Explaining Misleading Claims Using Graphs of Entities

Kateřina Haniková[1][0009−0009−7162−5925]

Faculty of Informatics and Statistics, Prague University of Economics and Business, nám. W. Churchilla 1938/4, Prague, Czech Republic katerina.hanikova@vse.cz

Abstract. This early-stage PhD paper introduces the intended research of the author's PhD study. The research is focused on explaining misleading claims, which have been fact-checked, using graphs of entities. This approach combines argumentation modeling with the fact-checking domain. Existing work seems to focus more on whole sentences concerning argumentation, while this research aims to go deeper at the level of entities and relationships between them. Preliminary studies have already revealed some repetitive argumentation elements that are based on an entity graph. The results from this study should help us understand the misleading claims better and suggest how knowledge behind unstructured data (meaning texts) can be expressed in formal representation.

Keywords: Explaining misinformation · Entity graph · Fact-checking · Argumentation.

1 Introduction

Fact-checking is an activity of checking claims that are spread through a public space. This activity is done by fact-checkers, such as PolitiFact.com, FullFact.org, LeadStories.com, or less-spoken-language one, the Czech Demagog.cz.[1] The task of the fact-checkers is to provide arguments supported by trustworthy resources, such as research papers by subject-matter experts, laws, etc., to verify if a claim is true, false, or a mixture of both (specific categories may vary depending on a fact-checker). Fact-checking reports tend to be detailed and thus long. Many readers do not want to read a full report. Therefore, some platforms include summarising parts (e.g., PolitiFact.com). Sometimes, it can be challenging to understand fact-checking reports and what is wrong with the claim. Modeling the core of fact-checked claims can reveal inaccurate information in a claim and help explain it better.

Fact-checked reports can also be a "fount of wisdom." Fact-checkers elaborate on a detailed argumentation that can be used for studying the underlying patterns and concepts. This can contribute to building some formal knowledge,

[1] The International Fact-Checking Network (IFCN) coordinated by Poynter.org provides a list of verified signatories of the IFCN code of principles, which is available at: https://ifcncodeofprinciples.poynter.org/signatories.

which can then be incorporated into ontologies. Additionally, we can learn more about argumentation techniques by decomposing fact-checkers' argumentation into argumentative elements.

Finally, this research builds a connection between knowledge graphs and psychology. The idea is to teach people critical thinking by explaining the core argumentation and misleading structures used in the fact-checked claims (leveraging on knowledge graphs). This follows on from the Inoculation theory [14], which was later applied to the misinformation domain [19] (more described in Section 2.1). This work hypothesizes that knowledge graphs can be effective inoculation tools.

2 State of the Art

2.1 Foundations of Inoculation Theory

Many researchers from various fields try to find ways to stop the dissemination of misinformation. A background theory for this research is the Inoculation theory, introduced in the 1960s by William J. McGuire [13,14]. Inoculation theory uses a biological metaphor referring to the phenomenon that when we are vaccinated by some weakened virus, our body develops immunity to face even stronger exposure to this virus. Analogically, when we want to build immunity against persuasive attacks, we should first "inject" people with small doses of such persuasive attacks [14]. This theory became popular, especially in countering the spread of misinformation.

One of the current experts on inoculation theory applied to misinformation is Sander van der Linden. He addresses the main trends such as *Inoculating against fake news about COVID-19* [12] or *Inoculating the public against misinformation about climate change* [11]. Traberg et al. [19] then focus on preventing misinformation from influencing people. They argue that prevention is the best protection against a virus (whether we talk about pandemics or infodemics). Fact-checkers are focused more on *debugging* misinformation – meaning that they verify claims *after* they are spread in public space. However, it seems more effective to teach people to think critically and recognize some misinformation techniques so that they avoid sharing potentially harmful content without verifying it first. Recently, the inoculation theory was also applied through different games. The most recent review on *tackling misinformation with games* [9] showed that this is a relatively new area, and articles on this topic started to be published in 2019.

To summarise, this PhD research hypothesis is that knowledge graphs can also be used as psychological "vaccines" to build resilience against misinformation.

2.2 Argumentation in the Fact-checking Domain

Different fact-checkers use different scales of verdicts or typologies of misinformation, which can lead to confusion because there is no harmonization or unified methodology. An approach considered in this PhD research is focused on the analysis and classification of fact-checked misleading claims with the help of graph structures, which provides a new perspective on how to look at those claims. The aim is to develop a formal ontological model that would allow to capture the argumentation of fact-checkers and also support explaining misleading claims. The following paragraphs are directed at an overview of existing work in this area, starting from argumentation models and continuing with models about fact-checking.

One of the ontological models from the domain of argumentation is the Argument Interchange Format (AIF) [3]. This model is focused primarily on the interchange of argumentation data between different software systems. The aim differs from this PhD project because it is not focused on end-users (people) but rather on machines. Another ontology is the Argument Model Ontology (AMO) [6]. The AMO builds upon Toulmin's theory[2] that defines six elements: claim, warrant, data (evidence), backing, qualifier, and rebuttal.

In the fact-checking domain, probably the most widely used model is ClaimReview[3]. The Duke University Reporters' Lab developed this model in cooperation with Schema.org.[4] The main three classes are *ClaimReview*, *Claim* and *Rating*. Additionally, many properties can be reused from Schema.org, allowing for the creation of interesting datasets – for example, ClaimsKG [18] uses ClaimReview.

Another model – The Open Claims conceptual model [1] elaborates more about different components of a claim. This model is based on three main classes: claim proposition, claim utterance, and claim context. The interesting part that this paper considers is a representation of the claim proposition. This representation can be textual or more formal. This is closely related to this PhD research. It is also focused on capturing claims in more formal representation. However, the research aims to go even further to study an argumentation provided by a fact-checker at the same formal level.

This research aims to connect the argumentation and fact-checking domains since they are closely related. The aim is to capture the core of fact-checking reports at a more granular level than current models offer.

2.3 Argumentation Mining

Another area of the research is the automated identification of argument components, which is a task of argumentation mining (AM). Argumentation mining aims at an automated extraction of structured arguments from unstructured information (such as textual documents) [10]. In this context, the subject of this PhD research can be classified as the extraction of the argumentation structure of a monologue [2], meaning there is an argumentative text from one side (provided by fact-checkers) and not a dialogue. Recent research highlights the

[2] Work describing Toulmin's theory [8]
[3] https://www.claimreviewproject.com/
[4] https://schema.org/

promising role of deep learning in the argumentation mining field, particularly transformer-based architectures like BERT [4]. For example, the study [7] used BERT and the argumentation context to classify argumentation components. All cited studies can serve as an inspiration for the task of automated classification of argumentation elements.

3 Problem Statement and Contributions

Fact-checkers' reports tend to be long, and people not involved in the topic covered by reports can have problems with understanding. Categorizing claims just by saying they are false or misleading does not say much about what is wrong with the claim. This research aims to improve critical thinking by transforming the argumentation developed by fact-checkers into graph structures. The main research questions are:

RQ1 Are there any repetitive patterns behind claims and argumentation developed by fact-checkers?
RQ2 How can we formalize argumentation developed by fact-checkers at the level of entities and relationships?
RQ3 How do people perceive and understand explanations based on graphs?
RQ4 Can machines learn to recognize argumentation elements in misleading claims?

RQ1 aims to explore whether there are some repetitive patterns behind fact-checked claims. The emphasis is on false or misleading claims and representing these claims in semi-formal models (graphs of entities). This is followed by the next question, RQ2, which aims to develop a formal representation of argumentation as a new kind of argumentation ontology. RQ3 tests whether explanations based on graphs can improve the understanding of fact-checked claims and if people can learn to think about claims in a more structured way. Answering RQ4 will have to rely on research in the domain of machine learning, namely, verifying whether machines can learn to recognize specific kinds of argumentation elements.

4 Research Methodology and Approach

With respect to the order of the research questions, the research will proceed as follows:

1. *Empirical research.* The research will start with manually analyzing fact-checked claims from different portals. The task is to represent the claims as semi-formal models and gather some repetitive argumentation patterns. Expected outcomes are an annotated dataset of fact-checked misleading claims and some catalog of graph entity models.
   A possible approach to modeling fact-checked misleading claims is to create semi-formal models in a tool called PURO Modeler [5]. PURO Modeler

distinguishes between *particulars* and *universals*, allowing to create models at the level of instances. Another advantage that can be useful in modeling fact-checked claims is that the tool supports creating n-ary relationships. Most ontology modeling tools allow just binary relationships, which is fine since the Web Ontology Language (OWL) is based on description logics that consider only binary relationships. However, for the first domain analysis and studying structures, it can be helpful not to be limited to binary relationships.
2. *Ontology development.* The semi-formal models will be explored in order to find some main concepts. After a conceptual model is created, the formal ontological model will be designed, considering the use of existing models. A state-of-the-art methodology to be possibly used for ontology development is Linked Open Terms (LOT) [15]. The expected outcome from this phase is a new kind of argumentation ontology tuned for the fact-checking domain.
3. *Cognitive experiments.* The experiments will require the elaboration of sophisticated questionnaires, allowing us to test if a misinformation explanation based on graphs can improve the understanding of misleading claims. The initial idea of the questionnaire is to let it consist of three parts. The first part will consist of claims and textual summaries from fact-checkers, and the participants will be asked questions revealing their degree of understanding. The second set will contain claims and explanations based on graphs of entities with similar questions. And finally, there will be just claims, and the participants will be asked to suggest what they think might be wrong or worth verifying. The expected output from this phase will be the analysis of the questionnaire results.
4. *Machine learning.* The task will be to develop predictive models for specific argumentation elements, most likely based on state-of-the-art LLMs. The annotated dataset from the first phase of this research will be used as training data. The expected outcome of this section is a set of models capable of predicting argumentation elements and experimental results about their application.

5 Evaluation Plan

The new ontology should help explain the misleading claims and suggest how the knowledge behind unstructured data (fact-checked claims) can be expressed in formal representation. The evaluation of the ontology model could be based on the LOT methodology [15] and the NeOn methodological guidelines [16]. Some corresponding evaluation criteria may be:

1. *Domain coverage.* To determine if the model can capture diverse types of fact-checked claims, it will be applied to a set of claims (from different fact-checkers), giving rise to a knowledge graph.
2. *Fit for purpose or application.* This will be determined with a user study verifying if explanations based on the knowledge graphs are more effective/ more understandable compared to explanations in natural language.

3. *Logical consistency checking.* The ontology model should be consistent and without logical errors, which can be partially achieved through reasoners.

Besides those three evaluation criteria, feedback from users, domain experts, and ontology experts will be needed. This feedback needs to be gained continuously as the research proceeds.

Apart from the formal ontological model and the knowledge graph, the other result will be an annotated dataset. The annotated dataset could serve to train a model that would automatically predict some argumentation elements. An evaluation of the prediction model will be based on an accuracy metric computed upon cross-validation on a ground truth corpus and ex-post user evaluation of random samples.

6 Preliminary Results

Preliminary studies started by modeling different examples of fact-checked misleading claims. Let me first show three examples to demonstrate the modeling process. The examples can be seen in Fig. 1. The models were created with PURO Modeler[5] and are at this phase rather semi-formal. The examples are taken from PolitiFact.org from the half-true category.

The first example[6] represents the claim *"The Earth just started spinning faster than ever before and scientists are gravely concerned."* The claim does not specify which scientist, so in the diagram, we can see a *Some objects* node representing an unspecified set of instances of *scientists*.[7] Afterward, there is a green diamond representing the relationship *are concerned about* between the *Some objects* and the event *the Earth is spinning faster*. From the fact-checker's report, we can learn that the event is actually true; however, scientists are *not* concerned about it. This means that the **relationship** is **false**, but the **event** is **true**. In the diagram, these are shown as grey notes.

The second example[8] represents the claim *"In the 1960s, liberals emptied our psych wards."*. In the diagram, we can see again *some objects* that are instances of *liberals*. We also have a relationship *played a role in*, which originates those *Some objects* and leads to the time specification *the 1960s* and *psych wards emptying*. What the fact-checker added is that in this claim, there is missing another group of people *conservatives* (we can call this group as *politicians*). So, in this claim, it was identified as **missing subset**.

The third example[9] represents the claim *"Two years ago this week, 18 million people were out of work needing unemployment benefits. Today, that number is under 1.6 million, the lowest in decades."* In this claim, two values are compared.

[5] https://protegeserver.cz/purom5/
[6] http://tinyurl.com/hd326hwk
[7] This specific PURO diagram primitive is based on the so-called MISO modeling pattern, addressing the problem of 'multiple indirectly specified objects' [17].
[8] http://tinyurl.com/yexwp7jf
[9] http://tinyurl.com/4jcf3u5c

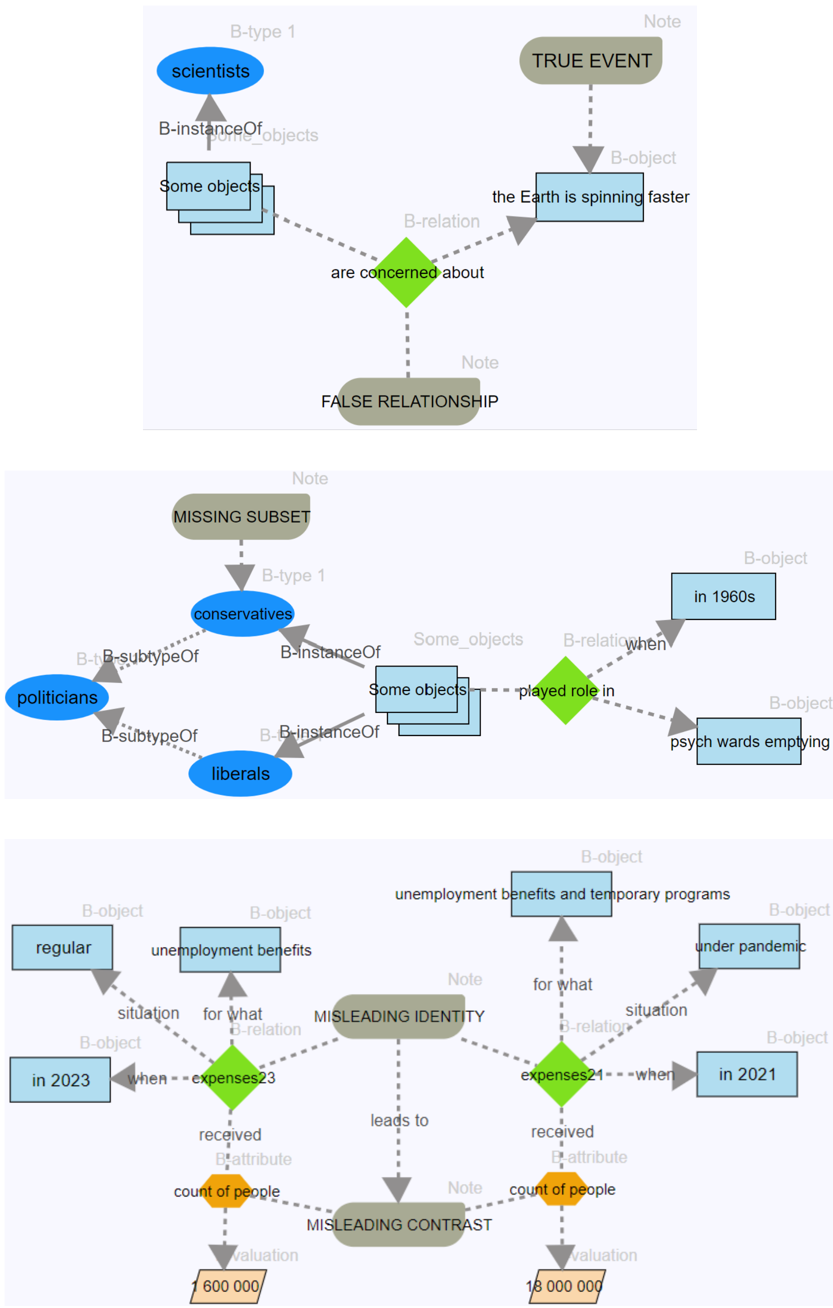

Fig. 1. Example models of misleading claims

However, if we look closer, we can see that those are not as comparable as it was claimed. On the right side of the diagram, there are *expenses* (green diamond) in 2021. On the left side, there are *expenses* in 2023. Now, the situation in 2021 was *under pandemic*, while in 2023 it was *regular*. Expenses in 2021 were for *unemployment benefits and temporary programs*; meanwhile, in 2023, they were just for *unemployment benefits*. So, if we look at both sides, we can see that those *expenses* are not identical, and this **misleading identity** leads to **misleading contrast**.

After explaining the examples, let us look at what has been discovered so far. Behind misleading claims that have been fact-checked, we can find some repetitive patterns, which can be expressed through argumentation elements. Those argumentation elements are tied to entity graphs (as shown in Fig. 1). Two types of argumentation elements can be recognized: 1) verdict argumentation elements (those in UPPER CASE) and 2) auxiliary argumentation elements (would be in Sentence case).

Verdict argumentation elements are created as a noun phrase (adjective and noun). The noun mainly refers to graphical primitives as 'relationship', 'object', 'type', 'attribute', and 'value'. There are also additional nouns not directly related to graphical primitives, such as 'identity' or 'contrast' in the third example. Six adjectives have also been defined: 'misleading', 'missing', 'false', 'unsubstantiated', 'exaggerated', and 'true'. The noun specifies *what* is wrong in the claim, and the adjective *how* it is wrong.

Sometimes, auxiliary argumentation elements are also needed to explain the claim. These were not in the presented examples. Often, they amount to notions such as 'Presumed justification' or 'Denial justification', representing evidence that supports the claim or evidence that refutes the claim. The preliminary studies were focused on 'half-true' claims. Some claims from this category can be true based on one viewpoint. However, the fact-checker adds another, less biased view, which denies the claim.

To conclude, up to this point, it was proved that misleading claims that have been fact-checked could be reformulated into graph structures, and via aggregating those graph structures, some repetitive argumentation elements can be found. The work will continue with an experiment to see if people already familiar with those argumentation verdict elements can, with some success, estimate them without having the fact-checker's explanation itself at their disposal, or, at least, if they can learn to think about claims in a more structured way through the graph-based thinking. So, they will be provided with some misleading claims, and the task will be to suggest what could be wrong in the claim based on their previous experience. The aim will be to capture just the *nouns* like relationship, value, subset, etc. If this experiment is successful and proves that people can learn to think about the claims in this way, we could possibly find a way to teach a machine (precisely speaking, some LLM) to perform the same task, namely, at least to suggest some Y/N questions that would lead the user to check the veracity of particular parts of the claim.

7 Conclusions

This study presents a novel view of misleading claims that have been fact-checked – a graph view. The preliminary results showed that the fact-checked claims can be transformed into entity graphs. These entity graphs can expose which part of the claim is wrong, misleading, or whether the claim is silent about some important piece of information that would affect its veracity. This research can reveal some interesting insights, either with respect to the misinformation domain or to the argumentation theory/ontology perspective. The claims are modeled in a structured way, so they should be independent of fact-checkers' verdict categorization. Here, emphasis is put directly on *what* is wrong rather than on the rate of truthfulness.

The main result should be the formalization of the knowledge gained from the study of graph structures behind fact-checked claims. This knowledge will be expressed as an ontological model. Apart from the formal ontological model, other results will be an annotated dataset and a knowledge graph. Both could be used with a combination of large language models to support explaining misinformation.

In conclusion, this research's findings should help explain misinformation and help people develop critical thinking by decomposing the argumentation behind fact-checked misleading claims into graph structures.

Acknowledgments. Special thanks go to my supervisor, Vojtěch Svátek, for his support and valuable feedback. The research is partially supported by CHIST-ERA within the CIMPLE project (CHIST-ERA-19-XAI-003).

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Boland, K., Fafalios, P., Tchechmedjiev, A., Dietze, S., Todorov, K.: Beyond facts – a survey and conceptualisation of claims in online discourse analysis. Semantic Web **13**(5), 793–827 (Jan 2022). https://doi.org/10.3233/SW-212838
2. Cao, L.: AutoAM: An End-To-End Neural Model for Automatic and Universal Argument Mining (Sep 2023), http://arxiv.org/abs/2309.09300
3. Chesñevar, C., Mcginnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., Willmott, S.: Towards an argument interchange format. The Knowledge Engineering Review **21**(4), 293–316 (2006). https://doi.org/10.1017/S0269888906001044
4. Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C.W., Choudhary, A., Agrawal, A., Billinge, S.J.L., Holm, E., Ong, S.P., Wolverton, C.: Recent advances and applications of deep learning methods in materials science. npj Computational Materials **8**(1), 1–26 (apr 2022). https://doi.org/10.1038/s41524-022-00734-6

5. Dudáš, M., Svátek, V., Vacura, M., Zamazal, O.: Starting Ontology Development by Visually Modeling an Example Situation - a User Study. Proceedings of the Second International Workshop on Visualization and Interaction for Ontologies and Linked Data pp. 114–119 (2016)
6. Fabio Vitali, Silvio Peroni: The Argument Model Ontology (May 2011), https://sparontologies.github.io/amo/current/amo.html
7. Hidayaturrahman, Dave, E., Suhartono, D., Arymurthy, A.M.: Enhancing argumentation component classification using contextual language model. Journal of Big Data **8**(1), 103 (Jul 2021). https://doi.org/10.1186/s40537-021-00490-2
8. Karbach, J.: Using toulmin's model of argumentation. Journal of Teaching Writing **6**(1), 81–92 (1987)
9. Kiili, K., Siuko, J., Ninaus, M.: Tackling misinformation with games: a systematic literature review. Interactive Learning Environments pp. 1–16 (2024). https://doi.org/10.1080/10494820.2023.2299999
10. Lawrence, J., Reed, C.: Argument Mining: A Survey. Computational Linguistics **45**(4), 765–818 (jan 2020), https://direct.mit.edu/coli/article/45/4/765/93362/Argument-Mining-A-Survey
11. van der Linden, S., Leiserowitz, A., Rosenthal, S., Maibach, E.: Inoculating the public against misinformation about climate change. Global Challenges **1**(2), 7 (2017). https://doi.org/10.1002/gch2.201600008
12. van der Linden, S., Roozenbeek, J., Compton, J.: Inoculating Against Fake News About COVID-19. Frontiers in Psychology **11** (2020). https://doi.org/10.3389/fpsyg.2020.566790
13. McGuire, W.J.: The Effectiveness of Supportive and Refutational Defenses in Immunizing and Restoring Beliefs Against Persuasion. Sociometry **24**(2), 184–197 (1961). https://doi.org/10.2307/2786067
14. McGuire, W.J.: Some contemporary approaches. In: Advances in Experimental Social Psychology, vol. 1, pp. 191–229. Academic Press (Jan 1964). https://doi.org/10.1016/S0065-2601(08)60052-0
15. Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., García-Castro, R.: LOT: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence **111** (May 2022). https://doi.org/10.1016/j.engappai.2022.104755
16. Sabou, M., Fernandez, M.: Ontology (Network) Evaluation, pp. 193–212. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24794-1_9
17. Svátek, V., Kluka, J., Vacura, M., Homola, M., Dudáš, M.: Patterns for referring to multiple indirectly specified objects (MISO): analysis and guidelines. In: Advances in Pattern-Based Ontology Engineering, the Workshop on Ontology Design and Patterns (WOP). Studies on the Semantic Web, vol. 51, pp. 1–24. IOS Press (2021). https://doi.org/10.3233/SSW210004
18. Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., Zloch, M., Zapilko, B., Dietze, S., Todorov, K.: Claimskg: A knowledge graph of fact-checked claims. In: ISWC 2019 - 18th International Semantic Web Conference. pp. 309–324. Springer, Auckland, New Zealand (2019). https://doi.org/10.1007/978-3-030-30796-7_20
19. Traberg, C.S., Roozenbeek, J., van der Linden, S.: Psychological Inoculation against Misinformation: Current Evidence and Future Directions. The ANNALS of the American Academy of Political and Social Science **700**(1), 136–151 (Mar 2022). https://doi.org/10.1177/00027162221087936