market strongly incentivizes abrupt and real-time evolution in content style of fake news to avoid being detected, often beyond what can be detected by current developments in style-based fake news studies. This constant evolution in content style demands a real-time representation and/or learning of news content style, where e.g., *deep learning* can be helpful [Gogate et al. 2017; Li et al. 2017b; Ren and Ji 2017; Wang et al. 2018]. Section 6 has more details on deep learning for fake news analysis.

Style-based Fake News Detection. The general deception detection strategy discussed can be utilized for style-based fake news detection. However, in addition to supervised learning, semi-supervised learning can also play an important role for two particular reasons. First, the number (and size) of the available datasets containing labeled (fake vs. normal) news articles are limited. Second, it is difficult to construct a "gold-standard" dataset for such studies as humans have been empirically proven to be poor deception detectors. Social psychology and communications studies demonstrate that human ability to detect deception is only slightly better than chance: typical accuracy rates are in the 55%-58% range, with a mean accuracy of 54% over 1,000 participants in over 100 experiments [Rubin 2010]. Furthermore, manual labeling does not scale with the volume of newly created information, especially on social media.

Independent of the learning framework used (supervised vs. semi-supervised), style-based fake news detection can complement knowledge-based fake news detection, which determines news authenticity, by assessing news intention. We conclude this section by providing a formal definition for news intention evaluation summarizing our discussion.

Problem 2 (News intention evaluation). *Assume a to-be-verified news article can be represented as a set of n content features denoted by feature vector $\vec{f} \in \mathbb{R}^n$. The task to evaluate the intention of the to-be-verified news article based on its content style is to identify a function $\mathcal{I}$, such that*

$$\mathcal{I} : \vec{f} \xrightarrow{TD} I \tag{5}$$

where $I \in [0, 1]$ is the intention index; $I = 1$ indicates a non-harmful intention for the news article and $I = 0$ indicates that the news article intends to deceive the public. $TD = \{(\vec{f}_k, I_k) : \vec{f}_k \in \mathbb{R}^n, I_k \in [0, 1], k = 1, 2, \ldots, m\}$ is the training dataset. The training dataset helps estimate the parameters within $\mathcal{I}$, consisting of a set of news articles represented by the same set of features ($\vec{f}_k$) with known intention indices (I_k).

4 PROPAGATION-BASED STUDY OF FAKE NEWS

Different from knowledge- and style-based perspectives that study fake news based on its content, when studying fake news from a propagation-based perspective, one takes advantage of the information related to the dissemination of fake news, e.g., how it propagates and users spreading it. Here, we first present (i) empirical patterns and (ii) mathematical models of fake news propagation in Sections 4.1 and 4.2, respectively. Next, we introduce, categorize, and compare techniques that utilize such patterns, models, or other news propagation information to detect fake news in Section 4.3.

However, before such patterns, models, or detection techniques are introduced, one has to answer a series of fundamental questions. For example, how can one describe or represent (fake) news propagation? What measures are appropriate for characterizing the propagation of (fake) news? Is there any difference between the propagation of fake news versus regular news (e.g., in terms of related measures)? Does fake news from various domains (e.g., politics, economy, and education), topics (e.g., natural disasters, presidential elections, and health), websites (e.g., Twitter, Reddit, and WeChat), or languages (e.g., English, Chinese, and Russian) propagate differently? These are a few examples of many questions regarding fake news propagation, the answers to many of which are still unclear.

Hence, to provide a unified representation for fake news propagation, we first define *fake news cascade* in Definition 5, a formal representation of fake news propagation that has been adopted in many studies (e.g., [Ma et al. 2018b; Vosoughi et al. 2018; Wu et al. 2015]). With this formalization, studying fake news from a propagation-based perspective boils down to studying fake news cascades. Several basic measures are then introduced to characterize fake news cascades. Based on fake news cascades, fake news propagation can be qualitatively or quantitatively analyzed. When conducting a qualitative analysis, we provide fake news propagation patterns (Section 4.1), and when performing a quantitative analysis, we present mathematical models that can well explain and model fake news propagation.

Definition 5 (Fake News Cascade). *A fake news cascade is a tree or tree-like structure that represents the propagation of a certain fake news article on a social network of users (Figure 4 provides examples). The root node of a fake news cascade represents the user who first published the fake news (i.e., creator or initiator); Other nodes in the cascade represent users that have subsequently posted the article by forwarding/posting it after it was posted by their parent nodes, which they are connected to via edges. A fake news cascade can be represented in terms of the number of steps (i.e., hops) fake news has traveled (i.e., hop-based fake news cascade) or the times it was posted (i.e. time-based fake news cascade).*

***Hop-based fake news cascade**, often a standard tree, allowing natural measures such as*

- *Depth: the maximum number of steps (hops) fake news has travelled within a cascade.*
- *Breadth (at k hops): the number of users that have received the fake news k steps (hops) after it was initially posted within a cascade.*
- *Size: the total number of users in a cascade.*

***Time-based fake news cascade**, often a tree-like structure, allowing natural measures such as*

- *Lifetime: the longest interval during which fake news has been propagating.*
- *Real-time heat (at time t): the number of users posting/forwarding the fake news at time t.*
- *Overall heat: the total number of users that have forwarded/posted the fake news.*

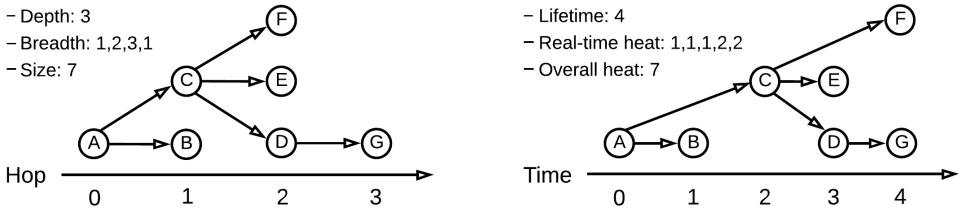
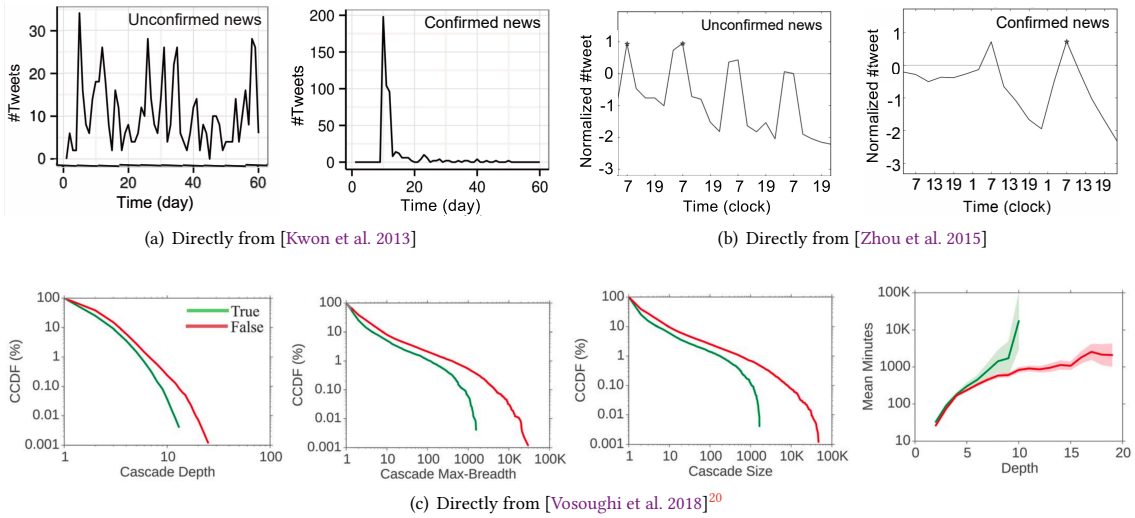


Fig. 4. Illustration of Fake News Cascades. On the left, we provide a hop-based fake news cascade, where the x-axis provides the number of steps fake news has travelled. On the right, we provide a time-based fake news cascade, where the x-axis denotes the of times at which fake news has been posted/forwarded.

Note that a specific fake news can lead to multiple simultaneous cascades due to multiple initiating users. Furthermore, often within a fake news cascade, nodes (users) are represented with a series of attributes and additional information, e.g., whether they (support or oppose) the fake news, their profile information, previous posts, and their comments.

4.1 Fake News Propagation Patterns

Fake news propagation patterns can divided into (1) patterns that only describe the propagation of fake news and (2) those that compare fake news propagation to that of regular news. Both types of patterns often provide measurements obtained from fake news cascades, e.g., *fake news travels farther and faster than true news*. As discussed, such measurements depend on how fake news propagation is represented (i.e., using a hop-based vs. time-based fake news cascade).



(a) Directly from [Kwon et al. 2013]

(b) Directly from [Zhou et al. 2015]

(c) Directly from [Vosoughi et al. 2018][20]

Fig. 5. Fake News Propagation Patterns

Studying both types of fake news propagation patterns have specific benefits. In particular, when comparing fake news propagation to the propagation of regular news (or other types of information), one can obtain patterns that can help distinguish between fake news and other types of news. These pattern will further facilitate effective modeling, detecting, and blocking of fake news. On the other hand, patterns that focus only on how fake news propagates are invaluable for understanding the variations that are expected in fake news propagations, e.g., how fake news propagations vary within different domains, topics, languages, or websites. Such patterns allow one to speedup fake news early detection and to identify check-worthy contents (see Section 6 for details).

We present fake news propagation patterns identified in recent studies. However, due to limited public data on fake news, only a few studies exist, especially under the narrow definition of fake news (i.e., intentionally false news). Hence, recent studies have extended pattern discovery to *news rumors*, comparing the propagation of confirmed and unconfirmed news. Others have focused on confirmed news, and compared the propagations of false and true confirmed news, or only the propagation of false news within various domains. We briefly review the major patterns identified:

- *Unconfirmed news often gets renoticed.* As shown in Figure 5(a), unconfirmed news tends to exhibit multiple and periodic discussion spikes, whereas confirmed news typically has a single prominent spike. The pattern becomes even more clearer after controlling for the total number of posts within a day (see Figure 5(b)).
- *False news spreads farther, faster, and more widely than true news,* as the cascade depth, max-breadth and size of false news cascades are generally greater than that of true news, while the time taken for false news cascades to reach any depth and size is less than that for true news cascades[19] (see Figure 5(c)).
- *Political false news spreads farther, faster, and more widely than false news within other domains*, where the trends between false political news and false news within other domains are similar to the ones between false news and true news in Figure 5(c). More details can be found in the study by Vosoughi et al. [Vosoughi et al. 2018].

[19]The study also adopts a measure called structural virality [Goel et al. 2015] and shows that false news has greater structural virality than true news.
[20]CCDF: Complementary Cumulative Distribution Function

4.2 Models for Fake News Propagation

Fake news propagation patterns are the outcome of a qualitative analysis of fake news propagation, while a quantitative analysis is often achieved by introducing mathematical models for fake news propagation. An accurate and explainable fake news model can play an important role in realistically describing, quantifying and predicting fake news. Given the time series describing fake news propagation (e.g., the times fake news was posted within a cascade), a general approach to model (or forecast) such propagation is through regression analysis, e.g., linear regression, Poisson regression, and regression trees [Du et al. 2014; Najar et al. 2012]. Regression modeling has been standard in many disciplines and hence we do not provide further details here (see e.g., [Draper and Smith 2014] for an introduction). In addition to regression modeling, some classical models in (I) epidemics and (II) economics are also suitable candidates to capture propagation dynamics. In the following, we will peruse how such models can be modified for modeling fake news propagation.

► *A modified epidemic diffusion model.* An epidemic model is a proper candidate for modeling and predicting the overall heat (i.e., number of spreaders) for fake news. This is due to the fact that fake news propagation shares many similarities to how infectious diseases evolve or spread. Hence, by analyzing the dynamics of fake news propagation one could glean insight into how fake news spreads online [Kucharski 2016]. Here, we aim to concretely illustrate how one can mathematically connect fake news propagation with classic epidemic models.

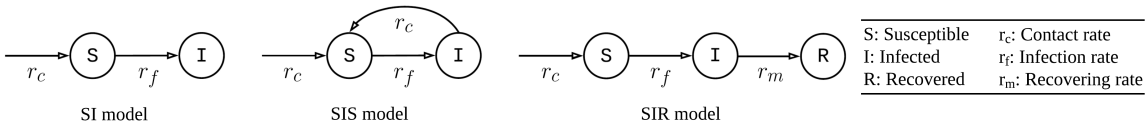


Fig. 6. Illustration of Classic Epidemic Models

Generally, there are three states for individuals within an epidemic model: **S** (Susceptible, refers to those who are potential candidates, but are not yet infected with the disease), **I** (Infected, refers to those who have been infected with the disease) and **R** (Recovered, refers to those who were infected with the disease but have recovered). A user within one state can transition to another with a certain rate. By allowing various states and state transitions, one can have the three classical epidemic models: SI (Susceptible-Infected), SIS (Susceptible-Infected-Susceptible) and SIR (Susceptible-Infected-Recovered), which are illustrated in Figure 6. In order to model fake news propagation using such epidemic models, the following steps should be subsequently taken.

Step 1: *Specifying States and Transition Rates.* The first step towards utilizing an epidemic model for fake propagation is to specify user states and transitions rates. One simple approach is to consider the following,

User states:
- **S**: Users who can potentially, but have not yet spread fake news.
- **I**: Users that have spread fake news.
- **R**: Users that have spread fake news, but after which they have removed the posts.

Transition rates:
- r_c: The rate at which a user receives or reads fake news.
- r_f: The rate at which a user spreads fake news after he or she has received or read it.
- r_m: the rate at which a user removes his or her fake news posts after it has been posted/forwarded.

Step 2: *Model Construction.* When selecting an epidemic model for fake news propagation, the corresponding assumptions of the model should align with the characteristics of fake news propagation. For instance, one can reasonably assume that users cannot recover [from posting fake news] once they are infected (i.e., have posted fake news) as

backfire effect, *conservatism bias*, and *Semmelweis reflex* reveal that individuals often reject ideas that are against their established beliefs; a recent empirical study has also shown that fake news is incorrect but hard to correct for readers [Roets et al. 2017]. In such case, the diffusion model of fake news can be simply defined as

$$\begin{aligned} s_{t_{k+1}} &= s_{t_k} + (N \cdot r_c)(r_f \cdot \frac{s_{t_k}}{N} \cdot (1 - \frac{s_{t_k}}{N})) \\ &= s_{t_k} + r_c \cdot r_f \cdot s_{t_k} \cdot (1 - \frac{s_{t_k}}{N}), \end{aligned} \tag{6}$$

where $s_{t_k}(s_{t_{k+1}})$ is the number of users that have spread fake news up to time $k(k+1)$ among a total of N users.

Step 3: *Identifying Transition Rates.* Transition rates should also be determined based on the real-world diffusion characteristics of fake news, where fundamental theories listed in Table 2 can be helpful. For example, r_f can be positively correlated to the number of infected users (i.e., users that have spread fake news) supported by, e.g., *normative influence theory* that implies individuals tend to conform to the behavior and attitudes of others.

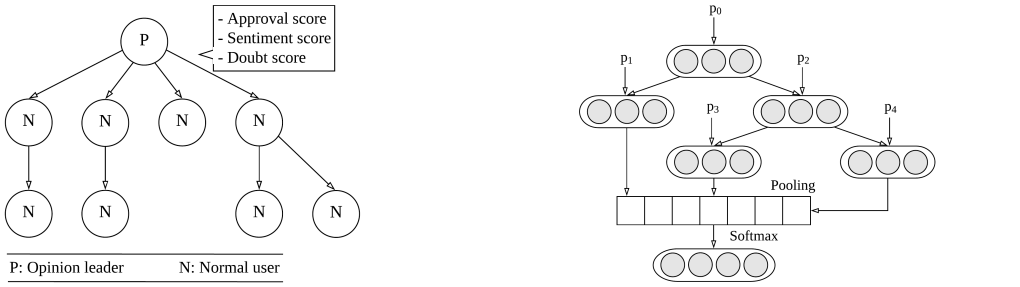
▶ *An economic-related model.* One can utilize economic models to captures and predict individuals' decision making and behavior towards fake news, e.g., when to forward or delete fake news. One such model is a two-player (publishers and consumers) strategy game [Shu et al. 2017a]. In this model, each player makes a decision or behaves by trading off two kinds of utilities: publishers correspond to a long-term utility g_p (i.e., reputation) and a short-term utility b_p (i.e., profit); consumers correspond to an information utility g_c (i.e., they prefer truth) and a psychological utility b_c (i.e., they prefer the information to confirm their preexisting beliefs, i.e., *confirmation bias*). When the short-term utility of a publisher dominates its overall utility, i.e., $\mathcal{U}(g_p, b_p) < b_p$, where $\mathcal{U}(\cdot)$ is some overall utility function, we conclude that the publisher will create fake news; if psychological utility of a consumer dominates his or her overall utility, i.e., $\mathcal{U}(g_c, b_c) < b_c$, we conclude that the consumer will spread fake news.

4.3 Propagation-based Fake News Detection

We have discussed related patterns and models that characterize or can possibly characterize fake news propagations. In the following, we will discuss how the aforementioned patterns and models can help detect fake news. We group current studies based on flexible architectures and typical strategies that one can take to detect fake news based on propagation information. These studies can be classified into (1) *cascade-based fake news detection* techniques, which take direct advantage of news propagation paths and news cascades to identify fake news, and (2) *network-based fake news detection* methods, which construct a flexible network from cascades, using which fake news is indirectly predicted.

4.3.1 Cascade-based Fake News Detection. When using cascades to detect fake news, one either distinguishes fake news by (1) computing the similarity of its cascade to that of other true/false news or (2) properly representing its cascade using an informative representation that facilitates distinguishing fake news from true news.

▶ *Utilizing Cascade-Similarity.* A common strategy to compute the similarity between cascade of some news (i.e., a graph) to cascades of other news (i.e., another graph) is to utilize graph kernels [Vishwanathan et al. 2010]. Such cascade similarities can be utilized as features within a supervised learning framework to detect fake news. For example, Wu et al. [2015] propose a graph-kernel based hybrid SVM classifier which captures the high-order propagation patterns (i.e., similarity between cascades) in addition to semantic features such as topics and sentiments. Specifically, they introduce user roles (*opinion leader* or *normal user*) as well as approval, sentiment, and doubt scores among user posts towards the to-be-verified fake news into news cascades. Figure 7(a) illustrates the structure. By assuming that fake news cascades are different from true ones, the authors detect fake news using a random walk (RW) graph kernel $\mathcal{K}_{RW}(\cdot, \cdot)$, which



(a) Utilizing Cascade-Similarity (e.g., Graph-Kernels [Wu et al. 2015]) (b) Utilizing Cascade-Representation (e.g., Deep nets [Ma et al. 2018b])

Fig. 7. Framework Architectures of Cascade-based Fake News Detection

can quantity the difference (distance) between any two cascades $\boldsymbol{T}_1, \boldsymbol{T}_2$. Mathematically,

$$\mathcal{K}_{RW}(\boldsymbol{T}_1, \boldsymbol{T}_2) = \sum_{i,j=1}^{|\mathbf{V}_\times|} [\sum_{n=0}^{\infty} \lambda_n \mathbf{A}_\times^n]_{ij}, \tag{7}$$

where $\mathbf{V}_\times$ denotes the vertex set of the direct product $\mathbf{V}_\times = \mathbf{V}(\boldsymbol{T}_1 \times \boldsymbol{T}_2)$, $\mathbf{A}_\times$ denotes the adjacency matrix of the direct product $\mathbf{A}_\times = \mathbf{A}(\boldsymbol{T}_1 \times \boldsymbol{T}_2)$, and $\lambda_0, \lambda_1, \cdots$ is a sequence of weights satisfying $\lambda_i \in [0, 1]$ for all $i \in \mathbb{N}$.

► *Utilizing Cascade-Representation.* When designing cascade representations, one seeks informative representations that can be utilized as features within a supervised learning framework. Such representation can be developed using *feature-engineering*, e.g., by representing a cascade using the number of nodes a feature; however, such techniques are not automatic. As an alternative one can conduct *representation learning*, often achieved via deep learning. For example, Ma et al. [2012] use deep learning by constructing Recursive Neural Networks (RNNs), a tree-structured neural network, based on fake news cascades [Ma et al. 2018b]. A top-down RNN model with Gated Recurrent Unites (GRUs) [Cho et al. 2014] is shown in Figure 7(b). Specifically, for each node j with a post on a certain news report represented as a TF-IDF vector $\mathbf{p}_j$, its hidden state $\mathbf{h}_j$ is recursively determined by $\mathbf{p}_j$ itself and the hidden state of its parent node $\mathcal{P}(j)$, denoted as $\mathbf{h}_{\mathcal{P}(j)}$. Mathematically, $\mathbf{h}_j$ can be calculated by

$$\mathbf{h}_j = \mathbf{z}_j \odot \sigma_h(\mathbf{W}_h \mathbf{p}_j \mathbf{V} + \mathbf{U}_h(\mathbf{h}_{\mathcal{P}(j)} \odot \mathbf{r}_j)) + (1 - \mathbf{z}_j) \odot \mathbf{h}_{\mathcal{P}(j)}\,, \tag{8}$$

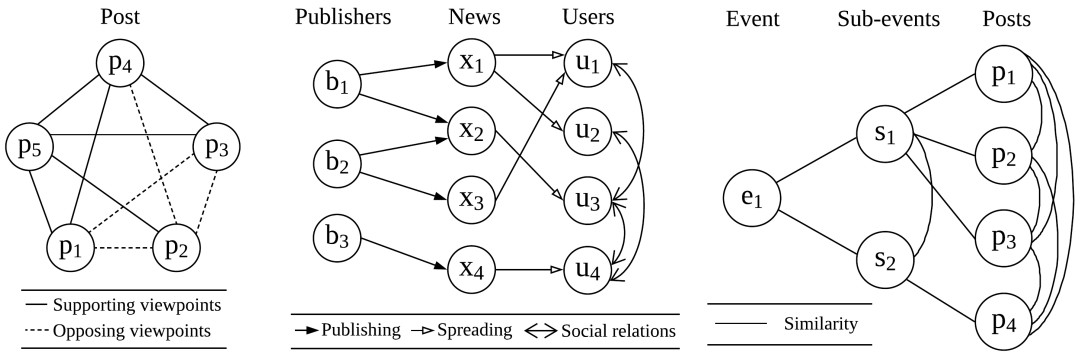
$$\mathbf{z}_j = \sigma_g(\boldsymbol{W}_z \mathbf{p}_j \mathbf{V} + \mathbf{U}_z \mathbf{h}_{\mathcal{P}(j)})\,, \tag{9}$$

$$\mathbf{r}_j = \sigma_g(\boldsymbol{W}_r \mathbf{p}_j \mathbf{V} + \mathbf{U}_r \mathbf{h}_{\mathcal{P}(j)})\,, \tag{10}$$

where $\mathbf{z}_j$ is an update gate vector, $\mathbf{r}_j$ is a reset gate vector, $\mathbf{W}_*$, $\mathbf{U}_*$, and $\mathbf{V}$ denote parameter matrices, σ_g is a sigmoid function, σ_h is a hyperbolic tangent, and $\odot$ denotes entry-wise product. In this way, the learned representations are computed for all leaf nodes of a cascade, denoted as $n_{l_1}, n_{l_2}, \cdots, n_{l_m}$ for $m \in \mathbb{N}_+$, which are inputs to the pooling layer which computes the final representation for the to-be-verified news. The pooling output $\mathbf{h}$ is obtained by $\mathbf{h}_i = \max[\mathbf{h}_{l1}, \mathbf{h}_{l2}, \cdots, \mathbf{h}_{lm}]_{ik}, k = 1, 2, \cdots, m$. Finally, the label of to-be-verified news report is predicted as

$$\tilde{y} = \text{Softmax}(\mathbf{Q}\mathbf{h} + \mathbf{b}), \tag{11}$$

where $\mathbf{Q}$ and $\mathbf{b}$ are parameters. The model (parameters) can be further trained (estimated) by minimizing some cost function, e.g., squared error [Ma et al. 2018b] or cross-entropy [Zhang et al. 2018].



(a) Homogeneous Network [Jin et al. 2016] (b) Heterogeneous Network [Shu et al. 2017b] (c) Hierarchical Network [Jin et al. 2014]

Fig. 8. Common Architectures for Network-based Fake News Detection

Comparing these two typical cascade-based fake news detection strategies, similarity-based studies allow on to additionally consider the roles that users play in fake news propagation; however, computing similarities between two cascades can be computationally expensive. One the other hand, representation-based methods can automatically represent to-be-verified news; however, the depth of cascades may challenge such methods as it is equal to the depth of the neural network, to which e.g., deep learning methods are often sensitive. Indeed, we can see from Figure 5(c) that the depth of fake news cascades can be near twenty, which can negatively impact performance of deep nets.

4.3.2 Network-based Fake News Detection. Network-based fake news detection constructs flexible networks to indirectly capture fake news propagation. The constructed networks can be homogeneous, heterogeneous, or hierarchical.

► *Homogeneous Network.* Homogeneous networks are networks containing a single type of node and a single type of edge [Shu et al. 2018]. A typical homogeneous network is a *stance network* [Jin et al. 2016], where nodes are news-related posts by users, and edges represent supporting (+) or opposing (-) relations among each pair of posts, e.g., the similarity between each pair of tweets that can be calculated using a distance measure such as Jensen-Shannon [Jin et al. 2016] or Jaccard distance [Jin et al. 2014]. The network is illustrated in Figure 8(a). Fake news detection using a stance network boils down to evaluating the credibility of news-related posts (i.e., lower credibility = fake news), which can be further cast as a graph optimization problem. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ denote the adjacency matrix of the aforementioned stance network with n nodes and $\mathbf{c} \in \mathbb{R}^n$ denote the vector of node credibility scores. By assuming that supporting posts have similar credibility values, the cost function in [Zhou et al. 2004a] can be adopted and the problem can be defined as

$$\arg\min_{\mathbf{c}} \underbrace{\mu||\mathbf{c}-\mathbf{c}_0||^2}_{\text{Fitting constraint}} + \underbrace{(1-\mu)\sum_{i,j=1}^{n} \mathbf{A}_{ij}\left(\frac{\mathbf{c}_i}{\sqrt{\mathbf{D}_{ii}}} - \frac{\mathbf{c}_j}{\sqrt{\mathbf{D}_{jj}}}\right)^2}_{\text{Smoothness constraint}} \tag{12}$$

where $\mathbf{c}_0$ refers to true credibility scores of training posts, $\mathbf{D}_{ij} = \sum_k \mathbf{A}_{ik}$, and $\mu \in [0, 1]$ is a regularization parameter.

► *Heterogeneous Network.* Heterogeneous networks have multiple types of nodes or edges. A major example in fake news analysis is the tri-relationship network among news publishers, news articles, and news spreaders (i.e., users) shown in Figure 9(a). For such a network, a hybrid framework [Shu et al. 2017b] with three main components can help detect fake news: (I) entity embedding and representation, (II) relation modeling, and (III) semi-supervised learning:

I. *Entity Embedding and Representation.* The first step is to learn a latent representation for news articles and spreaders, where Nonnegative Matrix Factorization (NMF) can be adopted. Mathematically,

$$\min \left\| \mathbf{X} - \mathbf{D}\mathbf{V}^T \right\|_F^2 \quad s.t.\ \mathbf{D}, \mathbf{V} \geq 0, \tag{13}$$

$$\min \left\| \mathbf{Y} \odot \mathbf{A} - \mathbf{U}\mathbf{T}\mathbf{U}^T \right\|_F^2 \quad s.t.\ \mathbf{U}, \mathbf{T} \geq 0, \tag{14}$$

where $\mathbf{X} \in \mathbb{R}_+^{m \times t}$ is the given article-word matrix for m news articles. $\mathbf{X}$ will be factorized as $\mathbf{D} = [\mathbf{D}_L, \mathbf{D}_U]^T \in \mathbb{R}_+^{m \times d}$ (i.e., article latent feature matrix) and $\mathbf{V} \in \mathbb{R}_+^{t \times d}$, where $\mathbf{D}_L \in \mathbb{R}_+^{r \times d}$ is the article latent feature matrix for r labeled articles with label vector $\mathbf{y}_L \in \{-1, 1\}^r$ (-1 indicates true news and 1 indicates fake news) and $\mathbf{D}_U \in \mathbb{R}_+^{(m-r) \times d}$ is the one for unlabeled articles; $\mathbf{A} \in \{0, 1\}^{n \times n}$ is the known spreader-spreader adjacency matrix to be factorized as $\mathbf{U} \in \mathbb{R}_+^{n \times f}$ (i.e., spreader latent feature matrix) and $\mathbf{T} \in \mathbb{R}_+^{f \times f}$, and $\mathbf{Y} \in \mathbb{R}^{n \times n}$ controls the contribution of $\mathbf{A}$. As Shu et al. [Shu et al. 2018] point out, one can design other user embedding methods that can preserve various network properties of the corresponding friendship networks, e.g., community structure.

II. *Relation Modeling.* Assume the partisan bias of publishers are known: left (-1), least-biased (0), right ($+1$), denoted by $\mathbf{b} \in \{-1, 0, 1\}^l$ for l publishers. To model publisher-article relationships, one can assume that partisan bias of publishers $\mathbf{b}$ can be represented using the learned latent features of articles that they publish. Mathematically,

$$\min \left\| \bar{\mathbf{P}}\mathbf{D}\mathbf{q} - \mathbf{b} \right\|_2^2 \tag{15}$$

where $\bar{\mathbf{P}} \in \mathbb{R}^{l \times m}$ is the normalized publisher-article relation matrix, and $\mathbf{q} \in \mathbb{R}^d$ is the weighting vector.

To model spreader-article relationships, we can reasonably assume that non-credible (credible) users spread fake news (true news), which leads to the following mathematical formulation

$$\min \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{r} \mathbf{W}_{ij}\mathbf{c}_i \left(1 - \frac{1 + \mathbf{y}_{L_j}}{2}\right) \left\| \mathbf{U}_i - \mathbf{D}_{Lj} \right\|_2^2}_{\text{True news}} + \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{r} \mathbf{W}_{ij}\left(1 - \mathbf{c}_i\right) \left(\frac{1 + \mathbf{y}_{L_j}}{2}\right) \left\| \mathbf{U}_i - \mathbf{D}_{Lj} \right\|_2^2}_{\text{Fake news}}, \tag{16}$$

where $\mathbf{W} \in \{0, 1\}^{n \times r}$ is the spreader-article relation matrix.

III. *Semi-supervised Learning.* Given embeddings, one can perform supervised learning by learning a weight vector $\mathbf{w} \in \mathbb{R}^d$ for article latent features by solving the following optimization problem:

$$\min \left\| \mathbf{D}_L\mathbf{w} - \mathbf{y}_L \right\|_2^2. \tag{17}$$

With w, one can predict whether a news is fake or true by computing $\text{sign}(w\mathcal{N})$, where $\mathcal{N} \in \mathbb{R}^d$ is some to-be-verified news article represented using latent features, i.e., $\mathcal{N}$ is a row of $\mathbf{D}_U$.

Other fake news detection studies based on heterogeneous networks can be seen in, e.g., [Gupta et al. 2012], where the authors establish a *user-post-news event* network for news verification; they design a PageRank-like algorithm and further obtain news event credibility through a similar optimization formulation to Equation (12). Another recent example can be seen in [Zhang et al. 2018], where a Recurrent Neural Network (RNN) model is designed to detect fake news through exploring news creators, articles, subjects and their relationships.

► *Hierarchical Network.* In hierarchical networks, various types of nodes and edges form set-subset relationships (i.e., a hierarchy). An example is shown in Figure 9(c), which contains relationships across (i.e., *hierarchical relationships*) and within (i.e., *homogeneous relationships*) news events, sub-events and posts. In such networks, news verification is also transformed into a graph optimization problem [Jin et al. 2014], extending the optimization in Equation (12).

5 CREDIBILITY-BASED STUDY OF FAKE NEWS

When studying fake news from a credibility-based perspective, one studies fake news based on news-related and social-related information. For instance, intuitively, a news article published on unreliable website(s) and forwarded by unreliable user(s) is more likely to be fake news than news posted by authoritative and credible users. Hence, studying fake news study from a credibility perspective thus overlaps with a propagation-based study of fake news, where studies have explored the relationships between news articles and components such as publishers [Shu et al. 2017b], users [Gupta et al. 2012; Shu et al. 2017b; Zhang et al. 2018] and posts [Gupta et al. 2012; Jin et al. 2014, 2016; Ma et al. 2018b; Wu et al. 2015]. Here, we separate a credibility-based study of fake news from that based on propagation as, at times, detecting fake news can be achieved using only auxiliary information and without considering news content or social relationships. For example, "All the false news stories identified in BuzzFeed News analysis came from either fake news websites that only publish hoaxes or from hyperpartisan websites that present themselves as publishing real news." [Silverman 2016] Such observations imply that fake news detection can in some cases be simplified to detecting an unreliable website source. Similarly, credibility of other sources information such as comments on fake news can help detect fake news. Based on current studies, we review how fake news can be detected by assessing the credibility for (1) news headlines, (2) news source, (3) news comments, and (4) news spreaders.

5.1 Assessing News Headline Credibility

Assessing news headline credibility often reduces to detecting *clickbaits*, headlines whose main purpose is to attract the attention of visitors and encourage them to click on a link to a particular web page. Examples of such clickbaits include "*You'll never look at Barbie dolls the same once you see these paintings*" and "*23 things parents should never apologize for.*" Though some clickbaits are "good" (or say, clever) for product advertising or marketing, few should be allowed in news articles. First, clickbaits do attract eyeballs but are rarely newsworthy [Pengnate 2016]. Second, as the readers keep switching to new articles after being baited by the headlines, the attention residue from these constant switches result in a cognitive overload, deterring them from reading more informative and in-depth news stories [Mark 2014]. Finally, clickbaits are often paired with fake news articles as they are powerful tools for fake news to gain high click-rate and public trust, as explained by, e.g., information-gap theory[21] and validity effect [Boehm 1994] (see Table 2 for details). The success of clickbaits on social networks has led to many social media sites such as Facebook[22] to take immediate actions against them. It should be noted that while news articles with clickbaits are generally unreliable, not all such news articles are fake news. While studying clickbaits is similar to a style-based study of fake news (see Section 3), here, we focus on specific methods that assess news headline credibility through clickbait detection and regard it as an indirect way to detect fake news. Current clickbait detection studies use linguistic features (e.g., term frequencies, readability, and forward references [Biyani et al. 2016]) and non-linguistic features (e.g., webpage links [Potthast et al. 2016], user interests [Chakraborty et al. 2016; Zheng et al. 2017] and headline stance [Bourgonje et al. 2017]) within a supervised learning framework, e.g., gradient boosted decision trees [Biyani et al. 2016; Zheng et al. 2017], to detect or block clickbaits [Chakraborty et al. 2016]. In addition to such studies, empirical studies have shown that clickbaits can be characterized by a cardinal number, easy readability, strong nouns and adjectives to convey authority and sensationalism [Vijgen 2014]. Deep learning-based clickbait detection has also emerged recently to avoid feature engineering (see recent studies, e.g., in [Anand et al. 2017; Rony et al. 2017; Zhou 2017]).

[21] "Information-gap theory views curiosity as arising when attention becomes focused on a gap in one's knowledge. Such information gaps produce the feeling of deprivation labeled curiosity. The curious individual is motivated to obtain the missing information to reduce or eliminate the feeling of deprivation." (p. 87) [Loewenstein 1994]
[22] https://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/

5.2 Assessing News Source Credibility

There has been evidence that most fake news stories come from either fake news websites that only publish hoaxes, or from hyperpartisan websites that present themselves as publishing real news [Silverman 2016]. Such observations indicate that the quality, credibility, and political bias of source websites, to some extent, determine the quality and credibility of news. Web credibility analysis has been an active research area, developing many practical techniques such as web ranking algorithms. Traditional web ranking algorithms such as PageRank [Page et al. 1998] and HITS [Kleinberg 1999] assess website credibility with the goal to improve search engines responses to user search queries. However, the weaknesses of these traditional web ranking algorithms provide opportunities for *web spam*, a major indicator of unreliable websites, to improve website rankings unjustifiably and thus motivate the development of web spam detection. A comprehensive survey can be seen in [Spirin and Han 2012]. Web spam can be categorized as (i) *content spam*, which leads to a spam webpage appearing among normal search results primarily due to fake word frequencies (e.g., TF-IDF scores). Content spam includes spamming of title, body, meta-tags, anchors and URLs; (ii) [outgoing and incoming] *link spam*, where the former targets mostly HITS-like algorithms to achieve high hub scores and the latter enhances website authority scores by attacking PageRank-like algorithms; and (iii) other types of spam such as *cloaking*, *redirection* and *click spams*. Algorithms to detect web spam thus can be classified into (i) content-based algorithms, which analyze web content features, such as word counts and content duplication [Fetterly et al. 2005; Ntoulas et al. 2006]; (ii) link-based algorithms, which detect web spam by utilizing graph information [Zhou and Pei 2009], learning statistical anomalies [Dong et al. 2015], and performing techniques such as (dis)trust propagation [Gyöngyi et al. 2004], link pruning [Bharat and Henzinger 1998] and graph regularization [Abernethy et al. 2010]; and (iii) other algorithms that are often based on click stream [Dou et al. 2008] or user behavior [Liu et al. 2015]. While not many, some website credibility assessment techniques with a special focus on fake news detection have been developed. For instance, the assessment of web credibility in [Esteves et al. 2018] is based on a set of content and link features within a machine learning framework, and that in [Dong et al. 2015] uses joint inference in a multi-layer probabilistic model.

5.3 Assessing News Comments Credibility

In Section 4, we have shown that user posts [Dungs et al. 2018; Jin et al. 2014, 2016; Ma et al. 2018a,b; Wu et al. 2015] have been utilized to analyze and detect fake news, where one can explore stance and opinion of users towards news articles. User comments on news websites and social media carry invaluable information on stances and opinions as well; however, comments are often ignored. Furthermore, current fake news studies that glean such user stance and opinions from comments, have paid little attention to comment credibility; a news article with many opposing viewpoints can be maliciously attacked by others and one can receive many compliments by recruiting "fake supporters", which is often the case with products on e-commerce websites. Hence, we discuss comment credibility of fake news along with comment credibility for products on e-commerce websites, which has long been an active research area known as *review spam detection*. Models to evaluate comment credibility can be classified into (I) content-based, (II) behavior-based and (III) graph(network)-based models. Below, we review these models and compare these models in Table 8.

I. Content-based Models. Content-based models assess comment credibility using series of language features extracted from user comments, and follow a strategy similar to that of style-based fake news detection (see Section 3 for details). Related studies can be seen in, e.g., [Jindal and Liu 2008; Li et al. 2014, 2017b; Mukherjee et al. 2013b; Ott et al. 2011; Popoola 2018; Ren and Ji 2017; Shojaee et al. 2013; Zhang et al. 2016].

Table 8. Model Comparison for Review Spam Detection

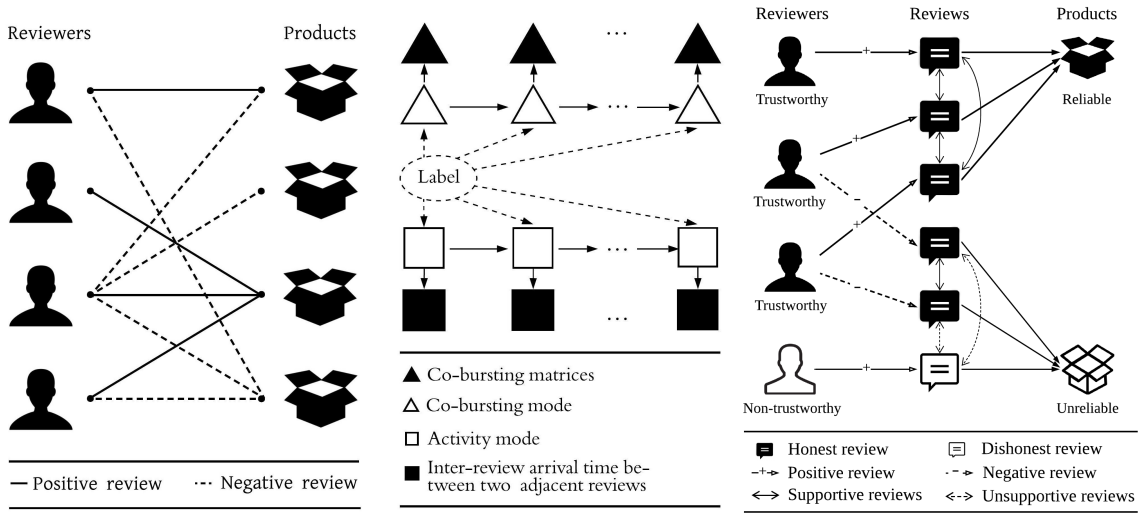
	Content-based models	**Behavior-based models**	**Graph-based models**
Domain Sensitivity	Comparatively sensitive	Insensitive	Insensitive
Assumptions/Constraints	Few (i.e., feature-based)	Often necessary	Often necessary
Model Explanability	Comparatively low	Comparatively high	Comparatively high
Relationships among Entities	Excluded	Excluded	Included

II.Behavior-based Models. Behavior-based models often leverage indicative features of unreliable comments extracted from the metadata associated with user behavior. Reviewing review spam detection studies, we organize these related behavioral features into five categories: burstiness, activity, timeliness, similarity, and extremity. Table 9 has the details. For instance, Mukherjee et al. [2013a] propose Author Spamicity Model (ASM), containing behavioral features from all five categories used within a Bayesian setting. Behavioral studies have shown that normal reviewers' arrival times are stable and uncorrelated to their temporal rating patterns, while spam attacks are usually bursty and either positively or negatively correlated to the rating (see Xie et al. [2012] for details). This allows one to view review spam detection as the detection of a co-bursty multi-dimensional time series.

III. Graph-based Models. Graph-based models take into account the relationships among reviewers, comments, products, etc. To assess credibility these model often adopt (1) Probabilistic Graphical Models (PGMs), (2) web ranking algorithms and centrality measures, or (3) matrix decomposition techniques.

Table 9. Behavioral Features for Review Spam Detection

Category	**Features**	[Mukherjee et al. 2013a]	[Hooi et al. 2016]	[Xie et al. 2012]	[Lim et al. 2010]	[Feng et al. 2012b]	[Wu et al. 2010]
Burstiness	Measuring the sudden promotion or descent of average rating, number of reviews, etc. for a product. This category of features emphasize on the *collective* behavior among reviewers	✓		✓			✓
Activity	Measuring the total or maximum number of reviews a reviewer writes for a single product or products in a fixed time interval. This category of features emphasize on the *individual* behavior of reviewers	✓			✓		✓
Timeliness	Measuring how early a product has received the review(s), or one reviewer has posted the reviews for products	✓			✓		
Similarity	Measuring the (near) duplicate reviews written by a single reviewer or for a product, or measuring the rating deviation of one reviewer from the others for a product	✓			✓		
Extremity	Measuring the ratio or number of extreme positive or negative reviews of a product, or for a reviewer among products	✓	✓			✓	✓



(a) [Akoglu et al. 2013] (b) [Li et al. 2017a] (c) [Wang et al. 2011]

Fig. 9. Illustration of Graph-based Models for Review Spam Detection

(1) **Probabilistic Graphical Models (PGMs)** are frequently adopted either in directed setting (Bayesian Network, BN) [Hooi et al. 2016; Rayana and Akoglu 2016] or undirected setting (Markov Random Field, MRF) [Rayana and Akoglu 2015]. For instance, based on MRF and Loopy Belief Propagation (LBP), Fei et al. [2013] detect review spams via constructing the network whose nodes represent reviewers and edges indicate the co-occurrence of reviews (i.e., reviewers write reviews in the same burst); the network in [Akoglu et al. 2013] is bipartite, where reviewers and products are connected in terms of positive (+) or negative (−) reviews a reviewer has posted for a product (see Figure 9(a) for the illustration). Li et al. [2017a] propose a supervised Coupled Hidden Markov Model (CHMM) based on the bimodal distribution and co-bursting pattern among review spams, where the model architecture has been illustrated in Figure 9(b).

(2) **Web ranking algorithms and centrality measures** are used to assign credibility scores to nodes (e.g., comments or users) within a graphs. Examples include studies by Wang et al. [2011] and Mukherjee et al. [2012], both proposing graph-based review spam detection models using PageRank- or HITS-like algorithms: the former construct a reviewers-reviews-products graph, and defines scores for trustiness of users, honesty of reviews, and reliability of products (see Figure 9(c) for the illustration), while the latter explores the relationships among group spams, member spams and products.

(3) **Matrix Decomposition** is used as an alternative, often to better represent reviews in some latent dimensions that can help better distinguish spam from non-spam. For example, Wang et al. [2016], in addition to defining eleven relations between reviewers and products, employ tensor decomposition for review spam detection.

5.4 Assessing News Spreader Credibility

Users play the most important role in fake news propagation. They are able to engage in fake news dissemination in multiple ways such as sharing, forwarding, liking and reviewing. In this process, all users can be broadly grouped

into (a) *malicious users*, with low credibility and (b) *normal users*, with relatively high credibility. User credibility has been used to study fake news, particularly, from a propagation-based perspective (Section 4), either directly [Gupta et al. 2012], or indirectly through their posts [Dungs et al. 2018; Jin et al. 2014, 2016; Ma et al. 2018a,b; Wu et al. 2015] and interconnections [Shu et al. 2017b]. Nevertheless, only a few studies have considered user vulnerability as the borderline between malicious and normal becomes unclear – normal users can frequently and unintentionally participate in fake news activities as well. This special phenomenon does not appear in many other social activities, e.g., the aforementioned product rating and reviewing. Here, we reclassify users into *participants* and *non-participants* in fake news activities. Participants are further grouped based on their intentions into (I) *malicious users* and (II) *naïve users*, each group being characterized with various strategies for fake news detection and intervention.

I. Malicious Users. Malicious users intentionally spread fake news, driven often by monetary and/or non-monetary benefits (e.g., power and popularity). Broadly speaking, each malicious user belongs to one of the following three categories, being either a (1) *bot*, a *software application* that runs automated tasks (scripts) over the Internet[23], (2) *troll*, a *person* who quarrels or upsets users to distract and sow discord by posting inflammatory and digressive, extraneous, or off-topic messages with the intent of provoking others into displaying emotional responses and normalizing tangential discussion,[24] or (3) *cyborg*, an account registered by a human as a camouflage to execute automated programs performing online activities [Shu et al. 2017a]. As it has been suggested, millions of malicious accounts have participated in online discussions around 2016 U.S. presidential election.[25] Hence, identifying and removing malicious users online is critical for detecting and blocking fake news. Recent studies can be seen in, for example, in [Cheng et al. 2017; Shao et al. 2017], that analyze behavioral patterns of bots and trolls in fake news propagation, in [Chu et al. 2012], that automatically classifies humans, bots and cyborgs in terms of their profiles, tweeting behavior and content posted, in [Cai et al. 2017], that detects social bots by jointly modeling behavior and content posted, and [Morstatter et al. 2016], that proposes a bot detection approach achieving both comparatively high precision and recall.

II. Naïve Users. Naïve users are vulnerable normal users who unintentionally engage in fake news propagation – they mistake fake news as truth. Humans are known to be poor at detecting fake news [Rubin 2010]; however, few studies have paid attention to identifying naïve users or have assessed the impact of such users on fake news propagation. To facilitate such studies, we have analyzed theories that explain the motivations of naïve users to engage in fake news propagation, and summarize all such influential factors which stem from either (i) social influence or (ii) self-influence.

- *Social influence* refers to environmental and exogenous factors such as *network structure* or *peer pressure* that can influence the dynamics of fake news. For example, network structure physically defines the potential exposure time to fake news for social network users, which is positively correlated to (i) how trustworthy individuals consider a piece of information (i.e., *validity effect*) and (ii) how likely is for that information to be forwarded. Similarly, peer pressure psychologically impacts user behavior towards fake-news-related activities (as indicated by the *bandwagon effect*, *normative influence theory* and *social identity theory*).
- *Self influence* refers to the internal and inherent attributes of users that can impact how they engage with fake news. For example, as *confirmation bias* and *naïve realism* imply, users will have a higher likelihood of trusting fake news or engage in its related activities when it confirms their preexisting knowledge. Note that preexisting knowledge of online users can often be approximated by assessing their generated content, e.g., their posts.

[23] https://en.wikipedia.org/wiki/Internet_bot
[24] https://en.wikipedia.org/wiki/Internet_troll
[25] http://comprop.oii.ox.ac.uk/research/public-scholarship/resource-for-understanding-political-bots/

6 DISCUSSION AND FUTURE WORK

In Sections 2 to Section 5, four perspectives to study fake news have been reviewed, summarized and evaluated: knowledge-based, style-based, propagation-based and credibility-based. An orthogonal approach to study fake news is from a technique perspective, where the aforementioned fake news studies can be generally grouped into (I) feature-based fake news studies, and/or (II) relation-based fake news studies.

I. Feature-based Fake News Studies. Feature-based fake news studies focus on manually generating or automatically learning a set of observed and/or latent features to well represent fake news. The goal is often to use these features to detect or block fake news within a machine learning framework. When manually engineering features, they are often inspired by related patterns and observed phenomena that are potentially useful. Feature selection for such engineered features is often time-consuming and labor-intensive. Related fake news studies can be seen in, e.g., [Bond et al. 2017; Potthast et al. 2017; Volkova et al. 2017]. When automatically learning features, deep learning techniques or matrix factorization are often utilized. Deep learning techniques can skip feature engineering and automatically represent features; thus, have been widely applied to many problems including fake news detection [Liu and Wu 2018; Ruchansky et al. 2017; Wang et al. 2018; Yang et al. 2018; Zhang et al. 2018]. Similarly, matrix factorization can facilitate latent feature extraction to detect fake news [Shu et al. 2017b]. Reviewing the aforementioned perspectives in Sections 2-5, style-based fake news studies (Section 3) are mostly feature-based studies. Regarding fake news detection as a feature-based classification or regression problem, similar to the framework outlined in Problem 2, content features discussed in Section 3 can be combined with image features [Jin et al. 2017a,b; Wang et al. 2018], user features [Jin et al. 2017a; Tacchini et al. 2017], cascade/network features [Kwon et al. 2013; Wu et al. 2015], temporal features [Kwon et al. 2013; Zhou et al. 2015], among other features to form a comprehensive set of features that can be used within a supervised learning framework to predict fake news.

II. Relation-based Fake News Studies. Relation-based fake news studies, on the other hand, emphasize on the relation among objects and features, and aim to study fake news through these relationships that can be explicit or implicit, sequential or non-sequential, and single- or multi-dimensional. Reviewing the aforementioned perspectives in Sections 2-5, knowledge-based and propagation-based fake news studies often depend on relationships, where the knowledge-based studies rely on multi-dimensional relationships between subjects and objects extracted from news contents, and propagation-based studies rely on relationships among, e.g., news articles, user posts (in particular, user stance [Dungs et al. 2018; Jin et al. 2016; Ma et al. 2018a]), and publishers. Probabilistic graph models [Dungs et al. 2018], tensor decomposition [Socher et al. 2013; Wang et al. 2016], multi-task learning [Ma et al. 2018a], PageRank-like algorithm [Gupta et al. 2012], among similar techniques have played an important role in analyzing such relationships.

6.1 Potential Research Opportunities for Fake News Studies

Based on fake news characteristics and current state of fake news research, we highlight the following potential research tasks that can facilitate a deeper understanding of fake news, as well as help improve the performance and efficiency of current fake news detection studies.

I. Fake News Early Detection. Fake news early detection aims to detect fake news at an early stage before it becomes wide-spread so that one can take early actions for fake news mitigation and intervention. Early detection is especially important for fake news as the more fake news spreads, the more likely for people to trust it (i.e., *validity effect* [Boehm 1994]). Meanwhile, it is difficult to correct users' perceptions after fake news has gained their trust [Roets et al. 2017]. To

detect fake news at an early stage during its lifespan one has to primarily rely on news content and limited social-related information and face multiple challenges. First, newly emerged events often generate new and unexpected knowledge that has not been stored in existing knowledge-bases or knowledge graphs, or is difficult to be inferred. Second, features that have well represented the style of fake news in the past may not be as useful in the future, especially due to the constant evolution of deceptive writing style. Finally, limited information may adversely impact the performance of machine learning techniques. To address these challenges and detect fake news early, one can focus on

(1) *timeliness of ground truth*, for example, technologies related to dynamic (real-time) knowledge-base construction should be developed to realize timely updates of ground truth;
(2) *feature compatibility*, specifically, features that can capture the generality of deceptive writing style across topics, domains, language, and the like, as well as the evolution of deceptive writing style, where Generative Adversarial Networks (GANs) [Wang et al. 2018] and Recurrent Neural Networks (RNNs) [Chen et al. 2017] have played or can play to their strengths;
(3) *verification efficiency*, for example, by identifying check-worthy content and topics [Hassan et al. 2017, 2015] one can improve the efficiency of fake news detection, which we will discuss as follows.

II. Identifying Check-worthy Content. With new information created and circulated online at an unprecedented rate, identifying check-worthy content can improve the efficiency of fake news detection and intervention by prioritizing content or topics that are check-worthy. Whether a given content or topic is check-worthy can be measured by, e.g., (i) its newsworthiness or potential to influence the society, for example, if it is related to national affairs and can lead to public panic, and (ii) its historical likelihood of being fake news. Thus a content or topic that is newsworthy, can potentially influence others and is generally favored by fake news' creators is more check-worthy. Evaluating the potential influence of a certain topic or event can rely on cross-topic fake news analysis, which we will discuss as a potential research task in this section as well. Additionally, as we have specified in Section 2.1, in addition to providing the authenticity assessments on news, fact-checking websites often provide (i) additional information that can be invaluable for identifying check-worthy content, for example, "the PolitiFact scorecard" in PolitiFact presents statistics on the authenticity distribution of all the statements related to a specific topic (see Figure 2(a) for an illustration), and (ii) detailed expert-based analysis for checked contents (e.g., what is false and why is it false), both of which to date has not been taken good advantages of.

III. Cross-domain (-topic, -website, -language) Fake News Studies. We highlight this potential research task for two reasons. First, current fake news studies emphasize on distinguishing fake news from truth with experimental settings that are generally limited to a certain social network and a language. Second, analyzing fake news across domains, topics, websites, and languages allows one to gain a deeper understanding of fake news and identify its unique non-varying characteristics, which can further assist in fake news early detection and the aforementioned identification of check-worthy content which we have discussed in Section 4 and this section.

IV. Deep Learning for Fake News Studies. The developments in deep learning can potentially benefit fake news research, with recent studies demonstrating such benefits. For instance, recent fake news studies have adopted either Recurrent/Recursive Neural Networks (RNNs) to represent *sequential* posts and user engagements [Ma et al. 2018b; Ruchansky et al. 2017; Zhang et al. 2018], or Convolutional Neural Networks (CNNs) to capture local features of texts and images [Yang et al. 2018], or both [Liu and Wu 2018]. Generative Adversarial Networks (GANs) have also been used and extended to obtain a "general feature set" for fake news across events to achieve fake news early detection. We

highlight deep learning for fake news studies as, first, deep learning techniques have shown their strength, in particular, in processing text, images, and speech [LeCun et al. 2015], all heavily observed in fake news. Second, deep learning bypasses feature engineering, which can be one of the most time-consuming but necessary parts of a machine learning framework. Third, a deep learning architecture can be relatively easily adapted to a new problem, e.g., using CNNs, RNNs, or Long Short-Term Memory (LSTM), which is valuable for fake news early detection. On the other hand, it should be noted that deep learning techniques often requires massive training data and time for model training, and are generally weak on providing interpretable models., i.e., explaining what it is learned.

V. Fake News Intervention. Fake news studies, e.g., Lazer et al. [2018], have emphasized the importance of business models adopted by social media sites to address fake news intervention, which suggests shifting the emphasis from maximizing user engagement to that on increasing information quality, e.g., using self- or government regulations. In addition to formulating policies and regulations, efficiently blocking and mitigating the spread of fake news also requires technical innovations and developments. Technically, a fake news intervention strategy can be based on network structure, which has been discussed in [Shu et al. 2018], or based on users, which we will discuss here. When intervening based on network structure, one aims to stop fake news from spreading by blocking its propagation paths, which relies on analyzing the network structure of its propagations and predicting how fake news is going to further spread. Here, we point out the possibility to achieve fake new intervention by taking advantage of users involved. From a user perspective, fake news intervention relies on specific roles users play in fake news activities. One such role is being an (i) *influential user* (i.e., *opinion leader*). When blocking a certain fake news in a social network, handling these influential spreaders first, leads to a more efficient intervention compared to handling users that may have a negligible social influence on others. Another beneficial role is being a (ii) *corrector*, users on social networks who take an active role in mitigating the spread of fake news by attaching links that debunk the fake news in their posts or comments [Vo and Lee 2018]. Furthermore, the intervention strategy for (iii) malicious users and (iv) naïve users should be different, while they both spread fake news; malicious users should be removed or penalized, while naïve users should be [actively or passively] assisted to improve their ability to distinguish fake news. To enhance a user's ability to differentiate fake news from true news, for example, *personal recommendation* of true news articles and/or related links for users can be helpful. The recommendation should not only cater to the topics that users want to read, but should also capture topics and events that users are most gullible to due to their political biases or preexisting knowledge.

7 CONCLUSION

The goal of this survey has been to comprehensively and extensively review, summarize, compare and evaluate the current research on fake news, which includes (1) the qualitative and quantitative analysis of fake news, as well as detection and intervention strategies for fake news from four perspectives: the false knowledge fake news communicates, its writing style, its propagation patterns, and its credibility; (2) main fake news characteristics (authenticity, intention, and being news) that allow distinguishing it from other related concepts (e.g., misinformation, disinformation, or rumors); (3) various news-related (e.g., headline, body-text, creator, and publisher) and social-related (e.g., comments, propagation paths and spreaders) information that can exploited to study fake news across its lifespan (being created, published, or propagated); (4) feature-based and relation-based techniques for studying fake news; and (5) available resources, e.g., fundamental theories, websites, tools, and platforms, to support fake news studies. A summary and comparison of various perspectives to study fake news is provided in Table 10. The open issues and challenges are also presented in this survey with potential research tasks that can facilitate further development in fake news research.

Table 10. Summary and Comparison of Perspectives to study Fake News

	Knowledge-based	**Style-based**	**Propagation-based**	**Credibility-based**
Potential Research Task(s)	Fake news analysis and detection	Fake news analysis and detection	Fake news analysis, detection, and intervention	Fake news analysis, detection, and intervention
Fake News Stage(s) Studied	Creation, publication and propagation	Creation, publication and propagation	Propagation	Creation, publication and propagation
Information Utilized	News-related	News-related	Primarily social-related	News-related and social-related
Objective(s)	News Authenticity Evaluation	News Intention Evaluation	News Authenticity and Intention Evaluation	News Authenticity and Intention Evaluation
Techniques	Relation-based	Feature-based	Primarily Relation-based	Relation-based and Feature-based
Resources	Knowledge graphs, e.g., Knowledge Vault	Theories, e.g., reality monitoring; however, not many theories focus on fake news	Theories in Table 2	Theories in Table 2.
Related Topic(s)	Fact-checking	Deception analysis and detection	Epidemic modeling, rumor analysis and detection.	Clickbait analysis and detection, (review and Web) spam detection.

REFERENCES

Jacob Abernethy, Olivier Chapelle, and Carlos Castillo. 2010. Graph regularization methods for web spam detection. *Machine Learning* 81, 2 (2010), 207–225.

Mohamed Abouelenien, Verónica Pérez-Rosas, Bohan Zhao, Rada Mihalcea, and Mihai Burzo. 2017. Gender-based multimodal deception detection. ACM, Proceedings of the Symposium on Applied Computing, 137–144.

Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting hoaxes, frauds, and deception in writing style online. In *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 461–475.

Leman Akoglu, Rishi Chandy, and Christos Faloutsos. 2013. Opinion Fraud Detection in Online Reviews by Network Effects. *ICWSM* 13 (2013), 2–11.

Mohammed Ali and Timothy Levine. 2008. The language of truthful and deceptive denials and confessions. *Communication Reports* 21, 2 (2008), 82–91.

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

Yasser Altowim, Dmitri V Kalashnikov, and Sharad Mehrotra. 2014. Progressive approach to relational entity resolution. *Proceedings of the VLDB Endowment* 7, 11 (2014), 999–1010.

Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. 2017. We used Neural Networks to Detect Clickbaits: You wonâĂŹt believe what happened Next!. In *European Conference on Information Retrieval*. Springer, 541–547.

Eric T Anderson and Duncan I Simester. 2014. Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research* 51, 3 (2014), 249–269.

Blake E Ashforth and Fred Mael. 1989. Social identity theory and the organization. *Academy of management review* 14, 1 (1989), 20–39.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The semantic web*. Springer, 722–735.

Péter Bálint and Géza Bálint. 2009. The Semmelweis-reflex. *Orvosi hetilap* 150, 30 (2009), 1430.

Sudipta Basu. 1997. The conservatism principle and the asymmetric timeliness of earnings1. *Journal of accounting and economics* 24, 1 (1997), 3–37.

Dan Berkowitz and David Asa Schwartz. 2016. Miley, CNN and The Onion: When fake news becomes realer than real. *Journalism Practice* 10, 1 (2016), 1–17.

Krishna Bharat and Monika R Henzinger. 1998. Improved algorithms for topic distillation in a hyperlinked environment. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 104–111.

Indrajit Bhattacharya and Lise Getoor. 2007. Collective entity resolution in relational data. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1, 1 (2007), 5.

Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. 2016. "8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality.. In *AAAI*. 94–100.

Lawrence E Boehm. 1994. The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin* 20, 3 (1994), 285–293.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. AcM, 1247–1250.

Gary D Bond, Rebecka D Holman, Jamie-Ann L Eggert, Lassiter F Speller, Olivia N Garcia, Sasha C Mejia, Kohlby W Mcinnes, Eleny C Ceniceros, and Rebecca Rustige. 2017. 'Lyin'Ted','Crooked Hillary', and 'Deceptive Donald': Language of Lies in the 2016 US Presidential Debates. *Applied Cognitive Psychology* 31, 6 (2017), 668–677.

Gary D Bond and Adrienne Y Lee. 2005. Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology* 19, 3 (2005), 313–329.

Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. 2017. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*. 84–89.

Chloé Braud and Anders Søgaard. 2017. Is writing style predictive of scientific fraud? *arXiv preprint arXiv:1707.04095* (2017).

Michael T Braun and Lyn M Van Swol. 2016. Justifications offered, questions asked, and linguistic patterns in deceptive and truthful monetary interactions. *Group decision and negotiation* 25, 3 (2016), 641–661.

Cody Buntain and Jennifer Golbeck. 2017. Automatically Identifying Fake News in Popular Twitter Threads. In *Smart Cloud (SmartCloud), 2017 IEEE International Conference on*. IEEE, 208–215.

Chiyu Cai, Linjing Li, and Daniel Zeng. 2017. Detecting Social Bots by Jointly Modeling Deep Behavior and Content Information. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1995–1998.

Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning.. In *AAAI*, Vol. 5. Atlanta, 3.

Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 9–16.

Tong Chen, Lin Wu, Xue Li, Jun Zhang, Hongzhi Yin, and Yang Wang. 2017. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. *arXiv preprint arXiv:1704.05973* (2017).

Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *CSCW: proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work*, Vol. 2017. NIH Public Access, 1217.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).

Peter Christen. 2008. Automatic record linkage using seeded nearest neighbour and support vector machine classification. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 151–159.

Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing* 9, 6 (2012), 811–824.

Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. *PloS one* 10, 6 (2015), e0128193.

Sarah Cohen, James T Hamilton, and Fred Turner. 2011. Computational journalism. *Commun. ACM* 54, 10 (2011), 66–71.

Aron Culotta and Andrew McCallum. 2005. Joint deduplication of multiple record types in relational data. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 257–258.

Douglas C Derrick, Thomas O Meservy, Jeffrey L Jenkins, Judee K Burgoon, and Jay F Nunamaker Jr. 2013. Detecting deceptive chat-based communication using typing behavior and message cues. *ACM Transactions on Management Information Systems (TMIS)* 4, 2 (2013), 9.

Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology* 51, 3 (1955), 629.

Thomas G Dietterich. 2000. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*. Springer, 1–15.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 601–610.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. 2015. Knowledge-based trust: Estimating the trustworthiness of web sources. *Proceedings of the VLDB Endowment* 8, 9 (2015), 938–949.

Zhicheng Dou, Ruihua Song, Xiaojie Yuan, and JiRong Wen. 2008. Are click-through data adequate for learning web search rankings?. In *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 73–82.

Norman R Draper and Harry Smith. 2014. *Applied regression analysis*. Vol. 326. John Wiley & Sons.

Nan Du, Yingyu Liang, Maria Balcan, and Le Song. 2014. Influence function learning in information diffusion networks. In *International Conference on Machine Learning*. 2016–2024.

Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can Rumour Stance Alone Predict Veracity?. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3360–3370.

David Dunning, Dale W Griffin, James D Milojkovic, and Lee Ross. 1990. The overconfidence effect in social prediction. *Journal of personality and social psychology* 58, 4 (1990), 568.

Diego Esteves, Aniketh Janardhan Reddy, Piyush Chawla, and Jens Lehmann. 2018. Belittling the Source: Trustworthiness Indicators to Obfuscate Fake News on the Web. *arXiv preprint arXiv:1809.00494* (2018).

Geli Fei, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Exploiting Burstiness in Reviews for Review Spammer Detection. *Icwsm* 13 (2013), 175–184.

Song Feng, Ritwik Banerjee, and Yejin Choi. 2012a. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics, 171–175.

Song Feng, Longfei Xing, Anupam Gogar, and Yejin Choi. 2012b. Distributional Footprints of Deceptive Product Reviews. *ICWSM* 12 (2012), 98–105.

Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res* 15, 1 (2014), 3133–3181.

Dennis Fetterly, Mark Manasse, and Marc Najork. 2005. Detecting phrase-level duplication on the world wide web. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 170–177.

Nico H Frijda. 1986. *The emotions*. Cambridge University Press.

Christie M Fuller, David P Biros, and Rick L Wilson. 2009. Decision support for determining veracity via linguistic-based cues. *Decision Support Systems* 46, 3 (2009), 695–703.

Lise Getoor and Ashwin Machanavajjhala. 2012. Entity resolution: theory, practice & open challenges. *Proceedings of the VLDB Endowment* 5, 12 (2012), 2018–2019.

Sharad Goel, Ashton Anderson, Jake Hofman, and Duncan J Watts. 2015. The structural virality of online diffusion. *Management Science* 62, 1 (2015), 180–196.

Mandar Gogate, Ahsan Adeel, and Amir Hussain. 2017. Deep learning driven multimodal fusion for automated deception detection. In *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*. IEEE, 1–6.

Ralph Grishman. 2015. Information extraction. *IEEE Intelligent Systems* 30, 5 (2015), 8–15.

Manish Gupta, Peixiang Zhao, and Jiawei Han. 2012. Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 153–164.

Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen. 2004. Combating web spam with trustrank. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment, 576–587.

Jeffrey T Hancock, Lauren E Curry, Saurabh Goorha, and Michael Woodworth. 2007. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes* 45, 1 (2007), 1–23.

Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1803–1812.

Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1835–1838.

Valerie Hauch, Iris Blandón-Gitlin, Jaume Masip, and Siegfried L Sporer. 2015. Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review* 19, 4 (2015), 307–342.

Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence* 194 (2013), 28–61.

Bryan Hooi, Neil Shah, Alex Beutel, Stephan Günnemann, Leman Akoglu, Mohit Kumar, Disha Makhija, and Christos Faloutsos. 2016. Birdnest: Bayesian inference for ratings-fraud detection. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 495–503.

Carl I Hovland, OJ Harvey, and Muzafer Sherif. 1957. Assimilation and contrast effects in reactions to communication and attitude change. *The Journal of Abnormal and Social Psychology* 55, 2 (1957), 244.

Sean L Humpherys, Kevin C Moffitt, Mary B Burns, Judee K Burgoon, and William F Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50, 3 (2011), 585–594.

Kathleen Hall Jamieson and Joseph N Cappella. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017a. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 795–816.

Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. 2014. News credibility evaluation on microblog with a hierarchical propagation model. In *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 230–239.

Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs.. In *AAAI*. 2972–2978.

Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. 2017b. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia* 19, 3 (2017), 598–608.

Nitin Jindal and Bing Liu. 2008. Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 219–230.

Marcia K Johnson and Carol L Raye. 1981. Reality monitoring. *Psychological review* 88, 1 (1981), 67.

Edward E Jones and Daniel McGillis. 1976. Correspondent inferences and the attribution cube: A comparative reappraisal. *New directions in attribution research* 1 (1976), 389–420.

Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.

Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.

Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160 (2007), 3–24.

Pigi Kouki, Christopher Marcum, Laura Koehly, and Lise Getoor. 2016. Entity resolution in familial networks. In *Proceedings of the 12th workshop on mining and learning with graphs Google Scholar*.

Nir Kshetri and Jeffrey Voas. 2017. The Economics of "Fake News". *IT Professional* 6 (2017), 8–12.

Adam Kucharski. 2016. Post-truth: Study epidemiology of fake news. *Nature* 540, 7634 (2016), 525.

Timur Kuran and Cass R Sunstein. 1999. Availability cascades and risk regulation. *Stanford Law Review* (1999), 683–768.

Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, et al. 2013. Prominent features of rumor propagation in online social media. In *International Conference on Data Mining*. IEEE.

Ni Lao and William W Cohen. 2010. Relational retrieval using a combination of path-constrained random walks. *Machine learning* 81, 1 (2010), 53–67.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

Harvey Leibenstein. 1950. Bandwagon, snob, and Veblen effects in the theory of consumers' demand. *The quarterly journal of economics* 64, 2 (1950), 183–207.

Huayi Li, Geli Fei, Shuai Wang, Bing Liu, Weixiang Shao, Arjun Mukherjee, and Jidong Shao. 2017a. Bimodal distribution and co-bursting in review spam detection. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1063–1072.

Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. 2014. Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1566–1576.

Luyang Li, Bing Qin, Wenjing Ren, and Ting Liu. 2017b. Document representation and feature combination for deceptive spam review detection. *Neurocomputing* 254 (2017), 33–41.

Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 939–948.

Xin Liu, Radoslaw Nielek, Paulina Adamska, Adam Wierzbicki, and Karl Aberer. 2015. Towards a highly effective and robust Web credibility evaluation system. *Decision Support Systems* 79 (2015), 99–108.

Yang Liu and Yi-fang Brook Wu. 2018. Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks.. In *AAAI*.

George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018a. Detect rumor and stance jointly by neural multi-task learning. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 585–593.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018b. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1980–1989.

Colin MacLeod, Andrew Mathews, and Philip Tata. 1986. Attentional bias in emotional disorders. *Journal of abnormal psychology* 95, 1 (1986), 15.

Amr Magdy and Nayer Wanas. 2010. Web-based statistical fact checking of textual documents. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM, 103–110.

Gloria Mark. 2014. Click bait is a distracting affront to our focus. *nytimes.com/roomfordebate/2014/11/24/you-wont-believe-whatthese-people-say-about-click-bait/click-bait-is-a-distracting-affrontto-our-focus* (2014).

David Matsumoto and Hyisung C Hwang. 2015. Differences in word usage by truth tellers and liars in written statements and an investigative interview after a mock crime. *Journal of Investigative Psychology and Offender Profiling* 12, 2 (2015), 199–216.

Fred Morstatter, Liang Wu, Tahora H Nazer, Kathleen M Carley, and Huan Liu. 2016. A new approach to bot detection: striking the balance between precision and recall. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 533–540.

Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013a. Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 632–640.

Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 191–200.

Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013b. What yelp fake review filter might be doing?. In *ICWSM*.

Anis Najar, Ludovic Denoyer, and Patrick Gallinari. 2012. Predicting information diffusion on social networks with partial knowledge. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 1197–1204.

Ndapandula Nakashole, Martin Theobald, and Gerhard Weikum. 2011. Scalable knowledge harvesting with high precision and high recall. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 227–236.

Ndapandula Nakashole, Gerhard Weikum, and Fabian Suchanek. 2012. PATTY: a taxonomy of relational patterns with semantic types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 1135–1145.

Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin* 29, 5 (2003), 665–675.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016. A review of relational machine learning for knowledge graphs. *Proc. IEEE* 104, 1 (2016), 11–33.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2012. Factorizing yago: scalable machine learning for linked data. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 271–280.

Raymond S Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* 2, 2 (1998), 175.

Feng Niu, Ce Zhang, Christopher Ré, and Jude Shavlik. 2012. Elementary: Large-scale knowledge-base construction via machine learning and statistical inference. *International Journal on Semantic Web and Information Systems (IJSWIS)* 8, 3 (2012), 42–73.

Alexandros Ntoulas, Marc Najork, Mark Manasse, and Dennis Fetterly. 2006. Detecting spam web pages through content analysis. In *Proceedings of the 15th international conference on World Wide Web*. ACM, 83–92.

Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 309–319.

Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, et al. 1998. The pagerank citation ranking: Bringing order to the web. (1998).

Jinie Pak and Lina Zhou. 2015. A comparison of features for automatic deception detection in synchronous computer-mediated communication. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on*. IEEE, 141–143.

Sachin Pawar, Girish K Palshikar, and Pushpak Bhattacharyya. 2017. Relation Extraction: A Survey. *arXiv preprint arXiv:1712.05191* (2017).

Supavich Fone Pengnate. 2016. Measuring emotional arousal in clickbait: eye-tracking approach. (2016).

Verónica Pérez-Rosas, Mohamed Abouelenien, Rada Mihalcea, Yao Xiao, CJ Linton, and Mihai Burzo. 2015. Verbal and nonverbal clues for real-life deception detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2336–2346.

Verónica Pérez-Rosas and Rada Mihalcea. 2014. Cross-cultural deception detection. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 2. 440–445.

Verónica Pérez-Rosas and Rada Mihalcea. 2015. Experiments in open domain deception detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1120–1125.

D Pisarevskaya. 2015. Rhetorical Structure Theory as a Feature for Deception Detection in News Reports in the Russian Language. In *Artificial Intelligence and Natural Language & Information Extraction, Social Media and Web Search (AINL-ISMW) FRUCT Conference, Saint-Petersburg, Russia*.

David Pogue. 2017. How to stamp out fake news. *Scientific American* 316, 2 (2017), 24–24.

Olu Popoola. 2018. Detecting Fake Amazon Book Reviews using Rhetorical Structure Theory. (2018).

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A Stylometric Inquiry into Hyperpartisan and Fake News. *arXiv preprint arXiv:1702.05638* (2017).

Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *European Conference on Information Retrieval*. Springer, 810–817.

Emily Pronin, Justin Kruger, Kenneth Savtisky, and Lee Ross. 2001. You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology* 81, 4 (2001), 639.

K Rapoza. 2017. Can âĂŸfake newsâĂŹ impact the stock market? (2017).

Shebuti Rayana and Leman Akoglu. 2015. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*. ACM, 985–994.

Shebuti Rayana and Leman Akoglu. 2016. Collective opinion spam detection using active inference. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 630–638.

Yafeng Ren and Donghong Ji. 2017. Neural networks for deceptive opinion spam detection: An empirical study. *Information Sciences* 385 (2017), 213–224.

Arne Roets et al. 2017. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* 65 (2017), 107–110.

Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, 232–239.

Victoria L Rubin. 2010. On deception and deception detection: Content analysis of computer-mediated stated beliefs. *Proceedings of the Association for Information Science and Technology* 47, 1 (2010), 1–10.

Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 797–806.

Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592* (2017).

Baoxu Shi and Tim Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. *Knowledge-Based Systems* 104 (2016), 123–133.

Somayeh Shojaee, Masrah Azrifah Azmi Murad, Azreen Bin Azman, Nurfadhlina Mohd Sharef, and Samaneh Nadali. 2013. Detecting deceptive reviews using lexical and syntactic features. In *Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on*. IEEE, 53–58.

Kai Shu, H Russell Bernard, and Huan Liu. 2018. Studying Fake News via Network Analysis: Detection and Mitigation. *arXiv preprint arXiv:1804.10233* (2018).

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017a. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

Kai Shu, Suhang Wang, and Huan Liu. 2017b. Exploiting tri-relationship for fake news detection. *arXiv preprint arXiv:1712.07709* (2017).

Michael Siering, Jascha-Alexander Koch, and Amit V Deokar. 2016. Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems* 33, 2 (2016), 421–455.

Craig Silverman. 2016. This analysis shows how viral fake election news stories outperformed real news on Facebook. *BuzzFeed News* 16 (2016).

Alexander Smith and Vladimir Banic. 2016. Fake News: How a partying Macedonian teen earns thousands publishing lies. *NBC News* 9 (2016).

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*. 926–934.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. Association for Computational Linguistics, 1201–1211.

Nikita Spirin and Jiawei Han. 2012. Survey on web spam detection: principles and algorithms. *Acm Sigkdd Explorations Newsletter* 13, 2 (2012), 50–64.

Rebecca C Steorts, Rob Hall, and Stephen E Fienberg. 2016. A bayesian approach to graphical record linkage and deduplication. *J. Amer. Statist. Assoc.* 111, 516 (2016), 1660–1672.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 697–706.

Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. 2017. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506* (2017).

Rakshit Trivedi, Bunyamin Sisman, Xin Luna Dong, Christos Faloutsos, Jun Ma, and Hongyuan Zha. 2018. LinkNBed: Multi-Graph Representation Learning with Entity Linkage. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 252–262.

Udo Undeutsch. 1967. Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der psychologie* 11 (1967), 26–181.

Bram Vijgen. 2014. THE LISTICLE: AN EXPLORING RESEARCH ON AN INTERESTING SHAREABLE NEW MEDIA PHENOMENON. *Studia Universitatis Babes-Bolyai, Ephemerides* 59, 1 (2014).

S Vichy N Vishwanathan, Nicol N Schraudolph, Risi Kondor, and Karsten M Borgwardt. 2010. Graph kernels. *Journal of Machine Learning Research* 11, Apr (2010), 1201–1242.

Nguyen Vo and Kyumin Lee. 2018. The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News. *arXiv preprint arXiv:1806.07516* (2018).

Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 647–653.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

M Mitchell Waldrop. 2017. News Feature: The genuine problem of fake news. *Proceedings of the National Academy of Sciences* 114, 48 (2017), 12631–12634.

Amy B Wang. 2016. Post-truth'named 2016 word of the year by Oxford Dictionaries. *Washington Post* (2016).

Guan Wang, Sihong Xie, Bing Liu, and S Yu Philip. 2011. Review graph based online store review spammer detection. In *Data mining (icdm), 2011 ieee 11th international conference on*. IEEE, 1242–1247.

Xuepeng Wang, Kang Liu, Shizhu He, and Jun Zhao. 2016. Learning to represent review with tensor decomposition for spam detection. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 866–875.

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 849–857.

Andrew Ward, L Ross, E Reed, E Turiel, and T Brown. 1997. Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and knowledge* (1997), 103–135.

Steven Euijong Whang and Hector Garcia-Molina. 2012. Joint entity resolution. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 294–305.

Guangyu Wu, Derek Greene, and Pádraig Cunningham. 2010. Merging multiple criteria to identify suspicious reviews. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 241–244.

Ke Wu, Song Yang, and Kenny Q Zhu. 2015. False rumors detection on sina weibo by propagation structures. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 651–662.

Sihong Xie, Guan Wang, Shuyang Lin, and Philip S Yu. 2012. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 823–831.

Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S Yu. 2018. TI-CNN: Convolutional Neural Networks for Fake News Detection. *arXiv preprint arXiv:1806.00749* (2018).

Dongsong Zhang, Lina Zhou, Juan Luo Kehoe, and Isil Yakut Kilic. 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems* 33, 2 (2016), 456–481.

Jiawei Zhang, Limeng Cui, Yanjie Fu, and Fisher B Gouza. 2018. Fake News Detection with Deep Diffusive Network Model. *arXiv preprint arXiv:1805.08751* (2018).

Hai-Tao Zheng, Xin Yao, Yong Jiang, Shu-Tao Xia, and Xi Xiao. 2017. Boost clickbait detection based on user behavior analysis. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*. Springer, 73–80.

Bin Zhou and Jian Pei. 2009. OSD: An online web spam detection system. In *In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, Vol. 9.

Denny Zhou, Olivier Bousquet, Thomas N Lal, Jason Weston, and Bernhard Schölkopf. 2004a. Learning with local and global consistency. In *Advances in neural information processing systems*. 321–328.

Lina Zhou, Judee K Burgoon, Jay F Nunamaker, and Doug Twitchell. 2004b. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation* 13, 1 (2004), 81–106.

Lina Zhou and Azene Zenebe. 2008. Representation and reasoning under uncertainty in deception detection: A neuro-fuzzy approach. *IEEE Transactions on Fuzzy Systems* 16, 2 (2008), 442–454.

Xing Zhou, Juan Cao, Zhiwei Jin, Fei Xie, Yu Su, Dafeng Chu, Xuehui Cao, and Junqiang Zhang. 2015. Real-Time News Cer tification System on Sina Weibo. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 983–988.

Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake News: Fundamental Theories, Detection Strategies and Challenges. In *The Twelfth ACM International Conference on Web Search and Data Mining*. ACM. https://doi.org/10.1145/3289600.3291382

Yiwei Zhou. 2017. Clickbait detection in tweets using self-attentive network. *arXiv preprint arXiv:1710.05364* (2017).

Miron Zuckerman, Bella M DePaulo, and Robert Rosenthal. 1981. Verbal and Nonverbal Communication of Deception1. In *Advances in experimental social psychology*. Vol. 14. Elsevier, 1–59.