Fact-checking, False Narratives, and Argumentation Schemes

Giovanni Luca Ciampaglia
glc3@mail.usf.edu
University of South Florida, Tampa, FL, USA

John Licato
licato@usf.edu
University of South Florida, Tampa, FL, USA

ABSTRACT

False narratives, with their frequent omission of crucial argument components and their reliance on cherry-picking accurate but misleading facts, pose significant challenges to NLP tools that aim to automated parts of the human intelligence task of fact-checking. Fortunately, these forms of enthymemes can be overcome using methods from argumentation theory, which has refined over several decades a repertoire of argumentation scheme that can help us reason and model these forms of weaponized disinformation. In this position paper we argue for a new approach to computational fact-checking based on normative patterns of argumentation.

KEYWORDS

fact-checking, argumentation mining, Walton-style arguments, enthymemes

1 INTRODUCTION

In the arms race against disinformation, the job of fact checkers is made significantly more difficult by the use of factual statements, misleadingly presented. For example, a recent email sent to supporters of President Trump stated that

> "Millions of mail-in ballots were sent to people who didn't ask for them."

In its investigation of this claim, Fact-checking website *Snopes* determined that, taken at a literal level, this claim is factually correct. But "it was a mischaracterization of states' election laws to frame that fact as evidence of impropriety or fraud," which was the email's intended but unstated conclusion. "Rather, the ballots were sent to registered voters in accordance with state laws."[1] Such tactics are not limited to one political side, as shown by an article that appeared in the *Independent* with the following headlines:

> "Trump 'haemorrhaging' Twitter followers in wake of election defeat;
> President has lost followers every day this week, figures show."[2]

Though the subtitle is factually correct, President Trump's Twitter account only lost approximately 0.25% of its followers over the week in question, and may have actually gained followers since the election. Such claims are used to plant or strengthen certain beliefs (for example, that the 2020 US Presidential Election was marred by electoral fraud) in their target audience by use of *enthymemes*,[3] through content which: (1) omit crucial parts of the argument's structure, such as premises or conclusions; (2) tend to make use of cherry-picked but verifiably correct facts; and (3) can often be easily refuted by locating and examining the missing components.

Fact-checking these and other types of claims is a complex human intelligence task. First, a fact-checker needs to identify the underlying belief the claim is trying to plant, along with how it is meant to connect to the facts the claim presents. Second, she needs to collect the underlying facts. Third, she needs to evaluate the collected evidence and come up with an overall determination. This last step may play out in different ways, depending on the claim. The evidence may turn out to be false, in which case the claim can be refuted as is. However, if the evidence is true, as is typically the case for enthymematic claims, it may still fail to provide sufficient argumentative support of the claim, or it may fail to hold when additional facts are considered.

Of the various steps that comprise the workflow described above, journalists are generally quite effective at the first two, i.e. identifying the underlying belief and collecting the underlying facts. As far as the third task is concerned, emerging research shows that simple factual claims can be evaluated against a database of facts [1, 2, 3]. However, the task of evaluating misleading claims based on true evidence, i.e. enthymematic claims, still present considerable challenges and is hard to scale. This task challenging since it requires knowledge of *normative* patterns of argumentation — knowledge about what types of argument schemes constitute "good" arguments, what their refutation conditions are, and which (if any) argument components are implied but not verified by a claim.

To overcome this problem and make it easier for fact-checkers to deal with complex claims in general, and more specifically with enthymematic claims, there is a strong need for manually curated argument schemes, organized into premises and conclusions. Each scheme should be detailed by flexible analyses of its strengths, weaknesses, conditions of refutation/strengthening, and applicability conditions. Fortunately, such a corpus of argument schemes already exists; it has been compiled, studied, and published over the past near-four decades by Douglas Walton and colleagues [4, 5].

For example, consider the *Argument from Verbal Classification* scheme [5]:

> **Individual Premise:** a has property F.
> **Classification Premise:** For all x, if x has property F, then x can be classified as having property G.
> **Conclusion:** a has property G.
> **Critical Questions:**
> - *CQ1:* What evidence is there that a definitely has property F, as opposed to evidence indicating room for doubt about whether it should be so classified?
> - *CQ2:* Is the verbal classification in the classification premise based merely on an assumption about word usage that is subject to doubt?

[1] https://www.snopes.com/fact-check/millions-mail-in-ballots/
[2] https://www.independent.co.uk/news/world/americas/us-election-2020/trump-losing-twitter-followers-election-defeat-b1762502.html
[3] An enthymeme is "an argument in which one premise is not explicitly stated." (source: Oxford Languages).

KnOD'21 Workshop - April 14, 2021

This scheme may be applicable to this paper's opening example, by substituting 'the 2020 elections' for a, 'had millions of mail-in ballots sent unsolicited' for F, and 'was fraudulent' for G. Stated in this way, it is extremely clear that although the individual premise is true, the classification premise is extremely weak, and can be struck down soundly.

Walton-style argument schemes have never been effectively applied to journalistic workflows such as fact-checking. This raises a host of questions. For example: are current taxonomy of argument schemes capable of describing all claims currently in circulation? And, given a claim expressed in natural language, can we detected efficiently the presence of one or more argument schemes from this taxonomy in it? Does knowledge of the detected argument scheme help to identify and verify the underlying facts of the claim? We argue that more research is needed combining Natural Language Processing (NLP), Network Science, and Machine Learning to seek an answer to these questions.

2 A NEW RESEARCH AGENDA

We seek to understand the role that argument patterns and schemes play in enthymematic disinformation, with the ultimate aim of developing tools for more effective fact checking. This involves applying state-of-the-art NLP in order to solve the following tasks:

(1) Given a carrier of enthymematic disinformation (e.g. the headline of an article shared on social media), identify its *intended conclusion*—i.e., the belief that the disinformation's creator intends to instill or reinforce in the reader. This will likely use patterns identified automatically, rather than normatively crafted argument schemes. To systematically explore this, we propose the use of new benchmark datasets and tasks which build on two existing tasks in natural language processing: *natural language inference* [6, 7, 8] and *argument reasoning comprehension* [9, 10, 11].

(2) Given an intended conclusion, use argument schemes to identify its necessary and sufficient premises, to aid in repetitive tasks related to the research required to fact-check a claim. This will use normatively crafted argument schemes, such as Walton's, since the goal here is to enforce critical thinking, rather than to just identify uncritical associations. To tame the complexity of this task, we could focus on domain-specific claims, such as political news headlines, and their related argument schemes.

(3) Given a scheme and its identified premises, develop new knowledge graph verification techniques that make full use of Walton-style argument schemes. A promising approach to tackle this issue is to devise novel graph traversal schemes that exploit knowledge of the statistical property of the underlying knowledge topology. The goal here is to mine a reference knowledge base for underlying evidence to be checked.

(4) Last but not least, we propose to engage with specific fact-checking organizations to develop new tools to aid human fact-checkers analyze false narratives through the lens of argument schemes.

The Defeasible Inference Task. We here propose a new benchmark task which we believe will contribute towards the development of the kinds of NLP systems that are necessary to achieve the goals described thus far. It is based on the Argument Reasoning Comprehension Task (ARCS) [9, 10], which (after correcting for biases in the initial version of the task) has been shown to be very difficult even for state-of-the-art NLP systems [11]. Defeasible reasoning [12, 13, 14], sometimes called "nonmonotonic reasoning," involves arguments whose conclusions are subject to later defeat by additional evidence or arguments. For example, consider the following statements:

- p ="A woman refuses to bathe her child in tap water."
- h = "The woman is unnecessarily overprotective of her child."
- w_1 = "The woman refuses to let her baby eat GMOs or processed foods."
- w_2 = "The woman has read credible reports of flesh-eating bacteria in the local water supply."

Faced with p alone, which may be a newspaper headline, a reader might infer h on their own, and implying h may indeed be the goal of the headline writer all along. If further given w_1, then the reader might be more likely to believe that h is true. But if then given w_2, a typical reader is likely to infer that h no longer holds (or is at least not sufficiently supported by the available evidence).

The ARCS task consists of four statements similar to those above, where the two warrants (w_1 and w_2) are such that one of them combined with the premise p infers the hypothesis h, whereas the other combined with p infers the negation of h. The task is to determine which warrant is which. We propose to extend this task in two ways: First, the premise will consist of an article's headline, and the hypothesis will consist of an intended conclusion of that headline. Second, the warrants $\mathbf{w} = \{w_1, ..., w_n\}$ will consist of factual statements (not limited to two) either directly extracted from the article, or taken from some fact checking service or process. The task, then, is to determine whether (1) $p \rightarrow h$,[4] and (2) if there exists some $w_i \in \mathbf{w}$ such that $\neg((p \wedge w_i) \rightarrow h)$. If this is the case, then the article corresponding to this problem can be flagged as a possible piece of enthymematic disinformation, and can then be reviewed by a fact checker.

Currently, no benchmark dataset for this task exists. In prior work, Shao, Ciampaglia, Varol, Yang, Flammini, and Menczer analyze a sample of $N = 100$ articles from 'low credibility' sources and found that the majority of them (77%–82%, depending on sampling strategy) conformed to some form of misleading or inaccurate information [15]. Actually collecting such a dataset is a focus of our present research, and a starting point can be existing datasets of clickbait [16], false claims, and fact-checks [17]. However, the potential benefits of the resulting dataset are significant: a model trained to perform well on this task may be able to automatically determine the prevalence of enthymematic disinformation.

3 DISCUSSION

From a computational perspective, enthymematic disinformation poses a set of interdisciplinary challenges touching upon, for example, NLP, Network Science, and Artificial Intelligence, to name a few. The research tasks outlined above are extremely under-researched

[4]The implication relationship $\rightarrow$ here would be an approximation of the "commonsense" inference that tasks like NLI [6] attempt to capture. Such an inferential relationship is, at a minimum, defeasible and non-monotonic.

by the state of the art in these disciplines. In particular, we aim to bridge the gap between the statistical approach, which is typical of NLP and Network Science, and the normative approach, which is more typical of argumentation theory. We also contribute to NLP and to automated reasoning by introducing new datasets, benchmarks, and challenges for them, as well as solutions for them which will combine the latest advances in NLP with the underutilized resources provided by argument schemes.

Disinformation is not new, but its spread throughout recent years is unprecedented and disastrous [18, 19, 20, 21]. But since the amount of energy required to successfully refute a false claim tends to be significantly greater than that required to make or spread the false claim, tools for making the fact-checking process easier (by reducing the burden associated to repetitive tasks) are desperately needed. The partnership between computer scientists and journalists dates back a considerable time and in some cases (e.g. news recommendation, digital advertising) it has been deeply transformative for the news media industry, but it has impacted more the business aspect of running a news outlet and less the editorial aspect of the job. In contrast, we envision that the proposed work will directly lead to the development of public tools for the fact-checking community, an emerging form of journalism that is increasingly becoming an essential component of the socio-technical infrastructure of the internet.

About the authors

Dr. Giovanni Luca Ciampaglia is an expert in Network Science, Data Science, and Computational Social Science. He is interested in all problems arising from the interplay between people and computing systems, in particular the integrity of information in cyberspace against disinformation [22, 23, 24, 15, 25, 26, 2], and the trustworthiness and reliability of social computing systems [26, 27, 28, 29]. At USF, he leads the Computational Sociodynamics Laboratory (CSDL).

Dr. John Licato is director of the Advancing Machine and Human Reasoning (AMHR) Lab, and is an expert in Natural Language Processing, Cognitive Modeling, and computational models of both formal [30, 31, 32] and informal reasoning [33, 8, 34], including the productive use of Walton-style argument schemes in AI [35, 36, 37].

Acknowledgement

The authors would like to thank Alexios Mantzarlis, Dario Taraborelli, and David Corney for their feedback on the manuscript and insightful conversations.

REFERENCES

[1] 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), pp. 1–13. DOI: 10.1371/journal.pone.0128193.

[2] Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. "Finding Streams in Knowledge Graphs to Support Fact Checking". In: *2017 IEEE International Conference on Data Mining (ICDM)*. Extended Version. Piscataway, NJ: IEEE, Nov. 2017, pp. 859–864. ISBN: 2374-8486. DOI: 10.1109/ICDM.2017.105. arXiv: 1708.07239 [cs.AI].

[3] "REMOD: Relation Extraction for Modeling Online Discourse". In: (2021).

[4] Douglas Walton. *Arguer's Position: A Pragmatic Study of Ad Hominem Attack, Criticism, Refutation, and Fallacy*. Greenwood Press, 1985.

[5] Douglas Walton, Chris Reed, and Fabrizio Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.

[6] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. "A large annotated corpus for learning natural language inference". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2015.

[7] Adina Williams, Nikita Nangia, and Samuel R. Bowman. "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference". In: *CoRR* abs/1704.05426 (2017). arXiv: 1704.05426. URL: http://arxiv.org/abs/1704.05426.

[8] Zaid Marji, Animesh Nighojkar, and John Licato. "Probing the Natural Language Inference Task with Automated Reasoning Tools". In: *Proceedings of The 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33)*. Ed. by Eric Bell and Roman Barták. AAAI Press, 2020.

[9] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. "The Argument Reasoning Comprehension Task: Identification and Reconstruction of Implicit Warrants". In: *CoRR* abs/1708.01425 (2018). arXiv: 1708.01425. URL: http://arxiv.org/abs/1708.01425.

[10] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. "SemEval-2018 Task 12: The Argument Reasoning Comprehension Task". In: *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018)*. Association for Computational Linguistics. 2018.

[11] Timothy Niven and Hung-Yu Kao. "Probing Neural Network Comprehension of Natural Language Arguments". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019, pp. 4658–4664.

[12] Roderick M. Chisholm. *Perceiving*. Cornell University Press, 1957.

[13] John L. Pollock. "Criteria and Our Knowledge of the Material World". In: *Philosophical Review* 76 (1967), pp. 28–62.

[14] John L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press, 1995.

[15] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. "The spread of low-credibility content by social bots". In: *Nature Communications* 9.1 (Nov. 2018), pp. 4787–. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06930-7.

[16] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. "Clickbait Detection". In: *Advances in Information Retrieval*. Ed. by Nicola Ferro, Fabio Crestani, Marie-Francine Moens, Josiane Mothe, Fabrizio Silvestri, Giorgio Maria Di Nunzio, Claudia Hauff, and Gianmaria Silvello. 6630 Cham, Switzerland: Springer International Publishing, 2016, pp. 810–817. ISBN: 978-3-319-30671-1.

[17] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthaus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. "ClaimsKG: A Knowledge Graph of Fact-Checked Claims". In: *International Semantic Web Conference*. Auckland, New Zealand, Oct. 2019, pp. 309–324. DOI: 10.1007/978-3-030-30796-7_20. URL: https://hal.archives-ouvertes.fr/hal-02404153.

[18] Leonardo Busztyn, Akaash Rao, Christopher Roth, and David Yanagizawa-Drott. *Misinformation During a Pandemic*.

[19] Binxuan Huang and Kathleen M. Carley. *Disinformation and Misinformation on Twitter during the Novel Coronavirus Outbreak*. 2020. arXiv: 2006.04278 [cs.SI].

[20] Areeb Mian and Shujhat Khan. "Coronavirus: the spread of misinformation". In: *BMC Medicine* 18.1 (2020), p. 89. DOI: 10.1186/s12916-020-01556-3. URL: https://doi.org/10.1186/s12916-020-01556-3.

[21] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, Aastha Dua, and Yan Liu. *Coronavirus on Social Media: Analyzing Misinformation in Twitter Conversations*. 2020. arXiv: 2003.12309 [cs.SI].

[22] Giovanni Luca Ciampaglia. "Fighting fake news: a role for computational social science in the fight against digital misinformation". In: *Journal of Computational Social Science* 1.1 (Jan. 2018), pp. 147–153. ISSN: 2432-2725. DOI: 10.1007/s42001-017-0005-6.

[23] Giovanni Luca Ciampaglia, Alexios Mantzarlis, Gregory Maus, and Filippo Menczer. "Research Challenges of Digital Misinformation: Toward a Trustworthy Web". In: *AI Magazine* 39.1 (Mar. 2018), pp. 65–74. DOI: 10.1609/aimag.v39i1.2783.

[24] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. "Hoaxy: A Platform for Tracking Online Misinformation". In: *Proceedings of the 25th International Conference Companion on World Wide Web*. WWW '16 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016, pp. 745–750. ISBN: 978-1-4503-4144-8. DOI: 10.1145/2872518.2890098.

[25] Chengcheng Shao, Pik-Mai Hui, Lei Wang, Xinwen Jiang, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. "Anatomy of an online misinformation network". In: *PLOS ONE* 13.4 (Apr. 2018), pp. 1–23. DOI: 10.1371/journal.pone.0196087.

[26] Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. "The production of information in the attention economy". In: *Scientific Reports* 5 (2015), p. 9452. DOI: 10.1038/srep09452.

[27] Giovanni Luca Ciampaglia, Azadeh Nematzadeh, Filippo Menczer, and Alessandro Flammini. "How algorithmic popularity bias hinders or promotes quality". In: *Scientific Reports* 8.1 (2018), pp. 15951–. ISSN: 2045-2322. URL: https://doi.org/10.1038/s41598-018-34203-2.

[28] D. Taraborelli and G.L. Ciampaglia. "Beyond Notability. Collective Deliberation on Content Inclusion in Wikipedia". In: *Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop (SASOW)*. Piscataway, NJ: IEEE, Sept. 2010, pp. 122–125. DOI: 10.1109/SASOW.2010.26.

[29] Shun Yamaya, Saumya Bhadani, Alessandro Flammini, Filippo Menczer, Brendan Nyhan, and Giovanni Luca Ciampaglia. "Political Audience Diversity and News Quality". In: *Proc. of Computation+Journalism*. Northeastern University. 2020. URL: https://cj2020.northeastern.edu/files/2020/02/CJ_2020_paper_59.pdf.

[30] John Licato, Naveen S. Govindarajulu, Selmer Bringsjord, Michael Pomeranz, and Logan Gittelson. "Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox". In: *Proceedings of the 23rd Annual International Joint Conference on Artificial Intelligence (IJCAI-13)* (2013).

[31] Elijah Malaby, Bradley Dragun, and John Licato. "Towards Concise, Machine-discovered Proofs of Gödel's Two Incompleteness Theorems". In: *Proceedings of The 33rd International Florida Artificial Intelligence Research Society Conference (FLAIRS-33)*. Ed. by Eric Bell and Roman Barták. AAAI Press, 2020.

[32] John Licato and Michael Cooper. "Assessing Evidence Relevance By Disallowing Direct Assessment". In: *Proceedings of the 12th Conference of the Ontario Society for the Study of Argumentation*. 2020.

[33] John Licato and Michael Cooper. "Evaluating Relevance in Analogical Arguments through Warrant-based Reasoning". In: *Proceedings of the European Conference on Argumentation (ECA 2019)*. 2019.

[34] Michael Cooper, Lindsay Fields, Marc Badilla, and John Licato. "WG-A: A Framework for Exploring Analogical Generalization and Argumentation". In: *Proceedings of the 42nd Cognitive Science Society Conference (CogSci 2020)*. 2020.

[35] John Licato and Zhitian Zhang. "Evaluating Representational Systems in Artificial Intelligence". In: *Artificial Intelligence Review* 52.2 (2019), pp. 1463 –1493.

[36] Ryan Quandt and John Licato. "Problems of Autonomous Agents following Informal, Open-textured Rules". In: *Proceedings of the AAAI 2019 Spring Symposium on Shared Context*. 2019.

[37] Ryan Quandt and John Licato. "Problems of Autonomous Agents following Informal, Open-textured Rules". In: *Human-Machine Shared Contexts*. Ed. by William F. Lawless, Ranjeev Mittu, and Donald A. Sofge. Academic Press, 2020.