EPFL

Thèse n° 8256

Combating Online Scientific Misinformation

Présentée le 1er juillet 2022

Faculté informatique et communications
Laboratoire de systèmes d'information répartis
Programme doctoral en informatique et communications

pour l'obtention du grade de Docteur ès Sciences

par

Panagiotis SMEROS

Acceptée sur proposition du jury

Prof. C. González Troncoso, présidente du jury
Prof. K. Aberer, C. Castillo, directeurs de thèse
Prof. F. Benevenuto, rapporteur
Dr Y. Mejova, rapporteuse
Prof. R. West, rapporteur

École polytechnique fédérale de Lausanne

2022

To change the world, my friend Sancho, is not madness nor utopia; it's justice.
– Miguel de Cervantes, Don Quixote (1605 - 1615)

Dedicated to the ones that never had the chance...

Acknowledgments

First and foremost, I would like to thank my advisor, **Karl Aberer**. From my very first moments in EPFL, I felt like LSIR was my secure place, and Karl was the person behind the scenes responsible for that. As Karl has a PhD in mathematics, is a talented teacher, and runs or is involved in several startups, he has a unique understanding of the theoretical and practical contributions of one's work. His constant challenging of my ideas helped me improve the quality of my thesis and be more confident when defending my positions. Furthermore, Karl never enforced his authority, while he trusted our opinions and respected our working hours and holidays – obvious things that are not taken for granted in the academic world. Finally, on a non-academic note, I really enjoy Karl's sense of humor – always with a touch of "British" sarcasm and nihilism.

I would also like to thank my co-advisor **Carlos Castillo**. Carlos jumped in my PhD journey in my second year and undoubtedly saved the day. During our very fruitful meetings, he would be able in a few seconds to understand, simplify, reframe, and finally question ideas that I would work on for weeks or months. Carlos encouraged me to be sincere about my work without hiding its weaknesses under the carpet and aim for soundness and completeness without caring about missing deadlines. Furthermore, Carlos was very supportive at conferences, introducing me to fellow academics and advertising our work while always being very humble. Finally, on a non-academic note, I admire Carlos' non-neutral and not purely-academic online presence and I genuinely believe he is the g.o.a.t. (I hope he now knows the meaning!).

Next, I would like to thank the outstanding researchers I had on my examining committee: **Carmela Troncoso** for the smooth coordination of my defense and the cheerful Friday afternoon meetups, **Bob West** for the thorough feedback on my thesis and for trusting me as a teaching assistant for his course, and **Fabrício Benevenuto** and **Yelena Mejova** for commenting with very warm words on my thesis despite my initial cold invitation email. Special thanks also to my external collaborators: **Maurício Gruppi** and **Sibel Adali**, with whom we co-authored the paper "SciLander", and to **Elena Kochkina**, **Marya Bazzi**, **Maria Liakata**, and **Arkaitz Zubiaga**, with whom we co-organized three editions of our ICWSM workshop MEDIATE.

In LSIR, I met great people I would like to thank: **Chantal** (the most cheerful and caring secretary in EPFL), **Jean-Eudes** (my one and only office-mate; every new year's day, I will be looking for you in random places in Athens), **Julia** (very grateful that you introduced me to Carlos; missed your random visits to my office and our crazy trip to Australia), **Rémi** (the true orchestrator of the lab), **Bekah** (the great founder of our Friday Sat sessions), **Negar** (you taught me more than I was able to teach you – Persian foods and traditions included!), **Angelika** (proud to be there at the beginning of your EPFL journey), **Marco** (grateful that you are always open

to new adventures with me, either these are Greek trash parties, lake activities, or Palestinian hip-hop events), **Tugrulcan** (a troll studying trolls with the best hat collection in EPFL; was a pleasure hosting you in Athens), **Thang** (your personal driver, furniture mover, electrician, and soon driving instructor; for you, it will always be "Panayiotis" and never "Panos"), and **Jérémie** (a phenomenal collaborator and a very supportive and inspiring friend; our dark and shaken photos from San Francisco speak for themselves). Special thanks also to: **Reza**, **Tam**, **Hamza**, **Alexandra**, **Alevtina**, **Semion**, **Martin**, **Amit**, and **Michele**.

I would also like to thank my fellow EPFL colleagues: **Ekaterina** (an always sunny office neighbor), **Kristina** (our common Balkan genes do not lie; I could see us as panelists-experts for Eurovision), **Manoel** (I had a gut instinct from our very first meeting), and **Tiziano** (a very altruistic and caring friend; we share the start and the end of our PhD journeys, many trips and gatherings, and... a selfie with a quokka!). Special thanks also to: **Marios**, **Arash**, **Akhil**, **Iraklis**, **Periklis**, **Viktor**, **Georgia**, **Zeinab**, **Panayiotis**, and **Konstantinos**.

Next, I would like to thank the Lausanne fam: **Manos**, **Eleni**, **Pavlos**, **Nathalie**, **Alexandros** and **Kyveli** – remember, "we'll always have trashformers"; special thanks to **Manos** (our friendship in a nutshell is Žižek eating consequently two hot-dogs) and to **Eleni** for impersonating *Giorgos Georgiou*. I would also like to thank: **Dimitra** (the party never stops – *unironically wearing sunglasses*), **Angeliki** (you bring panic in my life), **Chrystalleni** (you are the crystal in a pile of glasses), **Sylvia** (lucky that we [Chala]met), **Angelina** (you are the [cater]pillar of the group), **Ivi** (Antiparos is a state of mind), **Sofia** (nothing is given, we live in a borrowed time), **Konstantinos** (Konstantinho in the madness of fruit!), **George** (do you smell what George is cooking?), **Alexandra** (dance me to the end of my PhD), and **Amalia** (daba duba dum). Thanks for teaching me "où est la phase" – that's not even an expression, but please normalize it!

Next, the "Lads of Mahalas": **Panos** (if you know, you know bro), **Matt** (Antigone from Astypalaia is late today), **George** (my friend Maltese Corto), **Aggelos** (*Muah!*), **Kostas** (Kostas, lil Kostas, we got sickled), **Christos** (I am going to tell you), **Errikos** (Santé!), **Stelios** (if the first is yours, the second is mine), **Dimitris** (he's my friend), and **Prodromos** (in chestnut festivals, I'll be on the other hill). Thanks for reminding me that "we are not brothers with everyone".

Next, as an Athenian from the historic neighborhood of Kolonos, I would like to thank **Aggelos**, **Christos**, and **Vasilis** for having my back, and **Lefteris** for asking me science on his podcast. Also, special thanks to **Tasos** and **Eirini** – on our next road trip, do NOT talk to ME. Finally, I would like to thank my *Aveggzers*, **Kosmas**, **Thanos**, and **Nikos** that whatever they *hold is gold* – you're the *Genie*, I'm *Aladdin*; for every wish I rub that lamp; "I'll never have friends like you".

At this point, I would like to express my admiration for my family: my brother **Yannis**, for being a breathtaking performer, my sister **Natassa** for being the bravest person in the world, my mother **Anthoula**, for being extraordinarily humorous, and my father **Giorgos**, for being a true hard-worker; while ~~crying~~ writing this, an old Greek song from the 50s was ringing in my ears: "that holy day will come, I will hug my old mother, and with my old father I will sit and drink".

Oh yeah, before you go, there is another person in my life I would like to thank, **Stella**, that believes in me more than I would ever believe in myself. A few days after finishing these lines, I will have the chance to: "walk the streets of New Orleans, with the girl of my dreams".

Lausanne, June 1, 2022 Panayiotis Smeros

Abstract

The drastic shift towards digital communication in our mediasphere has caused a profound change in the production and consumption of information, which in turn has substantial implications on the social and political landscape. Misinformation, as a side effect of mass information diffusion, has become a fundamental problem for governments, platforms, and the general public in light of critical events such as elections, pandemics, and wars.

In this thesis, we focus on the problem of online scientific misinformation. As a starting point, we survey the evolution of misinformation and present the main characteristics and approaches against it, framing the high-level positioning of this thesis with respect to related literature. Then, we discuss three major scientific contributions of this thesis: our methods for combating claim-based, article-based, and source-based scientific misinformation.

For combating claim-based scientific misinformation, we introduce SciClops, a method for detecting and contextualizing scientific claims for assisting manual fact-checking. Our method involves three steps: (1) extracting scientific claims using a domain-specific, fine-tuned transformer model, (2) clustering similar claims together with related scientific literature using a method that exploits their content and the connections among them, and (3) highlighting check-worthy claims broadcasted by popular yet unreliable sources. Our experiments show that SciClops effectively assists non-expert fact-checkers in verifying complex scientific claims, facilitating them to outperform commercial fact-checking systems.

For combating article-based scientific misinformation, we introduce SciLens, a method for evaluating the quality of scientific news articles. Our method involves a series of quality indicators for news articles that derive from: (1) their content, including the use of attributed quotes, (2) their scientific context, including their semantic similarity and web proximity to the scientific literature, and (3) their social context, including their social media reach and stance. Our experiments show that these indicators help non-experts evaluate the quality of articles more accurately compared to non-experts that do not have access to these indicators. Moreover, SciLens can also produce completely automated quality scores for articles, which agree more with expert evaluators than manual evaluations done by non-experts.

For combating source-based scientific misinformation, we introduce SciLander, a method for learning representations of news sources reporting on scientific topics. Our method involves heterogeneous source indicators that capture: (1) the copying of news stories between sources, (2) the semantic shift of terms across sources, (3) the usage of jargon, and (4) the stance towards specific citations. SciLander uses these indicators as signals of source agreement to train unsupervised source embeddings. Our experiments show that the learned source

representations outperform state-of-the-art baselines on the task of news veracity classification while encoding information about the reliability, political leaning, and partisanship bias of these sources.

In the last part of this thesis, we introduce NewsTeller, a real-time news analytics platform that runs operationally, handling daily thousands of news articles, social media reactions, and references.

Keywords: *Misinformation, Scientific Misinformation, COVID-19 Misinformation, Fact-Checking, Computational Journalism, Scientific Journalism, Science Communication, Scientific News.*

Résumé

La numérisation de notre sphère médiatique a provoqué un profond changement dans la production et la consommation de l'information qui a d'importantes conséquences sur le paysage social et politique. La désinformation, qui représente l'un des effets néfastes de la diffusion massive d'informations, est devenue un problème fondamental pour les gouvernements, les plateformes et le grand public, à la lumière d'événements critiques tels que les élections, les pandémies et les guerres.

Dans cette thèse, nous nous concentrons sur le problème de la désinformation scientifique en ligne. Pour commencer, nous examinons l'évolution du phénomène et présentons les principales caractéristiques et les approches de la lutte contre la désinformation en définissant le positionnement de cette thèse par rapport à la littérature connexe. Ensuite, nous discutons les contributions scientifiques principales de cette thèse : trois méthodes de lutte contre la désinformation scientifique, la première basée sur les allégations, la seconde sur les articles et la dernière sur les sources.

Pour combattre la désinformation scientifique basée sur les allégations, nous présentons SciClops, une méthode de détection et de contextualisation des allégations scientifiques pour aider à la vérification manuelle des faits. Notre méthode comporte trois étapes : (1) l'extraction d'affirmations scientifiques à l'aide d'un modèle d'apprentissage automatique spécifique au domaine et est finement ajusté, (2) le regroupement d'affirmations similaires avec la littérature scientifique à l'aide d'une méthode qui exploite leurs contenu et les connexions entre elles, et (3) la mise en évidence d'affirmations diffusées par des sources populaires mais non fiables nécessitant une vérification. Nos expériences montrent que SciClops aide efficacement les vérificateurs non experts à éprouver des affirmations scientifiques complexes, ce qui leur permet de surpasser les systèmes commerciaux de vérification des faits.

Pour combattre la désinformation scientifique basée sur des articles, nous présentons SciLens, une méthode d'évaluation de qualité des articles d'actualité scientifique. Notre méthode comprend une série d'indicateurs de qualité des articles qui découlent de : (1) leur contenu, y compris l'utilisation de citations attribuées, (2) leur contexte scientifique, y compris leurs similarités sémantiques et leurs proximité avec la littérature scientifique, et (3) leur contexte social, y compris leur portée et leur position dans les médias sociaux. Nos expériences montrent que ces indicateurs aident les non-experts à évaluer la qualité des articles avec plus de précision que les non-experts qui n'ont pas accès à ces indicateurs. En outre, SciLens peut également produire, pour les articles, des scores de qualité entièrement automatisés qui s'accordent d'avantage avec des évaluations d'experts qu'avec des évaluations de non-experts.

Pour combattre la désinformation scientifique basée sur les sources, nous présentons SciLander, une méthode d'apprentissage automatique générant une représentation des sources d'information traitant de sujets scientifiques. Notre méthode fait appel à des indicateurs de sources hétérogènes qui capturent : (1) la copie d'articles d'actualité entre les sources, (2) le changement sémantique de termes entre les sources, (3) l'utilisation du jargon et (4) la position à l'égard de citations spécifiques. SciLander utilise ces indicateurs pour générer une représentation vectorielle des sources de façon non supervisée. Nos expériences montrent que cette représentation est plus prédictive que les méthodes existantes dans des tâches de classification de la véracité d'articles d'actualité. De plus, cette représentation encode des informations sur la fiabilité, l'orientation politique et le parti pris de ces sources.

Dans la dernière partie de cette thèse, nous présentons NewsTeller, une plateforme d'analyse d'actualité en temps réel pleinement opérationnelle, traitant quotidiennement des milliers d'articles d'actualité, de réactions sur les réseaux sociaux et de références.

Keywords *: Désinformation, Désinformation Scientifique, Désinformation COVID-19, Vérification des Faits, Journalisme Informatique, Journalisme Scientifique, Communication Scientifique, Actualité Scientifique.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

According to one of the most prevalent theories in physical cosmology, the space in the early stages of the universe expanded exponentially and faster than the speed of light; a phenomenon called *cosmic inflation*. Similarly to the cosmic inflation, the singularity moment for the information sphere arrived with the rise of web technologies, which enabled most individuals in technologically-developed countries to instantly diffuse information (particles), but also misinformation (antiparticles), to large audiences with little-to-no regulation or quality control. This evolution of the means of communication led to an *information inflation*, giving online media an unprecedented role in influencing political, economic, and social ecosystems [137].

1.1 News Media in the Digital Era

The digital era changed the news industry radically; news consumers shifted from traditional offline and passive media (e.g., newspapers, TV) to online and interactive media (e.g., news portals) (Figure 1.1). Hence, traditional publishers lost part of their power due to emerging independent publishers, which often provide alternative but sometimes also controversial views compared to the mainstream media. As most news consumers focus on the news stories rather than the sources, these alternative hubs of information gain more and more power [146].

Three cognitive effects of online media differentiate them from offline media: non-linearity, continuous-availability, and crowd-involvement [132, 151]. Online media are non-linear in the sense that the temporal nature of narratives is customized in a way that the users have control over it. Furthermore, online media create expectations for fresh content that is always available for on-demand consumption. Finally, in many cases, formerly passive news consumers are able to actively produce or correct already published news.

The digital shift presents benefits to both the media industry and the consumers. In particu-

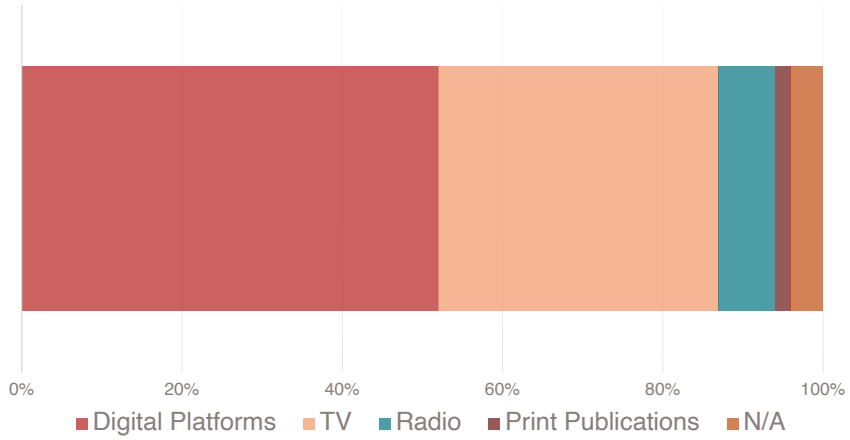


Figure 1.1: A 2020 survey on the news medium preference of US adults. Digital platforms (including news portals and podcasts) are preferred by 52% of the population, in contrast with print publications preferred only by 5% of the population. Data from Pew Research Center [43].

lar, digitalization reduces the cost of publishing, builds new bridges between media outlets and their audiences, supports the pluralism of public opinions, and generally facilitates information accessibility. However, these opportunities come at a price: digital information diffusion tends to amplify misinformation and polarization phenomena [66], making it hard to distinguish credible information from misleading content. This change has already led multiple disciplines to re-examine the notions of "truth" and "trust" online [153].

1.2 News Media in the Social Media Era

After digitalization, the news industry changed radically for a second time with the growth of social media (Figure 1.2). The majority of people using social media platforms started to actively generate, republish, interact, and comment on the news [147, 185]. In crisis events, such as the COVID-19 pandemic, governments and decision-makers utilize this power of social media to properly communicate crucial information and protect the public from falsehoods [25]. Nonetheless, the same social and news media work as a catalyst for the so-called *infodemic*, allowing misinformation to be dispersed on a large scale, regardless of the significant effort to hinder its spread [131].

On this occasion, traditional media companies and professional journalists are at a crossroads. The ephemeral, fast-paced nature of social media, the brevity of the messages circulating on them, the short attention span of their users, their preference for multimedia rather than textual content, and in general, the fierce competition for attention is forcing journalists to adapt in order to survive in the attention economy [142]. As a consequence, news outlets are increasingly using catchy headlines, as well as outlandish and out-of-context claims that perform well in terms of attracting eyeballs and clicks [174].

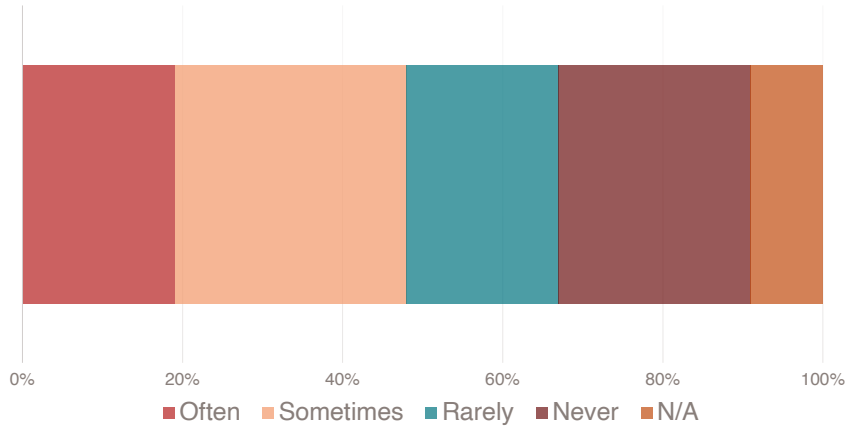


Figure 1.2: A 2021 survey on the frequency of news consumption of US adults in social media. Almost 50% of the population uses social media at least sometimes to consume news. Data from Pew Research Center [128].

On the other hand, social media platforms allow, if not amplify, this flood of fake or low-quality information. In such platforms, the curation of news flows is undertaken not only by conventional newsmakers but also by other factors such as online social contacts, recommendation algorithms, and individual preferences [199]. All these factors create a uniquely-personalized news repertoire, which increases the engagement on the platforms since their users prefer it over the generic editorial curation [200]. However, this personalization seems to amplify pre-existing biases by increasing the ideological segregation of the users and creating so-called *echo-chambers* [55], or what Negroponte et al. [145] describe as "*The Daily Me*".

1.3 Science Communication via News Media

Scientific literacy is broadly defined as a knowledge of basic scientific facts and methods. Deficits in scientific literacy are endemic in many societies [71], which is why understanding, measuring, and furthering the public understanding of science is essential to many scientists [11]. Mass media can be a potential ally in fighting scientific illiteracy. Reading scientific content has been shown to help align public knowledge of scientific topics with the scientific consensus, although, in highly politicized topics, it can also reinforce pre-existing biases [82].

Scientific journalism, as practiced by professional journalists as well as science communicators and bloggers from various backgrounds, can be seen as a translation from a discourse inside scientific institutions (i.e., highly specialized findings reported in scientific reports, journals, and books) to a discourse understandable to a non-specialized, broad audience [149]. By necessity, this process involves negotiating several trade-offs between desirable goals that sometimes enter into conflict, including appealing to the public and using accessible language, while accurately representing research findings, methods, and limitations [152].

However, there are many nuances that make this process much more than a simple translation. For instance, Myers [141], among others, notes that i) in many cases the gulf between experts and the public is not as large as it may seem, as many people may have some information on the topic; ii) there is a continuum of popularization through different genres, i.e., science popularization is a matter of degree; and iii) the scientific discourse is intertwined with other discourses, including the discussion of political and economic issues.

Producing high-quality news material presenting scientific findings to the general public is unquestionably a challenging task, and often there is much to criticize about the outcome. In the process of writing an article, "information not only changes textual form, but is simplified, distorted, hyped up, and dumbed down" [141], while the scientific evidence that may support or refute it remains absent or locked behind pay-walled journals [178]. Misrepresentation of scientific knowledge by journalists has been attributed to several factors, including "a tendency to sensationalize news, a lack of analysis and perspective when handling scientific issues, excessive reliance on certain professional journals for the selection of news, lack of criticism of powerful sources, and lack of criteria for evaluating information" [37].

In many cases, these issues can be traced to journalists adhering to journalistic rather than scientific norms. According to Dunwoody [40], this includes i) a tendency to favor conflict, novelty, and similar news values; ii) a compromise of accuracy by lack of details that might be relevant to scientists, but that journalists consider uninteresting or hard to understand for the public; and iii) a pursuit of "balance" that mistakenly gives similar coverage to consensus viewpoints and fringe theories. Journalists tend to focus on events or episodic developments rather than long-term processes, which results in preferential coverage to initial findings even if they are later contradicted, and little coverage if results are disconfirmed or shown to be wrong [39]. Furthermore, news articles typically do not include caveat/hedging/tentative language, i.e., they tend to report scientific findings using a language expressing certainty, which may have the opposite effect from what is sought, as tentative language makes scientific reporting more credible to readers [97].

1.4 Scientific Misinformation

Misinformation, in the form of propaganda, was a well-known political weapon used extensively in World War I and II [114]. Particularly the term "Fake News" came to prominence after the presidential elections in the United States in 2016, when the former United States president started a rhetorical war on established media outlets by labeling them as fake news media [50]. Following this paradigm, there were countless narratives expressing skepticism and decline of trust in previously dependable sources. According to these narratives, scientific evidence is no longer trusted, medical evidence is sidestepped, and patients search for their own truth online. Under this threat, the traditional keepers of truth, such as editors, journalists, and public intellectuals, have lost their monopoly on public issues, while malicious actors and misinformed

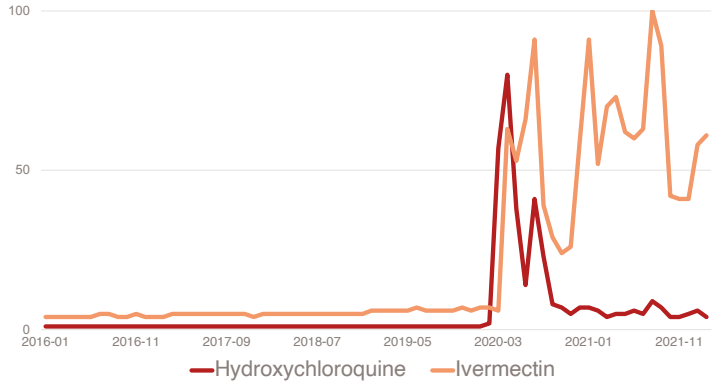


Figure 1.3: Relative search interest of two drugs presented as "safe alternatives" to the vaccines against COVID-19. As the debunking of Hydroxychloroquine was more viral than the debunking of Ivermectin, we observe two different behaviors after the initial attention shift of the audience. Data from Google Trends.

citizens continue spreading their own hate, propaganda, and fake information on a previously unseen scale. Thus, the current era has been characterized as a "Post-Truth Era", in which truth and reason have been replaced by alternative facts and individual gut feelings [51].

Throughout the pandemic, there were numerous viral news stories on possible treatments against COVID-19, which are still presented as "safe alternatives" to the vaccines (Figure 1.3). Hydroxychloroquine and chloroquine were the two most popular such drugs in the early stages of the pandemic, with the US Food and Drug Administration (FDA) cautioning against the usage outside of a hospital setting [52]. Recently, there was a new rumor regarding Ivermectin, a drug used to treat parasites in humans and livestock. Indeed, pharmacies in many countries ran out of the drug, despite the lack of evidence of its effectiveness in treating COVID-19, with the FDA again warning against its usage for that purpose [53]. As Carey et al. [24] note and we can see in Figure 1.3, reductions in COVID-19 misperception beliefs do not persist over time; thus, fact-checks and misinformation evidence has to be constantly reinforced to the people.

Scientific misinformation, especially in an emerging topic with constant updates on the scientific consensus, is affecting not only scientifically-illiterate people but also high-profile stakeholders and decision-makers. For instance, on March 11th, 2020, an article in *The Lancet Respiratory Medicine* theorized that nonsteroidal anti-inflammatory drugs such as Ibuprofen could worsen COVID-19 symptoms [49]. Without referencing explicitly to this article, but motivated by it, the Minister of Health of France posted on Twitter, advising people to avoid Ibuprofen when possible.[1] His message was re-posted nearly $43K$ times and liked nearly $40K$ times. In contrast, a World Health Organization's message posted four days later, which insisted Ibuprofen was safe, was re-posted only $7.5K$ times and liked only $8.5K$ times.[2]

[1] https://twitter.com/olivierveran/status/1238776545398923264
[2] https://twitter.com/WHO/status/1240409217997189128

1.5 Misinformation Taxonomy

Misinformation as a research problem can be sliced and diced in multiple ways, which complies with the multifaceted nature of the problem. The proposed taxonomy overlaps with related taxonomies proposed by two surveys conducted by Kumar et al. [112] and Zannettou et al. [223]. However, since misinformation has been studied in the context of different disciplines, from computer and social science to psychology and journalism, in this thesis, we cover a subset of all the aspects of misinformation studied in the scientific literature.

1.5.1 Characteristics

The main superficial characteristics that we discuss in this thesis are the modalities (formats) and topics in which misinformation emerges, the main actors and their intent when spreading misinformation, and the textual granularity levels in which we can identify misinformation.

Modalities

The digital shift of information led misinformation to appear in all the available digital formats [3]. Initially, false information appeared in textual form in social media platforms [27, 133], online encyclopedias [113], and instant messaging apps [169]. As multimedia formats are becoming more prevalent and easier to consume, misinformation appears in other formats such as audio [127], images [16, 77], and videos [204].

However, research efforts remain heavily concentrated on detecting and mitigating the effect of textual misinformation for two main reasons. First, most ground-truth knowledge bases as well as debunking sources are in textual format (e.g., Wikipedia, which is used by almost 40% of its readers for fact look-ups [190], and dedicated fact-checking portals such as Snopes.com). Furthermore, in many cases, we can detect misinformation in other formats using text-based techniques by introducing an intermediate "translation" step that decouples joint modalities (e.g., audio-to-text [126] and video-to-text [92] transcription).

Topics

Misinformation spans almost every topic for which there exists news, from politics and science to history and sports. Indeed, these topics are not exclusively disjoint; e.g., misinformation regarding COVID-19 can be both political and scientific. Since, in fact-checking platforms, the debunking process involves discovering viral potentially-false stories, a proxy to quantify the misinformation per topic is to quantify the respective debunking effort per topic.

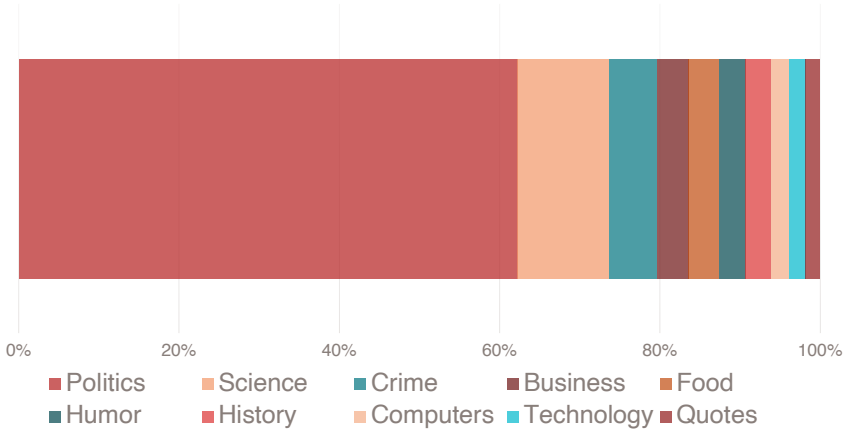


Figure 1.4: Distribution of topics in Snopes.com. Even after more than two years in a pandemic, fact-checks related to politics are almost six times more than fact-checks related to science. Data crawled from Snopes (February 2022 snapshot).

According to Snopes.com, one of the most well-known fact-checking portals with more than 10 million monthly visitors, the most prevalent topic is *Politics* with almost six times more fact-checks than *Science* and more than two times more fact-checks than all the other topics combined. Figure 1.4 confirms that, even after more than two years in a pandemic, the attention of the fact-checking community and the effort to debunk possibly misleading information is mainly focused on political news.

Actors

There are various actors that produce or help on the diffusion of misinformation. Essentially, all the actors that can create news-worthy content can potentially act as channels of false information. These actors can be:

- *Authorities* responsible for informing the public, i.e., politicians, stakeholders, experts;
- *News Industry Practitioners*, i.e., professional journalists or independent bloggers;
- *Bots* and *Sockpuppet Accounts*, i.e., instrumented accounts that often ignite but mainly amplify already spread misinformation;
- *Individuals*, i.e., social media and private messaging users, who, as we see next, can either maliciously or unwittingly share or interact with false information.

Intent

Intentional misinformation, often called *disinformation*, is fake information spread maliciously with the intent to deceive [162]. The most prominent types of intentional misinformation are:

- *Propaganda*, which usually has political motives [102];
- *Hoaxes* and *Rumors*, which usually target politicians and in general public figures [160];
- *Satire*, which usually consists of fabricated or fictional stories described with a clear sense of irony and humor [213].

On the other hand, *unintentional misinformation* emerges after the misinterpretation of accurate information due to lack of attention or due to an attempt of simplification or lack of knowledge on a complex topic [88]. The boundaries between intentional and unintentional misinformation are often blurry, as their only distinguishable characteristics are mainly hidden in auxiliary metadata such as the type of the distributing medium [90].

Granularity

Online misinformation appears at three distinct granularity levels with respect to the verbosity of the text containing it: at the level of claims, the level of articles, and the level of sources [20]:

- *Claims.* At the most granular level, we find claims, i.e., short passages containing check-worthy information such as messages exchanged on social media and messaging apps or sentence-wise fragments of news articles;
- *Articles.* At the middle level, we find articles in news outlets and personal blogs, which are longer passages that contain one or more claims and typically provide more contextual information to support, dispute, or satirize these claims;
- *Sources.* At the more abstract level, we find news sources, e.g., broadcasting syndications, news portals, or independent journalists, that regularly produce news-worthy content.

1.5.2 Approaches

As misinformation has been studied from multiple disciplines, the solutions to mitigate the problem are also multifaceted. The consensus in the scientific community is that there is no "silver bullet" to combat misinformation; thus, all the solutions listed below must be developed and combined to reduce or at least diminish the impact of misinformation in the world.

Education & Media Literacy

There are many initiatives against media illiteracy, either coming from educators and non-governmental organizations or from social and news platforms. From the side of the educators, there are attempts to empower schoolchildren and their teachers by "equipping school communities to fact-check online content, understand news media, make informed choices and resist peer pressure as they assemble their worldview" [103]. Indeed, according to their experience, technology currently plays a limited role in increasing media literacy; however, young people nowadays are more acutely aware of the danger of misinformation because of COVID-19 [154]. Similarly, video platforms organize internet safety workshops for teenagers, [222] and social media providers [134], news outlets [212], and fact-checking organizations [47] inform their customers on ways to identify misinformation.

Fact-Checking

Fact-Checking Portals can be divided into general-purpose (Snopes.com), political (PolitiFact.com), or scientific (ScienceFeedback.co) portals, working closely with domain experts and scientists to debunk misinformation and bring nuance to potentially misleading news. These portals employ specialized journalists who manually: i) detect suspicious claims and articles that "go viral" on social and news media, ii) investigate whether these claims and articles (or variants of both) have also been published by other sources, and iii) find the appropriate prism, consisting of diverse and reliable sources, under which they assess their credibility. However, the described process remains a labor-intensive and time-consuming task, and, despite all the efforts, it is incapable of scaling with the abundance of misinformation [84].

Platform Interventions

With the rise of high-profile cases of the negative real-world impact of misinformation, social media providers and search engines have significantly increased their efforts in debunking false or inaccurate information. Indeed, in collaboration with the academic community, they have proposed both automatic and human-in-the-loop methods for platform interventions to mitigate the problem of misinformation.

On automatic interventions, platforms algorithmically interfere to prevent the spread of misinformation, e.g., by tweaking the underlying, typically black-boxed, recommendation algorithms [23]. On human-in-the-loop interventions, authorities, fact-checking agencies as well as the audience of social media platforms contribute to mitigating misinformation by:

- moderating public discussions and applying strict rules for content removal [129];
- rating and flagging misleading or offensive content [69, 139];

- highlighting and promoting expert opinions [109].

However, unregulated interventions on social media are often regarded as a form of censorship imposed by private companies, governments, and other stakeholders [186].

Automated & Semi-automated Methods

Despite the fact that misinformation circulating online exceeding the capacity of manual effort, traditional news outlets are skeptical towards adopting fully-automated methods for detecting misinformation [192]. Their main concern is that such methods: i) provide poorly-interpretable evidence (according to the journalistic standards), ii) are non-generalizable to arbitrary news topics, and iii) can reinforce potential algorithmic biases and false judgments, leading to a downfall of the reputation of the outlet. As discussed above, this lack of transparency could lead to (unintended) censorship, amplification of extremist views, and silencing of "alternative" views, directly threatening democracy and contributing to political turmoil. Indeed, even big tech companies were forced to suspend automated fact-checking features due to similar criticism from news outlets [62]. Hence, the consensus on the usage of automation in journalism is that it should assist but not replace journalists and individuals when they validate the veracity of news, enabling the movement onward the era of citizen journalism [144].

Below we describe the main categories of automated and semi-automated methods; a detailed review of the related work, also discussing in what direction our methods advance the state-of-the art, is provided in Chapter 2.

User Modeling. This category includes models typically oriented for social media users. These models estimate the likelihood of a user being a certain type of malicious actor spreading misinformation. Hence, there are methods for detecting:

- recurring patterns and orchestrated behaviors, typically observed in *Social Media Bots* and *Sockpuppet Accounts* [44, 184];
- social media users that are prone to share certain types of low-quality information (e.g., health misinformation [67], spam and content promotion [13]);
- groups of users with certain characteristics that are vulnerable to misinformation (e.g., polarized communities, also known as "echo chambers" [65]).

The common element among these models is that they combine user-related features derived from: i) the structure of the social network (e.g., node degrees, PageRank), ii) the content generated by the users (e.g., writing style, sentiment), iii) the interactions among the users (e.g., their re-posting or friendship network).

Propagation Modeling. This category includes methods for modeling information propagation in social networks. Interestingly, properly-adapted epidemic models that were designed to reduce the rapid spread of diseases are also able to predict misinformation diffusion patterns, justifying the usage of the well-known term *infodemic* [196]. Such models are able to detect the key users that need to be "healed" (i.e., to be corrected regarding some false beliefs about a controversial topic) to limit the spread of misinformation [22].

Content Classification. The following categories focus more on a content-wise analysis of misinformation. There are several groups of classification problems for which the academic community has designed appropriate shared tasks and released benchmark datasets. These problems include, among others, news classification [148], rumor evaluation [70], and claim verification [10]. The typical supervised learning models proposed in such tasks are:

- *Deep Linguistic Models*, which are dedicated models designed specifically for narrow domains (e.g., for COVID-19 [106]);
- *Shallow Linguistic Models*, which focus more on generic textual characteristics that often correlate with the diffusion of misinformation (e.g., writing style, sentiment [90, 168]).

Content Retrieval. This category of methods utilizes the knowledge derived from fact-checking portals and ground-truth knowledge bases to detect recycled or repurposed, already-debunked information. Specifically, the task of these methods is to perform semantic textual matches between fact-checking repositories and information published in news and social media [85, 163]. Indeed, models in this category are neither fine-tuned nor tightly coupled with particular topics, and their coverage evolves together with the respective coverage of the fact-checking portals. However, this semantic textual matching, that combines the semantic similarity and the entity matching between two pieces of text, has been repeatedly criticized for its robustness in commercial real-world systems [62].

Content Understanding. The last category of methods goes beyond the detection misinformation to a deeper decomposition and contextualization of its components. Instead of computing boolean flags and reducing the credibility of a piece of information into a single dimension, these models decompose credibility to multidimensional and heterogeneous indicators [225]. Given a certain topic, means of communication, and users profile, such indicators may or may not correlate with the overall credibility of the underlying information.

These indicators occurred as a means of communication between journalists, fact-checking groups, research labs, and social and annotation platforms. Indeed, these bodies formed a broader alliance called *Credibility Coalition*[3] which drafted over 100 indicator suggestions that were later grouped into 12 major categories, including reader behavior, revenue models, publication metadata, and inbound and outbound references. Methods exploiting this methodology

[3] https://credibilitycoalition.org

operationalize these indicators using state-of-the-art language models and information extraction techniques to automatically detect them in news pieces [194].

1.6 Thesis Position & Contributions

In this thesis, we tackle several slices of the taxonomy defined above. We visualize these slices in Figure 1.5 and discuss them below.

Modalities. The primary focus of this thesis is misinformation in textual form. As we explain above: i) textual is nonetheless the most prevalent form of misinformation despite the uprising popularity of multimedia formats; and ii) text is the most wildly-used format for ground-truth information, used in online knowledge bases and fact-checking portals. Hence, the common element among all the contributions of this thesis is the deep content-based analysis on news as well as on social media postings sharing news and scientific references cited by news.

Topics. As we show in Figure 1.4, even after more than two years in a pandemic, science remains a rather unpopular topic in fact-checking portals with evidence that may support or refute scientific news often locked behind pay-walled journals and hardly-accessible academic repositories. Furthermore, as we explain above, this "translation" that scientific journalists perform, from a discourse inside scientific institutions to a discourse understandable to a non-specialized broad audience, often introduces linguistic complications that lead to misinterpretation or amplification of scientific findings.

In this thesis, we focus on science-related misinformation. We tackle the problems of the sparsity of ground-truth and the usage of specialized terminology in scientific news: i) by considering the referenced scientific literature as a potential ground-truth knowledge base for scientific news, and ii) by fine-tuning specialized language models and defining specialized vocabularies that are suitable for interconnecting news and science (e.g., *SciNewsBERT* §4.2.2).

Actors. Since we perform a content-based analysis of misinformation, it is out of our scope to analyze actor-specific characteristics and behaviors that make them prone to diffuse misinformation. However, we showcase that the outcomes of our analysis can help non-expert news consumers accurately evaluate the quality of news material in highly specialized scientific domains (e.g., the gene-editing technique CRISPR (Table 5.3), or the effects of therapeutic cannabis in Post-Traumatic Stress Disorder (Table 4.5)).

Intent. As we explain above, in terms of content, intentional misinformation is frequently indistinguishable from unintentional misinformation, with minor differences hidden in auxiliary metadata. Hence, in the context of this thesis, we treat both types of intent identically.

Granularity. We propose methods for all three aforementioned granularity levels: claims (§4), articles (§5), and sources (§6). We propose dedicated models that deal with the

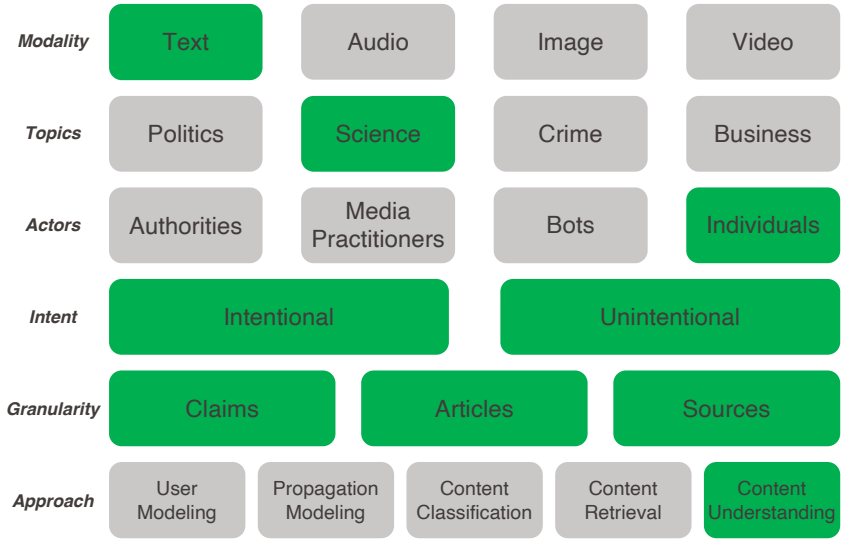


Figure 1.5: Thesis Position: We propose a *Content Understanding* approach for *Textual*, *Scientific*, *Intentional and Unintentional* misinformation which aims to help *Individuals* evaluate the quality of news, at the level of *Claims, Articles, and Sources*.

peculiarities of each level in terms of verbosity and redundancy of information. Particularly we propose models: i) for discovering claims and connecting them with related scientific literature; ii) for extracting quality indicators from the content and the social and scientific context of articles; and iii) for clustering sources based on their writing style and citation behavior.

Approach. As we explain above, for content-based analysis there are three categories of approaches: *Content Classification*, *Content Retrieval*, and *Content Understanding*. Below we explain why the first two categories are insufficient for the case of scientific misinformation and why we propose a *Content Understanding* approach.

Our approach is by design not a *Content Classification* approach, i.e., we do not propose a "fake news classifier", adapted for the case of scientific news. The reason is that such approaches introduce either *Deep Linguistic Models*, which are dedicated classifiers, fine-tuned for specific topics, or *Shallow Linguistic Models*, which are generic-featured classifiers, scratching the surface of the problem without truly understanding it. Indicatively, the most common feature used from such *Shallow Linguistic Models* is the number of misspelled words in a news piece. This feature, as well as other stylistic and grammatical features, may often correlate with the credibility of the classified news piece; nonetheless, it is not the actual signal of misinformation. Hence, recent approaches are moving from *Content Classification* to *Content Understanding* by proposing models with more interpretable features that are able to encode genuine signals of misinformation without being over-engineered for niche domains.

Our approach is also by design not a *Content Retrieval* approach, i.e., we do not search and retrieve debunking information from fact-checking portals and ground-truth knowledge bases. The justification behind this choice is two-fold:

- Retrieval and matching of debunked claims and articles has been shown to be a challenging task that state-of-the-art techniques cannot robustly tackle, forcing big tech companies to remove related functionality from commercial platforms [62]. The nature of scientific news, e.g., the usage of highly-specialized vocabulary and the sensitivity to minor details that determine the overall credibility, makes this task more convoluted and existing natural language processing techniques incapable of tackling it.
- As we explain above, the topic of science is relatively underrepresented in fact-checking portals and knowledge bases; thus, in most cases, debunking information does not even exist. In fact, this hypothesis is confirmed in our experimental evaluation, where debunking information for controversial scientific claims is only available in related scientific literature, not accessible even to commercial automated fact-checking systems (§4.5.3).

As derived from the above, we propose a *Content Understanding* approach. Since the scientific community and media practitioners have moved from the concept of boolean flagging and single-dimensional ranking [159], our approach breaks down the credibility of a news piece into multiple heterogeneous indicators that derive both from its content and its related context. Hence, our approach takes a step toward *Content Understanding*, proposing interpretable indicators that both experts and laypeople, as well as automated classifiers, can understand. As these indicators are more upstream than typical "features", they can be used in different tasks, including retrieval (§4), classification (§5), and high-level modeling (§6).

Indicatively, one such indicator is the *clickbaitness* of titles, which, as we show in our experimental evaluation, in the context of scientific news, does not correlate with the quality of news articles, in contrast with the indicator of the *stance* of social media postings (Figure 5.5). Furthermore, we showcase that these indicators indeed encode misinformation since they are able to cluster together reliable and unreliable sources with similar writing styles and citation behavior and, additionally, distinguish between different types of conspiracy theories that are based on Covid-19 misinformation (§6.5.5).

Thesis Statement

Online scientific misinformation is a crucial problem, especially in the times of a pandemic, where different disciplines have to cooperate in combating it. Among other approaches, existing content-based methods propose classification models which are either poorly interpretable or optimized for narrow domains. In this thesis, we propose methods for extracting explainable indicators from the content as well as the social and scientific context of news that: i) help non-experienced laypeople evaluate news similarly to proficient fact-checkers, and ii) reveal deep misinformation patterns among news sources.

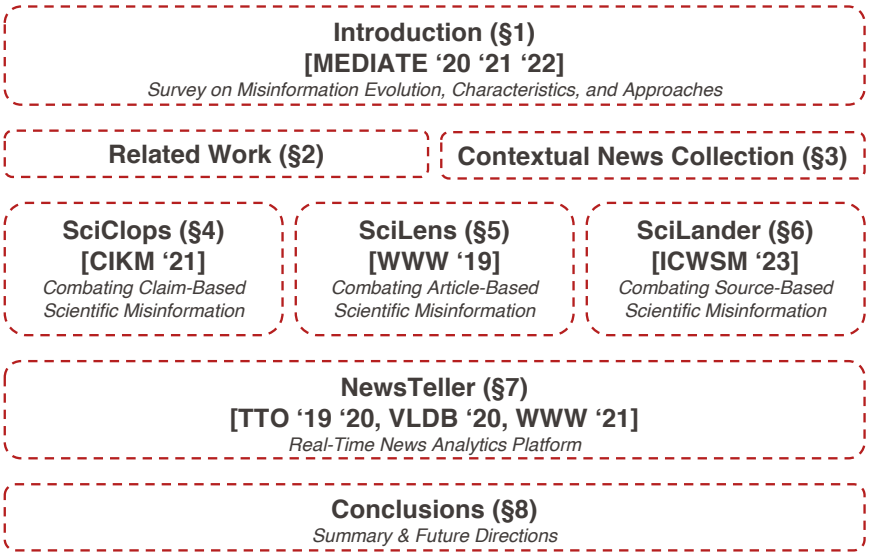


Figure 1.6: Thesis Roadmap: In Chapter 1, we present a survey on misinformation, in Chapter 2, we provide the related work for this thesis, in Chapter 3, we describe our news collection process, in Chapters 4, 5, and 6, we present our methods for combating misinformation at different granularity levels, in Chapter 7, we describe a system that implements these methods, and in Chapter 8, we conclude this thesis.

Thesis Contributions

The technical contributions as well as the roadmap of this thesis are summarized below and highlighted in Figure 1.6.

Chapter 1: Survey on Misinformation Evolution, Characteristics, and Approaches
Discussed in MEDIATE '20 '21 '22 [42, 153, 154]

In this chapter, we survey the evolution of misinformation and describe its main characteristics and categories of approaches. This chapter is partially inspired by the endeavors of the Special Interest Group: *Media in the Digital Age*[4] (among others, the three renditions of the workshop MEDIATE) to bring together media practitioners and technologists to discuss new opportunities and obstacles that arise in the modern era of information diffusion.

Chapter 2: Related Work

In this chapter, we analyze the related work and describe how this thesis advances the state-of-the-art on combating claim-based, article-based, and source-based scientific misinformation.

[4]https://digitalmediasig.github.io

Chapter 3: Contextual News Collection

In this chapter, we describe our contextual news collection process to obtain related social media postings, news articles, and scientific papers.

Chapter 4: Combating Claim-Based Scientific Misinformation (SciClops)
Originally published in the Proceedings of CIKM'21 [193]

In this chapter, we introduce SciClops, a method for detecting and contextualizing scientific claims for assisting manual fact-checking. SciClops involves three steps to process scientific claims found in news articles and social media postings: extraction, clustering, and contextualization. The technical contributions that we achieve with SciClops are summarized as follows:

- We pretrain and fine-tune a domain-specific transformer model (BERT) to facilitate the extraction of scientific claims;
- We cluster claims extracted from heterogeneous sources together with related scientific literature using a method that exploits their content and the connections among them;
- We highlight check-worthy claims, broadcasted by popular yet unreliable sources, together with an enhanced fact-checking context that includes related verified claims, news articles, and scientific papers.

In our experimental evaluation, we show that SciClops tackles sufficiently these three steps and effectively assists non-expert fact-checkers in the verification of complex scientific claims, facilitating them to outperform commercial fact-checking systems.

Chapter 5: Combating Article-Based Scientific Misinformation (SciLens)
Originally published in the Proceedings of WWW'19 [194]

In this chapter, we introduce SciLens, a method for evaluating the quality of scientific news articles using heterogeneous indicators. These indicators derive from the content of articles as well as their social and scientific context. The technical contributions that we achieve with SciLens are summarized as follows:

- We compute quality indicators from the content of articles such as clickbaitness, sentiment, and readability, and distinguish between attributed and unattributed quotes;
- We compute quality indicators from the scientific context of articles, measuring the semantic textual similarity and the web-graph proximity to related scientific literature;

- We compute quality indicators from the social context of articles, measuring the reach and the stance of their social media audience.

In our experimental evaluation, we show that these indicators help non-expert individuals evaluate the quality of a scientific news article more accurately than non-expert individuals who do not have access to these indicators. Furthermore, we show that we can use SciLens to produce a completely automated quality score for an article, which agrees more with expert evaluators than manual evaluations done by non-experts.

Chapter 6: Combating Source-Based Scientific Misinformation (SciLander)
Originally published in the Proceedings of ICWSM'23 [76]

In this chapter we introduce SciLander, a method for mapping the scientific news sources landscape. With SciLander, we learn unsupervised representations of scientific news sources by extracting and combining writing-style and citation-behavior indicators. The technical contributions that we achieve with SciLander are summarized as follows:

- We extract two writing-style indicators that capture (1) the copying of news stories between sources, and (2) the usage of the same terms to mean different things (i.e., the semantic shift of terms);
- We extract two citation-behavior indicators that capture (1) the usage of jargon, and (2) the stance towards specific references;
- We use these indicators as signals of source agreement, sampling pairs of positive (similar) and negative (dissimilar) samples, and combine them in a unified framework to train unsupervised news source embeddings with a triplet margin loss objective.

In our experimental evaluation, we show that the features learned by our model outperform state-of-the-art baseline methods on the task of news veracity classification. Furthermore, our clustering analysis suggests that the learned representations encode information about the reliability, political leaning, and partisanship bias of these sources.

Chapter 7: Real-Time News Analytics Platform (NewsTeller)
Originally presented in TTO '19 '20
Use Cases published in the Proceedings of VLDB '20 and WWW '21 [172, 173]

In this chapter, we introduce NewsTeller, a real-time news analytics platform. NewsTeller provides a wide variety of tools for mapping the media landscape, monitoring the reach of news consumers, and providing quality indicators for millions of news articles collected in real-time. The technical contributions that we achieve with NewsTeller are summarized as follows:

- We retrieve, process, store, and index a wide range of multilingual news articles, social media reactions, and references in real-time;
- We create an enhanced context for news articles combining heterogeneous content, social, and source indicators;
- We provide a multidimensional fact-checking environment for news articles to foster and highlight expert evaluation.

Our platform is intended for three types of audiences: i) the general public, for which we provide access to news with improved context, ii) media practitioners, for which we provide tools to monitor reactions and trends in real-time, and iii) researchers, for which we provide historical access to news data. A live version of NewsTeller is publicly available here: *https://newsteller.io*.

Chapter 8: Conclusions

Finally, in the last chapter, we summarize the contributions of this thesis, discuss its limitations, and propose future research directions.

Chapter 2

Related Work

In this chapter, we analyze the related work and describe how this thesis advances the state-of-the-art on combating scientific misinformation. Specifically, we present related claim-based (§2.1), article-based (§2.2), and source-based (§2.3) approaches.

2.1 Claim-Based Approaches

As described in §1.5.2, fact-checking portals employ specialized journalists who manually: i) detect suspicious claims (extraction), ii) discover variants of these claims published in social and news media (clustering), and iii) find the appropriate prism under which they assess their credibility (contextualization). We summarize some automated methods tackling these steps in Table 2.1 and discuss them below.

2.1.1 Claim Extraction

In the related literature, we find extraction techniques based on text segmentation. Particularly, these techniques, depending on the use-case, are able to detect *quotes*, *arguments*, *rumors*, or *claims* within documents. We observe that all these four types of text segments have similar syntactic structures; thus, we can train models and transfer knowledge across them (e.g., a model trained on an argument corpus can also be fine-tuned to detect claims). Below, we categorize techniques based on the learning scheme they use and not necessarily on the type of segment they are able to detect.

On *weakly supervised models*, Pavllo et al. [157] and Smeros et al. [194] generate complex rule-based heuristics to extract quotes from, respectively, general and scientific news articles.

Table 2.1: Approaches for Extraction, Clustering, and Contextualization in selected references

	Fact-Checking Portals	*Hassan et al. [85]*	*Popat et al. [163]*	*Jaradat et al. [96]*	*Shaar et al. [182]*	*Hansen et al. [81]*	*Zlatkova et al. [229]*	*Karagiannis et al. [105]*	*Pinto et al. [161]*	*Pavllo et al. [157]*	*Smeros et al. [194]*	*Levy et al. [116]*	*Stab et al. [195]*	*Patwari et al. [156]*	*Lippi et al. [121]*	*Jiang et al. [98]*	*Reimers et al. [167]*	*Yao et al. [220]*	*Zhou et al. [228]*	*Hamilton et al. [79]*	*Wang et al. [210]*	*Duong et al. [41]*	*Kochkina et al. [107]*	*Shao et al. [183]*	*Ciampaglia et al. [33]*	*Nadeem et al. [143]*	*Gad-Elrab et al. [63]*	*Chen et al. [29]*	**SciClops**
Extraction																													
Weak Supervision	✓	✗	✗	✗	✗	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	-	-	-	-	-	-	-	-	-	-	-	-	✓
Traditional ML Model	✗	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗	-	-	-	-	-	-	-	-	-	-	-	-	✓
Neural ML Model	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	✓
Clustering																													
Text Modality	✓	✓	✗	✗	✗	✗	✗	✗	✓	-	-	-	-	-	-	-	✓	✓	✓	✓	✗	✗	-	-	-	-	-	-	-	✓
Graph Modality	✗	✗	✗	✗	✗	✗	✓	✗	✓	-	-	-	-	-	-	-	✗	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	✓
Bipartite Clusters	✗	✗	✗	✗	✗	✗	✗	✗	✗	-	-	-	-	-	-	-	✗	✗	✗	✗	✗	✓	-	-	-	-	-	-	-	✓
Contextualization																													
Ground-Truth KBs	✓	✓	✓	✗	✓	✗	✓	✓	✗	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓
Priority Ranking	✗	✗	✗	✓	✓	✓	✗	✓	✗	-	-	-	-	-	-	-	-	-	-	-	-	✗	✗	✗	✓	✗	✗	✗	✓
Scientific Context	✓	✗	✗	✗	✗	✗	✗	✗	✓	-	-	-	-	-	-	-	-	-	-	-	-	✗	✗	✗	✗	✗	✗	✗	✓

On *traditional ML models*, Levy et al. [116] and Stab et al. [195] propose learning models for claim detection and argument mining and introduce publicly available datasets, which we utilize to train our extraction models (details in §4.5.1). Hassan et al. [85] and Popat et al. [163] propose claim classification models that use the aforementioned fact-checking portals to verify political claims, while Patwari et al. [156] and Lippi et al. [121] propose, respectively, an ensemble and a context-independent model for claim extraction. Finally, Zlatkova et al. [229] propose a claim extraction model for images, Karagiannis et al. [105] propose a framework for statistical claims verification, and Pinto et al. [161] propose a method for identifying pairwise relationships between scientific entities.

On *neural ML models*, Jaradat et al. [96] and Shaar et al. [182] detect and rank previously fact-checked claims using deep neural models, while Hansen et al. [81] also train a neural ranking model for check-worthy claims using weak supervision. Furthermore, Jiang et al. [98] use contextualized embeddings to factor fact-checked claims, while Reimers et al. [167] use also contextualized embeddings for claim extraction and clustering. Finally, `CheckThat!` Lab [10] features claim extraction and check-worthiness tasks which are oriented towards political debates in social media platforms.

While the other approaches cover the cases of political, statistical, and visual claims, our approach provides the first dedicated solution for scientific claims. Given the complex nature of the scientific claims in terms of structure and vocabulary, our approach is based on advanced language models with contextualized embeddings that are fine-tuned with domain-specific knowledge. Furthermore, our approach works with arbitrary input text, e.g., from social media postings, blog posts, or news articles.

2.1.2 Claim-Paper Clustering

Since our news collection contains multimodal information (the textual representation of claims and papers and the interconnections between them), we present multimodal clustering approaches that combine text and graph data modalities.

Yao et al. [220] propose a unified convolutional network of terms and documents which is used for document classification, while Zhou et al. [228] use weighted graphs that encode the attribute similarity of the clustered nodes. Hamilton et al. [79] introduce a methodology for jointly training embeddings based on text and graph information, while Reimers et al. [167] apply a numerical clustering on top of such embeddings. Finally, Wang et al. [210] propose a technique for training network embeddings that preserves the communities (clusters) of a graph, while Duong et al. [41] provide interpretable such embeddings.

In our approach, we jointly cluster scientific claims and referenced papers, using both content and graph information. To the best of our knowledge, this is the first approach that deals with heterogeneous passages in terms of length and vocabulary type, which are also interconnected through a bipartite graph.

2.1.3 Claim Contextualization

In addition to the extraction methods described above, the majority of which also provide contextualization/verification techniques (details in Table 2.1). Kochkina et al. [107] and Shao et al. [183] propose methods for automatic rumor verification using well-known fact-checking portals. Ciampaglia et al. [33], Nadeem et al. [143], and Chen et al. [29] use Wikipedia for fact-validation, while Gad-Elrab et al. [63] use custom knowledge graphs for generating interpretable explanations for candidate facts.

While other approaches describe this step as "verification", since essentially they lookup a claim in a ground-truth knowledge base, we consider the general case in which claims rarely appear in such knowledge bases. As we observe in §4.5.3, this is a pragmatic assumption since the majority of the fact-checking effort targets non-scientific topics. As the verification of scientific claims is typically more demanding than other types of claims (e.g., ScienceFeedback.co has built an entire peer-reviewing system for this purpose), we propose a methodology that contextualizes claims based on related scientific literature and ranks them based on the prevalence and the reliability of the broadcasting medium.

2.2 Article-Based Approaches

Our method of evaluating the quality of news articles relies on a series of indicators, computed automatically, and intersects previous literature describing related indicators. In this section, we summarize previous work on manual, automatic, and semi-automatic news evaluation (details in Table 2.2) as well as related methods to extract quality indicators from news articles.

2.2.1 Manual News Article Evaluation

The simplest approach for evaluating news article quality relies on the manual work of domain experts. This is a highly subjective task, given that quality aspects such as credibility are to a large extent perceived qualities, made of many dimensions [57]. In the health domain, evaluations of news article quality have been undertaken for both general health topics [177] and specific health topics such as Pancreatic Cancer [197].

As described in §1.5.2, fact-checking portals perform manual content verification by employing a mixture of professional and volunteer staff. They cover news articles on general topics (e.g., Snopes.com) or specific topics such as politics (e.g., PolitiFact.com). In the case of science news, ClimateFeedback.org is maintained by a team of experts on climate change with the explicit goal of helping non-expert readers evaluate the quality of news articles reporting on climate change. Each evaluated article is accompanied by a brief review and an overall quality score. Reviews and credibility scores from fact-checking portals have been recently integrated with search results [110] and social media posts [124] to help people find accurate information. Furthermore, they are frequently used as ground-truth to build systems for rumor tracking [183], claim assessment [163], and fake multimedia detection [15, 205]. Articles considered by fact-checking portals as misinformation have been used as "seeds" for diffusion-based methods studying the spread of misinformation [196].

Our approach differs from previous work because it is entirely automated and does not need to be initialized with labels from expert- or crowd-curated knowledge bases.

2.2.2 Automatic and Semi-Automatic News Article Evaluation

Recent work has demonstrated methods to automate the extraction of signals or indicators of article quality. These indicators are either expressed at a conceptual level [201] (e.g., *balance of view points, respect of personal rights*) or operationalized as features that can be computed from an article [225] (e.g., *expert quotes* or *citations*). Shu et al. [187] describe an approach for detecting fake news on social media based on social and content indicators. Kumar et al. [113] describe a framework for finding hoax Wikipedia pages mainly based on the author's behavior and social circle, while Ciampaglia et al. [33] use Wikipedia as ground-truth for testing the valid-

Table 2.2: Summary of selected references describing techniques for evaluating news articles

	Fact-Checking Portals	*Shao et al. [183]*	*Boididou et al. [15]*	*Vishwakarma et al. [205]*	*Popat et al. [163]*	*Tambuscio et al. [196]*	*Ciampaglia et al. [33]*	*Urban et al. [201]*	*Zhang et al. [225]*	*Shu et al. [187]*	*Kumar et al. [113]*	*Sbaffi et al. [177]*	*Taylor et al. [197]*	*Yang et al. [218]*	*Horne et al. [90]*	*Reis et al. [168]*	**SciLens**
Automatic assessment	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓
No ground-truth needed	✓	✗	✗	✗	✗	✗	✗	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓
Uses article content	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
Uses reactions from social media	✗	✓	✓	✗	✓	✓	✗	✗	✓	✓	✗	✗	✗	✓	✗	✗	✓
Uses referenced scientific literature	✓	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✗	✗	✗	✓
Domain-agnostic	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Web-scale	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓

ity of dubious claims. Baly et al. [8] describe site-level indicators that evaluate an entire website instead of individual pages, while Yang et al. [218] propose a probabilistic model based on social media indicators. Finally, Horne et al. [90] and Reis et al. [168] use stylistic and grammatical content indicators to detect low-credible news articles.

Our work differs from these by being, to the best of our knowledge, the first work that analyzes the quality of a news article on the web, combining its own content with context that includes social media reactions and referenced scientific literature. We provide a method that is generally applicable to any technical or scientific context at any granularity (from a broad topic such as "health and nutrition" to more specific topics such as "gene-editing techniques").

2.2.3 Quote Extraction and Attribution

The most basic approach to quote extraction is to consider that a quote is a "block of text within a paragraph falling between quotation marks" [45, 164]. Simple regular expressions for detecting quotes can be constructed [150, 176]. Pavllo et al. [157] leverages the redundancy of popular quotes in large news corpora (e.g., highly controversial statements from politicians that are intensely discussed in the press) for building unsupervised bootstrapping models, while Pareti et al. [155] and Muzny et al. [140] train supervised machine learning models using corpora of political and literary quotes (e.g., Quotebank [203] is such a corpus that contains general quotes).

Our work does not rely on simple regular expressions, such as syntactic patterns combined with quotations marks, which in our preliminary experiments performed poorly in quote extraction from science news; instead, we use regular expressions based on classes of words. We also do not use a supervised approach as there is currently no annotated corpus for scientific quote extraction. In the context of this thesis, we built an information extraction model specif-

ically for scientific quotes from scratch, i.e., a "bootstrapping" model, which is based on word embeddings. This is a commonly used technique for information extraction when there is no training data, and we can manually define a few high-precision extraction patterns [99].

2.2.4 Semantic Textual Similarity

One of the quality indicators that we use is the extent to which the content of a news article represents the scientific paper(s) it is reporting about. The Semantic Textual Similarity task in Natural Language Processing determines the extent to which two pieces of text are semantically equivalent. Three approaches that are part of many proposed methods over the last few years include: i) *surface-level similarity* (e.g., similarity between sets or sequences of words or named entities in the two documents); ii) *context similarity* (e.g., similarity between document representations); and iii) *topical similarity* [80, 120].

In our work, we adopt these three types of similarity, which we compute at the document, paragraph, and sentence level. The results we present suggest that the combination of different similarity metrics at different granularities results in notable improvements over using only one metric or only one granularity.

2.2.5 Social Media Stance Classification

Our analysis of social media postings to obtain quality indicators considers their *stance*, i.e., the way in which posting authors position themselves with respect to the article they are posting about. Stance can be binary ("for" or "against"), or be described by more fine-grained types (supporting, contradicting, questioning, or commenting) [83], which is what we employ in our work. Stance classification of social media postings has been studied mainly in the context of online marketing [108] and political discourse, and rumors [230].

In our work, we build a new stance classifier based on textual and contextual features of social media postings and replies, annotated by crowdsourcing workers. To the best of our knowledge, there is no currently available corpus covering the scientific domain. As part of our work, we release such corpus.

2.3 Source-Based Approaches

Source-based approaches are holistic approaches that evaluate the quality of a news source as a whole, without focusing on individual claims or articles extracted from it. Below, we describe some of these approaches as well as our proposal.

Baly et al. [8, 9] and Li et al. [118] highlight the importance of features beyond text to evaluate the veracity of news sources, such as the presence in social media and the existence of a Wikipedia page about a source. Furthermore, Shu et al. [188] explore the interactions between users, authors, and sources, while Gruppi et al. [75] observe content sharing trends among news publishers. Finally, Bourgeois et al. [17] and Rappaz et al. [166] study the selection bias in the topic coverage of news sources by exploring the co-references of these sources to the same news events, while Ribeiro et al. [171] infer the biases of news sources by utilizing their advertiser insights into the demographics of their social media audience.

Both claim- and article-level veracity assessments require data labeling at a very large scale (e.g., individual claims or articles labeled as *reliable* or *unreliable*) and heavily rely on text-specific features these short pieces of text provide. Our approach is, to the best of our knowledge, the first approach that aggregates information about the writing style and citation behavior of news sources to learn unsupervised source representations, that is aware of the science-related content published by them.

Furthermore, as our methodology is applied to a COVID-19 themed dataset, we observe that recently, there has been a significant interdisciplinary effort on detecting and mitigating the effects of misinformation related to the pandemic in social media [4, 28, 68, 191, 202, 217, 227]. Our approach complements the aforementioned approaches by providing a methodology for detecting source-based misinformation patterns.

Chapter 3

Contextual News Collection

The contextual news collection in our work seeks to capture all relevant content for evaluating news quality, including referenced scientific papers and reactions in social media. This methodology can be applied to any specialized or technical domain covered in the news, as long as: i) media coverage in the domain involves "translating" from primary technical sources, ii) such technical sources can be characterized by known domain names on the web, and iii) social media reactions can be characterized by the presence of certain technical terms. Examples where this type of contextual news collection could be applied beyond scientific news include news coverage of specialized topics such as law or finance.

Below, we present a "bottom-up" and a "middle-up" variant of our news collection. In the bottom-up collection, our starting point is a set of seed keywords that we use to retrieve related social media postings, then shared news articles, and finally, referenced scientific papers. On the other hand, in the middle-up collection, our starting point is an already established news collection, which we contextualize with the related scientific literature.

3.1 Bottom-Up Collection

Social media postings containing scientific claims are usually motivated by scientific news articles or scientific papers; thus, we use these postings as our starting point (Figure 3.1). We harvest social media postings from DataStreamer.io (formerly known as Spinn3r.com), covering a 5-year period from June 2013 through June 2018, using a set of seed keywords that we describe below. Additionally to the text of each posting, we collect the number of interactions the posting has received, i.e., other users *re-postings* and *likes*. Finally, we discard postings without outgoing URLs or with spam/unreachable URLs.

In the second step, we use a standard crawling method in which we visit, download, and

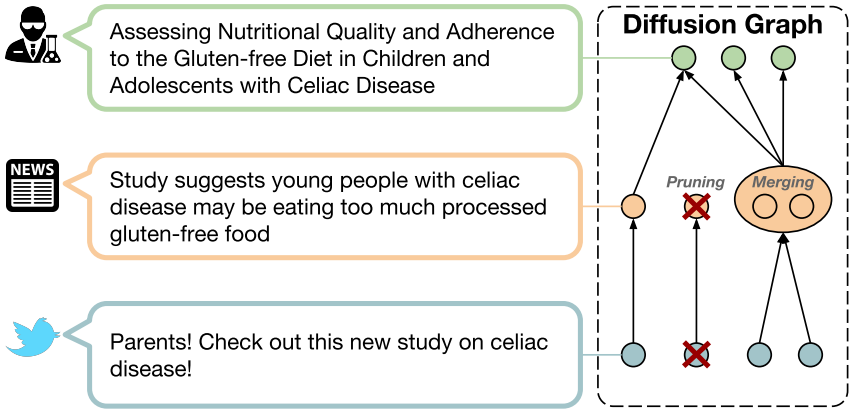


Figure 3.1: Contextual data collection, including social media postings, which reference a series of news articles, which cite one or more scientific papers. In our diffusion graph, paths that do not end up in a scientific paper or paths that contain unparsable nodes (e.g., malformed HTML pages) are *pruned*, and articles with the same content in two different outlets (e.g., produced by the same news agency) are *merged*.

parse the pages pointed by the URLs found in the social media postings. The majority of these pages comes from mainstream news outlets (e.g., theguardian.com or popsci.com), as well as from alternative blogging platforms (e.g., mercola.com or foodbabe.com). In the following chapters, we will refer to this middle-layer of our data collection simply as *news articles*.

In the last step of our procedure, we search within the news articles for references to scientific papers. The scientific papers are peer-reviewed or gray literature[1] papers hosted at universities, academic publishers, or scientific repositories. We use a large list of academic sources consisting of: i) the top-1000 universities in the world (retrieved from CWUR.org), and ii) about 150 academic databases (retrieved from Wikipedia[2]), including Scopus, PubMed, and JSTOR, among many others. For these scientific papers, we extract their title and full content. Finally, we discard from our data collection unparsable and pay-walled scientific papers, as well as news articles that do not contain any reference to scientific papers.

Seed Keywords. The procedure we described is domain-agnostic. The theme and the language of our data collection depend only on the selection of the seed keywords. In the context of this thesis, we choose English keywords from the vocabulary of *CDC A-Z Index*[3], which includes health terms used by laypeople and professionals such as names of food families, nutrients, conditions, and diseases.

Final Collection. Our data collection forms a directed graph, from social media postings to news articles to scientific papers, where edges denote a hyperlink connection. For the narrow

[1]https://en.wikipedia.org/wiki/Grey_literature
[2]https://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines
[3]https://www.cdc.gov/az

domain of *health and nutrition* and before the outbreak of *COVID-19* (in late 2019), we acquired ~$49K$ social media postings, referencing ~$12K$ news articles from ~$3.5K$ news outlets, referencing ~$24K$ scientific papers.

Subsets of this data collection are used to evaluate our methods for combating claim-based and article-based scientific misinformation (§4 & §5). Furthermore, this methodology has evolved and operationalized in our real-time news analytics platform (§7).

3.2 Middle-Up Collection

The following news collection is used to evaluate our method for combating source-based scientific misinformation (§6). More specifically, in this collection, we use a corpus of news articles targeted on the emerging topic of COVID-19 and a corpus of scientific references also targeted on COVID-19. We summarize the basic statistics of both corpora in Table 3.1.

NELA-GT-2020. The collection of news articles contains a total of 1.78 million articles published by 519 sources [74]. Each article in the dataset contains a title, full text, name of the publishing source, and publication timestamp. We use a subset containing only articles related to COVID-19, resulting in $991{,}116$ news articles from 493 sources, published over 18 months, between January 1st 2020 and July 1st 2021. We obtain this subset by applying keyword-based filtering using the COVID-19 terminology from Shugars et al. [189], selecting articles that contain at least one COVID-related keyword in the title or body text.

Media Bias/Fact Check Labels. We retrieve labels for sources in the corpus from the news assessment agency Media Bias/Fact Check[4]. We obtain the *political leaning* of news sources, represented by direction (left or right) and magnitude (mild, moderate, extreme). These are encoded as integer numbers in $[-3, 3]$, negative values indicate left-bias, positive values indicate right bias, and 0 represent center sources. Furthermore, we obtain a *conspiracy-theory* label, a binary indicator denoting whether a source publishes conspiracy theories or pseudoscience content. These are often highly unreliable sources and may or may not exhibit political leaning. Finally, we obtain *factual reporting*, an integer score from 0 to 5 assigned to each source, where 0 indicates the least credible score and 5 is the most credible score. A source that constantly publishes misleading content, fails to fact-check its publications, and does not disclose an editorial board tends to be associated with a lower factual reporting score.

Based on the *factual reporting* score, we divide news sources into two reliability classes, namely the *Reliable News Sources* and the *Uneliable News Sources*. The rules defining each class are described as follows:

- *Reliable News Sources*: sources whose *factual reporting* score is greater than 2.

[4]https://mediabiasfactcheck.com

Table 3.1: Summary of the used corpora. We see that more than half of the articles in NELA-GT-2020 are related to the topic of COVID-19. The labels for reliable, unreliable, and partisan sources are obtained from Media Bias/Fact Check.

NELA-GT-2020	
Total Articles	~1.8M
COVID-19 Articles	~1M
Total Sources	493
Labeled Sources	316
Reliable Sources	122
Unreliable Sources	194
Partisan Sources	162

Scientific References	
COVID-19 Papers (CORD-19)	~300K
Scientific Domains (SciLens)	~1K
References in NELA-GT-2020	~200K

- *Unreliable News Sources*: sources flagged as conspiracy-theory news producers or sources whose *factual reporting* score is less than or equal to 2.

Scientific References. We enhance the news collection described above by extracting the external scientific references of news articles, i.e., the outgoing hyperlinks from the main body of the news articles. We also extract the context of each reference, i.e., the passage of the news article that surrounds this reference. We consider the following two repositories of references:

- One of the most prominent collection of papers related to *COVID-19*, consisting of peer-reviewed papers as well as preprints and other historical coronavirus research, is *CORD-19* [208]. We use the *2021-06-14* release of *CORD-19*, containing a total of $310{,}833$ papers.
- The second source of scientific references is the list of academic sources described above. This list consists of the top-1000 university domains (as indicated by CWUR.org), enhanced with a manually curated list of open-access publishers and grey literature databases. Indeed, these scientific references are more prevalent in news than the *CORD-19* papers because their writing style and terminology used is typically more oriented towards a non-expert audience.

Chapter 4

Combating Claim-Based Scientific Misinformation

This chapter describes SciClops, a method to help combat online scientific misinformation. Although automated fact-checking methods have gained significant attention recently, they require pre-existing ground-truth evidence, which, in the scientific context, is sparse and scattered across a constantly-evolving scientific literature. Existing methods do not exploit this literature, which can effectively contextualize and combat science-related fallacies. Furthermore, these methods rarely require human intervention, which is essential for the convoluted and critical domain of scientific misinformation.

SciClops involves three main steps to process scientific claims found in online news articles and social media postings: extraction, clustering, and contextualization. First, the extraction of scientific claims takes place using a domain-specific, fine-tuned transformer model. Second, similar claims extracted from heterogeneous sources are clustered together with related scientific literature using a method that exploits their content and the connections among them. Third, check-worthy claims, broadcasted by popular yet unreliable sources, are highlighted together with an enhanced fact-checking context that includes related verified claims, news articles, and scientific papers. Extensive experiments show that SciClops tackles sufficiently these three steps, and effectively assists non-expert fact-checkers in the verification of complex scientific claims, facilitating them to outperform commercial fact-checking systems.

4.1 Introduction

Although the amount of news at our disposal seems to be ever-expanding, traditional media companies and professional journalists remain the key to the production and communication of news. The way in which news is disseminated has become more intricate than in the past,

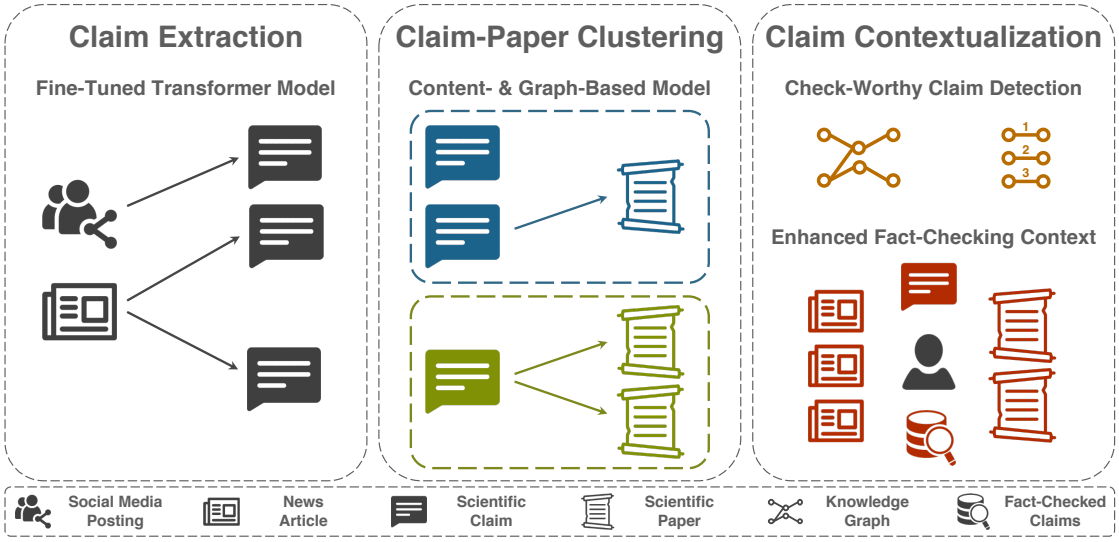


Figure 4.1: Overview of SciClops including the three methods for extraction (§4.2), clustering (§4.3), and contextualization (§4.4) of scientific claims.

with social media playing a fundamental role [128]. The ephemeral, fast-paced nature of social media, the brevity of the messages circulating on them, the short attention span of their users, their preference for multimedia rather than textual content, and in general the fierce competition for attention, has forced journalists to adapt in order to survive in the attention economy [142]. As a consequence, news outlets are increasingly using catchy headlines, as well as outlandish and out-of-context claims that perform well in attracting eyeballs and clicks [174].

When mainstream news media communicate scientific content to the public, the situation is by no means different. Oversimplified scientific claims are rapidly shared in social media, while the scientific evidence that may support or refute them remains absent or locked behind pay-walled journals [178].

Fact-checking portals such as ScienceFeedback.co, among others, work closely with domain experts and scientists to debunk misinformation and bring nuance to potentially misleading claims. This remains, however, a labor-intensive and time-consuming task [84]. On the other hand, despite misinformation circulating online exceeding the capacity of manual fact-checking, traditional news outlets are skeptical towards adopting fully-automated methods [192]. Their main concern is that such tools provide poorly-interpretable evidence (according to the journalistic standards), and any false judgment can lead to a downfall of the outlet's reputation. Indeed, even big tech companies were forced to suspend automated fact-checking features due to similar criticism from news outlets [62]. Hence, the consensus regarding the usage of automation in journalism is that it should assist but not replace journalists and news consumers when they validate the veracity of news, enabling the movement onward the era of citizen journalism [144].

Our work focuses on scientific claims in news articles and social media postings. As scientific claims, we consider *sentence-level segments that involve one or more scientific entities and are eligible for fact-checking*. For example, the sentence *"Ibuprofen can worsen COVID-19 symptoms"* is a scientific claim because it involves two scientific entities (*Ibuprofen* and *COVID-19*) and implies a causal relation between them. To increase the coverage of our definition, we bound neither the number of entities nor the type of relation between them. Such non-deterministic definition makes the detection of scientific claims a challenging task, even for human annotators (details in §4.5.1). To address this task and enable the discovery of complex-structured claims, there is a need for advanced language models which are fine-tuned with domain-specific knowledge.

Once we identify candidate scientific claims, we seek evidence that proves or contradicts them via *contextualization*, i.e., via building an enhanced context of trustworthy information. In the scientific domain, the appropriate context consists of related scientific papers. Grouping similar claims and linking them to related scientific literature is a complex task, to a large extent because of the different nature of the items that we are seeking to connect (i.e., social media postings, news articles, and scientific papers). These contain key passages that determine such connections, but are fundamentally different in terms of: i) verbosity, ranging from character-limited postings to extended scientific papers, and ii) complexity, ranging from a "social media friendly" style of writing to the more formal registry of journalism and academic writing.

Finally, since there is a plethora of controversial claims (especially in the times of a pandemic), there is a need for a check-worthiness ranking that considers the prevalence and the reliability of the broadcasting medium. Providing a scientific context enables non-expert fact-checkers to verify claims with more precision than commercial fact-checking systems, and more confidence since the provided context is fully-interpretable (details in §4.5.3).

Our Contribution. In this chapter we describe *SciClops* (Figure 4.1), a method to assist manual verification of dubious claims, in scientific fields with open-access literature and limited fact-checking coverage. The technical contributions we introduce are the following:

- pretrained and fine-tuned transformer-based models for scientific claim extraction from news and social media (§4.2);
- multimodal, joint clustering models for claims and papers that utilize both content and graph information (§4.3);
- methods for ranking check-worthy claims using a custom knowledge graph, and methods for creating enhanced scientific contexts to assist manual fact-checking (§4.4);
- extensive experiments involving expert and non-expert users, strong baselines and commercial fact-checking systems (§4.5).

4.2 Claim Extraction

We address claim extraction as a classification problem at the sentence level, i.e., we want to distinguish between claim-containing and non-containing sentences. Below, we present the baseline and the advanced extractors that we evaluate in §4.5.1.

4.2.1 Baseline Extractors

We implement several baseline extractors that cover most of the related work on claim extraction described in §2.1: i) two complex heuristics which are used by state-of-the-art *weakly supervised models* [157, 194]; ii) an off-the-shelf classifier trained with standard textual features which is used by state-of-the-art *traditional ML models* [85, 121]; and iii) a transformer model which is used by state-of-the-art *neural ML models* [167, 182].

Grammar-Based Heuristic

The usage of reporting verbs such as "say," "claim," or "report," is a typical element of pattern-matching heuristics for finding claims. Another element is the usage of domain-specific vocabulary; in the scientific context, common verbs in claims include "prove" and "analyze." Thus, we compile a seed set of such verbs, which we extend with synonyms from WordNet [136]. In the following, we refer to this set of reporting verbs as RV.

Scientific claims fundamentally refer to scientific studies, scientists or, more generally, scientific notions. Thus, we employ a shortlist of nouns related to studies and scientists (including "survey" or "researcher"). In the following, we refer to this set of nouns, together with the set of Person and Organization entities, as E.

Finally, to capture the syntactic structure of claims, we obtain part-of-speech tags from the candidate claim-containing sentences. Using this information, we construct a series of complex expressions over *classes of words* such as the following:

$$(root(s) \in RV) \wedge ((nsubj(s) \in E) \vee (dobj(s) \in E)) \implies (s \in Claims)$$

where s is a sentence, *root*$(.)$ returns the root verb of the syntactic tree of a sentence, *nsubj*$(.)$ returns the nominal subject, and *dobj*$(.)$ the direct object of a sentence.

Context-Based Heuristic

This heuristic is based on a frequent non-syntactic pattern, which is quite evident in our data: if an article is posted on social media, then its central claim is typically re-stated or minimally paraphrased in the postings. We investigate pairs (s,p) of candidate sentences s, extracted from news articles, and postings p, referencing these news articles. Our heuristic has the form:

$$(\exists p : \textit{sim}(s,p) \cdot \textit{pop}(p) \geq \textit{threshold}) \implies (s \in \textit{Claims})$$

where $\textit{sim}(s,p)$ denotes the *cosine similarity* between the embeddings representations of s and p, and $\textit{pop}(p)$ denotes the normalized popularity of p, i.e., the raw popularity of p over the sum of the popularity of all the p's that refer to s. As popularity, we consider the sum of the *re-postings* and *likes*. Finally, *threshold* is a hyper-parameter of our heuristic, which in our implementation is fixed to 0.9, yielding a good compromise of precision and recall. We note that this is the only proposed extractor that is not purely content-based as it also requires contextual information.

Random Forest Classifier

To train this classifier, we apply a standard text-preprocessing pipeline, including stop-words removal and part-of-speech tagging. Then, we transform the candidate claim-containing sentences into embeddings by averaging the word embeddings provided by GloVe [158]. As we see in our evaluation (§4.5.1), this classifier performs better than the aforementioned baselines; we also note that, compared to the complex transformer models, it is substantially less intensive in terms of computational resources and training time needed.

BERT Model

One of the most successful state-of-the-art approaches to several NLP tasks, including classification, is the *transformer model* [38]. In our implementation we use the well-known model *BERT* and particularly its version named *bert-base-uncased* [215]. The configuration parameters of the model are those suggested in a widely used software release of this model.[1]

As the last layer of the transformer architecture of *BERT* (and the variants we introduce next), we add a standard binary classification layer with two output neurons, which we train using the datasets described in §4.5.1. During the training, we keep the rest of the layers of the model frozen at their initial parameters.

[1] https://huggingface.co/bert-base-uncased

4.2.2 Fine-Tuned Transformer Extractors

Since *BERT* is originally trained on the generic corpus of Wikipedia, the word representations it generates are also generic. However, scientific claim extraction is a downstream task, where the model has to recognize patterns of a more narrow domain. Thus, we introduce three variants of *BERT* with domain-specific fine-tuning namely, *SciBERT*, *NewsBERT* and *SciNewsBERT*:

- *SciBERT* is pretrained on top of *BERT* with a corpus acquired from *Semantic Scholar* containing ~$1M$ papers [12]. *SciBERT* has a modified vocabulary, compared to the basic vocabulary of *BERT*, that is built to best match the scientific domain.
- *NewsBERT* is a new model that we introduce, built on top of *BERT* and pretrained on a corpus of ~$1M$ headlines published by the Australian Broadcasting Corporation [111].
- *SciNewsBERT* is also a new model that is pretrained like *NewsBERT*, albeit, it is built on top of *SciBERT* instead of *BERT*.

For training *NewsBERT* and *SciNewsBERT* we employ the standard tasks for training *BERT*-like models: i) *Masked Language Modeling*, where the model has to predict the randomly masked words in a sequence of text, and ii) *Next Word Prediction*, where the model has to predict the next word, given a set of preceding words. The hyper-parameters used for training the models are the default proposed by the software release referenced above. Since both *NewsBERT* and *SciNewsBERT* need substantial computational power and training time, we make them publicly available for research purposes.

4.3 Claim-Paper Clustering

Contextualizing scientific claims requires to connect them with related scientific papers. To achieve this, our approach employs a clustering methodology. The clusters, composed of a mixture of claims and papers, must have high semantic coherence and ideally maintain the connections that exist between some of these claims and papers. These implicit connections are hyperlinks starting from news articles and social media postings containing these claims and ending on referenced papers, forming a sparse bipartite graph.

The clustering methods that we employ are: i) *Content-Based* methods on top of either the raw text or an embeddings representation of the passages, ii) *Graph-Based* methods on top of the bipartite graph between the claims and the papers, or iii) *Hybrid* methods that combine the *Content-Based* and the *Graph-Based* methods. Furthermore, we consider both soft (overlapping) clustering (i.e., passages can belong to more than one cluster), and hard (non-overlapping) clustering (i.e., passages must belong to exactly one cluster). The notation used in this section is summarized in Table 4.1.

Table 4.1: Clustering notation. The embeddings dimension (*dim*) of our models is *300*. Matrix L has a *1* in position (*c*, *p*), iff a news article or a social media posting containing claim *c* has a hyperlink to paper *p*. Each row of the clustering matrices (C' and P') contains the probability of a claim or a paper to belong to a cluster; for hard clustering it is "one-hot", i.e., it has a single non-zero element, and for soft clustering it is a general probability distribution.

Symbol	Description
$C \in \mathbb{R}^{\lvert \text{claims} \rvert \times \text{dim}}$	initial claims matrix
$P \in \mathbb{R}^{\lvert \text{papers} \rvert \times \text{dim}}$	initial papers matrix
$L \in \{0,1\}^{\lvert \text{claims} \rvert \times \lvert \text{papers} \rvert}$	interconnection matrix
$C' \in [0,1]^{\lvert \text{claims} \rvert \times \lvert \text{clusters} \rvert}$	final claims clustering matrix
$P' \in [0,1]^{\lvert \text{papers} \rvert \times \lvert \text{clusters} \rvert}$	final papers clustering matrix
f_C: $C \rightarrow C'$	non-linear neural transformation
f_P: $P \rightarrow P'$	non-linear neural transformation
$\lVert . \rVert_F$	*Frobenius Norm*

4.3.1 Content-Based Clustering

Our baseline is content-based (topic) clustering. According to this approach, we assume that claims and papers are represented in the same latent space, in which we compute topical joint clusters. This approach does not consider the interconnections (i.e., the bipartite graph) between the claims and the papers.

For topic modeling, we use *Latent Dirichlet Allocation* (*LDA*), an unsupervised statistical model that computes a soft topic clustering of a given set of passages [14]. We also use *Gibbs Sampling Dirichlet Mixture Model* (*GSDMM*), which assumes a hard topic clustering and is more appropriate for small passages such as claims [119]. When the passages are projected in an embeddings space, we use either the generic *Gaussian Mixture Model* (*GMM*), which computes a soft clustering by combining multivariate Gaussian distributions [170], or *K-Means* [122], which computes a hard clustering. Finally, we test these methods with and without reducing the embeddings dimensions using *Principal Component Analysis* (*PCA*) [48].

4.3.2 Graph-Based Clustering

Since our data is multimodal, an alternative to pure *Content-Based* clustering is pure *Graph-Based* clustering. We define this problem as an optimization problem, introducing an appropriate loss function that we want to minimize. Our goal is to compute the optimal clusters C' and P', and our evaluation criterion is the extent to which C' and P' fit with the interconnection matrix L. Hence, we propose the following loss function:

$$loss = \left\| C' - LP' \right\|_F$$

This loss function is also known as the *Reconstruction Error* and is commonly used in *Linear Algebra* for factorization and approximation problems. By applying this loss function, we force C' and P' to be aligned with L: the claims that appear in a news article should belong to the same cluster as the papers referenced by this article.

A degenerate solution to the problem, if we use only this loss function, is a uniform clustering for both claims and papers. The loss is minimized, but the clustering is useless, because the probability of any claim and any paper to belong to any cluster is uniform. To overcome this problem, we exploit the following technique that is widely used in image processing [115].

In row-stochastic matrices (i.e., matrices that each row sums to 1), a uniform soft clustering has lower *Frobenius Norm* than a non-uniform clustering. Consequently, any hard clustering has the maximum possible *Frobenius Norm*. Thus, we introduce a regularizer that imposes non-uniformity on the clusters by penalizing low *Frobenius Norms* for C' and P':

$$\textit{regularizer} = \begin{cases} -\beta \left(\left\| C' \right\|_F + \left\| P' \right\|_F \right) & C', P' \in V \\ -\beta \left\| P' \right\|_F & C' \notin V \\ -\beta \left\| C' \right\|_F & P' \notin V \end{cases}$$

where V is the set of optimizable variables of our model, and β a hyper-parameter that in our experiments defaults to $\beta = 0.3$. We use a different regularizer in each alternative version of the model that we describe below. These alternative versions have varying flexibility, i.e., either both C' and P' are optimizable variables ($C', P' \in V$), or one of them is fixed, thus not optimizable ($C' \notin V$ or $P' \notin V$). If both of them are fixed ($C', P' \notin V$) then the model has no optimizable variables ($V = \emptyset$). Below we present the alternative versions of the model.

Graph-Based Adaptation

In this alternative (entitled *GBA-CP*), we start with arbitrary cluster assignments for C' and P', which we both optimize based on the loss function. This approach completely ignores the semantic information of C and P and adapts arbitrarily the clusters to the interconnection matrix L. This behavior of *GBA-CP* is confirmed in our experiments (§4.5.2).

In a less aggressive approach, we fix either C' or P' using one of the *Content-Based* algorithms explained above, and optimize only one clustering (the non-fixed) based on the loss function. We entitle these alternatives *GBA-C* for optimizing C', and *GBA-P* for optimizing P'.

Graph-Based Transformation

In this alternative (entitled *GBT-CP*), instead of optimizing directly C' and P', we optimize the weights of the non-linear neural transformations f_C and f_P. The architecture of f_C and f_P consists

of a hidden layer of neurons with a rectified linear unit (*ReLU*), and a linear *Softmax* classifier that computes the overall cluster-membership distribution. We use the same loss function as above where $C' = f_C(C)$ and $P' = f_P(P)$.

Similarly as above, in a less aggressive approach, we fix C' or P' using a *Content-Based* algorithm, and optimize only the weights of one transformation (f_C or f_P). We entitle these alternatives as *GBT-C* for optimizing f_C, and *GBT-P* for optimizing f_P.

4.3.3 Hybrid Clustering

The last clustering model that we propose is a *Hybrid* model that combines a *Content-Based* and a *Graph-Based* model. As we point out in our experimental evaluation (§4.5.2), there is a trade-off between these two approaches in terms of the semantic and interconnection coherence of the computed clusters. Thus, we introduce a tunable model that controls this trade-off.

Our model initializes the clusters C'_{init} and P'_{init} using a *Content-Based* model. Then, it uses an *Alternate Optimization* (*AO*) approach to jointly compute the final C' and P' that adjust best to L. More specifically, it iteratively freezes one of the two clusters and adjusts the other, until they both converge to an optimal state. The loss function of this model is the following:

$$loss = \begin{cases} \gamma \left\| C' - LP' \right\|_F + (1-\gamma) \left\| C' - C'_{init} \right\|_F & C'\text{-optim.} \\ \gamma \left\| C' - LP' \right\|_F + (1-\gamma) \left\| P' - P'_{init} \right\|_F & P'\text{-optim.} \end{cases}$$

where γ is a hyper-parameter that controls the trade-off between *Content-Based* and *Graph-Based* clustering. In our experiments for brevity we present results for three values: *AO-Content* for $\gamma = 0.1$, *AO-Balanced* for $\gamma = 0.5$, and *AO-Graph* for $\gamma = 0.9$.

4.4 Claim Contextualization

In the previous section, we explain how we construct claim-paper clusterings in an unsupervised fashion. These clusterings give already an initial context for claims since they relate them with relevant scientific literature. In this section, we describe how we rank claims within clusters based on their check-worthiness and how we complement their fact-checking context by discovering (when available) previously verified related scientific claims.

4.4.1 Check-Worthy Claim Ranking

The check-worthiness of a scientific claim depends on its intent (e.g., whether it implies a causal relation or describes a particular aspect of an entity) and its prevalence (e.g., in news

and social media). We construct a custom in-cluster knowledge graph in which we encode the intent of the claims into the topology of the graph and the prevalence of the claims into the weighting of the graph.

In-Cluster Knowledge Graph

We construct a knowledge graph by using terms from a domain-specific vocabulary as nodes. The edges of the graph denote the co-occurrence of two terms in the same claim (e.g., the claim *"Ibuprofen can worsen COVID-19 symptoms"* contributes the edge (*Ibuprofen – COVID-19*)).

Since the dataset we use in our evaluation is health-related (details in §4.5), we use the vocabulary of *CDC A-Z Index*[2] that includes health terms used by laypeople and professionals. We note that the rest of the methodology is independent of the domain of the dataset, and can be simply adapted by selecting an appropriate vocabulary.

Graph Topology

From all the possible graph topologies, we particularly focus on the following two types:

- *Causality-Based* topologies which contain nodes from distinct classes such as: i) *"Diseases and Disorders"* (e.g., *Depression*, *Influenza*, and *Cancer*), and ii) *"Conditions, Symptoms, Medications, and Nutrients"* (e.g., *Pregnancy*, *Fever*, and *Red Meat*). A directed edge between two nodes of a different class denotes, to a certain degree, a causal relation between these nodes in the underlying claims [31].
- *Aspect-Based* topologies which focus on the "ego-network" for one particular node (e.g., *"COVID-19"*) and the different aspects regarding this node (e.g., *"Origin"*, *"Mortality Rate"* or *"Common Symptoms"*) [125].

We note that these two types characterize only the topologies and not the underlying claims; thus, they only help us conceptually choose the appropriate graph ranking metric for detecting check-worthy claims without eliminating claims that do not fall under these two types.

Graph Weighting

The weighting scheme that we employ combines two check-worthiness criteria, namely the *popularity* and the *reputation* of the primary sources (i.e., the social media postings and the news articles) from which the claims were extracted.

[2] https://www.cdc.gov/az

The *popularity* of a posting is computed as the sum of the number of re-postings and likes. If multiple postings share the same claim, then their *popularity* is aggregated. Then, Box-Cox transformation ($\lambda = 0$) [18], to diminish the effect of the long-tail distribution, and Min-Max normalization in the interval $[0, 1]$ are applied.

On the other hand, the *reputation* of a news article is entailed from the reputation of the news outlet that publishes the article. In the context of our work, we use the outlet scores compiled by the *American Council on Science and Health* (*ACSH*) [5], which we also normalize in the interval $[0, 1]$. News outlets that are not on *ACSH*'s list (i.e., "long-tail" outlets hosting only 13.5% of the total articles in our collection) are assigned a neutral score (0.5).

Since we want to discover claims that are popular and come from low-reputable sources, we linearly combine the two metrics for each edge e, using a tuning parameter θ as follows:

$$\textit{weight}(e) = \theta\ \textit{popularity}(e) + (1 - \theta)\ (1 - \textit{reputation}(e))$$

In our implementation, we slightly favorite low reputation over popularity; thus, we use $\theta = 0.4$.

Claim Ranking

We rank the edges, and consequently the claims, of the *Causality-Based* topologies using the *Betweenness Centrality* metric [19], and the *Aspect-Based* topologies using the *in-Degree* metric. Examples of check-worthy claims in our data include the term pairs: (*Autism – Vaccines*), (*Breast Cancer – Abortion*), and (*Chemotherapy – Cannabis*) (details in §4.5.3).

4.4.2 Enhanced Fact-Checking Context

The final step for contextualizing the claims is to relate them (when available) with previously verified claims. To retrieve such claims, we use *ClaimsKG* [198], a knowledge graph that aggregates claims and reviews published using *ClaimReview*[3]. After filtering out, based on the mentioned entities, claims with non-scientific content (i.e., 62.3% of the total claims), we end up with a final set of ~$4K$ scientific claims, out of which 79.8% has been determined to be *False*, and 20.2% has been determined to be *True*. We relate claims by computing their Semantic Textual Similarity [120] and setting an appropriate threshold (0.9 in our experiments).

Our final fact-checking context for scientific claims consists of related scientific papers and news articles from the same cluster, and, if available, related verified claims. As we see in our experiments (§4.5.3), this enhanced context improves the verification accuracy and confidence of non-expert fact-checkers, facilitating them to outperform commercial fact-checking systems.

[3]https://www.claimreviewproject.com

4.5 Experimental Evaluation

In this section we evaluate the methods for extraction (§4.5.1), clustering (§4.5.2), and contextualization (§4.5.3) of scientific claims.

Raw Dataset. The main dataset that we use in our evaluation is the dataset we introduced in §3, built with the "bottom-up" methodology. This dataset has the form of a directed graph, from *social media postings* to *news articles* to *scientific papers*, where edges denote a hyperlink connection. The ~$50K$ *social media postings* of the dataset include the text of the postings as well as popularity indicators such as the number of *re-postings* and *likes*. The ~$12K$ *news articles* of the dataset include articles from mainstream news outlets (e.g., theguardian.com or popsci.com), as well as from alternative blogging platforms (e.g., mercola.com or foodbabe.com). Finally, the ~$24K$ *scientific papers* of the dataset include peer-reviewed or gray literature papers hosted at universities, academic publishers, or scientific repositories (e.g., Scopus, PubMed, JSTOR, and CDC). We note that the overall volume of the dataset simulates the typical news coverage on health-related topics for a period of four months.

4.5.1 Evaluation of Claim Extraction

The evaluation of the extractors is two-fold; first, we validate their accuracy using a widely-used clean and labeled dataset, and then, we use them in a real-world scenario where we apply them on the raw dataset described above, and evaluate them via crowdsourcing.

Training

Since there is no specific training dataset for the task of scientific claim extraction, we use two datasets mainly used for argumentation mining, namely *UKP* [195] and *IBM* [116]. The *UKP* dataset includes sentences from controversial search engine results with three labels (non-argument/supporting argument/opposing argument); for the purposes of our task, we consider non-argument sentences and negative examples of claims, and supporting/opposing argument sentences as positive examples of claims. The *IBM* dataset includes context-dependent claims from controversial Wikipedia articles, which, for the purposes of our task, we consider as positive examples of claims.

We train our classifiers using the balanced union of the two datasets (~$11K$ positive and negative samples). In the following, we refer to this dataset as the *Generic Dataset* of claims. We also train our classifiers with a "science-flavored" dataset derived from the *UKP* and *IBM* datasets. Specifically, in this dataset, we oversample claims regarding, e.g., "abortion" and downsample claims regarding, e.g., "school uniforms". We apply this data augmentation by manually processing based on the "general topic" field that exists in both *UKP* and *IBM* datasets. The

Table 4.2: Cross validation of scientific claim extractors. Since, as we explain in §4.5.1, both datasets are balanced, the evaluation metric that we use is *Accuracy* (*ACC*).

		Generic Dataset	**Scientific Dataset**
		ACC	*ACC*
Baseline	*Grammar-Based*	50.4%	52.3%
	Context-Based	49.5%	50.2%
	Random Forest	74.7%	75.6%
	BERT	**82.2%**	81.0%
SciClops	*SciBERT*	81.5%	80.6%
	NewsBERT	82.0%	80.0%
	SciNewsBERT	81.1%	**81.2%**

described dataset is also balanced, containing ~$16K$ positive and negative samples, and in the following, we refer to it as the *Scientific Dataset* of claims.

Cross Validation

We perform a 5-fold cross validation over the datasets described above; the results are shown in Table 4.2. We observe that the *Heuristic-Based* extractors perform poorly for this task, which confirms that it is a demanding task with many corner cases. Remarkably, the *Context-Based* heuristic, which is domain-agnostic, achieves identical *accuracy* with the *Grammar-Based* heuristic, which contains manually curated grammar rules. We also observe that the *Random Forest* classifier does not perform extremely worse than the *Transformer-Based* models, while being more eco-friendly in terms of resources and training time needed.

The performance of the transformer-based models confirms the fact that they are state-of-the-art in most NLP tasks. However, from this task, we do not see the benefits of the domain-specific pretraining. On the *Generic Dataset*, *BERT*, which is pretrained on a generic corpus, performs better, while on the *Scientific Dataset*, *SciNewsBERT*, which is pretrained on a scientific and a news corpus, performs better; nonetheless, their difference is negligible. The real difference among these models is shown in the next experiment.

Crowd Evaluation

We collect boolean labels for 700 sentences extracted from the raw dataset described above by asking the crowd workers a simple classification question (i.e., whether a given sentence contains a scientific claim or not). We use the platform *Mechanical Turk*, asking input from three independent crowd workers per sentence (57 in total). To ensure high-quality annotations, we employ what the platform calls *Master Workers*, i.e., the most experienced workers with ap-

Table 4.3: Crowd Evaluation of scientific claim extraction. Results reported for weak (2 out of 3) annotator agreement (*125* claims - *174* non-claims) and strong (3 out of 3) annotator agreement (*82* claims - *242* non-claims). Since, especially the second set is highly unbalanced, the evaluation metrics that we use are *Precision* (*P*), *Recall* (*R*), and *F1 Score* (*F1*).

		Weak Agreement			**Strong Agreement**		
		P	*R*	*F1*	*P*	*R*	*F1*
Baseline	*Grammar-Based*	51.8%	70.4%	59.9%	40.4%	28.0%	33.1%
	Context-Based	44.6%	49.6%	47.0%	24.5%	45.1%	31.8%
	Random Forest-gen	52.1%	70.4%	59.9%	43.7%	**80.5%**	56.7%
	Random Forest-sci	56.7%	54.4%	55.5%	43.3%	44.8%	44.1%
	BERT-gen	50.8%	50.4%	50.6%	33.5%	68.3%	45.0%
	BERT-sci	78.7%	38.4%	51.6%	79.2%	51.2%	62.2%
SciClops	*NewsBERT-gen*	55.0%	48.8%	51.7%	38.9%	62.2%	47.9%
	NewsBERT-sci	76.9%	40.0%	52.6%	74.2%	56.1%	63.9%
	SciBERT-gen	48.8%	66.4%	56.3%	32.8%	72.0%	45.0%
	SciBERT-sci	48.8%	66.4%	56.2%	**86.5%**	39.1%	53.8%
	SciNewsBERT-gen	49.8%	**80.0%**	**61.3%**	38.8%	78.0%	51.8%
	SciNewsBERT-sci	**84.4%**	30.4%	44.7%	82.7%	52.4%	**64.2%**

proval rate greater than 80%. Finally, we consider *Strong Agreement* among crowd-workers, the 3 out of 3 agreement, and *Weak Agreement* the 2 out of 3 agreement.

We note that there are 77 out of the 700 sentences for which the majority of the annotators answered *N/A*, because they could not distinguish whether these sentences contain a claim or not. For example, interrogative sentences like *"What? Ibuprofen Can Make You Deaf?"* confused the annotators, while similar affirmative sentences like *"Tylenol PM Causes Brain Damage"* were easily identified as scientific claims. The remaining 623 sentences are divided into two subsets; i) sentences having *Strong Agreement* among annotators, with 82 claims (positive examples) and 242 non-claims (negative examples), and ii) sentences having *Weak Agreement* among annotators, with 125 claims and 174 non-claims.

We observe that especially the subset with *Strong Agreement* is highly unbalanced, which is indeed a realistic scenario considering the ratio of claim and non-claim containing sentences in typical news articles. Furthermore, annotators fully agree that a sentence contains a scientific claim for less than 12% of total the sentences, which confirms it is a highly confusing task.

Results

The overall results of the comparison of the extraction models are summarized in Table 4.3. For all the models, we use the following naming convention: the suffix *-gen* is used to denote that models are trained on the *Generic Dataset* explained in §4.5.1, while suffix *-sci* is used to denote

that models are trained on the *Scientific Dataset* also explained in §4.5.1. This convention does not apply to heuristic models that do not require training.

We observe that all the *-gen* models have better or equally good recall as the respective *-sci* models. This happens because *-gen* models have been trained equally towards all the labeled claims and have learned to better recognize the structure of a claim. After analyzing the errors of the models, we noticed that claims with simple structure like *"Repetitive behaviors in autism show sex bias early in life"* were identified more from *-gen* than from *-sci* models. On the other hand, *-sci* models, which have been optimized for the narrow scientific domain, are more selective, hence they show in general better precision than the respective *-gen* models.

Focusing more on the variants of *BERT*, we observe that task-specific pretraining boosts the performance of the model, which is not visible in the first experiment. Specifically, we see that pretraining on both scientific and news domain gives the best results. One illuminative example is the claim *"Galactosides Treat Urinary Tract Infections Without Antibiotics"*, where *Galactosides* is a word that does not appear in the basic vocabulary of *BERT*[4], however, it appears in the extended vocabulary of *SciBERT*[5] and *SciNewsBERT*.

Finally, it is noteworthy that the *Random Forest* model provides quite comparable results to the transformer-based models, while being a much lighter and faster-to-train model.

4.5.2 Evaluation of Claim-Paper Clustering

Since we construct a bimodal clustering of claims and papers, we evaluate its quality with respect to two axes; a good-quality clustering must contain clusters of semantically related claims and papers (*Semantic Coherence*), and adhere to the implicit connections between these claims and papers (*Interconnection Coherence*).

Semantic Coherence

To measure the semantic coherence of a clustering, we compute a modified version of the *Average Silhouette Width* (*ASW*) [175]. The first modification is that the distance used is not a metric distance (e.g., Euclidean distance) but a semantic distance (Semantic Textual Similarity (*STS*)). The second modification is that we generalize the metric for two (or more) joint clusterings. The original metric computes the average distance between the centroid of each cluster and its elements. In our case, since we have two joint clusterings for claims and papers, we compute the metric for all the combinations of centroids ($\bar{c}$) and elements (e) of each cluster. Thus, the modified *ASW* is computed as follows:

[4]https://cdn.huggingface.co/bert-base-uncased-vocab.txt
[5]https://cdn.huggingface.co/allenai/scibert_scivocab_uncased/vocab.txt

$$ASW(cluster) = \frac{1}{|\text{ centroids }| \cdot |\text{ cluster }|} \sum_{\substack{e \in \text{cluster} \\ \bar{c} \in \text{centroids}}} STS(e, \bar{c})$$

where *centroids* consists of the claims centroid and the papers centroid of each cluster. Finally, we report the mean *ASW* across all clusters. This cross-computation of the metric allows capturing the semantic coherence of the clusters both individually and jointly.

Interconnection Coherence

To measure the interconnection coherence of the clusterings (i.e., the adaptivity of the clusterings towards the interconnection matrix *L*), we use ideas from link-based recommendation. First, we compute a hard clustering for claims and papers:

$$\begin{aligned} C'_{comp} &= argmax_x(C') \\ P'_{comp} &= argmax_x(P') \end{aligned}$$

Since, as we explain in Table 4.1, each row of C' and P' contains the probability of a claim or a paper to belong to a cluster, when we compute $argmax$ over rows we obtain a hard clustering, while when we compute $argmax$ over columns we obtain the cluster centroids. For example, given a single claim c and three clusters cl_0, cl_1, cl_2:

$$c' = [0.1, 0.8, 0, 1] \Rightarrow c'_{comp} = cl_1$$

Next, we use one clustering (e.g., of claims) to recommend possible instances of the other clustering (e.g., of papers). The recommendation is content-agnostic and exploits only the interconnection matrix *L*. Formally:

$$\begin{aligned} C'_{rec} &= argsort_x(sum_y(L \odot P')) \\ P'_{rec} &= argsort_x(sum_y(L^T \odot C')) \end{aligned}$$

where $\odot$ is the Hadamard (element-wise) product. For the same claim c, papers $p1$ and $p2$, and clusters cl_0, cl_1, cl_2 we have:

$$c \begin{array}{l} \nearrow p_1\,[0.5,0.1,0.4] \\ \searrow p_2\,[0.1,0.8,0.1] \end{array} \Rightarrow c'_{rec} = argsort_x(0.6, 0.9, 0.5) = [cl_1, cl_0, cl_2]$$

To compute the recommendation quality, we utilize the metric of *Recall@k* (*R@k*), which measures the ratio in which the correct cluster is recommended among the top-k results. We report the mean of the *R@k* for the claims and the papers clustering.

Table 4.4: Clustering Evaluation. *Semantic Coherence* is measured using the *Average Silhouette Width* (*ASW*), and *Interconnections Coherence* is measured using *Recall@3* (*R@3*).

		clusters=10		**clusters=50**		**clusters=100**	
		ASW	R@3	ASW	R@3	ASW	R@3
Content-Based	*LDA*	44.5%	86.8%	63.2%	69.4%	66.6%	69.5%
	GSDMM	42.1%	98.9%	48.5%	86.2%	48.7%	72.4%
	GMM	55.5%	68.9%	67.7%	52.4%	72.8%	45.2%
	PCA/GMM	51.3%	90.0%	66.6%	34.2%	71.7%	28.4%
	K-Means	53.2%	97.9%	**68.9%**	83.4%	73.2%	74.2%
	PCA/K-Means	52.0%	97.6%	66.8%	87.8%	71.2%	75.1%
Graph-Based	*GBA-CP*	38.2%	**100.0%**	40.9%	**100.0%**	44.5%	99.5%
	GBA-C	38.1%	96.7%	44.5%	93.2%	48.7%	92.0%
	GBA-P	40.0%	96.5%	43.0%	93.6%	47.3%	92.3%
	GBT-CP	26.5%	99.6%	27.1%	98.9%	32.1%	71.8%
	GBT-C	37.9%	92.5%	45.0%	59.8%	47.2%	53.8%
	GBT-P	36.4%	88.4%	42.3%	62.4%	43.7%	65.9%
Hybrid	*AO-Content*	54.8%	96.7%	67.9%	90.0%	**73.3%**	92.1%
	AO-Balanced	**56.0%**	99.8%	67.6%	99.6%	72.1%	99.5%
	AO-Graph	55.6%	99.8%	67.3%	**100.0%**	71.8%	**99.8%**

Results

The results of the evaluation are shown in Table 4.4. As we observe, the *Content-Based* (baseline) clustering techniques that use a textual representation of claims and papers (i.e., *LDA* and *GSDMM*), generate clusters with lower *Semantic Coherence* than the ones that use an embeddings representation (i.e., *GMM* and *K-Means*). This is partially explained by a vocabulary mismatch: the language used in papers is more complex and contains more scientific terms than the one used in social and news media (where the claims derive from). Thus, embeddings representations have the advantage of capturing the semantic proximity of topics, even if these topics occur from two heterogeneous vocabularies. Furthermore, we observe that soft clustering techniques (i.e., *LDA* and *GMM*) generate, in general, clusters with higher *Semantic Coherence* than the respective hard clustering techniques (i.e., *GSDMM* and *K-Means*), indicating that the theme of claims and papers is usually multifaceted. Finally, we observe that the dimensionality reduction, performed by *PCA*, is not helpful in the context of this task.

Regarding the *Graph-Based* techniques, we see that they construct clusters with high *Interconnections Coherence* but the lowest *Semantic Coherence*. Not surprisingly, *GBA-CP* achieves the maximum *Interconnections Coherence* since, as we explain in §4.3.2, it arbitrarily adapts the clusters to the interconnection matrix *L*.

Overall, we observe that the most robust technique in terms of balance between *Semantic* and *Interconnections Coherence* is the *Hybrid* technique (*AO-Balanced*), which computes a soft

clustering based on an embeddings representation and considers both the text and the graph modality of the dataset equally.

4.5.3 Evaluation of Claim Contextualization

The overall evaluation of our method is performed with an experiment that involves expert and non-expert fact-checkers as well as two state-of-the-art commercial systems. In this experiment, we evaluate the ability of SciClops to contextualize controversial claims in order to facilitate their verification by non-expert fact-checkers. Hence, given the ground-truth provided by experts, we compare the accuracy of non-expert fact-checkers that have or do not have access to the context provided by SciClops as well as the accuracy of the commercial systems.

Claim Processing

Using SciClops, we extract, cluster, and finally select the *top-40* check-worthy scientific claims in the data collection. The topics of the claims are heterogeneous, covering controversial online discussions such as the usage of therapeutic cannabis in modern medicine, the consumption of small amounts of alcohol during pregnancy, and the effect of vaccines in disorders such as autism. We notice that in some of the claims, redundant information that could confuse the fact-checkers is mentioned (e.g., we find the claim "Donald Trump has said vaccines cause autism," in which the scientific question is whether "vaccines cause autism" and not whether Donald Trump made this statement). Thus, to avoid misinterpretations and to mitigate pre-existing biases for or against public figures, we replace from these claims all the *Person* and *Organization* entities with indefinite pronouns.

Non-Experts

We employ crowdsourcing workers using the same setup described in §4.5.1, and ask them to evaluate the *Validity* of each claim in a *Likert Scale* [100] (from "Highly Invalid" to "Highly Valid"). We also ask them to rate their *Effort* to find evidence and their *Confidence* that the evidence they found is correct.

We divide non-experts into a control group of *Non-Experts Without Context*, and two experimental groups of *Non-Experts With Partial Context* and *Non-Experts With Enhanced Context*:

- *Non-Experts Without Context* are shown a bare scientific claim with no additional information, as they would read it online in, e.g., a messaging app or a social media posting.
- *Non-Experts With Partial Context* are shown a scientific claim and its source news article, i.e., the news article from which the claim was extracted.

- *Non-Experts With Enhanced Context* are shown a scientific claim, its source news article, and: i) the top-k news articles where similar claims were found, ii) the top-k most relevant papers, and, if available, iii) the top-k most similar, previously verified claims. To avoid overwhelming this experimental group with redundant information, we set $k = 3$.

Experts

We ask two independent experts with health (a senior Pediatrician) and biology (a Postdoctoral researcher in Microbiology) backgrounds to evaluate the validity of the claims. Each expert evaluated all 40 claims independently, and was given the chance to cross-check the ratings by the other expert and revise their own ratings, if deemed appropriate. Overall, we use the average of the two expert ratings as ground-truth.

Commercial Systems

Finally, for the verification of the same scientific claims, we use two commercial systems for fact-checking, namely ClaimBuster [85] and Google Fact Check Explorer[6]:

- *ClaimBuster* is a system used massively by journalists which initially aimed at detecting important factual claims in political discourses; however, its current architecture allows for investigating any kind of check-worthy claims (details in §2.1).
- *Google Fact Check Explorer* is also an exploration tool used by journalists to verify claims published using the tagging system of *ClaimReview*; we note that *ClaimReview* is also exploited in the contextualization step of SciClops (details in §4.4.2).

To homogenize the scores of these systems with the scores of the fact-checkers, we quantize them to the aforementioned *Likert Scale*.

Results

Results are summarized in Table 4.5. Given the ground-truth provided by the experts, we measure the accuracy of the three aforementioned groups of non-experts and the two commercial systems using the *Root Mean Square Error* (*RMSE*).

We observe that *ClaimBuster* performs better than our control group of *Non-Experts Without Context* while providing a solution without human intervention. Furthermore, we observe that *Google Fact Check Explorer* performs poorly, mainly because only 20% of the queried

[6]https://toolbox.google.com/factcheck/explorer

Table 4.5: Left: Root Mean Square Error (*RMSE*) between the scores provided by the *Experts* (ground-truth) and the scores provided by *Non-Experts* and *Commercial Systems*; the last row shows the *RMSE* across *Experts* (lower is better).
Right: Verification of two contradictory claims from *CNN* and *MensJournal* by *Non-Experts* and *Commercial Systems*; the last row shows the ground-truth provided by the *Experts*.

	RMSE	**CNN Claim**	**MensJournal Claim**
Non-Experts			
Without Context	1.91	*Borderline*	*Borderline*
With Partial Context	1.73	***Valid***	*Valid*
With Enhanced Context (SciClops)	**1.54**	***Valid***	***Highly Invalid***
Commercial Systems			
ClaimBuster	1.74	***Valid***	*Borderline*
Google Fact Check Explorer	2.79	*N/A*	*N/A*
Experts	1.02	***Highly Valid***	***Highly Invalid***

claims were present in the fact-checking portals it monitors (e.g., the claim "*Vaccines cause Autism*" is present in the fact-checking section of *USA Today* [207], while the *Contradictory Claims* described next are absent from all the fact-checking portals).

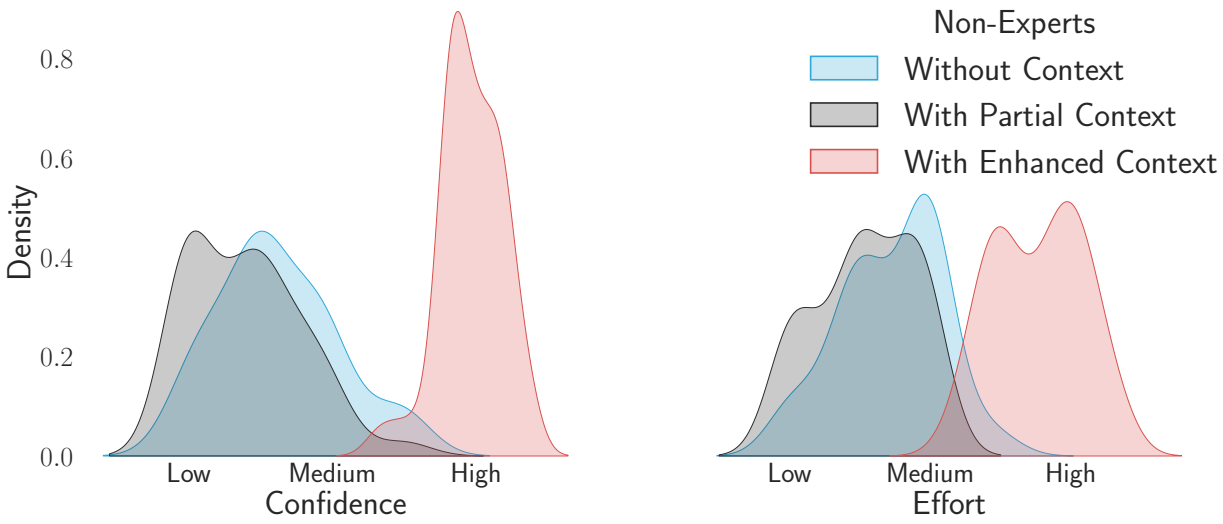
Finally, regarding the non-expert human fact-checkers, we observe that the more contextual information is available, the more accurately they rate the claims. Indicatively, the *RMSE* of *Non-Experts With Enhanced Context* is only *50%* greater than the *RMSE* across *Experts*. Overall, we see that, when the under-verification claims derive from a narrow scientific domain, **non-expert human fact-checkers, provided with the proper fact-checking context, may outperform state-of-the-art commercial systems**.

Case Study: Contradictory Claims

Within the set of under-verification claims, we noticed two contradictory claims. The first claim opposes the use of therapeutic cannabis for treating *Post-Traumatic Stress Disorder* (*PTSD*) and comes from a mainstream news outlet (*CNN*).[7] The second claim supports the use of cannabis for treating *PTSD* and comes from a popular health blog (*MensJournal*).[8] Current scientific understanding supports the first claim (from *CNN*), but not the second one (from *MensJournal*), as evidenced by a paper of the *Journal of Clinical Psychiatry* [214].

As we show in Table 4.5, *ClaimBuster* and all the groups of *Non-Experts* mostly support the claim from *CNN* as valid. Moreover, as discussed above, *Google Fact Check Explorer* provides no answer for these two claims since they are not present in the monitored fact-checking portals.

[7] *CNN*: "*Marijuana does not treat chronic pain or post-traumatic stress disorder.*" [181]
[8] *MensJournal*: "*Marijuana can help battle depression, anxiety, post-traumatic stress disorder, and even addictions to alcohol and painkillers.*" [104]



Average Work Time of *Non-Experts* (seconds)		
Without Context	*With Partial Context*	*With Enhanced Context*
191s	*388s*	***898s***

Figure 4.2: Kernel Density Estimation (*KDE*) of *Confidence* (left) and estimated *Effort* (right), and *Average Work Time* (bottom), of *Non-Experts* verifying claims. Best seen in color.

Indeed, only *Non-Experts With Enhanced Context* are able to indicate that the claim from *MensJournal* is invalid, mainly because **SciClops provides a fact-checking context that includes a paper from the *Journal of Clinical Psychiatry* which debunks the claim even in its title.**[9]

Case Study: Confidence & Effort

As we observe in Figure 4.2, *Non-Experts* that were shown the *Enhanced Context* of claims were more confident in their verification, additionally to being more accurate than the other two groups of users, which is partially explained by the fact that the provided context is fully-interpretable (as explained above), thus more trustworthy. However, the same users' self-assessment of their effort as well as their actual work time was higher than the other two groups of users, which is explained by the fact that they had to visit more potential verification sources.

[9] *Journal of Clinical Psychiatry*: "*Marijuana use is associated with worse outcomes in symptom severity and violent behavior in patients with posttraumatic stress disorder.*" [214]

4.6 Summary

In this chapter, we have described an effective method for assisting non-experts in the verification of scientific claims. We have shown that transformer models are indeed the state-of-the-art on scientific claim detection, however, they require domain-specific fine-tuning to perform better than other baselines. We have also shown that, by exploiting the text of a claim and its connections to scientific papers, we effectively cluster topically-related claims and papers, as well as that, by building an in-cluster knowledge graph, we enable the detection of check-worthy claims. Overall, we have shown that SciClops can build the appropriate fact-checking context to help non-expert fact-checkers verify complex scientific claims, outperforming commercial systems. We believe that our method complements these systems in domains with sparse or non-existing ground-truth evidence, such as the critical domains of science and health.

Chapter 5

Combating Article-Based Scientific Misinformation

This chapter describes SciLens, a method for evaluating the quality of scientific news articles. The starting point for our work is structured methodologies that define a series of quality aspects for manually evaluating news. Based on these aspects, we describe a series of news quality indicators, which derive from both the content and the context of articles, where context is provided by (1) explicit and implicit references on the article to scientific literature, and (2) reactions in social media referencing the article. SciLens introduces models for extracting such indicators, including models for quote extraction and attribution, semantic similarity between news articles and scientific papers, and social media stance classification. According to our experiments, these indicators help non-experts evaluate the quality of a scientific news article more accurately compared to non-experts that do not have access to these indicators. Furthermore, SciLens can also produce a completely automated quality score for an article, which agrees more with expert evaluators than manual evaluations done by non-experts.

5.1 Introduction

Scientific literacy is broadly defined as knowledge of basic scientific facts and methods. Deficits in scientific literacy are endemic in many societies, which is why understanding, measuring, and furthering the public understanding of science is essential to many scientists [11].

Mass media can be a potential ally in fighting scientific illiteracy. Reading scientific content has been shown to help align public knowledge of scientific topics with the scientific consensus, although, in highly politicized topics, it can also reinforce pre-existing biases [82]. There are many ways in which mass media approaches science, and even within the journalistic practice, there are several sub-genres. Scientific news portals include most of the categories of articles

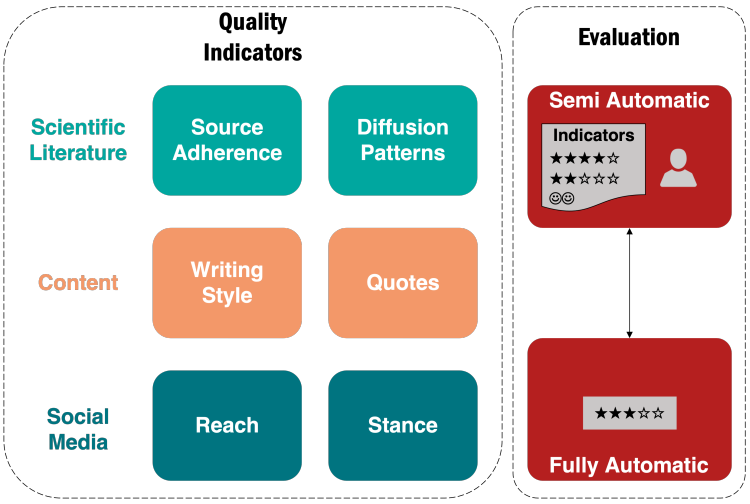


Figure 5.1: Overview of SciLens, including quality indicators from the content of articles and from their referencing social media postings and referenced scientific literature.

appearing traditionally in newspapers, such as *editorial* and *op-ed* [58]. However, the main category of articles is scientific *news* articles, where journalists describe scientific advances.

Scientific news articles have many common characteristics with other classes of news articles; for instance, they follow the well-known *inverted pyramid*[1] style, where the most relevant elements are presented at the beginning of the text. However, they also differ in important ways. Scientific news is often based on findings reported in scientific journals, books, and talks, which are highly specialized. The task of the journalist is then to *translate* these findings to make them understandable to a non-specialized, broad audience. By necessity, this involves negotiating several trade-offs between desirable goals that sometimes enter into conflict, including appealing to the public and using accessible language, while at the same time accurately representing research findings, methods, and limitations [152].

The resulting portrayal of science in news varies widely in quality. For example, the website "Kill or Cure?"[2] has reviewed over $1,200$ news stories published by The Daily Mail (a UK-based tabloid), finding headlines pointing to 140 substances or factors that cause cancer (including obesity, but also Worcestershire sauce), 113 that prevent it (including garlic and green tea), and 56 that both cause and prevent cancer (including rice). Evidently, news coverage of cancer research that merely seeks to classify every inanimate object into something that either causes or prevents cancer does not help to communicate effectively scientific knowledge on this subject.

The goal of SciLens is to help evaluate the quality of scientific news articles. Thus, we compute a series of quality indicators from the content of articles and from their referencing social media postings and referenced scientific literature (Figure 5.1).

[1] https://en.wikipedia.org/wiki/Inverted_pyramid_(journalism)
[2] http://kill-or-cure.herokuapp.com

Regarding the content of the articles, we begin by computing several baseline features described by previous work. Next, we perform an analysis of quotes in articles, which are quite prevalent in the case of scientific news. Given that attributed quotes are more telling of high quality than unattributed or "weasel" quotes, we also carefully seek to attribute each quote to a named entity which is often a scientist but can also be an institution.

Regarding the scientific literature, we would like to know the strength of the connection of articles to scientific papers. For this, we consider two groups of indicators: text-based and graph-based. Text-based indicators are built upon various metrics of text similarity between the content of an article and the content of scientific papers, considering that the technical vocabulary is unlikely to be preserved as-is in articles written for the general public. Graph-based indicators are based on a diffusion graph in which scientific papers and web pages in academic portals are nodes connected by links. High-quality articles are expected to be connected through many short paths to academic sources in this graph.

Regarding social media postings, we measure two aspects: reach and stance. Reach is measured through various proxies for attention that seek to quantify the impact that an article has on social media. The stance is the positioning of posting authors with respect to an article, which can be positive (supporting or commenting on an article without expressing doubts) or negative (questioning an article or directly contradicting what the article is saying).

We evaluate the extent to which the indicators computed in SciLens help determine the quality of a scientific news article. We consider that these indicators can be helpful in two ways. First, in a semi-automatic setting, we show the indicators to end-users and ask them to evaluate the quality of a scientific news article; if users who see these indicators are better at this task than users who do not see them, we claim that the indicators are useful. Second, in a fully automatic setting, we train a model based on all the computed indicators. In both cases, the ground-truth for evaluation is provided by experts in communication and science.

Our contribution. In this chapter, we describe SciLens, a method for evaluating the quality of scientific news articles. The technical contributions we introduce are the following:

- a series of automatically-computed quality indicators (Table 5.1) describing:
  - the content of a news article, where we introduce a method to use quotes appearing on it as quality indicators (§5.2);
  - the relationship of a news article with the scientific literature, where we introduce text-based and graph-based similarity methods (§5.3);
  - the social media reactions to the article, where we introduce a method to interpret stance as quality signal (§5.4);
- an experimental evaluation of our methods involving experts and non-experts (§5.5).

5.2 Content Indicators

In this section, we introduce our content indicators, i.e., the indicators which are based on the textual content of a news article. These indicators consider the writing style of an article as well as the usage of attributed or unattributed quotes.

5.2.1 Writing-Style (Baseline) Indicators

As a starting point, we adopt a set of content-based quality indicators described by previous work. These indicators are described below:

- *Title Deceptiveness* and *Sentiment*: we consider if the title is "clickbait" that oversells the contents of an article in order to pique interest [130, 211];
- *Article Readability*: we consider the level of education someone would need to easily read and understand the article [56];
- *Article Length* and presence of *Author Byline*: we consider the verbosity of an article and whether author details are available to the readers [225].

5.2.2 Quote-Based Indicators

Quotes are a common and essential element of many scientific news articles. While selected by journalists, they provide an opportunity for experts to directly present their viewpoints in their own words [34]. However, identifying quotes, in general, is challenging, as noted by previous work (§2.2.3). In the specific case of our corpus, we observe that they are seldom contained in quotation marks in contrast to other kinds of quotes (e.g., political quotes [164]). We also note that each expert quoted tends to be quoted once, which makes the problem of *attributing* a quote challenging as well. An illustrative example of our extraction and the attribution procedure is shown in Figure 5.2.

Quote Extraction Model

To extract quotes, we start by addressing a classification problem at the level of a sentence, i.e., we want to distinguish between quote-containing and non-containing sentences. To achieve this, we first select a random sample from our dataset, then manually identify quote patterns, and finally, we generalize automatically these patterns to cover the entire dataset. As we describe in the related work section (§2.2.3), this is known as a "bootstrapping" model oriented to detect high-precision patterns.

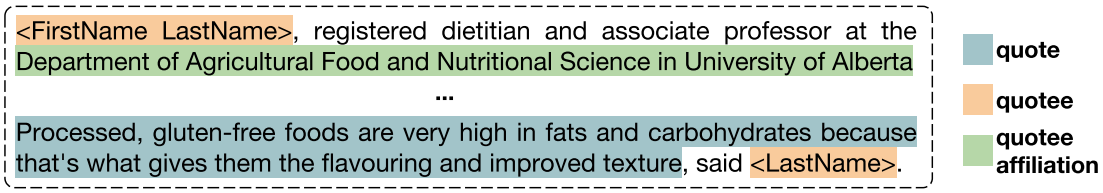


Figure 5.2: Example of quote extraction & attribution (quotee anonymized). Best seen in color.

The usage of *reporting verbs* is a typical element of quote extraction models [155]. Along with common verbs that are used to quote others (e.g., "say," "claim"), we used verbs that are common in scientific contexts, such as "prove" or "analyze." First, we manually create a seed set of such verbs. Next, we extend it with their nearest neighbors in a word embedding space; the word embeddings we use are the *GloVe* embeddings, which are trained on a corpus of Wikipedia articles [158]. We follow a similar approach for nouns related to studies (e.g., "survey," "analysis") and nouns related to scientists (e.g., "researcher," "analyst"). Syntactically, quotes are usually expressed using indirect speech. Thus, we also obtain part-of-speech tags from the candidate quote-containing sentences.

Using this information, we construct a series of regular expressions over *classes* of words ("reporting verbs," "study-related noun," and part-of-speech tags) which we evaluate in §5.5.1.

Quote Attribution

To evaluate article quality, it is essential to know not only that an article has quotes, but also their provenance: *who* or *what* is quoted. After extracting all the candidate quote-containing sentences, we categorize them according to the information available about their quotee.

A quotee can be an *unnamed scientist* or an *unnamed study* if the person or article being quoted is not disclosed (e.g., "researchers believe," "most scientists think" and other so-called "weasel" words). Sources that are not explicitly attributed, such as these ones, are, as a general rule, considered less credible than sources in which the quotee is named [225].

A quotee can also be a *named entity* identifying a specific person or organization. In this case, we apply several heuristics for quote attribution. If the quotee is a *named person*, if they are referred with their last or first name, we search within the article for the full name. When the full name is not present in the article, we map the partial name to the most common full name that contains it within our corpus of news articles. We also locate sentences within the article that mention this person together with a named organization. This search is performed from the beginning of the article as we assume they follow an *inverted pyramid* style. In case there are several, the most co-mentioned organization is considered as the affiliation of the quotee. An example of a detected quotee and quotee affiliation is shown in Figure 5.2.

If the quotee is an *organization*, then it can be either mentioned in full or using an acronym. We map acronyms to full names of organizations when possible (e.g., we map "WHO" to "World Health Organization"). If the full name is not present in an article, we follow a similar procedure as the one used to determine the affiliation of a researcher, scanning all the articles for co-mentions of the acronym and a named organization.

Scientific Mentions

News articles tend to follow journalistic conventions rather than scientific ones [40]; regarding citation practices, this implies they seldom include formal references in the manner in which one would find them in a scientific paper. Often there is no explicit link: journalists may consider that the primary source is too complex or inaccessible to readers to be of any value, or may find that the scientific paper is located in a "pay-walled" repository. However, even when there is no explicit link to the paper(s) on which an article is based, good journalistic practices still require identifying the information source (institution, laboratory, or researcher).

Mentions of academic sources are partially obtained during the quote extraction process (§5.2.2), and complemented with a second pass that specifically looks for them. During the second pass, we use the list of universities and scientific portals that we used during our contextual news collection (§3) to identify them as potential quotees in the article.

5.3 Scientific Literature Indicators

In this section, we describe text- and graph-based indicators measuring how articles are related to the scientific literature. Specifically, these indicators measure the semantic adherence and the web proximity to the primary scientific sources of the articles.

5.3.1 Source Adherence Indicator

When there is an explicit link from a news article to a scientific paper, we can measure the extent to which these two documents convey the same information. This is essentially a computation of the *Semantic Textual Similarity* between the news article and its source.

Supervised Learning for Semantic Textual Similarity

We construct a model using supervised learning; the features that we use as input to the model consist of the following text similarity metrics:

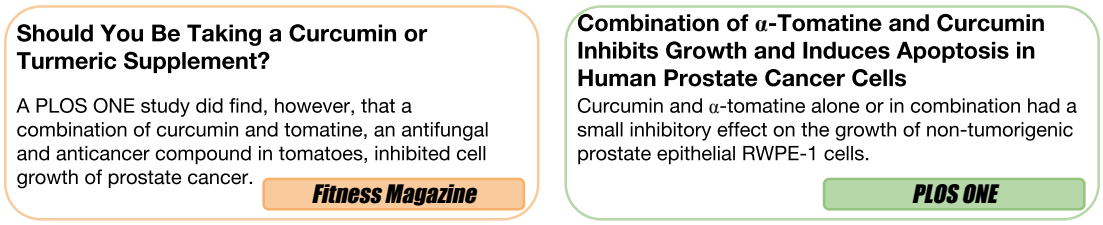


Figure 5.3: A news article (left) and a scientific paper (right) with Semantic Textual Similarity of 87.9%. Indicatively, two passages from these documents, whose conceptual similarity is captured by our method, are presented. We can see the effort of the journalist in translating from an academic to a less formal language without misrepresenting the results of the paper.

- the Jaccard similarity between the sets of named entities (persons and organizations), dates, numbers, and percentages of the two texts;
- the cosine similarity between the *GloVe* embeddings of the two texts;
- the Hellinger similarity [87] between the LDA topic vectors [14] of the two texts;
- the relative difference between the length in words of the two texts.

Each of the similarities is computed three times: i) considering the entire contents of the article and the paper; ii) considering one paragraph at a time, and then computing the average similarity between a paragraph in one document and a paragraph in the other; and iii) considering one sentence at a time, and then computing the average similarity between a sentence in one document and a sentence in the other. In other words, in (ii) and (iii), we compute the average of each similarity between the Cartesian product of the passages.

The training data that we use is automatically created from pairs of documents consisting of a news article and a scientific paper. Whenever a news article has exactly one link to a scientific paper, we add the article and the paper to training data in the positive class. For the negative class, we sample random pairs of news articles and papers. Details regarding the evaluation of these schemes are provided in §5.5.1. An example of a highly related pair of documents, as determined by this method, is shown in Figure 5.3.

Handling Multi-Sourced Articles

When an article has a single link to a scientific paper, we use the Semantic Textual Similarity of them as an indicator of the article quality. When an article has multiple links to scientific papers, we select the one that has the maximum score according to the model we just described. We remark that this is just an indicator of article quality, and we do not expect that by itself it is enough to appraise the quality of the article. Deviations from the content of the scientific paper are not always wrong, and indeed a good journalist might consult multiple sources and summarize them in a way that re-phrases content from the papers used as sources.

5.3.2 Diffusion Graph Indicators

We also consider that referencing scientific sources, or referencing pages that reference scientific sources, are good quality indicators. Figure 3.1 showing a graph from scientific papers to articles, and from articles to social media postings and from them to their reactions, suggests this can be done using graph-based indicators. We consider the following:

- personalized PageRank [86] on the graph having scientific articles and universities as root nodes and news articles as leaf nodes;
- betweenness and degree on the full diffusion graph [59, 60].

Additionally, we consider the traffic score computed by Alexa.com for the website in which each article is hosted, which estimates the total number of visitors to a website.

5.4 Social Media Indicators

We extract signals describing the quantity and characteristics of social media postings referencing each article. Quantifying the amount of reactions in various ways might give us signals about the interest in different articles (§5.4.1). However, this might be insufficient or even misleading, if we consider that false news may reach a larger audience and propagate faster than actual news [206]. Hence, we also need to analyze the content of these postings (§5.4.2).

5.4.1 Social Media Reach

Not every social media user posting the URL of a scientific news article agrees with the content of the article, and not all users have sufficient expertise to appraise its contents properly. Indeed, sharing articles and reading articles are often driven by different mechanisms [1]. However, and similarly to citation analysis and to link-based ranking, the volume of social media reactions to an article might be a signal of its quality, although the same caveats apply.

Given that we do not have access to the number of times a social media posting is shown to users, we extract several proxies of the *reach* of such postings. First, we consider the total number of postings including a URL, and the number of times those postings are "liked" in their platform. Second, we consider the number of followers and followees of posting users in the social graph. Third, we consider a proxy for international news coverage, which we operationalize as the number of different countries (declared by the users themselves) from which users posted about an article.

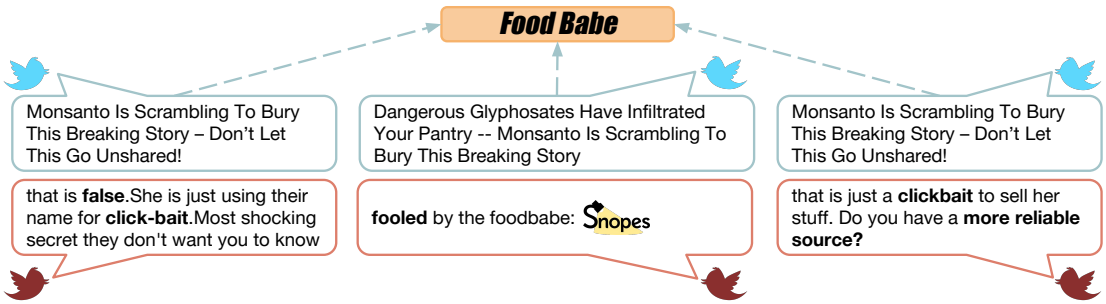


Figure 5.4: Example in which the stance of social media replies (bottom row) indicates the poor quality of an article promoted through a series of postings (top row).

Additionally, we assume that a level of attention that is sustained can be translated to a larger exposure and may indicate long-standing interest in a topic. Hence, we consider the temporal coverage, i.e., the length of the time window during which postings in social media are observed. To exclude outliers, we compute this period for 90% of the postings, which is also known in the literature as the "shelf life" of the article [26].

5.4.2 Social Media Stance

We consider the stance of social media postings with respect to the article they link to, as well as the stance of the responses (replies) to those postings. According to what we observe in this corpus, repliers sometimes ask for (additional) sources, express doubts about the quality of an article, and in some cases post links to fact-checking portals that contradict the claims of the article. These repliers are, indeed, acting as "social media fact-checkers," as the example in Figure 5.4 shows. Following a classification used for analyzing ideological debates [83], we consider four possible stances: supporting, commenting, contradicting, and questioning.

Retrieving Replies

Twitter's API does not provide a programmatic method to retrieve all the replies to a tweet. Thus, we use a web scraper that retrieves the text of the replies of a tweet from the page in which each tweet is shown on the web. The design of this web scraper is straightforward and allows us to retrieve all the *first-level* replies of a tweet.

Classifying Replies

To train our stance classifier, we use: i) a general-purpose dataset provided in the context of *SemEval 2016* [138], and ii) a set of 300 tweets from our corpus, which were annotated by crowd-

Table 5.1: Summary of all the quality indicators provided by SciLens

Context	Type	Indicator
Article	Baseline	Title [Clickbait, Subjectivity, Polarity], Article Readability, Article Word Count, Article Bylined
	Quote-Based	#Total Quotes, #Person Quotes, #Scientific Mentions, #Weasel Quotes
Sci. literature	Source Adherence	Semantic Textual Similarity
	Diffusion Graph	Personalized PageRank, Betweenness, [In, Out] Degree, Alexa Rank
Social media	Reach	#Likes, #Retweets, #Replies, #Followers, #Followees, [International News, Temporal] Coverage
	Stance	Tweets/Replies [Stance, Subjectivity, Polarity]

sourcing workers. From the first dataset, we discard tweets that are not relevant to our corpus (e.g., debates on *Atheism*); thus, we keep only debates on *Abortion* and *Climate Change*. The second set of annotated tweets is divided into 97 contradicting, 42 questioning, 80 commenting, and 71 supporting tweets. We also have 10 tweets that were marked as "not-related" by the annotators, and thus we exclude them from our training process. The combined dataset contains $1,140$ annotated tweets. The learning scheme we use is a Random Forest classifier based on features including the number of: i) total words, ii) positive/negative words (using the Opinion Lexicon [93]), iii) negation words, iv) URLs, and v) question/exclamation marks. We also computed the similarity between the replies and the tweet being replied to (using cosine similarity on *GloVe* vectors [158]) and the sentiment of the reply and the original tweet [123]. Details regarding the evaluation are provided in § 5.5.1.

5.5 Experimental Evaluation

In our experimental evaluation we use our "bottom-up" news collection (details in §3). We begin the evaluation by studying the performance of the methods we have described to extract quality indicators (§5.5.1). Then, we evaluate if these indicators correlate with scientific news quality. First, we determine if publications that have a good (bad) reputation or track record of rigor in scientific news reporting have higher (lower) scores according to our indicators (§5.5.2). Second, we use labels from experts (§5.5.3) to compare quality evaluations done by non-experts with and without access to our indicators (§5.5.4).

5.5.1 Evaluation of Indicator Extraction Methods

In this section, we evaluate individually the models introduced for quote extraction and attribution, source adherence, and social media stance.

Quote Extraction and Attribution

The evaluation of our quote extraction and attribution method (§5.2.2) is based on a manually-annotated sample of articles from our corpus. A native English speaker performed an annotation finding 104 quotes (37 quotes attributed to persons, 33 scientific mentions, and 34 "weasel" or unattributed quotes) in a random sample of 20 articles.

We compare three algorithms: i) a baseline approach based on regular expressions searching for content enclosed in quote marks, which is usually the baseline for this type of task; ii) our quote extraction method without the quote attribution phase, and iii) the quote extraction and attribution method, where we consider a quote as correctly extracted if there is no ambiguity regarding the quotee (e.g., if we extract only the last name, we consider it as incorrect).

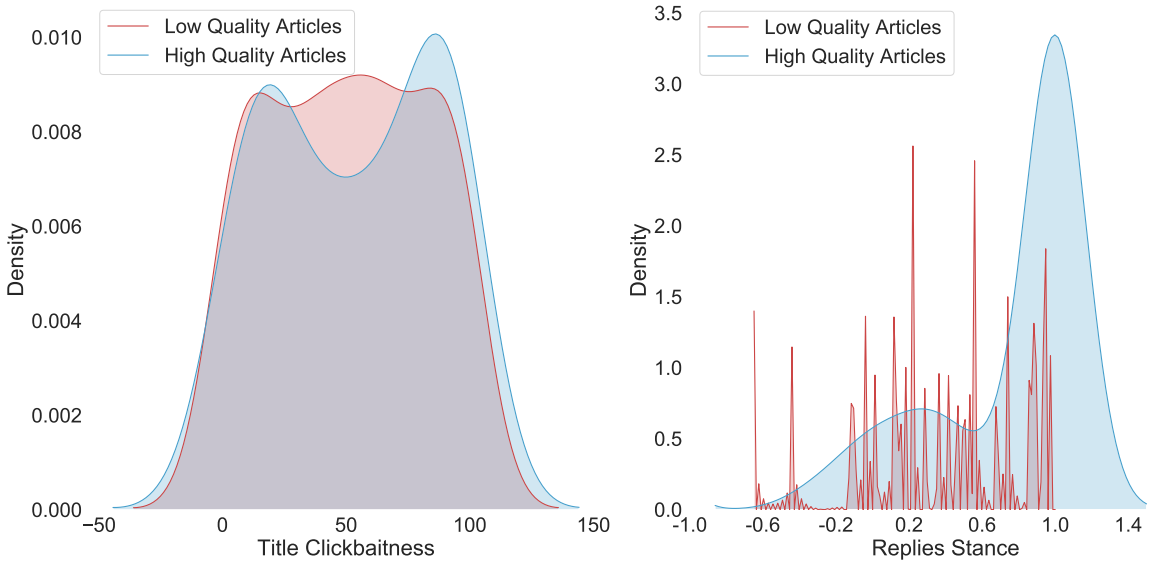
Although the baseline approach has the optimal precision, it is unable to deal with cases where quotes are not within quote marks, which are the majority (100% precision, 8.3% recall). Thus, our approach, without the quote attribution phase, improves significantly in terms of recall (81.8% precision, 45.0% recall). Remarkably, the heuristics we use for quote attribution work well in practice and increase both precision and recall (**90.9%** precision, **50.0%** recall).

Source Adherence

We use the supervised learning method described on §5.3.1 to measure Semantic Textual Similarity. We test three different learning models: Support Vector Machine, Random Forest, and Neural Network (double-layer, fully-connected perceptron). The three classifiers use similarities computed at the document, sentence, and paragraph level and combine all features from the three levels. Overall, the best accuracy (**93.5%**) was achieved by using a Random Forest classifier and all the features from the three levels of granularity, combined.

Social Media Stance

We evaluate the stance classifier described in §5.4.2 by performing 5-fold cross-validation over our dataset. When we consider all four possible categories for the stance (supporting, commenting, contradicting, and questioning), the accuracy of the classifier is **59.42%**. This is mainly due to confusion between postings expressing mild support for the article and postings just commenting on the article, which also tend to elicit disagreement between annotators. Hence, we merge these categories into a "supporting or commenting" category comprising postings that do not express doubts about an article. Similarly, we consider "contradicting or questioning" as a category of postings expressing doubts about an article; previous work has observed that indeed false information in social media tends to be questioned more often (e.g., [27]). The problem is then reduced to binary classification.



(a) (b)

Figure 5.5: Kernel Density Estimation (KDE) of a traditional quality indicator (*Title Clickbaitness* on the left) and our proposal quality indicator (*Replies Stance* on the right). We observe that for both high and low quality articles, the distribution of *Title Clickbaitness* is similar; thus, the indicator is non-informative. However, most of the high quality articles have *Replies Stance* close to 1.0, which represents the *Supporting/Commenting* class of replies, whereas low quality articles span a broader spectrum of values and often have smaller or negative values representing the *Contradicting/Questioning* class of replies. Best seen in color.

To aggregate the stance of different postings that may refer to the same article, we compute their weighted average stance considering supporting or commenting as $+1$ (positive stance) and contradicting or questioning as -1 (negative stance). As weights, we consider the popularity indicators of the postings (i.e., the number of likes and retweets). This is essentially a text quantification task [64], and the usage of a classification approach for a quantification task is justified because our classifier has nearly identical pairs of true positive/negative rates (**80.65%** and **80.49%** respectively), and false positive/negative rates (**19.51%** and **19.35%** respectively).

5.5.2 Correlation of Indicators among Portals of Diverse Reputability

We use two lists that classify news portals into different categories by reputability. The first list, by the American Council on Science and Health [5] comprises 50 websites sorted along two axes: whether they produce evidence-based or ideologically-based reporting and whether their science content is compelling. The second list, by Climate Feedback [46], comprises 20 websites hosting 25 highly-shared stories on climate change, categorized into five groups by scientific credibility, from very high to very low.

Table 5.2: Top five discriminating indicators for articles sampled from pairs of outlets having different levels of reputability (p-value: < 0.005 ***, < 0.01 **, < 0.05 *)

The Atlantic vs. Daily Mail (very high vs. very low)	NY Times vs. Daily Mail (medium vs. very low)
Alexa Rank***	Alexa Rank***
#Scientific Mentions***	Article Bylined***
Article Readability**	#Scientific Mentions***
#Total Quotes*	Article Readability***
Title Polarity	#Total Quotes**
The Atlantic vs. NY Times (very high vs. medium)	**All Outlets (from very high to very low)**
Alexa Rank***	Alexa Rank***
Article Bylined***	Article Bylined***
Article Word Count*	Article Word Count***
#Replies*	#Scientific Mentions***
#Followers	Article Readability***

We sample a few sources according to reputability scores among the sources given consistent scores in both lists: high reputability (The Atlantic), medium reputability (New York Times), and low reputability (The Daily Mail). Next, we compare all of our indicators in the sets of articles in our collection belonging to these sources. Two example features are compared in Figure 5.5. We perform ANOVA [54] tests to select discriminating features. The results are shown in Table 5.2. Traffic rankings by Alexa.com, scientific mentions, and quotes are among some of the most discriminating features.

5.5.3 Expert Evaluation

We ask a set of four external experts to evaluate the quality of a set of articles. The experts include three people who work in communication of science in an academic context and one biologist. Two of them evaluated a random sample of 20 articles about the gene-editing technique CRISPR, a specialized topic discussed recently in mass media. The other two experts evaluated a random sample of 20 articles on the effects of Alcohol, Tobacco, and Caffeine (the "ATC" set in the following), which are frequently discussed in science news.

Experts read each article and gave it a score in a *Likert Scale*, from very low quality to very high quality. Each expert annotated the 20 articles independently and was given afterward a chance to cross-check the ratings by the other expert and revise their own ratings if deemed appropriate. The agreement between experts is distributed as follows:

- *Strong Agreement*, when the expert rates are the same (7/20 in ATC, 6/20 in CRISPR);

(a) Articles on Alcohol, Tobacco, and Caffeine

(b) Articles on CRISPR

Figure 5.6: Evaluation of two sets of 20 scientific articles. The line corresponds to expert evaluation, while the bars indicate fully automatic evaluation (red), assisted evaluation by non-experts (light blue), and manual evaluation by non-experts (dark blue). Best seen in color.

- *Weak Agreement*, when the rates differ by one point (12/20 in ATC, 10/20 in CRISPR);
- *Disagreement*, when the rates differ by two or more points (1/20 in ATC, 4/20 in CRISPR).

5.5.4 Expert vs. Non-Expert Evaluation

We perform a comparison of quality evaluations by experts and non-experts. Non-experts are workers in a crowdsourcing platform. We ask for five non-expert labels per article and employ what our crowdsourcing provider, *Figure Eight*, calls tier-3 workers, which are the most experienced and accurate. As a further quality assurance method, we use the agreement among workers to disregard annotators producing consistently annotations that are significantly different from the rest of the crowd. This is done at the worker level, and as a result, we remove on average about one outlier judgment per article.

We consider two experimental conditions. On the first condition, entitled *Non-Expert (No Indicators)*, non-experts are shown the exact same evaluation interface as experts. On the second condition, entitled *Non-Expert (Indicators)*, non-experts are shown 7 of the quality indicators we produced, which are selected according to Table 5.2. Each indicator (except the last two) is shown with stars, with ★☆☆☆☆ indicating that the article is in the lowest quintile according to that metric, and ★★★★★ indicating the article is in the highest quintile. The following legend is provided to non-experts to interpret the indicators:

Table 5.3: Differences among expert evaluations, evaluations provided by non-experts, and fully automated evaluations provided by SciLens, measured using RMSE (lower is better). ATC and CRISPR are two sets of 20 articles each. Strong agreement indicates cases where experts fully agree, weak agreement when they differed by one point, and disagreement when they differed by two or more points. No-Ind. is the first experimental condition for non-experts, in which no indicators are shown. Ind. is the second condition, in which indicators are shown.

	Experts by agreement	#	**Non-Experts** No ind.	Ind.	**Fully Automated**
ATC	Strong agreement	7	0.80	**0.45**	1.41
	Weak agreement	12	1.28	1.18	**0.76**
	Disagreement	1	0.40	1.30	**0.00**
	All articles	20	1.10	**1.00**	**1.00**
CRISPR	Strong agreement	6	1.40	1.17	**1.00**
	Weak agreement	10	0.86	0.76	**0.67**
	Disagreement	4	**0.96**	1.22	1.03
	All articles	20	1.96	0.96	**0.85**

Visitors per day of this news website (more visitors = more stars)
Mentions of universities and scientific portals (more mentions = more stars)
Length of the article (longer article = more stars)
Number of quotes in the article (more quotes = more stars)
Number of replies to tweets about this article) (more replies = more stars)
Article bylined by its author (✓= bylined, ✗= not bylined)
Sentiment of the article's title (☺☺= most positive, ☹☹= most negative)

Results of comparing the evaluation of experts and non-experts in the two conditions we have described are summarized in Figure 5.6. In the figure, the 20 articles in each set are sorted by increasing expert rating; assessments by non-experts differ from expert ratings, but this **difference tends to be reduced when non-experts have access to quality indicators**.

In Table ,5.3 we show how displaying indicators leads to a decrease in these differences, meaning that non-expert evaluations become closer to the average evaluation of experts, particularly when experts agree. In the ATC set, the improvement is small, but in CRISPR, it is large, **bringing non-expert scores about 1 point (out of 5) closer to expert scores**.

Table 5.3 and Figure 5.6 also include a fully automated quality evaluation built using a weakly supervised classifier over all the features we extracted. As weak supervision, we used the lists of sites in different tiers of reputability (§5.5.2) and considered that *all articles* on each site had the same quality score as the reputation of the site. Then, we used this classifier to annotate the 20 articles in each of the two sets. Results show that **this classifier achieves the lowest error with respect to expert annotations**.

5.6 Summary

In this chapter we have described a method for evaluating the quality of scientific news articles. We have introduced new quality indicators that consider quotes in the articles, the similarity and relationship of articles with the scientific literature, and the volume and stance of social media reactions. The approach is general and can be applied to any specialized domain where there are primary sources in technical language that are "translated" by journalists and bloggers into accessible language.

In the course of this work, we developed several quality indicators that can be computed automatically, and demonstrated their suitability for this task through multiple experiments. First, we showed several of them are applicable at the site level, to distinguish among different tiers of quality with respect to scientific news. Second, we showed that they can be used by non-experts to improve their evaluations of quality of scientific articles, bringing them more in line with expert evaluations. Third, we showed how these indicators can be combined to produce fully automated scores using weak supervision, namely data annotated at the site level.

Chapter 6

Combating Source-Based Scientific Misinformation

This chapter describes SciLander, a method for learning representations of news sources reporting on science-based topics. The COVID-19 pandemic has fueled the spread of misinformation on social media and the Web as a whole. The phenomenon dubbed 'infodemic' has taken the challenges of information veracity and trust to new heights by massively introducing seemingly scientific and technical elements into misleading content. Despite the existing body of work on modeling and predicting misinformation, the coverage of very complex scientific topics with inherent uncertainty and an evolving set of findings, such as COVID-19, provides many new challenges that are not easily solved by existing tools.

SciLander introduces four heterogeneous indicators for the news sources; two generic indicators that capture (1) the copying of news stories between sources, and (2) the use of the same terms to mean different things (i.e., the semantic shift of terms), and two scientific indicators that capture (1) the usage of jargon and (2) the stance towards specific citations. We use these indicators as signals of source agreement, sampling pairs of positive (similar) and negative (dissimilar) samples, and combine them in a unified framework to train unsupervised news source embeddings with a triplet margin loss objective. We evaluate our method on a novel COVID-19 dataset containing nearly 1M news articles from 500 sources spanning a period of 18 months since the beginning of the pandemic in 2020. Our results show that the features learned by our model outperform state-of-the-art baseline methods on the task of news veracity classification. Furthermore, a clustering analysis suggests that the learned representations encode information about the reliability, political leaning, and partisanship bias of these sources.

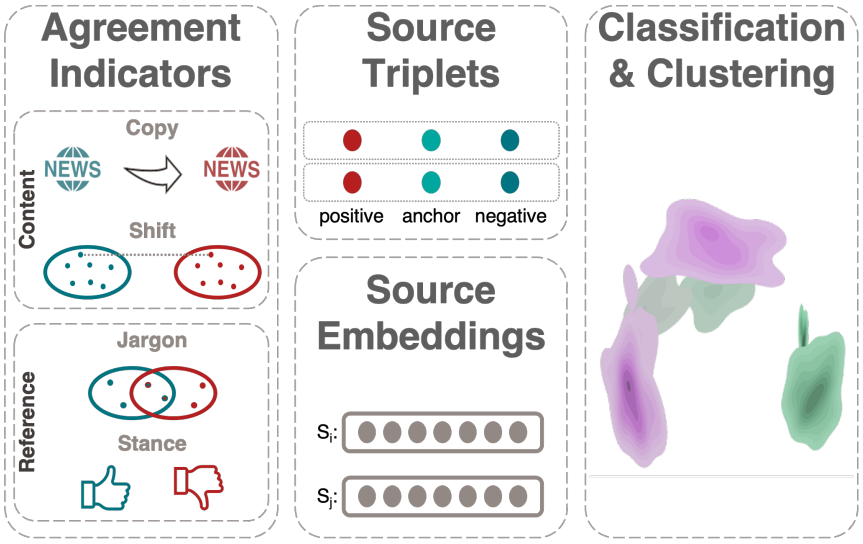


Figure 6.1: Overview of SciLander, including agreement indicator extraction (§6.2 & §6.3), triplet sampling and unsupervised source embeddings training (§6.4), and evaluation on the downstream tasks of classification and clustering (§6.5).

6.1 Introduction

The COVID-19 pandemic has resulted in a significant increase in information production and consumption at the same time. With this came a large increase in unreliable information, dubbed 'infodemic' [21]. This increase was also coupled with the growing scrutiny of media sources and purposeful amplification of any errors they made. As the readers sought correct, timely, and trustworthy information, many news and media sources worked hard to discredit others and create confusion [202].

Governments and public health agencies have the responsibility to respond to the crisis and protect the public from misinformation by utilizing the power of social and news media [25]. Yet, the same social and news media work as a catalyst for the infodemic, allowing disinformation to be widely dispersed, regardless of the significant effort to hinder its spread [131].

Despite the existing body of work on modeling and predicting misinformation, coverage of a complex scientific topic with inherent uncertainty and evolving set of findings, such as COVID-19, provides many new challenges that are not easily solved by existing tools [224]. On the article level, the evaluation of news stories may be challenging as they may contain information that cannot be easily verified. Moreover, many sources may not have the necessary staffing for the proper communication of scientific topics; they may be known to publish incorrect information, this information may change over time, or the source may correct it.

Often, language-based methods fail in such a task because different sources may use the same terms to mean different things. Furthermore, many sources may use scientific references

to back up their claims; however, the validity of these references is not easily verifiable. Being able to map out the consequential and systematic patterns of behavior of such sources in terms of both *content* and *references* would be particularly useful in such scenarios [32]. It would allow sources to be compared to other known sources in terms of their coverage, and develop explanations to the aspects in which they are similar to or different from each other.

To address these challenges, we introduce a novel method called SciLander. SciLander builds on a set of novel features, based on the deep processing of news articles published by a set of sources, producing a vector representation of these news sources. To build this, we incorporate measures of similarity and difference between the sources based on their citation behavior, the republishing of articles from each other, and their general language usage. In particular, we use the coverage of COVID-19 to show that this embedding has many desirable features that can help multiple downstream tasks.

Our Contribution. The technical contributions we introduce are the following:

- We propose four news agreement indicators for sources: i) the shared content or republished articles, ii) the semantic shift of terms in the common vocabulary, iii) the usage of scientific jargon, and iv) the citation stance of the news sources (§6.2 & §6.3);
- We combine these indicators in a unified framework for training unsupervised news source embeddings (§6.4);
- We evaluate our method using a dataset of news publications related to COVID-19. Sources in this dataset are labeled with respect to reliability and political leaning;
- We compare our method to strong baselines on the problem of veracity classification of news sources and show a significant gain in performance when combining the indicators proposed in our work;
- We test the applicability of our method in an online learning experiment, showing that it can be used to learn features from sources even if little data is available or if new coming sources are presented in the landscape;
- We show that the learned features encode information about the reliability level, partisanship bias, and political leaning of news sources through a clustering analysis experiment.

6.2 Content Indicators

In this section, we introduce two content-based indicators that we use to align news sources. Particularly, we introduce an indicator regarding the shared content and an indicator regarding the semantic shift of terms between sources.

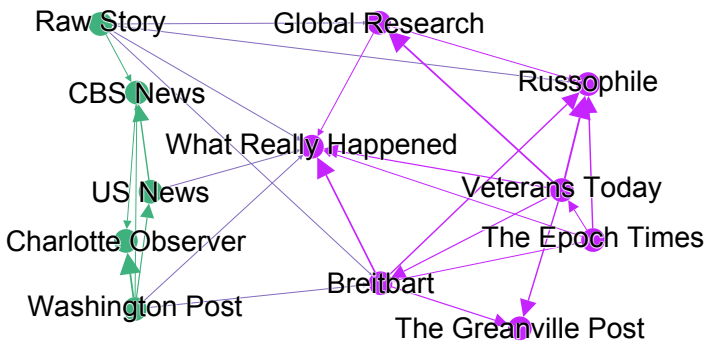


Figure 6.2: Example of a subgraph of the Content Sharing Network where nodes, representing sources, are connected by directed edges denoting the direction of the copied content between sources. Node color indicates the reliability class of the source (green for *Reliable*, purple for *Unreliable*), and edge width indicates the amount of content copied.

6.2.1 Copy Indicator

Content Sharing Network is a model of content replication by sources in the news landscape. The sharing of news articles has been shown to be a common factor between news sources that adopt similar narratives around certain topics, which also correlates with the credibility of these sources [91]. Figure 6.2 illustrates how sources are related in a Content Sharing Network, where articles are copied from source to source.

Content Sharing Network is modeled as a directed graph where nodes represent news sources and edges indicate sources that copy articles verbatim from one another. Edges weights are proportional to the amount of content copied between the connected sources. The adjacency matrix C of such network represents the affinity between the news sources. We obtain this matrix using the method proposed by Horne et al. [91] which consists of computing document vector representations for news articles using a TF-IDF bag-of-words representation. Articles are considered verbatim copies of each other if the cosine similarity between their vectors is greater than a threshold of 0.85, and the direction of the copying is determined by the publication date of the article. The similarity threshold is defined following the recommendations from Horne et al. [91].

The final adjacency matrix is obtained by aggregating all copied articles at the source level. Thus, a directed edge from node i to j exists if source j copies articles from source i. The complement of the degree of relatedness distance between sources i and j, is given as a function of the weight of the edge (i, j) and is defined as:

$$d_{cpy}(i, j) = 1 - \frac{|A_i \cap A_j|}{|A_j|}$$

where A_i and A_j are articles published by sources i and j; thus, their intersection should contain articles from source i copied by source j.

Table 6.1: Semantic shift of the term "antiviral". We observe a contextual shift of the word. In the top two cases, the term is used to describe alternative medicine with herbs, while in the bottom two cases, the term is used with its ordinary (scientific) connotation.

Source	**Usage**
Modern Alternative Mama	{...} **these specific herbs have strong antiviral actions**, including against other strains of coronavirus.
Healthy Holistic Living	{...} Garlic is known to have potent antibacterial, **antiviral**, antifungal and antiprotozoal abilities.
The Guardian	{...} overwhelming emergency departments and causing governments to overspend on **antiviral medications**.
The Washington Post	{...} although the **antiviral drug** remdesivir has been shown to help some patients {...}

6.2.2 Shift Indicator

We analyze how specific technical terms are used differently between news sources. Different uses of a certain term in two pieces of text can occur if that same term is used in a different context in each of the texts. Semantic shift is the process through which the usage of a given word drifts when compared across different sources. Specifically, we consider the lexical semantic shift, which posits the semantics of a word to be defined by its contextual relationships to other lexicons [35]. We argue that significant contextual shifts of topic-related words may serve as a signal of source disagreement, i.e., two sources using a certain target word in significantly different contexts may indicate that they use such words with different intents. An illustrative example is shown in Table 6.1. We note that, in both examples, the word antiviral is still used to indicate "something that is effective against viruses"; however, the contexts give different connotations to what the antiviral product is.

Semantic shift has been used extensively in computational linguistics studies of language evolution [78] and, more recently, in studies quantifying the linguistic differences across domains [179, 221]. In our method, we use semantic shift as an indicator of agreement among sources as it helps to uncover unique narratives created by unreliable sources, especially those based on conspiracy theories, deviating significantly from the narratives from reliable media.

The semantic shift between two sources i and j is measured by the deviation in the usage of words they have in common. Specifically, we define semantic shift as the aggregated distance between word embeddings for terms in the common vocabulary of sources i and j. However, because the word embeddings are trained independently from each other, they cannot be directly compared. For example, suppose that v_a and v_b denote word vectors for the word *virus* learned from the sources *The Washington Post* and *Global Research*, respectively. The cosine distance $d_{cos}(v_a, v_b)$ is not meaningful unless we first create a mapping between the embedding spaces of each source. This mapping can be achieved by applying an orthogonal transforma-

tion to one of the embedding spaces to minimize the sum of the pairwise Euclidean distances between word vectors of the common vocabulary. Being orthogonal means that this transformation preserves the inner product of the embeddings in the transformed space; for that reason, this mapping is also called embedding *alignment* [78, 101].

Finding the best alignment of two embedding spaces is not a trivial task. Learning a transformation from all the words in the common vocabulary is often undesired, as the objective of the mapping is to minimize the distance between every pair of word vectors, hence minimizing the distance between words that are potentially semantically distinct [221]. To learn alignments between word embeddings, we employ the state-of-the-art self-supervised semantic shift (S4) method [73], which is designed to select the best words for generating a mapping between two embeddings. This procedure is applied to embeddings trained using Word2Vec [135].

Once we train and align the embeddings, we compute the semantic distance between sources i and j as the average cosine distance between the top 10% most frequent words in i and j (stop words excluded). Thus, the distance between sources i and j is defined as:

$$d_{sem}(i,j) = \frac{\sum_{v \in V_i \cap V_j} cos(emb_i(v), emb_j(v))}{|V_i \cap V_j|}$$

where V_i and V_j are the vocabularies of sources i and j, $emb_i(v)$ and $emb_j(v)$ compute the embeddings representation of word v, and cos computes the cosine distance between the embeddings. Additionally, V_i and V_j may be replaced with subsets of the common vocabulary to avoid using every word in the analysis (e.g., filter for the most frequent words).

6.3 Reference Indicators

In this section, we introduce the reference indicators that we used to align news sources. Particularly, we introduce two dedicated scientific indicators, namely, the usage of scientific jargon and the citation stance. These indicators are *reference indicators*, i.e., they define a distance among sources given a common (scientific) reference.

6.3.1 Reference Context Extraction

To compute the *reference indicators*, we need the textual context of the references, i.e., the paragraph in which these references are cited. To extract this context, we: i) locate the references by parsing the raw HTML page of each news article of our data collection, and ii) traverse the structural tree of the page to discover the most fine-grained text passage that contains the reference. Currently, we do not support end-notes within articles, i.e., anchors at the bottom of articles where all the scientific references are listed, because it is a journalistic practice rarely appearing in our corpus.

Table 6.2: Usage of scientific jargon when citing a report by *CDC* [36]. We highlight that the citation context of *TheNewYorker* is semantically closer to the referenced CDC report than the citation context of *RedState*.

Source	**Reference Context**
TheNewYorker	In June, just three months into a historic health crisis, a survey by the Center for Disease Control and Prevention found that forty per cent of Americans were already struggling with at least one **mental-health issue**.
RedState	It is no wonder that many Americans have **lost their faith** throughout 2020. Too many leaders have been inconsistent in their actions minus their continued breaches of the public trust.

Reference	**Title**
CDC	**Mental Health**, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic — United States, June 24–30, 2020.

6.3.2 Jargon Indicator

This indicator quantifies the scientific nature of the context in which a reference is used. To estimate this indicator, we need a lexicon of terms ($jargon\_terms$ in the following) that are considered jargon in the scientific domain of our corpus. Since, as we explain in §3.2, our corpus contains news articles related to COVID-19, we use the vocabulary of *CDC A-Z Index*[1], manually enhanced with common COVID-19 terminology. After applying standard cleaning (e.g., punctuation removal), we compute the following distance:

$$d_{jar}(i,j) = |ctx_r(i) \cap ctx_r(j) \cap jargon\_terms|$$

where $ctx_r(i)$ and $ctx_r(j)$ are the terms in the citation contexts of sources i and j for each common reference r.

We note that we do not aggregate for all common references between sources i and j; hence, we do not limit to a single distance between these sources. In this way, we encode the co-citation volume between sources i and j, which is useful for our triplet sampling strategy (details in 6.4.1). After computing $d_{jar}(i,j)$, we apply *Min-Max Normalization* in the interval $[0,1]$ to comply with the previously-defined distances. As we observe in Table 6.2, even such a simplistic metric is able to capture cases in which news sources completely distort the scientific message of the cited reference.

[1]https://www.cdc.gov/az

Table 6.3: Stance of news sources when citing a webinar by *CDC* [6]. We highlight that *The Truth About Cancer* uses more emotionally loaded words than *FiveThirtyEight*.

Source	**Reference Context**
Five Thirty Eight	{...} But even as the guidelines were revised and the national death count — which includes probable as well as confirmed cases — shot upward, experts said that undercounting was still more likely than overcounting.
The Truth About Cancer	Perhaps worst, the **CDC has continued to lie** about the death count by artificially inflating it. CDC guidelines for determining COVID-19 deaths include: Anyone who tests positive, even if they died from other causes. Anyone who had COVID-19 symptoms, even if they aren't tested.

Reference	**Title**
CDC	Guidance for Certifying Deaths Due to Coronavirus Disease 2019 (COVID-19)

6.3.3 Stance Indicator

This indicator quantifies the sentiment charge of the context in which a reference is cited. To measure this sentiment charge, we use the *Multi-Genre Natural Language Inference* model *BART* for zero-shot classification [117]. This model[2] computes the probability that we infer a certain *hypothesis* given a *premise*. Thus, the model needs no explicit training on the downstream task of stance classification since the desired classes are provided implicitly in the *hypothesis*. After experimenting with various templates for *premise* and *hypothesis*, we report the ones that yield the most reliable results:

$$\textit{premise} = \textbf{\textit{reference context}}$$
$$\textit{hypothesis} = \textit{"The stance of this example is } \textbf{\textit{negative}}\textit{"}$$

The output of this model is a value in the interval $[0, 1]$, denoting the probability a given premise implies our hypothesis. We note that, by using this premise and hypothesis, we treat *neutral* and *positive* stances similarly, i.e., as *non-negative* stances because we want to highlight extremely negative stances (Table 6.3). Using this model we compute the following distance:

$$d_{ref}(i, j) = |stance(ctx_r(i)) \; - \; stance(ctx_r(j))|$$

where $stance(.)$ computes the stance of the citation contexts of sources i and j for each common reference r.

[2] https://huggingface.co/facebook/bart-large-mnli

6.4 Unsupervised Source Embeddings

In the previous section, we described the heterogeneous indicators that we extract from each news source. In this section, we describe how we combine these indicators in a unified framework capable of learning unsupervised representations of news sources. The triplets sampling and embeddings training methods employed in this framework are well-established methods [89] used mainly in learning-to-rank recommendation systems [30, 209].

6.4.1 Triplet Sampling

Our goal is, using the distances defined by the indicators, to discover pairs of similar sources and pairs of dissimilar sources. By joining these two sets of pairs, we create triplets of the form (*anchor, positive, negative*), where *anchor* is the common element of the pairs, *positive* is the element similar to the *anchor*, and *negative* is the element dissimilar to the *anchor*. For simplicity, in the following, we will refer to these triplets as (*a, p, n*).

We note that these triplets may not occur from the same indicator, i.e., the positive pair may occur from an indicator that is more appropriate for capturing the affinity between sources, and the negative pair may occur from an indicator that is more appropriate for capturing the disparity between sources. In our experimental evaluation (§6.5.1), we evaluate each indicator in its ability to produce good positive and negative pairs as well as full triplets.

Positive Pair Sampling

We use the distances computed for each indicator to generate pairs of similar sources. For all indicators we introduce in §6.2 & §6.3, short distance denotes similarity. Given an indicator f ($copy, shift, jargon,$ or $reference$), we generate a positive pair of similar sources i, j with a probability inversely proportional to the distance between i and j:

$$pp_f(i,j) = \frac{d_f^{-1}(i,j)}{\sum_k d_f^{-1}(i,k)} \; \forall j \neq i$$

We draw l positives samples from this distribution for each indicator and each source in the dataset, producing a total of l positive source pairs (a, p).

Negative Pair Sampling

For negative sampling, we employ two strategies. For some indicators (e.g., the stance indicator), a large distance between sources denotes opposing sentiment, thus disagreement (e.g.,

the sources in Table 6.3). Hence, we use the inverse distribution we used for generating positive pairs to generate negative pairs:

$$np_f(i,j) = 1 - pp_f(i,j)\ \forall j \neq i$$

Similarly as above, we draw l negative samples from this distribution for each indicator and each source in the dataset, producing a total of l negative source pairs (a,n).

Nonetheless, there are indicators (e.g., the copy indicator) for which a large distance between sources does not necessarily denote disagreement; it only denotes the absence of agreement. In these cases, we draw the negative pairs uniformly from the set of sources.

Finally, we employ a cleaning heuristic to increase the accuracy of our triplets (detailed experiment in §6.5.1). Specifically, we make sure that we do not select a negative pair (a,n) which we have already selected as positive pair (a,p):

$$(a,p) \wedge (a,n) \Rightarrow p \neq n$$

6.4.2 Embeddings Training

Once we extract all the triplets, we use them for training a dense representation model for news sources with the *Triplet Margin Loss* [7]. The learning objective of *Triplet Margin Loss* is to minimize the distance between an anchor and a positive sample while maximizing the distance between the anchor and the negative sample.

The procedure we employ is the following. First, we initialize the embeddings for all the sources into a low-dimensional, dense vector space by randomly setting the weights in the embedding layer following a normal distribution $\mathbb{N}(0,1)$. Then, given the input triplets (a,p,n), we train these embeddings by minimizing the loss function L:

$$L(a,p,n) = max\{d(a,p) - d(a,n) + M, 0\}$$

where d is the distance function, and M is the margin parameter that controls the gap between positive and negative distances. The larger M is, the larger is the gap between $d(a,p)$ and $d(a,n)$. We train the embeddings over several epochs until convergence and then use them as the representation of the news sources.

The parameters of this method are the margin M, the distance function d, and the size of the output vectors s. We release the optimal training parameters as well as the trained sources embeddings in our code release.

6.5 Experimental Evaluation

Our experimental evaluation is three-fold; first, we evaluate the indicators individually, then we evaluate the source embeddings on the downstream task of source reliability classification, and finally, we perform an unsupervised clustering where we analyze the patterns in the news sources captured by the learned features. In the following experiments, our "middle-up" news collection is used (details in §3), word embeddings for the semantic shift are trained using Word2Vec with dimension 100, context window of 10, and minimum word count of 20. The parameters for SciLander are margin $M = 1$, vector size $s = 50$, and d is the cosine distance.

6.5.1 Indicator Coverage

In our first experiment, we measure the overlap of the introduced indicators in terms of source and triplet coverage. We also measure the accuracy of the triplets computed by these indicators. We define the source coverage (sc) and the triplet coverage (tc) between two indicators i, j as follows:

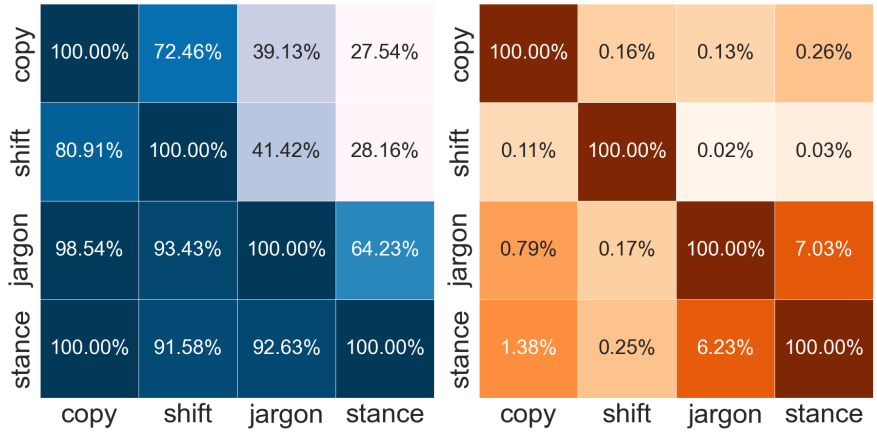
$$sc(i,j) = \frac{|src(i) \cap src(j)|}{|src(i)|},\ tc_(i,j) = \frac{|trpl(i) \cap trpl(j)|}{|trpl(i)|}$$

where $src(.)$ and $trpl(.)$ compute the distinct set of sources and triplets covered by a given indicator. We note that the metrics sc and tc are non-symmetric; consequently, the source and triplet coverage heatmaps in Figure 6.3 are also non-symmetric.

To measure the accuracy of the computed triplets, we use the metric *Area Under the Receiver Operating Characteristics* (AUROC), which measures the *True Positive Rate* over the *False Positive Rate*. We also break down the AUROC of the triplets into i) the $AUROC_p$ of the positive part of the triplets (a, p), ii) the $AUROC_n$ of the negative part of the triplets (a, n), and iii) the $AUROC_f$ of the full triplets (a, p, n). Specifically, for each individual AUROC, we consider the following as true positives:

$$AUROC_p : \{(a,p)\ s.t.\ label(a) = label(p)\}$$
$$AUROC_n : \{(a,n)\ s.t.\ label(a) \neq label(n)\}$$
$$AUROC_f : \{(a,p,n)\ s.t.\ label(a) = label(p) \wedge label(a) \neq label(n)\}$$

As we observe in Figure 6.3, although the sources covered by some indicators heavily overlap, the contributed triplets are quite unique. Indicatively, the stance indicator covers 27.5% of the sources, totally overlapping with the copy indicator. However, the contributed triplets of the stance indicator are different from the contributed triplets of all the other indicators and also more accurate. Indeed, we see that there is a trade-off between the source coverage of the indicators and the AUROC. Hence, the more specific the indicator is (e.g., the stance indicator), the better AUROC it has.



Indicator	**AUROC$_\mathbf{p}$**	**AUROC$_\mathbf{n}$**	**AUROC$_\mathbf{f}$**	**#sources**
copy	72.7%	51.0%	36.3%	257
shift	61.9%	60.8%	41.8%	308
stance	89.7%	73.3%	68.3%	87
jargon	81.9%	51.0%	42.9%	126
overall	77.0%	69.7%	57.5%	316

Figure 6.3: Overlap of indicators in terms of source coverage (top left) and triplet coverage (top right); AUROC of the positive part, negative part, and full triplets (bottom). Although the sources covered by most indicators heavily overlap, their triplets are quite unique. Also, there is a trade-off between the source coverage of the indicators and their AUROC.

Finally, we observe that the overall AUROC for positive and negative pairs (AUROC$_p$ and AUROC$_n$, respectively) are above the 50% baseline of a random positive (or negative) pair selection is truly positive (or negative).

It should be noted that the AUROC for complete triplets (AUROC$_f$) is lower than 50%. This happens because the choice of the final triplets involves two independent decisions: the choice of the positive sample, and the choice of the negative sample. As noted above, each choice has a chance of success of 50% if chosen at random. Thus, for a triplet to be correctly selected, the random baseline is that a correct positive pair is chosen *and* a correct negative pair is chosen, which results in a $0.5 \times 0.5 = 0.25$, or 25% baseline chance. As we see in the following experiments, the model for training source embedding is robust to noisy triplets as it yields highly accurate results in all the downstream tasks we use it.

6.5.2 Offline Source Classification

In this experiment, we evaluate the computed embeddings on a downstream classification task. We assume that, for all sources in our corpus, we have (offline) access to a significant fraction of their history of published articles.

Baselines

For this task, we implement baselines using *Stylistic Text Features*, *Contextualized Embeddings*, and *Co-citation Embeddings*, as well as combinations of the above.

Stylistic Text Features. We utilize stylistic text features from Horne et al. [90] aggregated at the source level as representations. These features include, among others, the number of: part of speech tags, punctuation symbols, and capitalized words, which are the features that are typically used in news classifiers.

Contextualized Embeddings. We compute BERT [38] embeddings for a total of 32 tokens from the title and the opening paragraph of the article, and average them for each source. Similarly, we compute SciBERT [12] instead of BERT embeddings, which have been shown to lead to better performance in tasks involving scientific text. The configuration parameters of both BERT and SciBERT are those suggested in a widely used release of this model [215].

Co-Citation Embeddings. We compute a co-citation graph of sources based on their scientific references. We weight this graph either uniformly for each common reference, or by emphasizing the uniquely used references, using their TF-IDF score. In the overall graph, we run node2vec [72] to extract source embeddings.

Joint Embeddings. The *Contextualized Embeddings* and the *Co-Citation Embeddings* capture two different modalities of news sources; their content and citation behavior. Thus, we create a joint representation by concatenating the two embeddings. Since the dimensionality of the joint embeddings is high, we apply Principal Component Analysis to reduce it.

Evaluation

We test the usefulness of the learned representations in the problem of source veracity classification. We use the embeddings computed by i) SciLander trained on all indicators, ii) SciLander trained only on content indicators (shift or copy), and iii) the aforementioned baseline models, to train a Nearest Neighbors classifier in a 10-fold cross-validation setting. Figure 6.4 shows the F1 score of each model for increasing values of k.

Relying uniquely on textual features limits classifiers to a restricted set of signals. Our framework combines stylistic, semantic, and behavioral indicators to produce a representation that improves the separation of reliable and unreliable sources. Thus, compared to traditional baselines such as stylistic features or features extracted by BERT, our embeddings show significant performance improvement. **Our method obtains the best F1 score (87%) for $\mathbf{k = 37}$.**

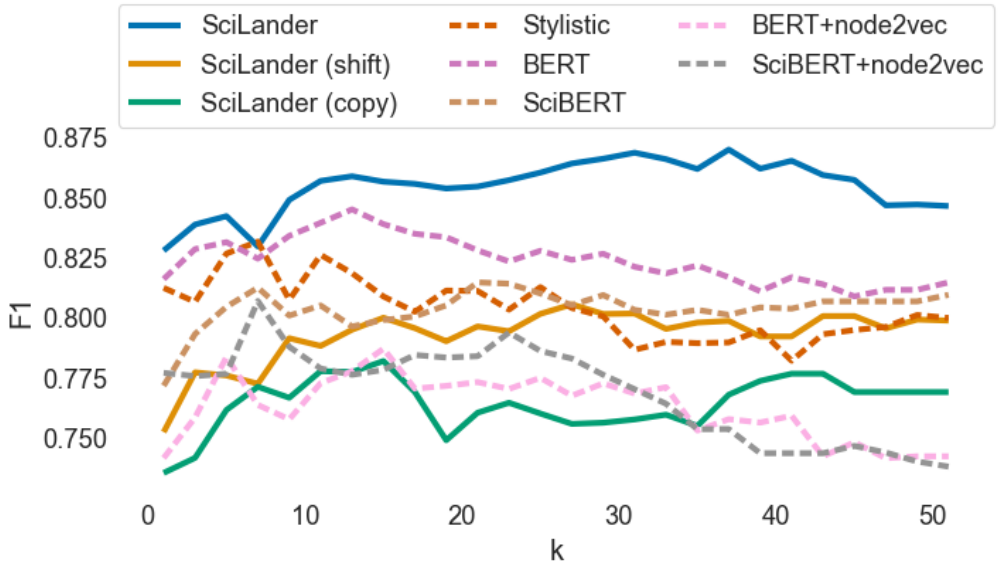


Figure 6.4: F1 scores using k-nearest neighbors classifiers over the source embeddings representations computed by SciLander and the various baselines described in §6.5.2. SciLander obtains the best F1 score (*87%*) for *k=37*.

6.5.3 Online Source Classification

In this experiment, we assume that we have two types of sources: i) offline (known) sources, for which we have access to a significant fraction of their publication history, and ii) online (newcomer) sources, for which we have access to a limited fraction of their publication history. As assessing articles from newcomer sources might be a time-consuming task, we inspect the lowest fraction of articles that is needed to accurately classify these sources.

The procedure that we employ is the following: i) we train embeddings for the *offline sources* (as explained in §6.4.2); ii) we freeze these embeddings for the *offline sources*; iii) we train embeddings for the *online sources*, in the already shaped by the *offline sources* embeddings space.

We conduct the experiment on a 10-fold cross-validation setting. In Figure 6.5, we report the learning curve (F1 score) for increasing fractions of articles from newcomer sources in the same classification task described in §6.5.2. We note that the temporal axis is not in chronological order but sampled randomly from the entire corpus (e.g., we sample articles representing a 3-month publishing activity of an online source from the entire publishing activity of that source). In that way, each temporal interval is independent of external events (e.g., the development of the vaccines), which affects the activity of most sources. As we observe in Figure 6.5, **SciLander is able to reliably classify sources, using only three months of their publishing activity**.

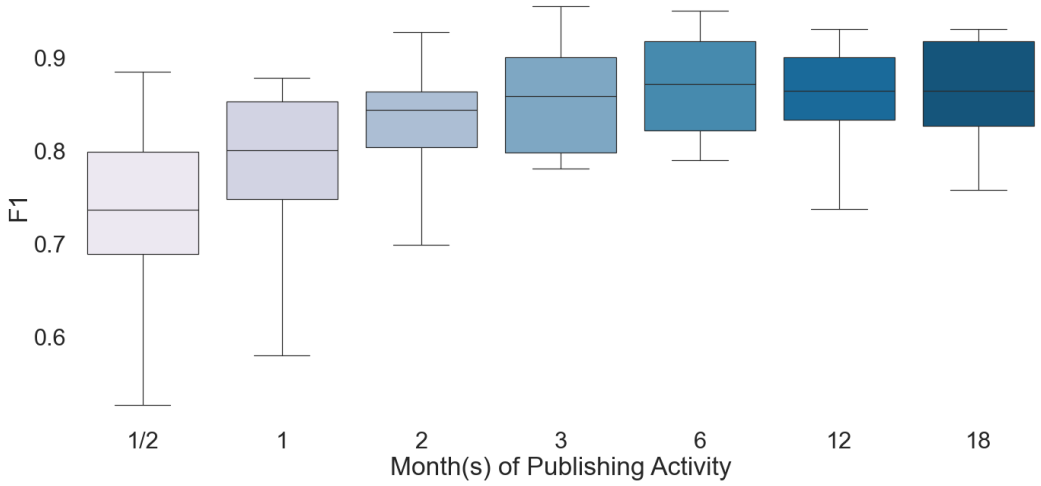


Figure 6.5: Learning curve (F1 score) for increasing fractions of articles from newcomer sources. SciLander reliably (*F1>85%*) classifies sources using only *3* months of their publishing activity.

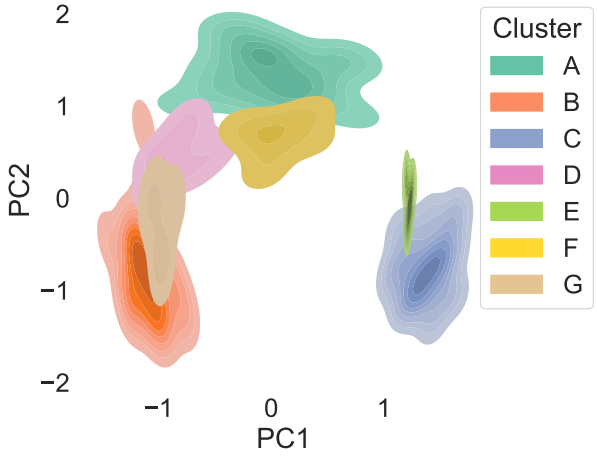
6.5.4 Source Clustering Analysis

We conduct an unsupervised clustering experiment to investigate potential trends revealed by the features learned by SciLander. Using the same embeddings from the previous experiments (50 dimensions, $M = 1$), we apply DBSCAN clustering to the source vectors with the cosine distance as distance metric, minimum distance parameter $\epsilon = 0.1$ and minimum cluster size $n = 1$. The resulting clusters are shown in Figure 6.6(a); each of the 7 clusters is shown in different color shades and labeled from A to G.

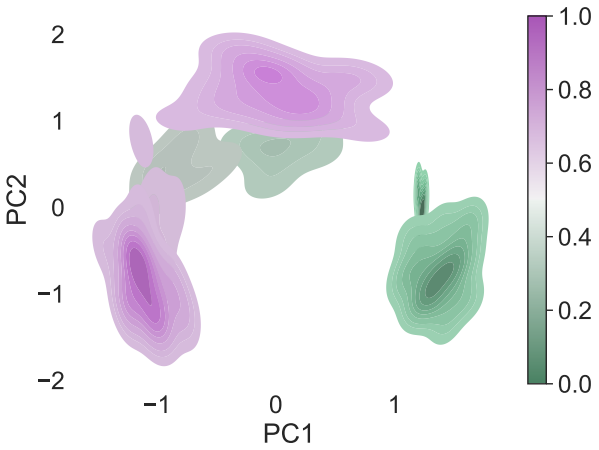
We characterize the clusters quantitatively with respect to the density of unreliable sources, political leaning, and the level of partisanship bias aggregated across the news sources within them. For each cluster, we compute the proportion of unreliable sources to the total number of sources in the cluster. Figure 6.6(b) shows the density of unreliable sources within each cluster. This result suggests that **the source embeddings carry information about source credibility when grouping them, even though credibility labels or related features were unknown to the model during training**.

Clusters C and E contain no unreliable sources and hold mostly mainstream news sources such as The Washington Post, Vox, National Public Radio (NPR), and Chicago Tribune. The clusters containing the largest proportions of unreliable sources are the clusters A, B, and G, and most sources in these clusters are websites that propagate conspiracy theories and promote pseudoscience. Details on the discovered clusters are shown in Table 6.4.

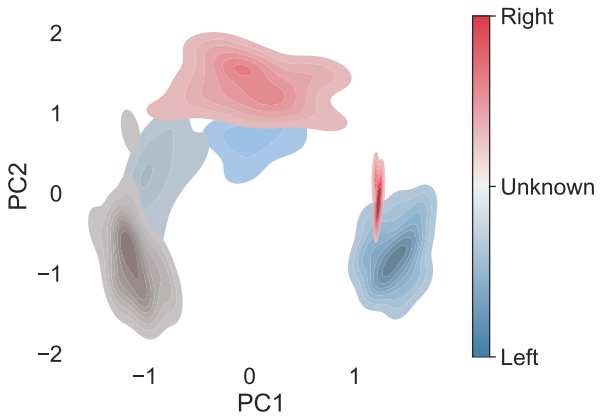
These results show that the SciLander embeddings are able to group sources based on similar reliability. Multiple clusters of relatively high purity with respect to reliability are created, some reliable (75%-100% reliable sources), some unreliable (0%-30% reliable sources).



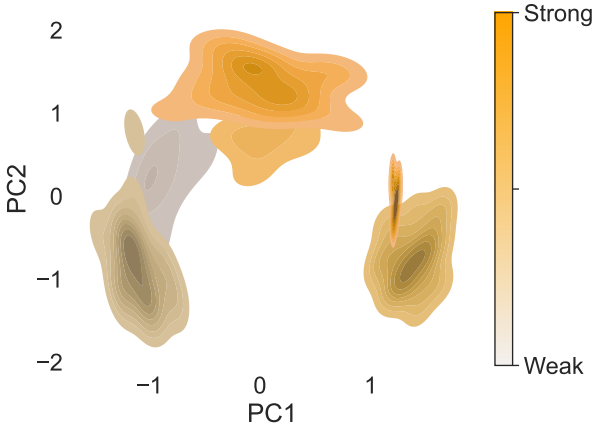
(a) Kernel Density Estimation of the clusters



(b) Density of Unreliable Sources



(c) Density of Political Leaning



(d) Density of Partisanship Bias

Figure 6.6: Density analysis of the clusters computed by SciLander. Components PC1 and PC2, obtained from Principal Component Analysis on the source embeddings, are the components with the highest explained variance ratio.

We also compute the overall political leaning of a cluster by averaging the political leaning scores of the sources within that cluster. Partisanship bias is obtained by the absolute value of leaning, scaled to a value in $[0, 1]$, with 0 indicating that there is no partisanship bias in the cluster, and 1 indicating the maximum partisanship bias, where all sources in the cluster exhibit a strong political leaning. The partisanship bias describes the agreement between the political leanings of sources within the cluster, and the magnitude of such leanings. The distribution of political leanings and partisanship bias are shown in Figures 6.6(c) and 6.6(d). There is a noticeable disparity between the partisanship bias found in the two biggest unreliable clusters A and B. Sources in cluster A exhibit a strong bias, which is nearly absent in cluster B. We explore the particularities of these clusters next.

Table 6.4: Unreliability score (proportion of unreliable sources), average Partisanship score, and core sources (nearest neighbors to the centroid) of the identified clusters.

Cl.	Unreliability	Partisanship	Core Sources
A	.70	.25	NewsWars, Veterans Today, The D.C. Clothesline
B	.84	.03	Mercola, Healthy Holistic Living, Vaccine Reaction
C	.00	.11	The Washington Post, Vox, NPR
D	.25	.00	The American Conservative, Roll Call
E	.00	.20	Chicago Tribune
F	.12	.03	Washington Monthly, FiveThirtyEight, Atlantic
G	.80	.00	Ice Age Now

6.5.5 Different Types of Conspiracy Theories

We observe two clusters with high density of unreliable sources (clusters A and B). Both clusters include many unreliable news sources, and there exist qualitative differences between them, which we describe in this section.

To uncover qualitative differences between sources in clusters A and B, we measure the shift in context each of these clusters has when compared to the mainstream cluster C. Specifically, we compute the semantic shift across clusters of sources by training Word2Vec models E_A, E_B, and E_C using articles from the core sources of each cluster and using the same hyperparameters as in the previous experiments. Then, we extract the words with the highest cosine distance between pairs (E_C, E_A) and (E_C, E_B) to find the terms that most contribute to the deviation in the news from sources in C to each of the unreliable clusters A and B.

Let S_A and S_B be the lists of the 100 words most shifted to C, from A and B, respectively. We find that there is only one word in common between the S_A and S_B: "natural". To characterize the words in both lists, we identify words that refer to people, entities and places, political issues, and health and nutrition. Examples of these words are given below and listed on Table 6.5.

The largest group of words shifted in cluster A are related to individuals, entities, places (25%), and political topics (12%). Almost all individuals found are political figures (with a few exceptions). There are only 1.5% of terms related to health and nutrition. Many of these news outlets are conspiracy theory websites such as NewsWars, Veterans Today, and InfoWars. According to a Media Bias/Fact Check analysis[3], these sites often publish hate-speech-filled content in addition to misleading or false information.

In contrast, the largest group of shifted words was detected in cluster B (21.5%), with only 2% people and 1% related to political topics. We note that the only person found as shifted in Cluster B is Bill Gates which does not appear as a top shifted word for Cluster A due to a common set of conspiracy theories surrounding Bill Gates and the Gates Foundation claiming

[3]https://mediabiasfactcheck.com/veterans-today

Table 6.5: List of words from clusters *A* and *B* that are most shifted from the mainstream cluster *C*. People, Places, and Political Terms appear as the most shifted words in cluster *A*, suggesting that its sources push politically-oriented misinformation, while sources in cluster *B* focus more on alternative health solutions.

Cluster A		**Cluster B**
People and Places	**Political Terms**	**Health**
Kamala Harris	BLM (Black Lives Matter)	Coronavirus
Bernie Sanders	Patriot	Food
Nancy Pelosi	Voting	Vaccines
Mike Pence	Abortion	Doctors
Alex Jones	Partisan	Mask

that he plotted to use the pandemic to seize power [61]. Two of the most prominent sources of this cluster are Mercola and Healthy Holistic Living. According to Media Bias/Fact Check journalists[4], these sources promote alternative health notions, sell questionable products and supplements, and promote anti-vaccination positions with pseudoscience-based arguments.

Based on this, we conclude that while cluster A is a cluster of mostly politically-unreliable news sources covering COVID-19 stories mixed with other political topics, cluster B is much more focused on covering alternative medicine-based misinformation with slight political leaning, presumably to appeal to individuals with different political opinions. On these sites, health-based information is often mixed with promotion and affiliate links to sites selling alternative medicine products and supplements. **Our method is able to properly distinguish these different types of COVID-19 misinformation, without explicitly training on related features**.

6.6 Summary

In this chapter, we have described a method for learning a representation of news sources reporting science-related content. Our method uses a combination of signals to estimate the similarity between news sources. We have shown that these signals complement each other, capturing relationships between distinct sets of sources from a dataset of news articles related to COVID-19. Furthermore, the features learned by our model demonstrated superior performance to baselines for the task of source credibility detection, both in an offline and an online setting, requiring as little as three months of publication activity to accurately classify news sources. Lastly, we have shown that the learned source representations encode information of credibility and political leaning, forming clusters of sources that show similar reliability and political bias. In particular, we discovered two large clusters of unreliable sources to which different types of conspiracy news sources flock. One of them concentrates on alternative health misinformation, and the other promotes hyper-partisan political conspiracies.

[4] https://mediabiasfactcheck.com/mercola

Chapter 7

Real-Time News Analytics Platform

In this chapter, we present NewsTeller, a real-time news analytics platform. NewsTeller retrieves, processes, stores, and indexes a wide range of multilingual news articles, social media reactions, and references in real-time. Furthermore, our platform automatically extracts, stores, and displays heterogeneous quality indicators which derive from: i) social media discussions regarding news articles, showcasing the reach and stance towards these articles, and ii) their content and their referenced sources, showcasing the journalistic foundations of these articles. Finally, NewsTeller provides a multidimensional fact-checking environment for news articles to foster and highlight expert evaluation. Our platform is built in a distributed and robust fashion and runs operationally, handling daily thousands of news articles. In the following sections, we present an overview of the system as well as three tangible use cases. A live version of NewsTeller is publicly available here: *https://newsteller.io*.

7.1 System Overview

NewsTeller incorporates three modules for processing articles (§7.1.1), extracting quality indicators from them (§7.1.2), and acquiring expert reviews for them (§7.1.3). The overall architecture of the system is presented in §7.1.4.

7.1.1 Article Processing

The pipeline for article processing involves the steps of *Text Extraction*, *Entity Extraction*, *Article Classification*, and *Reactions Tracking* [165]. First, the *Text Extraction* process downloads and parses the title and the text of an article given its URL. This process is parallelized to deal with the required throughput of around 1.2 articles per second. Second, the *Entity Extraction* process extracts named entities from the text (e.g., names, organizations, and locations), and the

Table 7.1: Monthly data collection of NewsTeller. Social Media reactions (i.e., Likes, Retweets, Replies, and Quotes) are collected from the *Twitter Streaming API*.

Tracked Sources	913
Articles	1.3M / month
Likes	121M / month
Retweets	28M / month
Replies	16M / month
Quotes	9M / month
External References	6M / month

Article Classification process infers the article category (e.g., science or politics). Third, the *Reactions Tracking* process, implemented on top of the *Twitter Streaming API*, processes around 25 reactions per second. Statistics on the volume of data collected are shown in Table 7.1.

7.1.2 Quality Indicators

We compute three heterogeneous sets of quality indicators, namely, content, news context, and social media indicators. Regarding the content of a news article, we consider various well-established metrics for the quality of news, such as the clickbaitness of its title, the subjectivity and readability of its body, and whether it is by-lined by its author.

As for the news context of an article, we investigate the strength of the connection between this article and its primary sources of information. Thus, we consider three types of references:

- internal references within the same news outlet; many news outlets, in order to increase their user engagement, introduce such references either in "see also" sections or in the main body of their articles;
- external references to potential primary sources of information (e.g., references from nation-wide news outlets to local news outlets);
- particularly for the case of scientific news, scientific references, i.e., references to a predefined list of academic repositories, grey-literature, and peer-reviewed journals and institutional websites; as we see in our first use-case (§7.1.3), articles from high-quality scientific outlets are expected to have more references pointing to academic sources than articles from low-quality scientific outlets.

Finally, regarding the social media context, we measure two aspects, specifically the reach and stance towards a news article. Reach is measured through the proxy of social media popularity, which quantifies the impact of an article on a social media platform. On the other hand, stance is the positioning of social media platform users towards an article. Stance can be posi-

tive (i.e., users support or comment on an article without expressing doubts), or negative (i.e., users question the quality of an article or directly contradict what the article is saying).

According to a thorough experimental evaluation which is presented by Smeros et al. [194], the aforementioned indicators help non-expert users evaluate more accurately the quality of news articles, compared to non-experts that do not have access to these indicators.

7.1.3 Expert Reviews

Along with the set of automated quality indicators, the system allows experts to annotate any article based on seven criteria: 1) Factual accuracy, 2) Scientific understanding, 3) Logic/Reasoning, 4) Precision/Clarity, 5) Sources quality, 6) Fairness, and 7) Clickbaitness on a *Likert Scale* [100], from very low quality to very high quality. These are standard criteria used in state-of-the-art fact-checking portals like ScienceFeedback.co.

Based on these evaluation scores, the system computes a weighted average and displays a final score of the criteria for each article. Optionally, expert users can provide extensive free-text reviews about the news articles, which are also displayed to non-expert users.

7.1.4 System Architecture

The architecture of NewsTeller (Figure 7.1) consists of three components which are responsible for the collection, storage, and segmentation of data as well as for the models training and the indicators serving to the web application.

Data Collection and Storage

NewsTeller uses a hybrid data storage scheme that supports both real-time computational operations (with an RDBMS) and ad-hoc querying on historical data and efficient data warehousing (with Distributed Storage). The main data entry point of the system is an outlet-based streaming pipeline wrapped around the *Twitter Streaming API*. This subsystem acts as a messaging queue and fetches, in real-time, postings from a specific set of social media accounts along with their reactions. These incoming data streams are processed, and the corresponding news articles are extracted. For these transformations, the system leverages the distributed file system of *Hadoop* and the distributed computational framework *Spark* for parallel data processing and storing. The data synchronization between the RDBMS and the Distributed Storage is made through a daily data migration process.

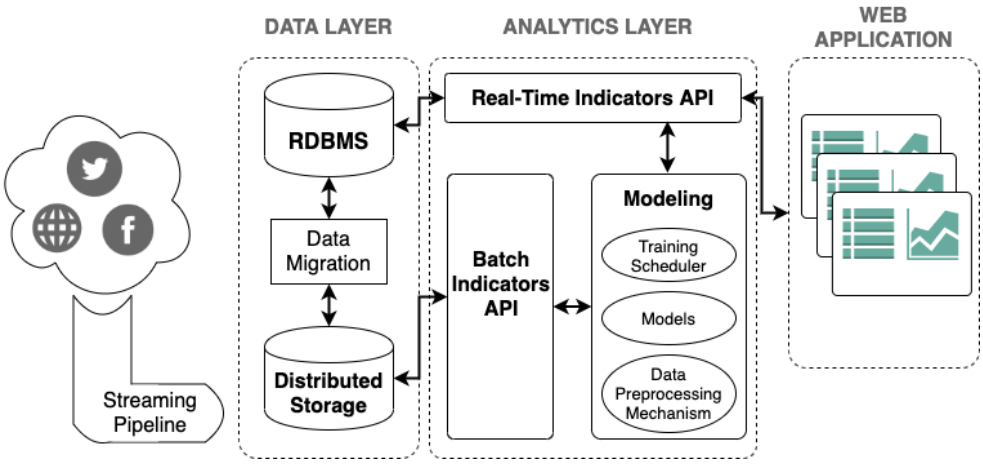


Figure 7.1: NewsTeller architecture. First, a streaming pipeline acts as the entry point of data collection. Then, a data layer, comprised of an RDBMS and a Distributed Storage, stores the incoming data. Lastly, the analytics layer manages the data, trains the models, and serves the extracted indicators to the web application.

Data Management and Model Training

As we show in our first use case (§7.2.1), an essential aspect of our system is the computation of analytics on top of particular segments of our data. These segments are combinations of content-based supervised topics of news and quality-based categories of news outlets.

More specifically, regarding the content-based segmentation, the system performs a probabilistic hierarchical clustering on the articles and assigns one or more topics to each one of them. These topics can be generic (e.g., Health) or specific (e.g., COVID-19). On the other hand, regarding the outlet quality-based segmentation, the system groups the articles by the news outlet that they are published and then groups the outlets with similar quality. The quality of an outlet is either computed using expert reviews or imported from external sources (e.g., in §7.2.1 we use a ranking published by the *American Council on Science and Health*).

Finally, our system periodically trains models on top of the Distributed Storage, accessing the entire history of our data collection. These models are used to extract the quality indicators that we describe in §7.1.2.

Indicators API

The last core component of our system is the Indicators API, which is responsible for the real-time article evaluation. Its architecture is based on micro-services, which are lightweight, loosely coupled services that support parallel execution. The main functionality of this component is to compute and serve quality indicators of articles to the web application.

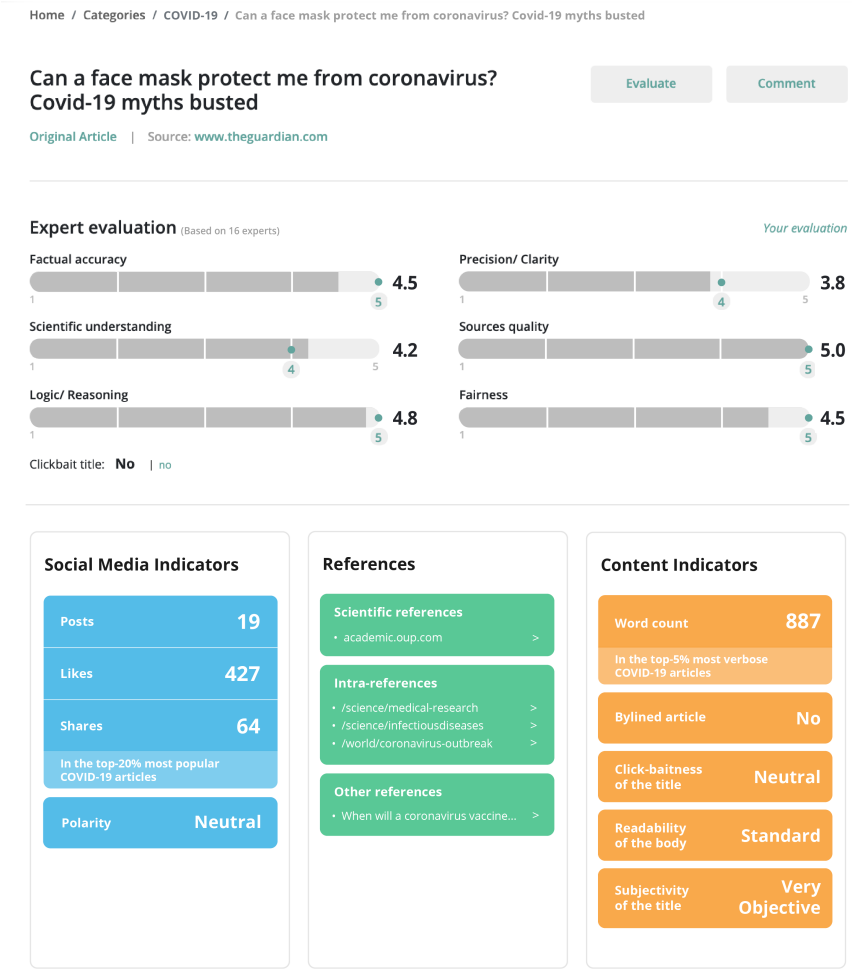


Figure 7.2: Enhanced article view of NewsTeller. A wide range of automatically extracted quality indicators combined with manually-operated expert reviews.

7.2 Use Cases

In this section, we present three applications and case studies conducted using the infrastructure of NewsTeller. Specifically, we present an early-stage study on the news coverage of COVID-19 (§7.2.1), a social bot to diversify the news consumption on Twitter (§7.2.2), and a reference prediction task (§7.2.3).

7.2.1 Early-Stage Study on COVID-19 News Coverage

In the first use case, we conducted a trial study using NewsTeller on the early stages of the COVID-19 pandemic. To prepare the data segment for the study, we used a shortlist, published by the *American Council on Science and Health*, containing 45 news outlets accompanied by

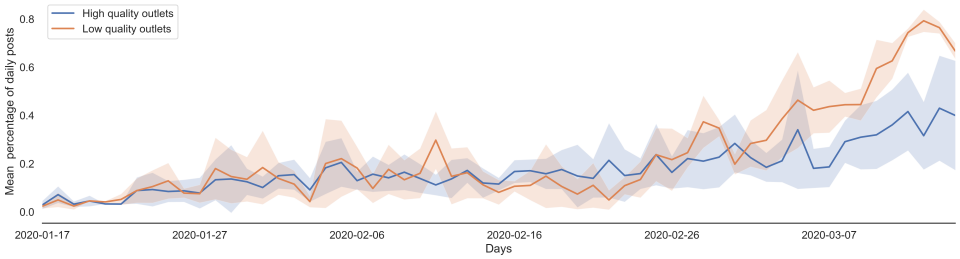


Figure 7.3: Mean percentage of daily posts referred to COVID-19 per rating category. Low-quality outlets seem to be driven by the breaking news, whereas high-quality outlets are more conservative on their publication rate.

their quality ranking, from very low to very high quality. The time frame of the data collection covers the 60-day period from *2020-01-15* to *2020-03-15*.

As COVID-19 is a topic with a highly trending nature, it triggers an abundance of news articles and social media discussions. Given such a prominent topic, the task of discerning between low and high-quality articles becomes very challenging for non-experts in the fields of medicine and epidemiology. On that end, we present how fused information retrieved from our system allows end-users to i) assess the quality of individual news articles, and ii) obtain aggregated insights for the topic of COVID-19.

Single Article Assessment

As we explain in §7.1, an end-user of the platform can explore, in real-time, a wide range of automatically extracted quality indicators combined with manually-operated expert reviews. A snapshot of this enhanced view of news articles is depicted in Figure 7.2. This functionality is available for all the articles in our news collection.

Aggregated Insights

Apart from the single article assessment, a user can interact with aggregated insights regarding a topic (in our case, COVID-19). The outlets published COVID-19 articles are evaluated based on three axes, namely their newsroom activity, evidence seeking, and social engagement.

Newsroom Activity. To study the newsroom activity, the system computes the distribution of daily posts for each outlet. Then, it groups all the media outlets by their quality ranking and creates a time series of the mean percentage of daily posts per rating class. The results an end-user can see are presented in Figure 7.3.

We observe that in the early stages of the discussion on the pandemic, both low and high-

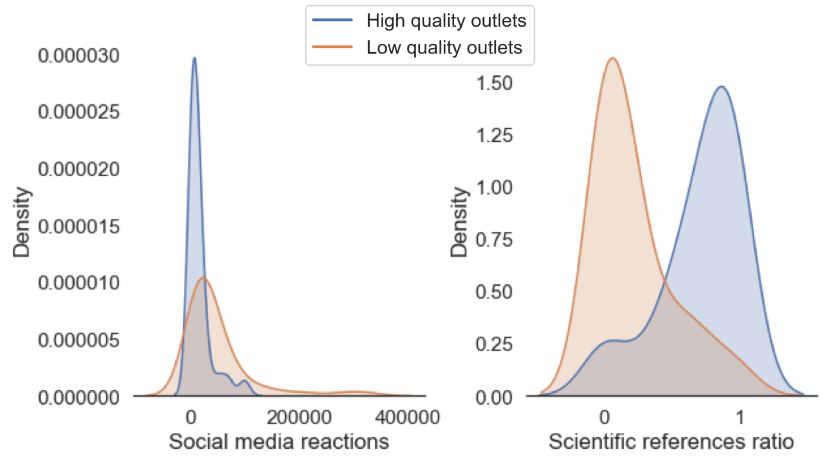


Figure 7.4: Kernel Density Estimation (KDE) of the number of Social Media Reactions (left) and Scientific References Ratio (right). The low-quality outlets tend to have a wider distribution of reactions but a lower number of scientific references, whereas the high-quality outlets tend to have a narrower distribution of reactions but a higher number of scientific references.

quality outlets posted with the same frequency. However, by the end of the first month, low-quality outlets started dedicating a larger percentage of their published articles on this topic. The latter implies a trade-off between the quantity and the quality of the articles. Low-quality outlets seem to be driven by the breaking news, whereas high-quality outlets are more conservative on their publication rate; however, as we see next, they have better scientific foundations.

Social Engagement & Evidence Seeking. Moreover, the system provides the end-user with insights regarding the social engagement (i.e., the number of social media reactions) and the evidence seeking (i.e., in our use-case, the ratio of scientific references used) of the news outlets. As shown in Figure 7.4, one can verify the assumption that low-quality outlets tend to publish more and thus acquire a higher amount of social media reach than high-quality outlets. Conversely, high-quality outlets base their findings more on well-established scientific references.

7.2.2 Social Bot for News Diversification

The second application on top of NewsTeller is NewsDiversifier, a social bot for diversifying the news consumption in social platforms. In our implementation, we focus on Twitter because of the popularity of the platform and the intuitive way of creating bot accounts. NewsDiversifier is comprised of an offline and an online component. The offline component is tasked with fetching and preprocessing articles from NewsTeller. The online component processes the input article, extracts articles from our database that are similar to this article, and runs a diversification algorithm to select articles that represent the most diverse perspectives. NewsDiversifer is topic agnostic; however, we treat the COVID-19 topic separately. Users can provide the bot with a particular input (the hashtag #covid), along with the input article, and instead of

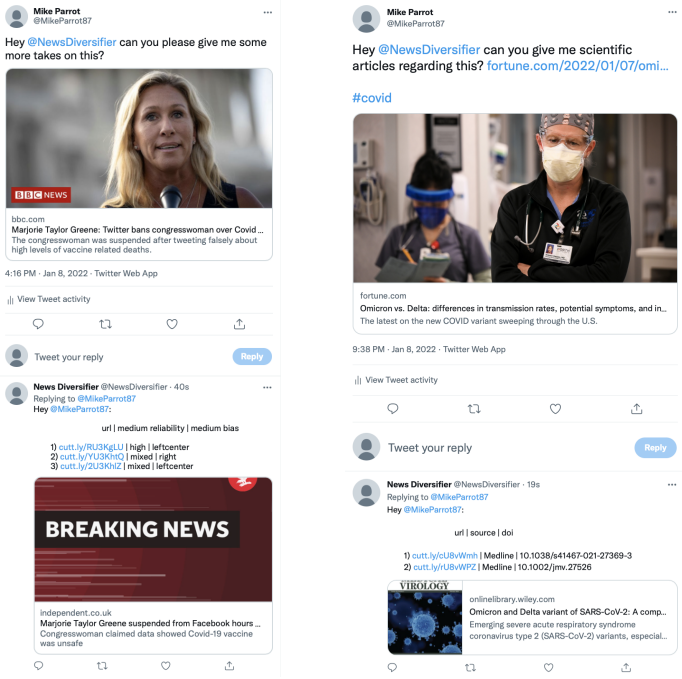


Figure 7.5: Examples of diversifying political (left) and scientific (right) news

receiving standard articles, receive scientific papers published in a constantly-updated COVID-19 paper collection, namely the CORD-19 dataset [2].

Offline Fetching and Data Preprocessing

Every day, NewsDiversifier fetches articles and metadata (e.g., publication time, author) that NewsTeller collects in real-time, keeping a sliding window of 7 days to reduce the search space and keep our news recommendations fresh. To optimize the speed at which the NewsDiversifier bot replies to user queries, our system performs heavy offline processing. Hence, the most expensive operations, i.e., the data cleaning (e.g., removal of punctuation and stopwords, lemmatization) and the computation of the article embeddings, are performed offline.

Similarity Calculation and Diversification

The similarity calculation component is tasked with discovering news articles that are the most similar to the input article. Hence, it calculates the embedding of the input and computes the similarity with the embeddings of the news collection, using the cosine similarity. Finally, it extracts the five most similar articles and feeds them into the diversification algorithm.

Out of the five most similar articles, the diversification algorithm, inspired by the work of

Indyk et al. [95], selects the articles that are the furthest away from each other in the embeddings space. By taking all possible 3-combinations of embeddings and calculating the area of the triangle that forms between them, the algorithm selects the triad with the biggest area.

Bot in Action

Users can tag @NewsDiversifier in a tweet thread, and the bot will automatically analyze the content of any news article shared. Then, the bot will find the articles in the database that cover the same topic and provide the most diverse perspectives. Finally, the bot will reply to the user with links to the articles and information regarding the sources, such as their political bias and reliability. Examples of diversifying political and scientific news are shown in Figure 7.5.

7.2.3 Reference Prediction Task

As a large number of scientific news articles do not explicitly cite the original references from the scientific literature, we utilize their content as well as web-graph information in order to predict missing links. We model the problem as a link prediction model where we want to discover edges connecting news articles with relevant scientific papers.

In the third use case of NewsTeller, we extract news articles related to scientific topics such as COVID-19, vaccination, healthcare, artificial intelligence, recycling. We filter these articles by considering the ones with at least one reference to the scientific literature. The time frame of the data collection covers the 4 month period from *2020-08-01* to *2020-11-31* containing 472 news articles and $1,242$ papers.

Methodology

To create graph representations for the news articles - scientific papers network, we implement a baseline graph neural network for relational graphs (R-GCN) as proposed by Schlichtkrull et al. [180]. For the link prediction task, R-GCN is comprised of a graph auto-encoder model. The encoder creates contextual representations for each entity, and a DistMult [216] decoder produces a score for every potential edge in the graph using these hidden node representations. We implement the R-GCN encoder with a single embedding layer. The encoder is regularized through edge dropout before normalization, with a dropout rate of 0.4. The model uses an *Adam* optimizer, and it is trained using full-batch gradient descent.

We test this model by comparing it with two content-aware methods that leverage the attention mechanism. The first method is a content-aware heterogeneous graph neural network model (HetGNN), as proposed by Zhang et al. [226]. The produced model is comprised of two

Table 7.2: Performance of different models on predicting connections between news articles and scientific papers. According to both evaluation metrics, we observe that the content-aware models (HetGNN and HGT) perform better than the content-agnostic model (R-GCN).

	MRR	Hits @		
Model		1	3	10
R-GCN	0.388	0.239	0.288	0.544
HetGNN	**0.412**	**0.317**	0.422	0.546
HGT	0.408	0.311	**0.426**	**0.589**

modules: the first, extracts a content embedding for each node using a recurrent neural network on the various attributes of the node, while the second, utilizes another recurrent neural network to aggregate these embeddings for each neighboring node, and applies an attention mechanism to measure the impact of heterogeneous node types, creating the final embedding.

The second content-aware model is a transformer-based model as proposed by Hu et al. [94]. This model uses dedicated representations for each different type of nodes and edges and produces a node- and edge-type dependent attention mechanism. Another contribution of this model is that it tackles the temporal nature of the graph by capturing the dynamic structural dependency with arbitrary window sizes. For this model, we use the published date as an additional feature of the news article entities.

Both content-aware models use an *Adam* optimizer and are trained using mini-batch gradient descent. For a fair comparison, we set the embedding size to 128 for all the approaches.

Experimental Evaluation

As described above, we assess the aforementioned models on the downstream task of link prediction, i.e., the task of filling the missing edges of a given network. We split the network edges into a training and a test set with a ratio set to 5:1 for all experiments. We evaluate the models in the test set with two commonly used metrics: Mean Reciprocal Rank (MRR) and Hits at n (Hits@n). In order to meet the original implementations of the models, we report the filtered MRR and the filtered Hits at 1, 3, and 10 positions.

For both content-aware models, we used as input features the titles of the scientific news article and the paper. Thus, for each node, we used pretrained XLNet [219] to get the representations of each word in its title. Next, we calculate the weighted average of words attention to get the title representation for each paper and news article, as proposed by Hu et al. [94]. The results of this experiment are summarized in Table 7.2. According to both evaluation metrics, we observe that the content-aware models (HetGNN and HGT) perform better than the content-agnostic model (R-GCN).

7.3 Summary

In this chapter, we presented NewsTeller, a real-time news analytics platform. NewsTeller builds a unique news collection consisting of a wide range of multilingual news articles, social media reactions, and references. Our platform is built in a distributed and robust fashion and runs operationally, handling daily thousands of news articles, in real-time. Moreover, in this chapter, we presented three applications and case studies conducted using the infrastructure of NewsTeller, namely, an early-stage study on the news coverage of COVID-19, a social bot, to diversify the news consumption on Twitter, and a reference prediction task.

Chapter 8

Conclusions

In this thesis, we have proposed methods for combating online scientific misinformation. As a starting point, we have surveyed the evolution of misinformation, showing the effects of digitalization and social media on the amplification and propagation of its impact. Furthermore, we have presented the main characteristics and approaches against misinformation and explained the high-level positioning of this thesis with respect to related scientific literature.

With this thesis, we have achieved three major scientific contributions; proposing methods for combating claim-based, article-based, and source-based scientific misinformation.

Regarding claim-based scientific misinformation, we have presented an effective method for assisting non-experts in the verification of scientific claims. We have shown that transformer models are indeed state-of-the-art on scientific claim detection; however, they require domain-specific fine-tuning to perform better than standard baselines. We have also shown that, by exploiting the text of a claim and its connections to scientific papers, we effectively cluster topically-related claims and papers, as well as that, by building an in-cluster knowledge graph, we enable the detection of check-worthy claims. Overall, we have shown that our method can build the appropriate fact-checking context to help non-expert fact-checkers verify complex scientific claims, facilitating them to outperform commercial systems. We believe that our method complements these systems in domains with sparse or non-existing ground-truth evidence, such as the critical domains of science and health.

Regarding article-based scientific misinformation, we have presented a method for evaluating the quality of scientific news articles. We have introduced new quality indicators that consider the quotes in articles, the similarity and relationship of articles with the scientific literature, and the volume and stance of social media reactions. We have shown that these indicators can distinguish among different tiers of quality with respect to scientific news. Moreover, we have shown that they can be used by non-experts to improve their evaluations on scientific news, bringing them more in line with expert evaluations. Finally, we have shown how these indicators can be combined to produce fully automated scores using weak supervision.

Regarding source-based scientific misinformation, we have presented a method for learning a representation of news sources reporting science-related content, using a combination of writing-style and citation-behavior signals. We have shown that the features learned by our model demonstrate superior performance to baselines for the task of source credibility detection, both in an offline and an online setting, requiring as little as three months of publication activity to accurately classify news sources. Furthermore, we have shown that the learned source representations encode information of credibility and political leaning, forming clusters of sources that show similar reliability and political bias, while distinguishing clusters with different conspiratorial narratives.

The last contribution of this thesis is a real-time news analytics platform that runs operationally, handling daily thousands of news articles. Our platform builds a unique collection consisting of a wide range of multilingual news articles, social media reactions, and references.

Overall, in this thesis, we have argued that the problem of scientific misinformation is a crucial problem, especially in the times of a pandemic, where different disciplines have to cooperate in combating it. We have proposed an approach for extracting explainable indicators from the content as well as the social and scientific context of news that: i) help non-experienced laypeople evaluate news similarly to proficient fact-checkers, and ii) reveal deep misinformation patterns among news sources.

8.1 Reproducibility

Our code uses the following `Python` libraries: i) `Pandas` and `Spark` for data management, ii) `NetworkX` for graph processing, iii) `scikit-learn` and `PyTorch` for training machine learning models, iv) `Simple Transformers`, `SpaCy`, `Beautiful Soup`, `Newspaper`, `TextSTAT` and `TextBlob` for natural language processing, and `seaborn` for data visualization. All the data, code, models, as well as the expert and crowd annotations used in this thesis, are publicly available for research purposes in ***http://scilens.epfl.ch***.

8.2 Discussion

In this section, we acknowledge the major limitations of this thesis and discuss potential workarounds that would help us overcome these limitations.

8.2.1 Ethics Statement

The methods proposed by this thesis, like most other AI/ML methods, are heavily data-driven; they use indicators found in the content and context of news to infer the credibility of claims, articles, and sources. The computation of these indicators can cause our methods to make biased decisions, especially if the input data is biased towards/against societal groups, such as underrepresented minorities and other vulnerable groups. Hence, we strongly suggest that all our methods (i.e., SciClops, SciLens, and SciLander) are used in human-in-the-loop settings, assisting but not replacing human decision-making.

8.2.2 Transparency and Explainability of Indicators

As we have already explained, our methods propose transparent and explainable indicators for news pieces at different granularity levels. We acknowledge that these indicators are vulnerable to adversarial attacks from predatory news portals that artificially construct seemingly credible (according to the indicators) news pieces. Nonetheless, as we describe above, we strongly suggest that all our methods are used in human-in-the-loop settings since human fact-checkers could both detect such adversarial attacks and propose additional indicators that cover these attacks. Transparency and explainability make our methods more robust and reliable rather than more vulnerable, as long as these methods are used to assist human decision-making.

8.2.3 Beyond English

SciClops, SciLens, and SciLander are currently applicable only on English corpora; extending these methods to other languages would require:

- *SciClops*: i) non-English training datasets for the claim classifier, ii) aligned, multilingual embeddings for clustering non-English claims with English scientific papers (given that most of the top-class scientific literature is in English), and iii) translation/adaptation of the domain-specific lexicon used to construct the knowledge graph.
- *SciLens*: i) translation/adaptation of the domain-specific lexicon used for quote extraction and attribution, ii) non-English training datasets or pretrained multilingual models for stance classification, and iii) aligned, multilingual embeddings for embedding non-English news articles together with English scientific papers.
- *SciLander*: translation/adaptation of the domain-specific lexicon used to compute the jargon indicator or simply skipping this indicator and training using the other three; all the other indicators, as well as the introduced embedding model, are based either on language-agnostic or already multilingual models.

8.2.4 Beyond Scientific News

SciClops, SciLens, and SciLander are currently applicable only on scientific news; extending these methods to other topics (e.g., to political news) would require: i) a general topic specified prior to the application of the methods; the chosen topic must include a set of keywords or entities referred to by news sources (e.g., political figures in the political news domain), ii) primary sources of information that would replace scientific references in the information diffusion graph; such sources might have a particular writing style, structure, or format, requiring our methods to adapt appropriately.

8.2.5 Accessibility of Social Media

The starting point of our *"Bottom-Up" Contextual News Collection* (§3) is social media; thus, news pieces not shared on social media are out of the scope of our collection. The latter is a strong limitation that compensates with the strong credibility signals extracted from social media (e.g., the stance indicator §5.4.2). Alternatively, we also support a *"Middle-Up" Contextual News Collection* (§3) in which we completely ignore the social media layer and process directly the full publication activity of news sources, independently of whether they have or do not have presence in social media (§6). We follow this two-fold approach because we have a limited bandwidth of requests to social media platforms. Ideally, we would be able to complement every news piece in our collection with social media metadata; however, this procedure is not feasible with the provided functionality from the platforms.

Regarding the social media platforms, we use a single API, namely the Twitter API. Twitter is the only mainstream social media platform providing a free academic API, with full access to real-time and historical data. One minor limitation of this API comes from the fact that we are able to collect only first-level replies to social media postings and not nested replies-to-replies.

8.2.6 Accessibility of News Media

When we build our news collection, we consider a broad definition of "news", covering mainstream media as well as other portals and blogs that often correspond to fringe news media. Under this assumption, we are able to collect not only high-quality news but also news containing misinformation and conspiracy theories. Hence, our news collection is closer to a real-world collection, with heterogeneous news sources of varying credibility.

8.2.7 Accessibility of Scientific Literature

When we build our news collection, we also consider a broad definition of "scientific literature", covering peer-reviewed but also preprints and gray-literature papers. Under this assumption, we extend our paper collection significantly, as almost 75% of the references to scientific literature consist of references to gray-literature papers. The latter mitigates the limitation of our methods on accessing and processing pay-walled papers and papers in unparsable formats (e.g., PDF) since, as we explain above, these papers are rarely referenced in news articles.

Furthermore, our methods support only explicit citations, i.e., direct outgoing links to scientific papers, and not implicit mentions of science-related entities (e.g., universities) because the latter design choice would introduce ambiguity and noise to our news collection.

8.3 Future Work

There are several future directions that could potentially be motivated by this thesis. Below we list several research opportunities and briefly discuss their applicability.

8.3.1 Open-Access News Collection

Recently, we have observed an extensive interest from the scientific community in research related to news and particularly phenomena such as misinformation. As we have explained in this thesis, misinformation has played an unprecedented role in influencing political, economic, and social ecosystems. However, there are not many public and open-access datasets available for research purposes to the academic community.

As in the context of this thesis we have introduced a real-time news analytics platform, handling millions of multilingual news articles, we plan to open access to our news collection for all the members of the academic community. Our collection could facilitate not only computer scientists (e.g., to analyze news and viewpoints related to particular events) but also social and political scientists and journalists, fostering interdisciplinary research on the field.

Apart from releasing our news collection, there is also metadata that could be beneficial for the community. Social media reactions to news are an important piece of auxiliary information that is currently missing from state-of-the-art news collections. This unified view of social and news media could support research on studying, e.g., the virality of news and its interplay with the credibility of the underlying information.

Finally, deeper information could be extracted and be continuously updated from our news collection, taking advantage of recent developments in language understanding models, such

as news-based word embeddings, spatiotemporal representations of entities, and knowledge graphs, enabling a series of related studies.

8.3.2 Multilingual News Analytics

As in this thesis we have focused mainly on research regarding English news, future directions could include studying the applicability of our methods in other languages, with particular interest in low-resource languages. Since science is itself a low-resource news topic, in the sense that, as we have explained, most of the fact-checking effort is focused on news related to politics, the sparsity of available debunking information increases drastically in low-resource languages. Furthermore, as scientific news is a "translation" of scientific findings using a more accessible-to-laypeople language, the further translation into a low-resource language is a highly challenging process. Thus, semi-automated methods as the ones we have introduced in this thesis could be developed, taking advantage of any knowledge transfer that can be achieved from high-resource languages and particularly English.

8.3.3 Crowd-Sourced Fact-Checking

Currently, the fact-checking field faces two main challenges: i) the scale of information on the web hinders the traditional editorial process of manually selecting and verifying information, and ii) the demand from the public and the potential effects of fact-checking are currently either unknown or partially explored (e.g., studies have explored the engagement of the public only into flagging or reporting activities). As our platform has already been deployed with fact-checking functionality, one potential research direction could be to perform a controlled field study to investigate whether crowd-sourced fact-checking can tackle sufficiently these challenges and explore the limits of citizen journalism.

8.3.4 Applications in Other Domains

The methodology that we introduced in this thesis is not tightly coupled with the evaluation of news. Indeed, we have observed similar quality signals in other formats (e.g., in campaigns published in science-based crowdfunding platforms). Such campaigns have many similarities with scientific news, such as: i) the abstracted technical terminology, ii) the appealing writing-style, iii) the clickbait titles, and iv) the credibility of the authors that is inherited to their publications. Future research directions could include studies on elements that predict the success or failure of such campaigns and how these elements correlate with the scientific soundness of a campaign, focusing particularly on deprecated/fringe/pseudo science.

Bibliography

[1] Deepak Agarwal, Bee-Chung Chen, and Xuanhui Wang. "Multi-faceted ranking of news articles using post-read actions". In: *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*. Ed. by Xue-wen Chen, Guy Lebanon, Haixun Wang, and Mohammed J. Zaki. ACM, 2012, pp. 694–703. DOI: 10.1145/2396761.2396850. URL: https://doi.org/10.1145/2396761.2396850.

[2] Allen Institute For AI. *COVID-19 Open Research Dataset Challenge (CORD-19)*. https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge. 2021.

[3] Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimiter Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, and Preslav Nakov. "A Survey on Multimodal Disinformation Detection". In: *CoRR* abs/2103.12541 (2021). arXiv: 2103.12541. URL: https://arxiv.org/abs/2103.12541.

[4] Firoj Alam, Fahim Dalvi, Shaden Shaar, Nadir Durrani, Hamdy Mubarak, Alex Nikolov, Giovanni Da San Martino, Ahmed Abdelali, Hassan Sajjad, Kareem Darwish, and Preslav Nakov. "Fighting the COVID-19 Infodemic in Social Media: A Holistic Perspective and a Call to Arms". In: *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media, ICWSM 2021, held virtually, June 7-10, 2021*. Ed. by Ceren Budak, Meeyoung Cha, Daniele Quercia, and Lexing Xie. AAAI Press, 2021, pp. 913–922. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/18114.

[5] Alex Berezow. "Infographic: The Best and Worst Science News Sites". In: *American Council on Science and Health* (Mar. 2017). URL: https://acsh.org/news/2017/03/05/infographic-best-and-worst-science-news-sites-10948.

[6] Robert N. Anderson, Margaret Warner, Lee Anne Flagg, and Farida. Ahmad. *Guidance for Certifying Deaths Due to Coronavirus Disease 2019 (COVID-19)*. 2020. URL: https://emergency.cdc.gov/coca/calls/2020/callinfo_041620.asp.

[7] Vassileios Balntas, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. "Learning local feature descriptors with triplets and shallow convolutional neural networks". In: *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*. Ed. by Richard C. Wilson, Edwin R. Hancock, and William A. P. Smith. BMVA Press, 2016. URL: http://www.bmva.org/bmvc/2016/papers/paper119/index.html.

[8] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James R. Glass, and Preslav Nakov. "Predicting Factuality of Reporting and Bias of News Media Sources". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii. Association for Computational Linguistics, 2018, pp. 3528–3539. DOI: 10.18653/v1/d18-1389. URL: https://doi.org/10.18653/v1/d18-1389.

[9] Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James R. Glass, and Preslav Nakov. "Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Ed. by Jill Burstein, Christy Doran, and Thamar Solorio. Association for Computational Linguistics, 2019, pp. 2109–2116. DOI: 10.18653/v1/n19-1216. URL: https://doi.org/10.18653/v1/n19-1216.

[10] Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, Shaden Shaar, and Zien Sheikh Ali. "Overview of CheckThat! 2020: Automatic Identification and Verification of Claims in Social Media". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22-25, 2020, Proceedings*. Ed. by Avi Arampatzis, Evangelos Kanoulas, Theodora Tsikrika, Stefanos Vrochidis, Hideo Joho, Christina Lioma, Carsten Eickhoff, Aurélie Névéol, Linda Cappellato, and Nicola Ferro. Vol. 12260. Lecture Notes in Computer Science. Springer, 2020, pp. 215–236. DOI: 10.1007/978-3-030-58219-7_17. URL: https://doi.org/10.1007/978-3-030-58219-7%5C_17.

[11] Martin W Bauer, Nick Allum, and Steve Miller. "What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda". In: *Public understanding of science* 16.1 (2007), pp. 79–95.

[12] Iz Beltagy, Kyle Lo, and Arman Cohan. "SciBERT: A Pretrained Language Model for Scientific Text". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Ed. by Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan. Association for Computational Linguistics, 2019, pp. 3613–3618. DOI: 10.18653/v1/D19-1371. URL: https://doi.org/10.18653/v1/D19-1371.

[13] Fabrício Benevenuto, Tiago Rodrigues, Virgílio A. F. Almeida, Jussara M. Almeida, and Marcos André Gonçalves. "Detecting spammers and content promoters in online video social networks". In: *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009*. Ed. by James Allan, Javed A. Aslam, Mark Sanderson, ChengXiang

Zhai, and Justin Zobel. ACM, 2009, pp. 620–627. DOI: 10.1145/1571941.1572047. URL: https://doi.org/10.1145/1571941.1572047.

[14] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". In: *J. Mach. Learn. Res.* 3 (2003), pp. 993–1022. URL: http://jmlr.org/papers/v3/blei03a.html.

[15] Christina Boididou, Symeon Papadopoulos, Lazaros Apostolidis, and Yiannis Kompatsiaris. "Learning to Detect Misleading Content on Twitter". In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ICMR 2017, Bucharest, Romania, June 6-9, 2017*. Ed. by Bogdan Ionescu, Nicu Sebe, Jiashi Feng, Martha A. Larson, Rainer Lienhart, and Cees Snoek. ACM, 2017, pp. 278–286. DOI: 10.1145/3078971.3078979. URL: https://doi.org/10.1145/3078971.3078979.

[16] Christina Boididou, Symeon Papadopoulos, Yiannis Kompatsiaris, Steve Schifferes, and Nic Newman. "Challenges of computational verification in social multimedia". In: *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume*. Ed. by Chin-Wan Chung, Andrei Z. Broder, Kyuseok Shim, and Torsten Suel. ACM, 2014, pp. 743–748. DOI: 10.1145/2567948.2579323. URL: https://doi.org/10.1145/2567948.2579323.

[17] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. "Selection Bias in News Coverage: Learning it, Fighting it". In: *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*. Ed. by Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis. ACM, 2018, pp. 535–543. DOI: 10.1145/3184558.3188724. URL: https://doi.org/10.1145/3184558.3188724.

[18] G. E. P. Box and D. R. Cox. "An Analysis of Transformations". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2 (1964), pp. 211–252. ISSN: 00359246. URL: http://www.jstor.org/stable/2984418.

[19] Ulrik Brandes. "A faster algorithm for betweenness centrality". In: *The Journal of Mathematical Sociology* 25.2 (2001), pp. 163–177. DOI: 10.1080/0022250X.2001.9990249. eprint: https://doi.org/10.1080/0022250X.2001.9990249. URL: https://doi.org/10.1080/0022250X.2001.9990249.

[20] J Scott Brennen, Felix Simon, Philip N Howard, and Rasmus Kleis Nielsen. "Types, sources, and claims of COVID-19 misinformation". In: *Reuters Institute* 7.3 (2020), p. 1.

[21] Mark Buchanan. "Managing the infodemic". In: 16.9 (2020), pp. 894–894.

[22] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. "Limiting the spread of misinformation in social networks". In: *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*. Ed. by Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar. ACM, 2011, pp. 665–674. DOI: 10.1145/1963405.1963499. URL: https://doi.org/10.1145/1963405.1963499.

[23] Philip Bump. “Google’s top news link for ‘final election results’ goes to a fake news site with false numbers”. In: *The Washington Post* (Nov. 2016). URL: https://www.washingtonpost.com/news/the-fix/wp/2016/11/14/googles-top-news-link-for-final-election-results-goes-to-a-fake-news-site-with-false-numbers.

[24] John M Carey, Andrew M Guess, Peter J Loewen, Eric Merkley, Brendan Nyhan, Joseph B Phillips, and Jason Reifler. “The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada”. In: *Nature Human Behaviour* (2022), pp. 1–8.

[25] Carlos Castillo. *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press, 2016.

[26] Carlos Castillo, Mohammed El-Haddad, Jürgen Pfeffer, and Matt Stempeck. “Characterizing the life cycle of online news stories using social media reactions”. In: *Computer Supported Cooperative Work, CSCW ’14, Baltimore, MD, USA, February 15-19, 2014*. Ed. by Susan R. Fussell, Wayne G. Lutters, Meredith Ringel Morris, and Madhu C. Reddy. ACM, 2014, pp. 211–223. DOI: 10.1145/2531602.2531623. URL: https://doi.org/10.1145/2531602.2531623.

[27] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. “Information credibility on twitter”. In: *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*. Ed. by Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar. ACM, 2011, pp. 675–684. DOI: 10.1145/1963405.1963500. URL: https://doi.org/10.1145/1963405.1963500.

[28] Emily Chen, Kristina Lerman, and Emilio Ferrara. “Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set”. In: *JMIR Public Health Surveill* 6.2 (May 2020), e19273. ISSN: 2369-2960. DOI: 10.2196/19273. URL: https://doi.org/10.2196/19273.

[29] Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyou Zhou, and William Yang Wang. “TabFact: A Large-scale Dataset for Table-based Fact Verification”. In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL: https://openreview.net/forum?id=rkeJRhNYDH.

[30] Xu Chen, Zheng Qin, Yongfeng Zhang, and Tao Xu. “Learning to Rank Features for Recommendation over Multiple Categories”. In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. Ed. by Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel. ACM, 2016, pp. 305–314. DOI: 10.1145/2911451.2911549. URL: https://doi.org/10.1145/2911451.2911549.

[31] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. “Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Comput-*

ing Systems, San Jose, CA, USA, May 7-12, 2016. Ed. by Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade. ACM, 2016, pp. 2098–2110. DOI: 10.1145/2858036.2858207. URL: https://doi.org/10.1145/2858036.2858207.

[32] Chung Joo Chung, Yoonjae Nam, and Michael A. Stefanone. "Exploring Online News Credibility: The Relative Influence of Traditional and Technological Factors". In: *Journal of Computer-Mediated Communication* 17.2 (Jan. 2012), pp. 171–186. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2011.01565.x. eprint: https://academic.oup.com/jcmc/article-pdf/17/2/171/19492784/jjcmcom0171.pdf. URL: https://doi.org/10.1111/j.1083-6101.2011.01565.x.

[33] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. "Computational Fact Checking from Knowledge Networks". In: *PLOS ONE* 10.6 (June 2015). Ed. by Alain Barrat, e0128193. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0128193. URL: http://dx.plos.org/10.1371/journal.pone.0128193.

[34] Peter Conrad. "Uses of expertise: Sources, quotes, and voice in the reporting of genetics in the news". In: *Public Uniderstanding of Science* 8 (1999).

[35] D Alan Cruse, David Alan Cruse, D A Cruse, and D A Cruse. *Lexical semantics*. Cambridge university press, 1986.

[36] Mark É Czeisler, Rashon I Lane, Emiko Petrosky, Joshua F Wiley, Aleta Christensen, Rashid Njai, Matthew D Weaver, Rebecca Robbins, Elise R Facer-Childs, Laura K Barger, et al. "Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24–30, 2020". In: *Morbidity and Mortality Weekly Report* 69.32 (2020), p. 1049.

[37] Vladimir De Semir. "Scientific journalism: problems and perspectives". In: *International Microbiology* 3.2 (2000), pp. 125–128.

[38] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Ed. by Jill Burstein, Christy Doran, and Thamar Solorio. Association for Computational Linguistics, 2019, pp. 4171–4186. DOI: 10.18653/v1/n19-1423. URL: https://doi.org/10.18653/v1/n19-1423.

[39] Estelle Dumas-Mallet, Andy Smith, Thomas Boraud, and François Gonon. "Poor replication validity of biomedical association studies reported by newspapers". In: *PLOS ONE* 12.2 (Feb. 2017), pp. 1–15. DOI: 10.1371/journal.pone.0172650. URL: https://doi.org/10.1371/journal.pone.0172650.

[40] Sharon Dunwoody. "Science journalism: prospects in the digital age". In: *Routledge handbook of public communication of science and technology*. Routledge, 2014, pp. 43–55.

[41] Chi Thang Duong, Quoc Viet Hung Nguyen, and Karl Aberer. “Interpretable node embeddings with mincut loss”. In: *Learning and Reasoning with Graph-Structured Representations Workshop - ICML 2019* (2019).

[42] Elena Kochkina, Panayiotis Smeros, Jérémie Rappaz, Marya Bazzi, Maria Liakata, Arkaitz Zubiaga. *Mediate Workshop*. https://digitalmediasig.github.io/Mediate2022. 2022.

[43] Elisa Shearer. “More than eight-in-ten Americans get news from digital devices”. In: *Pew Research Center* (Jan. 2021). URL: https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices.

[44] Tugrulcan Elmas, Rebekah Overdorf, Ahmed Furkan Özkalay, and Karl Aberer. “Ephemeral Astroturfing Attacks: The Case of Fake Twitter Trends”. In: *IEEE European Symposium on Security and Privacy, EuroS&P 2021, Vienna, Austria, September 6-10, 2021*. IEEE, 2021, pp. 403–422. DOI: 10.1109/EuroSP51992.2021.00035. URL: https://doi.org/10.1109/EuroSP51992.2021.00035.

[45] David K. Elson and Kathleen R. McKeown. “Automatic Attribution of Quoted Speech in Literary Narrative”. In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*. Ed. by Maria Fox and David Poole. AAAI Press, 2010. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1945.

[46] Emmanuel M. Vincent. “Most popular climate change stories of 2017 reviewed by scientists”. In: *Climate Feedback* (Jan. 2018). URL: https://climatefeedback.org/most-popular-climate-change-stories-2017-reviewed-scientists/.

[47] Eugene Kiely and Lori Robertson. *How to Spot Fake News*. https://www.factcheck.org/2016/11/how-to-spot-fake-news/. 2016.

[48] Karl Pearson F.R.S. “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), pp. 559–572. DOI: 10.1080/14786440109462720. URL: https://doi.org/10.1080/14786440109462720.

[49] Lei Fang, George Karakiulakis, and Michael Roth. “Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?” In: *The Lancet Respiratory Medicine* (2020).

[50] Johan Farkas and Jannick Schou. “Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood”. In: *Javnost - The Public* 25.3 (2018), pp. 298–314. DOI: 10.1080/13183222.2018.1463047. eprint: https://doi.org/10.1080/13183222.2018.1463047. URL: https://doi.org/10.1080/13183222.2018.1463047.

[51] Johan Farkas and Jannick Schou. *Post-truth, fake news and democracy: Mapping the politics of falsehood*. Routledge, 2019.

[52] FDA. “FDA cautions against use of hydroxychloroquine or chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems”. In: *FDA* (2020). URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or.

[53] FDA. “Why You Should Not Use Ivermectin to Treat or Prevent COVID-19”. In: *FDA* (2021). URL: https://www.fda.gov/consumers/consumer-updates/why-you-should-not-use-ivermectin-treat-or-prevent-covid-19.

[54] Ronald Aylmer Fisher. *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 2006.

[55] Seth Flaxman, Sharad Goel, and Justin M. Rao. “Filter Bubbles, Echo Chambers, and Online News Consumption”. In: *Public Opinion Quarterly* 80.S1 (Mar. 2016), pp. 298–320. ISSN: 0033-362X. DOI: 10.1093/poq/nfw006. eprint: https://academic.oup.com/poq/article-pdf/80/S1/298/17120810/nfw006.pdf. URL: https://doi.org/10.1093/poq/nfw006.

[56] Rudolph Flesch. “A new readability yardstick.” In: *Journal of applied psychology* 32.3 (1948), p. 221.

[57] B. J. Fogg and Hsiang Tseng. “The Elements of Computer Credibility”. In: *Proceeding of the CHI ’99 Conference on Human Factors in Computing Systems: The CHI is the Limit, Pittsburgh, PA, USA, May 15-20, 1999*. Ed. by Marian G. Williams and Mark W. Altom. ACM, 1999, pp. 80–87. DOI: 10.1145/302979.303001. URL: https://doi.org/10.1145/302979.303001.

[58] Jim Foust. *Online journalism: principles and practices of news for the Web*. Taylor & Francis, 2017.

[59] Linton C Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977), pp. 35–41.

[60] Linton C Freeman. “Centrality in social networks conceptual clarification”. In: *Social networks* 1.3 (1978), pp. 215–239.

[61] Richard A. Friedman. “Why Humans Are Vulnerable to Conspiracy Theories”. In: *Psychiatric Services* 72.1 (2021). PMID: 32703120, pp. 3–4. DOI: 10.1176/appi.ps.202000348. eprint: https://doi.org/10.1176/appi.ps.202000348. URL: https://doi.org/10.1176/appi.ps.202000348.

[62] Daniel Funke. “Google suspends fact-checking feature over quality concerns”. In: *Poynter* (Jan. 2018). URL: https://www.poynter.org/fact-checking/2018/google-suspends-fact-checking-feature-over-quality-concerns.

[63] Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. “ExFaKT: A Framework for Explaining Facts over Knowledge Graphs and Text”. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*. Ed. by J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman. ACM, 2019, pp. 87–95. DOI: 10.1145/3289600.3290996. URL: https://doi.org/10.1145/3289600.3290996.

[64] Wei Gao and Fabrizio Sebastiani. “From classification to quantification in tweet sentiment analysis”. In: *Social Netw. Analys. Mining* 6.1 (2016), 19:1–19:22. DOI: 10.1007/s13278-016-0327-z. URL: https://doi.org/10.1007/s13278-016-0327-z.

[65] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. “Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship”. In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*. Ed. by Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis. ACM, 2018, pp. 913–922. DOI: 10.1145/3178876.3186139. URL: https://doi.org/10.1145/3178876.3186139.

[66] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. “Political Polarization in Online News Consumption”. In: *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media, ICWSM 2021, held virtually, June 7-10, 2021*. Ed. by Ceren Budak, Meeyoung Cha, Daniele Quercia, and Lexing Xie. AAAI Press, 2021, pp. 152–162. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/18049.

[67] Amira Ghenai and Yelena Mejova. “Fake Cures: User-centric Modeling of Health Misinformation in Social Media”. In: *Proc. ACM Hum. Comput. Interact.* 2.CSCW (2018), 58:1–58:20. DOI: 10.1145/3274327. URL: https://doi.org/10.1145/3274327.

[68] Kristina Gligoric, Manoel Horta Ribeiro, Martin Müller, Olesia Altunina, Maxime Peyrard, Marcel Salathé, Giovanni Colavizza, and Robert West. “Experts and authorities receive disproportionate attention on Twitter during the COVID-19 crisis”. In: *CoRR* abs/2008.08364 (2020). arXiv: 2008.08364. URL: https://arxiv.org/abs/2008.08364.

[69] Ben Gomes. “Our latest quality improvements for Search”. In: *Google* (Apr. 2017). URL: https://blog.google/products/search/our-latest-quality-improvements-search.

[70] Genevieve Gorrell, Ahmet Aker, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. “SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours”. In: *Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, June 6-7, 2019*. Ed. by Jonathan May, Ekaterina Shutova, Aurélie Herbelot, Xiaodan Zhu, Marianna Apidianaki, and Saif M. Mohammad. Association for Computational Linguistics, 2019, pp. 845–854. DOI: 10.18653/v1/s19-2147. URL: https://doi.org/10.18653/v1/s19-2147.

[71] Alan G. Gross. “The roles of rhetoric in the public understanding of science”. In: *Public Understanding of Science* 3.1 (1994), pp. 3–23. DOI: 10.1088/0963-6625/3/1/001. eprint: https://doi.org/10.1088/0963-6625/3/1/001. URL: https://doi.org/10.1088/0963-6625/3/1/001.

[72] Aditya Grover and Jure Leskovec. “node2vec: Scalable Feature Learning for Networks”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. Ed. by Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi. ACM, 2016, pp. 855–864. DOI: 10.1145/2939672.2939754. URL: https://doi.org/10.1145/2939672.2939754.

[73] Maurício Gruppi, Pin-Yu Chen, and Sibel Adali. “Fake it Till You Make it: Self-Supervised Semantic Shifts for Monolingual Word Embedding Tasks”. In: *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, pp. 12893–12901. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17525.

[74] Maurício Gruppi, Benjamin D. Horne, and Sibel Adali. “NELA-GT-2020: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles”. In: *CoRR* abs/2102.04567 (2021). arXiv: 2102.04567. URL: https://arxiv.org/abs/2102.04567.

[75] Maurício Gruppi, Benjamin D. Horne, and Sibel Adalı. “Tell Me Who Your Friends Are: Using Content Sharing Behavior for News Source Veracity Detection”. In: *CoRR* abs/2101.10973 (2021). arXiv: 2101.10973. URL: https://arxiv.org/abs/2101.10973.

[76] Maurício Gruppi, Panayiotis Smeros, Sibel Adali, Carlos Castillo, and Karl Aberer. “SciLander: Mapping the Scientific News Landscape”. In: *CoRR* abs/2205.07970 (2022). DOI: 10.48550/arXiv.2205.07970. arXiv: 2205.07970. URL: https://doi.org/10.48550/arXiv.2205.07970.

[77] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. “Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy”. In: *22nd International World Wide Web Conference, WWW ’13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*. Ed. by Leslie Carr, Alberto H. F. Laender, Bernadette Farias Lóscio, Irwin King, Marcus Fontoura, Denny Vrandecic, Lora Aroyo, José Palazzo M. de Oliveira, Fernanda Lima, and Erik Wilde. International World Wide Web Conferences Steering Committee / ACM, 2013, pp. 729–736. DOI: 10.1145/2487788.2488033. URL: https://doi.org/10.1145/2487788.2488033.

[78] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. “Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*. Ed. by Jian Su, Xavier Carreras, and Kevin Duh. The Association for Computational Linguistics, 2016, pp. 2116–2121. DOI: 10.18653/v1/d16-1229. URL: https://doi.org/10.18653/v1/d16-1229.

[79] William L. Hamilton, Zhitao Ying, and Jure Leskovec. “Inductive Representation Learning on Large Graphs”. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. Ed. by Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett. 2017, pp. 1024–1034. URL: https://proceedings.neurips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html.

[80] Lushan Han, Justin Martineau, Doreen Cheng, and Christopher Thomas. “Samsung: Align-and-Differentiate Approach to Semantic Textual Similarity”. In: *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*. Ed. by Daniel M. Cer, David Jurgens, Preslav Nakov, and Torsten Zesch. The Association for Computer Linguistics, 2015, pp. 172–177. DOI: 10.18653/v1/s15-2031. URL: https://doi.org/10.18653/v1/s15-2031.

[81] Casper Hansen, Christian Hansen, Stephen Alstrup, Jakob Grue Simonsen, and Christina Lioma. “Neural Check-Worthiness Ranking with Weak Supervision: Finding Sentences for Fact-Checking”. In: *Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. Ed. by Sihem Amer-Yahia, Mohammad Mahdian, Ashish Goel, Geert-Jan Houben, Kristina Lerman, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia. ACM, 2019, pp. 994–1000. DOI: 10.1145/3308560.3316736. URL: https://doi.org/10.1145/3308560.3316736.

[82] P Sol Hart, Erik C Nisbet, and Teresa A Myers. “Public attention to science and political news and support for climate change mitigation”. In: *Nature Climate Change* 5.6 (2015), p. 541.

[83] Kazi Saidul Hasan and Vincent Ng. “Stance Classification of Ideological Debates: Data, Models, Features, and Constraints”. In: *Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013*. Asian Federation of Natural Language Processing / ACL, 2013, pp. 1348–1356. URL: https://aclanthology.org/I13-1191/.

[84] Naeemul Hassan, Bill Adair, James T Hamilton, Chengkai Li, Mark Tremayne, Jun Yang, and Cong Yu. “The quest to automate fact-checking”. In: *Proceedings of the 2015 Computation+ Journalism Symposium*. 2015.

[85] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. “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, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 2017, pp. 1803–1812. DOI: 10.1145/3097983.3098131. URL: https://doi.org/10.1145/3097983.3098131.

[86] Taher H. Haveliwala. “Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search”. In: *IEEE Trans. Knowl. Data Eng.* 15.4 (2003), pp. 784–796. DOI: 10.1109/TKDE.2003.1208999. URL: https://doi.org/10.1109/TKDE.2003.1208999.

[87] Ernst Hellinger. “Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.” In: *Journal für die reine und angewandte Mathematik* 136 (1909), pp. 210–271.

[88] Peter Hernon. “Disinformation and misinformation through the internet: Findings of an exploratory study”. In: *Government information quarterly* 12.2 (1995), pp. 133–139.

[89] Elad Hoffer and Nir Ailon. “Deep Metric Learning Using Triplet Network”. In: *Similarity-Based Pattern Recognition - Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015, Proceedings*. Ed. by Aasa Feragen, Marcello Pelillo, and Marco Loog. Vol. 9370. Lecture Notes in Computer Science. Springer, 2015, pp. 84–92. DOI: 10.1007/978-3-319-24261-3_7. URL: https://doi.org/10.1007/978-3-319-24261-3%5C_7.

[90] Benjamin D. Horne and Sibel Adali. “This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News”. In: *CoRR* abs/1703.09398 (2017). arXiv: 1703.09398. URL: http://arxiv.org/abs/1703.09398.

[91] Benjamin D. Horne, Jeppe Nørregaard, and Sibel Adalı. “Different Spirals of Sameness: A Study of Content Sharing in Mainstream and Alternative Media”. In: *Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019, Munich, Germany, June 11-14, 2019*. Ed. by Jürgen Pfeffer, Ceren Budak, Yu-Ru Lin, and Fred Morstatter. AAAI Press, 2019, pp. 257–266. URL: https://aaai.org/ojs/index.php/ICWSM/article/view/3227.

[92] Rui Hou, Verónica Pérez-Rosas, Stacy L. Loeb, and Rada Mihalcea. “Towards Automatic Detection of Misinformation in Online Medical Videos”. In: *International Conference on Multimodal Interaction, ICMI 2019, Suzhou, China, October 14-18, 2019*. Ed. by Wen Gao, Helen Mei-Ling Meng, Matthew A. Turk, Susan R. Fussell, Björn W. Schuller, Yale Song, and Kai Yu. ACM, 2019, pp. 235–243. DOI: 10.1145/1122445.3353763. URL: https://doi.org/10.1145/1122445.3353763.

[93] Minqing Hu and Bing Liu. “Mining and summarizing customer reviews”. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*. Ed. by Won Kim, Ron Kohavi, Johannes Gehrke, and William DuMouchel. ACM, 2004, pp. 168–177. DOI: 10.1145/1014052.1014073. URL: https://doi.org/10.1145/1014052.1014073.

[94] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. “Heterogeneous Graph Transformer”. In: *WWW ’20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*. Ed. by Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen. ACM, 2020, pp. 2704–2710. DOI: 10.1145/3366423.3380027. URL: https://doi.org/10.1145/3366423.3380027.

[95] Piotr Indyk, Sepideh Mahabadi, Mohammad Mahdian, and Vahab S. Mirrokni. “Composable core-sets for diversity and coverage maximization”. In: *Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS’14, Snowbird, UT, USA, June 22-27, 2014*. Ed. by Richard Hull and Martin Grohe. ACM, 2014,

pp. 100–108. DOI: 10.1145/2594538.2594560. URL: https://doi.org/10.1145/2594538.2594560.

[96] Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Márquez, and Preslav Nakov. "ClaimRank: Detecting Check-Worthy Claims in Arabic and English". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 2-4, 2018, Demonstrations*. Ed. by Yang Liu, Tim Paek, and Manasi S. Patwardhan. Association for Computational Linguistics, 2018, pp. 26–30. DOI: 10.18653/v1/n18-5006. URL: https://doi.org/10.18653/v1/n18-5006.

[97] Jakob D Jensen. "Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility". In: *Human communication research* 34.3 (2008), pp. 347–369.

[98] Shan Jiang, Simon Baumgartner, Abe Ittycheriah, and Cong Yu. "Factoring Fact-Checks: Structured Information Extraction from Fact-Checking Articles". In: *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*. Ed. by Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen. ACM, 2020, pp. 1592–1603. DOI: 10.1145/3366423.3380231. URL: https://doi.org/10.1145/3366423.3380231.

[99] Wei Jin, Hung Hay Ho, and Rohini K. Srihari. "OpinionMiner: a novel machine learning system for web opinion mining and extraction". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*. Ed. by John F. Elder IV, Françoise Fogelman-Soulié, Peter A. Flach, and Mohammed Javeed Zaki. ACM, 2009, pp. 1195–1204. DOI: 10.1145/1557019.1557148. URL: https://doi.org/10.1145/1557019.1557148.

[100] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. "Likert scale: Explored and explained". In: *Current Journal of Applied Science and Technology* (2015), pp. 396–403.

[101] Armand Joulin, Piotr Bojanowski, Tomás Mikolov, Hervé Jégou, and Edouard Grave. "Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii. Association for Computational Linguistics, 2018, pp. 2979–2984. DOI: 10.18653/v1/d18-1330. URL: https://doi.org/10.18653/v1/d18-1330.

[102] Garth S Jowett and Victoria O'donnell. *Propaganda & persuasion*. Sage publications, 2018.

[103] Juliane von Reppert-Bismarck. *Lie Detectors*. https://lie-detectors.org. 2020.

[104] Melaina Juntti. "Study: Marijuana Can Help Battle Depression, Anxiety, PTSD, and Addiction". In: *Men's Journal* (2017). URL: https://www.mensjournal.com/health-fitness/study-marijuana-can-help-battle-depression-anxiety-ptsd-and-addiction-w453012.

[105] Georgios Karagiannis, Mohammed Saeed, Paolo Papotti, and Immanuel Trummer. “Scrutinizer: A Mixed-Initiative Approach to Large-Scale, Data-Driven Claim Verification”. In: *Proc. VLDB Endow.* 13.11 (2020), pp. 2508–2521. URL: http://www.vldb.org/pvldb/vol13/p2508-karagiannis.pdf.

[106] Suleman Khan, Saqib Hakak, N Deepa, B Prabadevi, Kapal Dev, and Silvia Trelova. “Detecting COVID-19-Related Fake News Using Feature Extraction”. In: *Frontiers in Public Health* 9 (2021).

[107] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. “All-in-one: Multi-task Learning for Rumour Verification”. In: *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*. Ed. by Emily M. Bender, Leon Derczynski, and Pierre Isabelle. Association for Computational Linguistics, 2018, pp. 3402–3413. URL: https://www.aclweb.org/anthology/C18-1288/.

[108] Anand Konjengbam, Subrata Ghosh, Nagendra Kumar, and Manish Singh. “Debate Stance Classification Using Word Embeddings”. In: *Big Data Analytics and Knowledge Discovery - 20th International Conference, DaWaK 2018, Regensburg, Germany, September 3-6, 2018, Proceedings*. Ed. by Carlos Ordonez and Ladjel Bellatreche. Vol. 11031. Lecture Notes in Computer Science. Springer, 2018, pp. 382–395. DOI: 10.1007/978-3-319-98539-8_29. URL: https://doi.org/10.1007/978-3-319-98539-8%5C_29.

[109] Justin Kosslyn and Cong Yu. “Fact Check now available in Google Search and News around the world”. In: *Google* (Apr. 2017). URL: https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world.

[110] Justin Kosslyn and Cong Yu. “Fact Check now available in Google Search and News around the world”. In: *Google* (Apr. 2017). URL: http://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world.

[111] Rohit Kulkarni. *A Million News Headlines*. Version V5. 2018. DOI: 10.7910/DVN/SYBGZL. URL: https://doi.org/10.7910/DVN/SYBGZL.

[112] Srijan Kumar and Neil Shah. “False Information on Web and Social Media: A Survey”. In: *CoRR* abs/1804.08559 (2018). arXiv: 1804.08559. URL: http://arxiv.org/abs/1804.08559.

[113] Srijan Kumar, Robert West, and Jure Leskovec. “Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes”. In: *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*. Ed. by Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao. ACM, 2016, pp. 591–602. DOI: 10.1145/2872427.2883085. URL: https://doi.org/10.1145/2872427.2883085.

[114] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. “The science of fake news”. In: *Science* 359.6380 (2018), pp. 1094–1096. DOI: 10.1126/science.aao2998. eprint: https://www.

science.org/doi/pdf/10.1126/science.aao2998. URL: https://www.science.org/doi/abs/10.1126/science.aao2998.

[115] Stamatios Lefkimmiatis, Aurélien Bourquard, and Michael Unser. "Hessian-Based Norm Regularization for Image Restoration With Biomedical Applications". In: *IEEE Trans. Image Process.* 21.3 (2012), pp. 983–995. DOI: 10.1109/TIP.2011.2168232. URL: https://doi.org/10.1109/TIP.2011.2168232.

[116] Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. "Context Dependent Claim Detection". In: *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*. Ed. by Jan Hajic and Junichi Tsujii. ACL, 2014, pp. 1489–1500. URL: https://aclanthology.org/C14-1141/.

[117] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*. Ed. by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault. Association for Computational Linguistics, 2020, pp. 7871–7880. DOI: 10.18653/v1/2020.acl-main.703. URL: https://doi.org/10.18653/v1/2020.acl-main.703.

[118] Chang Li and Dan Goldwasser. "Encoding Social Information with Graph Convolutional Networks forPolitical Perspective Detection in News Media". In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. Ed. by Anna Korhonen, David R. Traum, and Lluís Màrquez. Association for Computational Linguistics, 2019, pp. 2594–2604. DOI: 10.18653/v1/p19-1247. URL: https://doi.org/10.18653/v1/p19-1247.

[119] Chenliang Li, Haoran Wang, Zhiqian Zhang, Aixin Sun, and Zongyang Ma. "Topic Modeling for Short Texts with Auxiliary Word Embeddings". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. Ed. by Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel. ACM, 2016, pp. 165–174. DOI: 10.1145/2911451.2911499. URL: https://doi.org/10.1145/2911451.2911499.

[120] Matthias Liebeck, Philipp Pollack, Pashutan Modaresi, and Stefan Conrad. "HHU at SemEval-2016 Task 1: Multiple Approaches to Measuring Semantic Textual Similarity". In: *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*. Ed. by Steven Bethard, Daniel M. Cer, Marine Carpuat, David Jurgens, Preslav Nakov, and Torsten Zesch. The Association for Computer Linguistics, 2016, pp. 595–601. DOI: 10.18653/v1/s16-1090. URL: https://doi.org/10.18653/v1/s16-1090.

[121] Marco Lippi and Paolo Torroni. "Context-Independent Claim Detection for Argument Mining". In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. Ed. by Qiang

Yang and Michael J. Wooldridge. AAAI Press, 2015, pp. 185–191. URL: http://ijcai.org/Abstract/15/033.

[122] Stuart P. Lloyd. "Least squares quantization in PCM". In: *IEEE Trans. Information Theory* 28.2 (1982), pp. 129–136. DOI: 10.1109/TIT.1982.1056489. URL: https://doi.org/10.1109/TIT.1982.1056489.

[123] Steven Loria. *Sentiment Analysis*. 2018. URL: http://textblob.readthedocs.io.

[124] Tessa Lyons. "Hard Questions: What's Facebook's Strategy for Stopping False News?" In: *Facebook* (May 2018). URL: http://newsroom.fb.com/news/2018/05/hard-questions-false-news.

[125] Yukun Ma, Haiyun Peng, and Erik Cambria. "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*. Ed. by Sheila A. McIlraith and Kilian Q. Weinberger. AAAI Press, 2018, pp. 5876–5883. URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16541.

[126] Alexandre Maros, Jussara M. Almeida, Fabrício Benevenuto, and Marisa Vasconcelos. "Analyzing the Use of Audio Messages in WhatsApp Groups". In: *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*. Ed. by Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen. ACM, 2020, pp. 3005–3011. DOI: 10.1145/3366423.3380070. URL: https://doi.org/10.1145/3366423.3380070.

[127] Alexandre Maros, Jussara M. Almeida, and Marisa Vasconcelos. "A Study of Misinformation in Audio Messages Shared in WhatsApp Groups". In: *Disinformation in Open Online Media - Third Multidisciplinary International Symposium, MISDOOM 2021, Virtual Event, September 21-22, 2021, Proceedings*. Ed. by Jonathan Bright, Anastasia Giachanou, Viktoria Spaiser, Francesca Spezzano, Anna George, and Alexandra Pavliuc. Vol. 12887. Lecture Notes in Computer Science. Springer, 2021, pp. 85–100. DOI: 10.1007/978-3-030-87031-7_6. URL: https://doi.org/10.1007/978-3-030-87031-7%5C_6.

[128] Mason Walker and Katerina Eva Matsa. "News Consumption Across Social Media in 2021". In: *Pew Research Center* (Sept. 2021). URL: https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021.

[129] Gina M. Masullo and Jiwon Kim. "Exploring "Angry" and "Like" Reactions on Uncivil Facebook Comments That Correct Misinformation in the News". In: *Digital Journalism* 9.8 (2021), pp. 1103–1122. DOI: 10.1080/21670811.2020.1835512. eprint: https://doi.org/10.1080/21670811.2020.1835512. URL: https://doi.org/10.1080/21670811.2020.1835512.

[130] Saurabh Mathur. *Clickbait Detector*. 2017. URL: http://github.com/saurabhmathur96/clickbait-detector.

[131] Spencer McKay and Chris Tenove. “Disinformation as a Threat to Deliberative Democracy”. In: *Political Research Quarterly* 74.3 (2021), pp. 703–717. DOI: 10.1177/1065912920938143. eprint: https://doi.org/10.1177/1065912920938143. URL: https://doi.org/10.1177/1065912920938143.

[132] Irene Costera Meijer and Tim Groot Kormelink. “Checking, Sharing, Clicking and Linking”. In: *Digital Journalism* 3.5 (2015), pp. 664–679. DOI: 10.1080/21670811.2014.937149. eprint: https://doi.org/10.1080/21670811.2014.937149. URL: https://doi.org/10.1080/21670811.2014.937149.

[133] Yelena Mejova and Kyriaki Kalimeri. “COVID-19 on Facebook Ads: Competing Agendas around a Public Health Crisis”. In: *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2020, Guayaquil, Ecuador, June 15-17, 2020*. ACM, 2020, pp. 22–31. DOI: 10.1145/3378393.3402241. URL: https://doi.org/10.1145/3378393.3402241.

[134] Meta. *Tips to Spot False News*. https://www.facebook.com/journalismproject/programs/third-party-fact-checking/tips-to-spot-false-news. 2020.

[135] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient Estimation of Word Representations in Vector Space”. In: *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2013. URL: http://arxiv.org/abs/1301.3781.

[136] George A. Miller. “WordNet: A Lexical Database for English”. In: *Commun. ACM* 38.11 (1995), pp. 39–41. DOI: 10.1145/219717.219748. URL: http://doi.acm.org/10.1145/219717.219748.

[137] Amy Mitchell, Jeffrey Gottfried, Michael Barthel, and Elisa Shearer. “The modern news consumer: News attitudes and practices in the digital era”. In: (2016).

[138] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. “Stance and Sentiment in Tweets”. In: *ACM Trans. Internet Techn.* 17.3 (2017), 26:1–26:23. DOI: 10.1145/3003433. URL: https://doi.org/10.1145/3003433.

[139] Adam Mosseri. “News Feed FYI: Addressing Hoaxes and Fake News”. In: *Facebook* (Dec. 2016). URL: https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news.

[140] Grace Muzny, Michael Fang, Angel X. Chang, and Dan Jurafsky. “A Two-stage Sieve Approach for Quote Attribution”. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*. Ed. by Mirella Lapata, Phil Blunsom, and Alexander Koller. Association for Computational Linguistics, 2017, pp. 460–470. DOI: 10.18653/v1/e17-1044. URL: https://doi.org/10.18653/v1/e17-1044.

[141] Greg Myers. “Discourse studies of scientific popularization: Questioning the boundaries”. In: *Discourse studies* 5.2 (2003), pp. 265–279.

[142] Merja Myllylahti. “An attention economy trap? An empirical investigation into four news companies’ Facebook traffic and social media revenue”. In: *Journal of Media Business Studies* 15.4 (2018), pp. 237–253. DOI: 10.1080/16522354.2018.1527521. URL: https://doi.org/10.1080/16522354.2018.1527521.

[143] Moin Nadeem, Wei Fang, Brian Xu, Mitra Mohtarami, and James R. Glass. “FAKTA: An Automatic End-to-End Fact Checking System”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Demonstrations*. Ed. by Waleed Ammar, Annie Louis, and Nasrin Mostafazadeh. Association for Computational Linguistics, 2019, pp. 78–83. DOI: 10.18653/v1/n19-4014. URL: https://doi.org/10.18653/v1/n19-4014.

[144] Preslav Nakov, David P. A. Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, and Giovanni Da San Martino. “Automated Fact-Checking for Assisting Human Fact-Checkers”. In: *CoRR* abs/2103.07769 (2021). arXiv: 2103.07769. URL: https://arxiv.org/abs/2103.07769.

[145] Nicholas Negroponte, Randal Harrington, Susan R. McKay, and Wolfgang Christian. “Being Digital”. In: *Computers in Physics* 11.3 (1997), pp. 261–262. DOI: 10.1063/1.4822554. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4822554. URL: https://aip.scitation.org/doi/abs/10.1063/1.4822554.

[146] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David Levy, and Rasmus Kleis Nielsen. “Reuters Institute digital news report 2017”. In: *Available at SSRN 3026082* (2017).

[147] Rasmus Kleis Nielsen and Kim Christian Schrøder. “The Relative Importance of Social Media for Accessing, Finding, and Engaging with News”. In: *Digital Journalism* 2.4 (2014), pp. 472–489. DOI: 10.1080/21670811.2013.872420. eprint: https://doi.org/10.1080/21670811.2013.872420. URL: https://doi.org/10.1080/21670811.2013.872420.

[148] Jeppe Nørregaard, Benjamin D. Horne, and Sibel Adali. “NELA-GT-2018: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles”. In: *Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019, Munich, Germany, June 11-14, 2019*. Ed. by Jürgen Pfeffer, Ceren Budak, Yu-Ru Lin, and Fred Morstatter. AAAI Press, 2019, pp. 630–638. URL: https://aaai.org/ojs/index.php/ICWSM/article/view/3261.

[149] Kurzgesagt - In a Nutshell. *Behind the Lies*. https://www.youtube.com/watch?v=XFqn3uy238E. 2021.

[150] Timothy O’Keefe, Silvia Pareti, James R. Curran, Irena Koprinska, and Matthew Honnibal. “A Sequence Labelling Approach to Quote Attribution”. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea*. Ed. by Jun’ichi Tsujii, James Henderson, and Marius Pasca. ACL, 2012, pp. 790–799. URL: https://aclanthology.org/D12-1072/.

[151] Scott Robert Olson and Timothy Pollard. “The Muse Pixeliope: Digitalization and Media Literacy Education”. In: *American Behavioral Scientist* 48.2 (2004), pp. 248–255. DOI: 10.1177/0002764204267272. eprint: https://doi.org/10.1177/0002764204267272. URL: https://doi.org/10.1177/0002764204267272.

[152] Chiara Palmerini. “Science reporting as negotiation”. In: *Journalism, Science and Society*. 2007. Chap. 11, pp. 113–122.

[153] Panayiotis Smeros, Jérémie Rappaz, Marya Bazzi, Elena Kochkina, Maria Liakata, Karl Aberer. *Mediate Workshop*. https://digitalmediasig.github.io/Mediate2021. 2021.

[154] Panayiotis Smeros, Jérémie Rappaz, Marya Bazzi, Karl Aberer. *Mediate Workshop*. https://digitalmediasig.github.io/Mediate2020. 2020.

[155] Silvia Pareti, Timothy O’Keefe, Ioannis Konstas, James R. Curran, and Irena Koprinska. “Automatically Detecting and Attributing Indirect Quotations”. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 2013, pp. 989–999. URL: https://aclanthology.org/D13-1101/.

[156] Ayush Patwari, Dan Goldwasser, and Saurabh Bagchi. “TATHYA: A Multi-Classifier System for Detecting Check-Worthy Statements in Political Debates”. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*. Ed. by Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li. ACM, 2017, pp. 2259–2262. DOI: 10.1145/3132847.3133150. URL: https://doi.org/10.1145/3132847.3133150.

[157] Dario Pavllo, Tiziano Piccardi, and Robert West. “Quootstrap: Scalable Unsupervised Extraction of Quotation-Speaker Pairs from Large News Corpora via Bootstrapping”. In: *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*. AAAI Press, 2018, pp. 231–240. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17827.

[158] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. “Glove: Global Vectors for Word Representation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. Ed. by Alessandro Moschitti, Bo Pang, and Walter Daelemans. ACL, 2014, pp. 1532–1543. DOI: 10.3115/v1/d14-1162. URL: https://doi.org/10.3115/v1/d14-1162.

[159] Gordon Pennycook and David G Rand. “The psychology of fake news”. In: *Trends in cognitive sciences* 25.5 (2021), pp. 388–402.

[160] Warren A Peterson and Noel P Gist. “Rumor and public opinion”. In: *American Journal of Sociology* 57.2 (1951), pp. 159–167.

[161] José María González Pinto, Janus Wawrzinek, and Wolf-Tilo Balke. “What Drives Research Efforts? Find Scientific Claims that Count!” In: *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*. Ed. by Maria Bonn, Dan Wu, J. Stephen Downie, and Alain Martaus. IEEE, 2019, pp. 217–226. DOI: 10.1109/JCDL.2019.00038. URL: https://doi.org/10.1109/JCDL.2019.00038.

[162] Peter Pomerantsev and Michael Weiss. *The menace of unreality: How the Kremlin weaponizes information, culture and money*. Vol. 14. Institute of Modern Russia New York, 2014.

[163] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. “Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media”. In: *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017*. Ed. by Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich. ACM, 2017, pp. 1003–1012. DOI: 10.1145/3041021.3055133. URL: https://doi.org/10.1145/3041021.3055133.

[164] Bruno Pouliquen, Ralf Steinberger, and Clive Best. “Automatic detection of quotations in multilingual news”. In: *Proceedings of Recent Advances in Natural Language Processing*. 2007, pp. 487–492.

[165] Jérémie Rappaz. *Media Observatory Initiative*. https://www.media-initiative.ch/wp-content/uploads/2019/10/moi-ofcomv11.pdf. 2019.

[166] Jérémie Rappaz, Dylan Bourgeois, and Karl Aberer. “A Dynamic Embedding Model of the Media Landscape”. In: *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. Ed. by Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia. ACM, 2019, pp. 1544–1554. DOI: 10.1145/3308558.3313526. URL: https://doi.org/10.1145/3308558.3313526.

[167] Nils Reimers, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. “Classification and Clustering of Arguments with Contextualized Word Embeddings”. In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. Ed. by Anna Korhonen, David R. Traum, and Lluís Márquez. Association for Computational Linguistics, 2019, pp. 567–578. DOI: 10.18653/v1/p19-1054. URL: https://doi.org/10.18653/v1/p19-1054.

[168] Julio C. S. Reis, André Correia, Fabricio Murai, Adriano Veloso, and Fabrício Benevenuto. “Explainable Machine Learning for Fake News Detection”. In: *Proceedings of the 11th ACM Conference on Web Science, WebSci 2019, Boston, MA, USA, June 30 - July 03, 2019*. Ed. by Paolo Boldi, Brooke Foucault Welles, Katharina Kinder-Kurlanda, Christo Wilson, Isabella Peters, and Wagner Meira Jr. ACM, 2019, pp. 17–26. DOI: 10.1145/3292522.3326027. URL: https://doi.org/10.1145/3292522.3326027.

[169] Gustavo Resende, Philipe F. Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara M. Almeida, and Fabrício Benevenuto. “(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures”. In: *The World Wide Web Con-*

ference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019. Ed. by Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia. ACM, 2019, pp. 818–828. DOI: 10.1145/3308558.3313688. URL: https://doi.org/10.1145/3308558.3313688.

[170] Douglas A. Reynolds. "Gaussian Mixture Models". In: *Encyclopedia of Biometrics*. Ed. by Stan Z. Li and Anil K. Jain. Springer US, 2009, pp. 659–663. DOI: 10.1007/978-0-387-73003-5_196. URL: https://doi.org/10.1007/978-0-387-73003-5%5C_196.

[171] Filipe Nunes Ribeiro, Lucas Henrique C. Lima, Fabrício Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P. Gummadi. "Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale". In: *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*. AAAI Press, 2018, pp. 290–299. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17878.

[172] Angelika Romanou, Panayiotis Smeros, and Karl Aberer. "On Representation Learning for Scientific News Articles Using Heterogeneous Knowledge Graphs". In: *Companion of The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*. Ed. by Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia. ACM, 2021, pp. 422–425. DOI: 10.1145/3442442.3451362. URL: https://doi.org/10.1145/3442442.3451362.

[173] Angelika Romanou, Panayiotis Smeros, Carlos Castillo, and Karl Aberer. "SciLens News Platform: A System for Real-Time Evaluation of News Articles". In: *Proc. VLDB Endow.* 13.12 (2020), pp. 2969–2972. DOI: 10.14778/3415478.3415521. URL: http://www.vldb.org/pvldb/vol13/p2969-romanou.pdf.

[174] Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. "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, Sydney, Australia, July 31 - August 03, 2017*. Ed. by Jana Diesner, Elena Ferrari, and Guandong Xu. ACM, 2017, pp. 232–239. DOI: 10.1145/3110025.3110054. URL: https://doi.org/10.1145/3110025.3110054.

[175] Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: http://www.sciencedirect.com/science/article/pii/0377042787901257.

[176] Andrew Salway, Paul Meurer, Knut Hofland, and Øystein Reigem. "Quote Extraction and Attribution from Norwegian Newspapers". In: *Proceedings of the 21st Nordic Conference on Computational Linguistics, NODALIDA 2017, Gothenburg, Sweden, May 22-24, 2017*. Ed. by Jörg Tiedemann and Nina Tahmasebi. Vol. 131. Linköping Electronic Conference Proceedings. Linköping University Electronic Press / Association for Computational Linguistics, 2017, pp. 293–297. URL: http://www.ep.liu.se/ecp/article.asp?issue=131%5C&article=041%5C&volume=.

[177] Laura Sbaffi and Jennifer Rowley. “Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research.” In: *Journal of medical Internet research* 19.6 (June 2017), e218. ISSN: 1438-8871. DOI: 10.2196/jmir.7579. URL: http://www.ncbi.nlm.nih.gov/pubmed/28630033%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5495972.

[178] Dietram A. Scheufele. “Communicating science in social settings”. In: *Proceedings of the National Academy of Sciences* 110.Supplement 3 (2013), pp. 14040–14047. ISSN: 0027-8424. DOI: 10.1073/pnas.1213275110. URL: https://www.pnas.org/content/110/Supplement_3/14040.

[179] Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. “A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains”. In: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*. Ed. by Anna Korhonen, David R. Traum, and Lluís Màrquez. Association for Computational Linguistics, 2019, pp. 732–746. DOI: 10.18653/v1/p19-1072. URL: https://doi.org/10.18653/v1/p19-1072.

[180] Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. “Modeling Relational Data with Graph Convolutional Networks”. In: *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*. Ed. by Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam. Vol. 10843. Lecture Notes in Computer Science. Springer, 2018, pp. 593–607. DOI: 10.1007/978-3-319-93417-4_38. URL: https://doi.org/10.1007/978-3-319-93417-4%5C_38.

[181] Susan Scutti. “Little evidence that marijuana helps chronic pain, PTSD, studies find”. In: *CNN* (Aug. 2017). URL: https://edition.cnn.com/2017/08/14/health/medical-marijuana-pain-ptsd-study/index.html.

[182] Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. “That is a Known Lie: Detecting Previously Fact-Checked Claims”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*. Ed. by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault. Association for Computational Linguistics, 2020, pp. 3607–3618. URL: https://www.aclweb.org/anthology/2020.acl-main.332/.

[183] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. “Hoaxy: A Platform for Tracking Online Misinformation”. In: *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume*. Ed. by Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao. ACM, 2016, pp. 745–750. DOI: 10.1145/2872518.2890098. URL: https://doi.org/10.1145/2872518.2890098.

[184] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. “The spread of low-credibility content by social bots”. In: *Nature communications* 9.1 (2018), pp. 1–9.

[185] Elisa Shearer and Amy Mitchell. *News use across social media platforms in 2020*. 2021.

[186] Qinlan Shen, Michael Miller Yoder, Yohan Jo, and Carolyn Penstein Rosé. “Perceptions of Censorship and Moderation Bias in Political Debate Forums”. In: *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*. AAAI Press, 2018, pp. 350–359. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17809.

[187] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. “Fake News Detection on Social Media: A Data Mining Perspective”. In: *SIGKDD Explorations* 19.1 (2017), pp. 22–36. DOI: 10.1145/3137597.3137600. URL: https://doi.org/10.1145/3137597.3137600.

[188] Kai Shu, Suhang Wang, and Huan Liu. “Beyond News Contents: The Role of Social Context for Fake News Detection”. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019, Melbourne, VIC, Australia, February 11-15, 2019*. Ed. by J. Shane Culpepper, Alistair Moffat, Paul N. Bennett, and Kristina Lerman. ACM, 2019, pp. 312–320. DOI: 10.1145/3289600.3290994. URL: https://doi.org/10.1145/3289600.3290994.

[189] Sarah Shugars, Adina Gitomer, Stefan McCabe, Ryan J. Gallagher, Kenneth Joseph, Nir Grinberg, Larissa Doroshenko, Brooke Foucault Welles, and David Lazer. “Pandemics, Protests, and Publics: Demographic Activity and Engagement on Twitter in 2020”. In: *Journal of Quantitative Description: Digital Media* 1 (Apr. 2021). DOI: 10.51685/jqd.2021.002. URL: https://journalqd.org/article/view/2570.

[190] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. “Why We Read Wikipedia”. In: *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*. Ed. by Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich. ACM, 2017, pp. 1591–1600. DOI: 10.1145/3038912.3052716. URL: https://doi.org/10.1145/3038912.3052716.

[191] Lisa Singh, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, Kornraphop Kawintiranon, Colton Padden, Rebecca Vanarsdall, Emily K. Vraga, and Yanchen Wang. “A first look at COVID-19 information and misinformation sharing on Twitter”. In: *CoRR* abs/2003.13907 (2020). arXiv: 2003.13907. URL: https://arxiv.org/abs/2003.13907.

[192] Julia Sittmann and Andrew Tompkins. “The strengths and weaknesses of automated fact-checking tools”. In: *Deutsche Welle* (July 2020). URL: https://www.dw.com/en/the-strengths-and-weaknesses-of-automated-fact-checking-tools/a-53956958.

[193] Panayiotis Smeros, Carlos Castillo, and Karl Aberer. “SciClops: Detecting and Contextualizing Scientific Claims for Assisting Manual Fact-Checking”. In: *CIKM ’21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*. Ed. by Gianluca Demartini, Guido

Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong. ACM, 2021, pp. 1692–1702. DOI: 10.1145/3459637.3482475. URL: https://doi.org/10.1145/3459637.3482475.

[194] Panayiotis Smeros, Carlos Castillo, and Karl Aberer. "SciLens: Evaluating the Quality of Scientific News Articles Using Social Media and Scientific Literature Indicators". In: *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. Ed. by Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia. ACM, 2019, pp. 1747–1758. DOI: 10.1145/3308558.3313657. URL: https://doi.org/10.1145/3308558.3313657.

[195] Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. "Cross-topic Argument Mining from Heterogeneous Sources". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Ed. by Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii. Association for Computational Linguistics, 2018, pp. 3664–3674. DOI: 10.18653/v1/d18-1402. URL: https://doi.org/10.18653/v1/d18-1402.

[196] Marcella Tambuscio, Giancarlo Ruffo, Alessandro Flammini, and Filippo Menczer. "Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks". In: *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*. Ed. by Aldo Gangemi, Stefano Leonardi, and Alessandro Panconesi. ACM, 2015, pp. 977–982. DOI: 10.1145/2740908.2742572. URL: https://doi.org/10.1145/2740908.2742572.

[197] Joseph W. Taylor, Marie Long, Elizabeth Ashley, Alex Denning, Beatrice Gout, Kayleigh Hansen, Thomas Huws, Leifa Jennings, Sinead Quinn, Patrick Sarkies, Alex Wojtowicz, and Philip M. Newton. "When Medical News Comes from Press Releases - A Case Study of Pancreatic Cancer and Processed Meat". In: *PLOS ONE* 10.6 (June 2015), pp. 1–13. DOI: 10.1371/journal.pone.0127848. URL: https://doi.org/10.1371/journal.pone.0127848.

[198] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. "ClaimsKG: A Knowledge Graph of Fact-Checked Claims". In: *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II*. Ed. by Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtech Svátek, Isabel F. Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon. Vol. 11779. Lecture Notes in Computer Science. Springer, 2019, pp. 309–324. DOI: 10.1007/978-3-030-30796-7_20. URL: https://doi.org/10.1007/978-3-030-30796-7%5C_20.

[199] Kjerstin Thorson and Chris Wells. "Curated Flows: A Framework for Mapping Media Exposure in the Digital Age". In: *Communication Theory* 26.3 (Nov. 2015), pp. 309–328. ISSN: 1050-3293. DOI: 10.1111/comt.12087. eprint: https://academic.oup.com/ct/article-pdf/26/3/309/21955206/jcomthe0309.pdf. URL: https://doi.org/10.1111/comt.12087.

[200] Neil Thurman, Judith Moeller, Natali Helberger, and Damian Trilling. "My Friends, Editors, Algorithms, and I". In: *Digital Journalism* 7.4 (2019), pp. 447–469. DOI: 10.1080/

21670811.2018.1493936. eprint: https://doi.org/10.1080/21670811.2018.1493936. URL: https://doi.org/10.1080/21670811.2018.1493936.

[201] Juliane Urban and Wolfgang Schweiger. "News Quality from the Recipients' Perspective". In: *Journalism Studies* 15.6 (2014), pp. 821–840. DOI: 10.1080/1461670X.2013.856670. eprint: https://doi.org/10.1080/1461670X.2013.856670. URL: https://doi.org/10.1080/1461670X.2013.856670.

[202] Jay J Van Bavel, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, et al. "Using social and behavioural science to support COVID-19 pandemic response". In: *Nature human behaviour* 4.5 (2020), pp. 460–471.

[203] Timoté Vaucher, Andreas Spitz, Michele Catasta, and Robert West. "Quotebank: A Corpus of Quotations from a Decade of News". In: *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*. Ed. by Liane Lewin-Eytan, David Carmel, Elad Yom-Tov, Eugene Agichtein, and Evgeniy Gabrilovich. ACM, 2021, pp. 328–336. DOI: 10.1145/3437963.3441760. URL: https://doi.org/10.1145/3437963.3441760.

[204] Luisa Verdoliva. "Media Forensics and DeepFakes: An Overview". In: *IEEE J. Sel. Top. Signal Process.* 14.5 (2020), pp. 910–932. DOI: 10.1109/JSTSP.2020.3002101. URL: https://doi.org/10.1109/JSTSP.2020.3002101.

[205] Dinesh Kumar Vishwakarma, Deepika Varshney, and Ashima Yadav. "Detection and veracity analysis of fake news via scrapping and authenticating the web search". In: *Cogn. Syst. Res.* 58 (2019), pp. 217–229. DOI: 10.1016/j.cogsys.2019.07.004. URL: https://doi.org/10.1016/j.cogsys.2019.07.004.

[206] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *Science* 359.6380 (2018), pp. 1146–1151. ISSN: 0036-8075. DOI: 10.1126/science.aap9559. eprint: http://science.sciencemag.org/content/359/6380/1146.full.pdf. URL: http://science.sciencemag.org/content/359/6380/1146.

[207] Bayliss Wagner. "Fact check: Autism diagnosis criteria changes have led to increased rates". In: *USA Today* (Apr. 2021). URL: https://www.usatoday.com/story/news/factcheck/2021/04/18/fact-check-autism-diagnosis-changes-over-years-account-high-rate/7102414002.

[208] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, Paul Mooney, Dewey A. Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Chris Wilhelm, Boya Xie, Douglas Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. "CORD-19: The Covid-19 Open Research Dataset". In: *CoRR* abs/2004.10706 (2020). arXiv: 2004.10706. URL: https://arxiv.org/abs/2004.10706.

[209] Ruoxi Wang, Rakesh Shivanna, Derek Zhiyuan Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed H. Chi. “DCN V2: Improved Deep & Cross Network and Practical Lessons for Web-scale Learning to Rank Systems”. In: *WWW ’21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*. Ed. by Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia. ACM, 2021, pp. 1785–1797. DOI: 10.1145/3442381.3450078. URL: https://doi.org/10.1145/3442381.3450078.

[210] Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang. “Community Preserving Network Embedding”. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*. Ed. by Satinder P. Singh and Shaul Markovitch. AAAI Press, 2017, pp. 203–209. URL: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14589.

[211] Wei Wei and Xiaojun Wan. “Learning to Identify Ambiguous and Misleading News Headlines”. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. Ed. by Carles Sierra. ijcai.org, 2017, pp. 4172–4178. DOI: 10.24963/ijcai.2017/583. URL: https://doi.org/10.24963/ijcai.2017/583.

[212] Deutsche Welle. *Fact check: How do I spot fake news?* https://www.dw.com/en/fact-check-how-do-i-spot-fake-news/a-59978706. 2022.

[213] Robert West and Eric Horvitz. “Reverse-Engineering Satire, or "Paper on Computational Humor Accepted despite Making Serious Advances"”. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 7265–7272. DOI: 10.1609/aaai.v33i01.33017265. URL: https://doi.org/10.1609/aaai.v33i01.33017265.

[214] Samuel T Wilkinson, Elina Stefanovics, and Robert A Rosenheck. “Marijuana use is associated with worse outcomes in symptom severity and violent behavior in patients with posttraumatic stress disorder”. In: *The Journal of clinical psychiatry* 76.9 (Sept. 2015), pp. 1174–1180. ISSN: 1555-2101. DOI: 10.4088/JCP.14m09475. URL: https://pubmed.ncbi.nlm.nih.gov/26455669%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6258013/.

[215] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. “HuggingFace’s Transformers: State-of-the-art Natural Language Processing”. In: *CoRR* abs/1910.03771 (2019). arXiv: 1910.03771. URL: http://arxiv.org/abs/1910.03771.

[216] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. “Embedding Entities and Relations for Learning and Inference in Knowledge Bases”. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2015. URL: http://arxiv.org/abs/1412.6575.

[217] Kai-Cheng Yang, Francesco Pierri, Pik-Mai Hui, David Axelrod, Christopher Torres-Lugo, John Bryden, and Filippo Menczer. “The COVID-19 Infodemic: Twitter versus Facebook”. In: *Big Data & Society* 8.1 (2021), p. 20539517211013861. DOI: 10.1177/20539517211013861. eprint: https://doi.org/10.1177/20539517211013861. URL: https://doi.org/10.1177/20539517211013861.

[218] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. “Unsupervised Fake News Detection on Social Media: A Generative Approach”. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 5644–5651. DOI: 10.1609/aaai.v33i01.33015644. URL: https://doi.org/10.1609/aaai.v33i01.33015644.

[219] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. “XLNet: Generalized Autoregressive Pretraining for Language Understanding”. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett. 2019, pp. 5754–5764. URL: https://proceedings.neurips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html.

[220] Liang Yao, Chengsheng Mao, and Yuan Luo. “Graph Convolutional Networks for Text Classification”. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 7370–7377. DOI: 10.1609/aaai.v33i01.33017370. URL: https://doi.org/10.1609/aaai.v33i01.33017370.

[221] Zi Yin, Vin Sachidananda, and Balaji Prabhakar. “The Global Anchor Method for Quantifying Linguistic Shifts and Domain Adaptation”. In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Ed. by Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett. 2018, pp. 9434–9445. URL: https://proceedings.neurips.cc/paper/2018/hash/80b618ebcac7aa97a6dac2ba65cb7e36-Abstract.html.

[222] Youtube. *Internet Citizens*. https://internetcitizens.withyoutube.com. 2017.

[223] Savvas Zannettou, Michael Sirivianos, Jeremy Blackburn, and Nicolas Kourtellis. “The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans”. In: *ACM J. Data Inf. Qual.* 11.3 (2019), 10:1–10:37. DOI: 10.1145/3309699. URL: https://doi.org/10.1145/3309699.

[224] John Zarocostas. “How to fight an infodemic”. In: *The Lancet* 395.10225 (2020), p. 676. ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(20)30461-X.

[225] Amy X. Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, Ed Bice, Sandro Hawke, David R. Karger, and An Xiao Mina. "A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles". In: *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*. Ed. by Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis. ACM, 2018, pp. 603–612. DOI: 10.1145/3184558.3188731. URL: https://doi.org/10.1145/3184558.3188731.

[226] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. "Heterogeneous Graph Neural Network". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*. Ed. by Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis. ACM, 2019, pp. 793–803. DOI: 10.1145/3292500.3330961. URL: https://doi.org/10.1145/3292500.3330961.

[227] Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. "ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research". In: *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*. Ed. by Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux. ACM, 2020, pp. 3205–3212. DOI: 10.1145/3340531.3412880. URL: https://doi.org/10.1145/3340531.3412880.

[228] Yang Zhou, Hong Cheng, and Jeffrey Xu Yu. "Graph Clustering Based on Structural/Attribute Similarities". In: *PVLDB* 2.1 (2009), pp. 718–729. DOI: 10.14778/1687627.1687709. URL: http://www.vldb.org/pvldb/2/vldb09-175.pdf.

[229] Dimitrina Zlatkova, Preslav Nakov, and Ivan Koychev. "Fact-Checking Meets Fauxtography: Verifying Claims About Images". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Ed. by Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan. Association for Computational Linguistics, 2019, pp. 2099–2108. DOI: 10.18653/v1/D19-1216. URL: https://doi.org/10.18653/v1/D19-1216.

[230] Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. "Discourse-aware rumour stance classification in social media using sequential classifiers". In: *Inf. Process. Manage.* 54.2 (2018), pp. 273–290. DOI: 10.1016/j.ipm.2017.11.009. URL: https://doi.org/10.1016/j.ipm.2017.11.009.

Panayiotis Smeros

Curriculum Vitae

EPFL IC IINFCOM LSIR
BC 142 (Bâtiment BC)
Station 14
CH-1015 Lausanne
Switzerland
☎ *+41.21.69.37573*
✉ *panayiotis.smeros@epfl.ch*
panayiotis.smeros.info

Education

2016 - 2022 PhD Student
École Polytechnique Fédérale de Lausanne
Thesis: *Combating Online Scientific Misinformation*
Advisors: *Karl Aberer, Carlos Castillo*

2007 - 2014 MSc & BSc Student
National and Kapodistrian University of Athens
MSc Specialization: *Advanced Information Systems*
MSc Thesis: *Discovering Spatial and Temporal Links among RDF Data*
BSc Specialization: *Computer Systems and Applications*
BSc Thesis: *Storing RDF Data and Evaluating SPARQL Queries using MapReduce*
Supervisor: *Manolis Koubarakis*

Selected Publications

ICWSM'23 M. Gruppi, P. Smeros, S. Adali, C. Castillo, K. Aberer. *SciLander: Mapping the Scientific News Landscape.*

CIKM'21 P. Smeros, C. Castillo, K. Aberer. *SciClops: Detecting and Contextualizing Scientific Claims for Assisting Manual Fact-Checking.*

VLDB'20 A. Romanou, P. Smeros, C. Castillo, K. Aberer. *SciLens News Platform: A System for Real-Time Evaluation of News Articles.*

WWW'19 P. Smeros, C. Castillo, K. Aberer. *SciLens: Evaluating the Quality of Scientific News Articles Using Social Media and Scientific Literature Indicators.*

AAAI'17 M.A. Sherif, K. Dreßler, P. Smeros, A.-C. Ngonga Ngomo. *Radon – Rapid Discovery of Topological Relations.*

LDOW'16 P. Smeros, M. Koubarakis. *Discovering Spatial and Temporal Links among RDF Data.*

ESWC'13 K. Bereta, P. Smeros, M. Koubarakis. *Representing and Querying the Valid Time of Triples for Linked Geospatial Data.*

Experience

Research & Development

2019 - 2022 NewsTeller: Real-Time News Analytics Platform

2019 - 2022 SciClops & SciLens: Methods to Combat Online Scientific Misinformation

2015 - 2016 Silk: Linked Data Integration Framework

2012 - 2015 Strabon: Semantic SpatioTemporal RDF Store

Teaching Assistance

2016 - 2020 École Polytechnique Fédérale de Lausanne
Distributed Information Systems, Applied Data Analysis

2011 - 2015 National and Kapodistrian University of Athens
Artificial Intelligence, Knowledge Technologies

Conferences & Journals

PC Member/Organizer: MEDIATE['22,'21,'20], LinkedGeo['15]

Reviewer: ICDE['21,'20,'19,'18], COLING['20], TWEB['20], EDBT['20,'19,'17], CIKM['19], WISE['19,'18], ESWC['19,'17,'16,'14,'13], BigData['18,'17], ISWC['14], JWS['13]

Attendee: ICWSM['22], CIKM['21], TTO['21,'20], VLDB['20], WWW['19,'18,'17,'16], AMLD['19,'18,'17], ESWC['15,'13]

Volunteer: RR['14], RW['14] and EDF['14]

Research Visits

2015 Data Science Group, University of Edinburgh

2015 Agile Knowledge Engineering & Semantic Web Group, University of Leipzig

2013 Database Architectures Group, Centrum Wiskunde & Informatica

EU & Swiss Projects

2020 - 2022 Media in the Digital Age (Special Interest Group)

2020 - 2022 The Importance of Journalism for the Digital Information Behaviour of Young Adults

2019 - 2020 Media Observatory Initiative

2019 - 2020 Evaluating the Quality of Scientific News

2013 - 2015 LEO: Linked Open Earth Observation Data for Precision Farming

2013 - 2015 MELODIES: Maximising the Exploitation of Linked Open Data In Enterprise & Science

2012 - 2013 TELEIOS: Virtual Observatory Infrastructure for Earth Observation Data

Awards

2019 2-year funding for SciLens (Open Science Fund & Swiss Academy of Engineering Sciences)

2016 2nd Rank in Kaggle Competition (ML Text Classification)

2016 6-month Fellowship (Distributed Information Systems Lab)