Interrupting the Propaganda Supply Chain

Kyle Hamilton
kyle.i.hamilton@mytudublin.ie
Technological University Dublin
Dublin, Ireland

Dr. Bojan Božić
bojan.bozic@tudublin.ie
Technological University Dublin
Dublin, Ireland

Dr. Luca Longo
luca.longo@tudublin.ie
Technological University Dublin
Dublin, Ireland

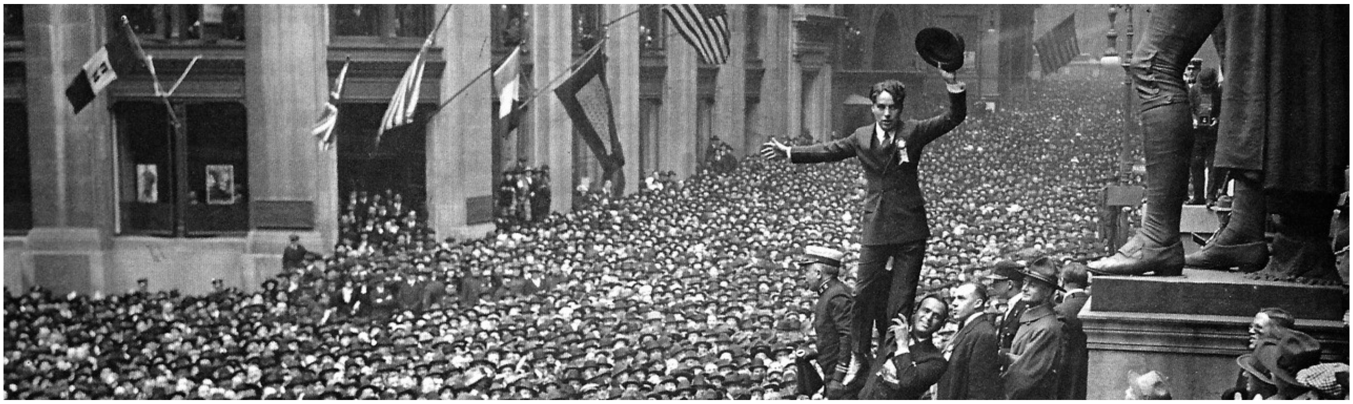

Figure 1: Charles Chaplin held up by Douglas Fairbanks, selling WWI War Bonds. Credit: The New York Times Photo Archives (1918).

ABSTRACT

In this early-stage research, a multidisciplinary approach is presented for the detection of propaganda in the media, and for modeling the spread of propaganda and disinformation using semantic web and graph theory. An ontology will be designed which has the theoretical underpinnings from multiple disciplines including the social sciences and epidemiology. An additional objective of this work is to automate triple extraction from unstructured text which surpasses the state-of-the-art performance.

CCS CONCEPTS

• **Computing methodologies** → **Ontology engineering**; *Semantic networks*; **Information extraction**.

KEYWORDS

Propaganda, Semantic Web, Ontological Computation, Machine Learning, Knowledge Extraction, Multidisciplinary.

1 INTRODUCTION

The word "infodemic" was coined by David Rothkopf in 2003, as a blend of the words information and pandemic at the time of the SARS breakout. "What exactly do I mean by the 'infodemic'? A few facts, mixed with fear, speculation and rumor, amplified and relayed swiftly worldwide by modern information technologies..." [20]. Rothkopf's prescient description depicts an almost naive audience of unwitting participants merely interfacing with a rapidly evolving misinformation ecosystem. By 2013, the World Economic Forum named "massive digital misinformation" as a global risk. [11]. And as recently as January 2021, journalist Chris Stirewalt put it this way: "Having worked in cable news for more than a decade after a wonderfully misspent youth in newspapers, I can tell you the result: a nation of news consumers both overfed and malnourished. Americans gorge themselves daily on empty informational calories, indulging their sugar fixes of self-affirming half-truths and even outright lies" [27].

To confirm Stirewalt's hypothesis, one only needs to look at the rise in popularity of fringe news networks, and the evolution of programming in mainstream media [1], from the lens of basic supply vs demand economics in the competition for viewership, and ultimately, advertising dollars. In the aftermath of the 2020 US presidential election, what were once fringe networks, like *OAN* and *Newsmax*, have grown in viewership by telling increasingly partisan, and outright fabricated stories, supporting the hypothesis that the demand is indeed substantial. So substantial in fact, that in early 2021, *Fox News* cancelled its prime-time news reporting replacing it with an "opinion" hour, in the hopes of winning back some of their most gluttonous viewers - to use Stirewalt's analogy. Add to that the miasma emanating from social media platforms like *Facebook*, *Twitter*, *Youtube*, *Parler* and *Gab* to name a few, and we end up with the perfect breeding ground for information disease, and a fertile soil for the sowing of propaganda like the Catholic Church couldn't have even dreamed of! [17]

Social media platforms like *Facebook* and *Twitter* have come under scrutiny for not doing enough to stop the flow of disinformation through their networks, while at the same time facing accusations of ideologically motivated censorship when users' posts have been either flagged, demonetised, or removed. How to strike the right balance between harm mitigation and censorship is not a new problem. What is unique today is the specific sociopolitical context and technological landscape. The premise of this work is that whatever solutions we come up with, we have to treat the present context as

KnOD'21 Workshop - April 14, 2021

a fundamental starting point, borrowing from multiple disciplines like the social sciences, economics, epidemiology, and more (See figure 2).

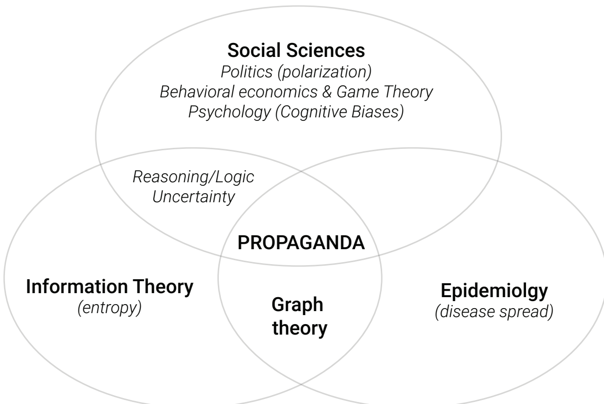

Figure 2: Multidisciplinary theoretical underpinnings. Each circle represents a discipline, at the intersection of which this work resides.

2 STATE OF THE ART

2.1 Propaganda Identification

For the purposes of this work, any information which is intended to influence beliefs or modify behaviors in order to further an agenda, will be considered propaganda. Often, this type of information is of questionable veracity, but this is not a requirement. The use of logical fallacies and/or emotional appeals are the hallmarks of propagandist messaging.

There doesn't appear to be very much literature dealing with propaganda detection using semantic web technologies specifically. Some closely related challenges are addressed in the following works: Mitzias et al [16] present a unified semantic infrastructure for information fusion of terrorism-related content, where propaganda techniques are often utilized; Castillo-Zúñiga et al [5] present a framework for generating an ontology in the related domain of cyber-terrorism using Natural Language Processing and Machine Learning.

2.2 False Information Identification

Although propaganda doesn't have to involve falsehoods, false or misleading information is often an instrument of the propagandist. Automatic false information identification has been tackled in a variety of ways which can be broken down into four major categories. This section briefly describes each one.

2.2.1 Style Based. These techniques take into consideration the linguistic content of the news piece. The premise being, that the writing style of a fake news item differs from a true one. The theory is that fake news items tend to appeal to emotions and be generally more sensationalist. These nuances can be picked up from the text using Natural Language Processing [2, 32]. Style based techniques can be used to annotate the nodes in the knowledge graph at the time of its generation.

2.2.2 Content/Knowledge Based. This is an approach which relies on fact checking against a large knowledge base, usually one which is publicly available, for example Wikipedia. This data is represented as a Knowledge Graph (KG) in the form of SOP (subject, object, predicate) triples. At a high level, such triples are extracted from the news item in question, and checked against the KG. This is typically posed as a link prediction problem. The challenge with content based approaches is that KGs are often incomplete, especially when it comes to recent events.[3, 6, 18]

2.2.3 Network based. Utilizing social context, how false information propagates through the network, who is spreading the false information, and how the spreaders connect with each other, is used to understand the patterns of false information through network theory. For example, one can perform a series of random walks to generate representations of the relationships between news items, publishers, and users, which can then be fed to a downstream supervised learning model. Here the challenge lies in the dependency on the social context which is only available after the news item has propagated through the network, thus early detection, a crucial component of the solution, is difficult [21, 26, 31].

2.2.4 Human-in-the-loop. To leverage both human insight and computational efficiency, hybrid approaches have been developed. In the "wisdom-of-the-crowd" approach [8, 12, 19, 22, 28, 29], no one individual has to score a news item correctly. Theoretically, the same number of people will underestimate a news item's veracity, as will overestimate it. If enough independent scores are averaged, this results in an unbiased estimator of the correct score of the item. Another hybrid technique involves the identification of "checkworthy content" automatically [10], and sending these items for human review. Again, these techniques can be helpful for annotation at the time of the KG generation.

2.3 Argument Mining

In 2018, the BBC launched *The Evidence Toolkit*[1] in response to growing concerns about people's ability to discern mis/disinformation. The tool is "designed to encourage users to dissect and critically appraise the internal reasoning structure of news reports." It draws from Argumentation Theory [9] in order to reach this objective.

In *Five Years of Argument Mining: a Data-driven Analysis*[4], Cabrio and Villata discuss current approaches to Argument Mining, ranging from SVM (Support Vector Machine) to RNN (Recurrent Neural Networks), and NB (Naive Bayes) among others. They list disinformation detection among the applications of Argument Mining, but they point out that more work is required to improve the performance of these systems. None of the techniques surveyed used semantic web techniques. Moreover, the majority of the work on propaganda detection is done at the document level, or even more coarsely, at the source level. There are drawbacks to such coarse classifications. For example, a news source can be generally propagandist, but that doesn't mean that all documents produced or published by that source are necessarily so. In fact, a technique of propaganda is to disseminate legitimate information in order to build trust, thus making it easier to manipulate the reader at a later

[1] https://cacm.acm.org/magazines/2020/11/248201-reason-checking-fake-news/fulltext

stage. Another downside is the lack of explain-ability as to exactly which fragments of a document are deemed propagandist and why. Hence it's important to aim to detect propagandist devices at a more granular level.

In 2019, De San Martino et al [7] created and released a data set for the detection of 18 propaganda devices in News articles, and invited the AI community to compete in the challenge of detecting said devices at a granular level. The competition [14] consists of two tasks: the first task is to identify spans of text containing propagandist devices, and the second task is a multilabel classification task to identify the specific device used, such as *name calling*, or *loaded language*. Looking at the leader-board [2] results, this is a difficult task rarely exceeding F1 scores above 0.5, leaving ample opportunity for further research. The authors have also provided code, an API, and a web app [3] for this purpose.

A natural fit for detecting propaganda devices such as faulty logic is to borrow from the field of argumentation. If logical arguments can be modelled using graphical representations [30] it may be worth exploring whether graphical representations (ontologies) can be used to model logical validity, for example by using description logic such as OWL, and Semantic Web Rule Language (SWRL). Indeed, in their survey of ontological modeling of rhetorical concepts for argument mining, Mitrovic et al [15], list "political discourse analysis" as a potential application.

3 PROBLEM STATEMENT AND CONTRIBUTIONS

Propaganda and disinformation in the media is a serious problem and has been the subject of research in multiple disciplines. Most of the work in automating the detection of propaganda and disinformation has focused on natural language processing and network analysis. While some efforts have been made to utilize semantic web techniques, especially as it pertains to fact-checking, a challenge unique to semantic web is the paucity of data suitable to analysis.

One of the objectives of this research is to develop an ontology for the media ecosystem to aid in the scholarship of the changing media landscape, and in particular, the evolution of propaganda in social discourse. For example, one might use PROV to analyse the origins of problematic ideologies. Another example might be to use an ontological rule set to evaluate sub-graphs of propaganda in order to predict if a particular media outlet will promote conspiracy theories.

Borrowing from network theory and epidemiology, a linked data knowledge graph can be used to measure the spread of propaganda and disinformation. One can model the spread in terms of the reproduction number R [4], for the purposes of making simulations and ultimately for the prevention of spread. In *Rules of Contagion*[13] Kucharsky decomposes the R number into four components:

$$R = Duration \times Opportunities \times Transmission\ probability \times Susceptibility. \quad (1)$$

[2]https://propaganda.qcri.org/semeval2020-task11/leaderboard.php

[3]https://www.tanbih.org/prta

[4]In the theory of epidemics, R represents the number of new infections we'd expect a typical infected person to generate on average.

Applying this formula to semantic web, each of the components can be mapped as per figure 3:

A challenge especially pertinent in the media domain, is the speed with which new events arrive, and thus the need for the knowledge base to be frequently updated. Another objective of the research is to develop a software framework for automatically extracting triples from unstructured text. Based on the literature search to date, this is an area which would benefit from improvement.

Having a robust and up to date knowledge base, which also contains historical information, will help to answer the following research question:

- **RQ** - To what extent can semantic web technology be used to model the spread of propaganda in a rapidly evolving media ecosystem?
- **Hypothesis** - If factors such as R-number, provenance, cognitive biases, and economic incentives are included in a linked data model based on a domain specific ontology which includes PROV, and a description logic for argument mining, then the spread of propaganda in the media ecosystem can be predicted exceeding state-of-the-art accuracy.

4 RESEARCH METHODOLOGY AND APPROACH

As this research is only in very early stages, the first step is to continue the investigation into the state-of-the-art, and to perform a comprehensive literature review. Other high-level tasks include:

- Continue to flesh out figure 5 by adding nodes and connections, which will help inform the schema/ontology of the news knowledge graph. Of particular importance are the relations between drivers and effects, and the paths to extremist ideology and adverse events.
- Generate News domain Knowledge Graph from unstructured text (news articles) using the above ontology
- Collect data on a regular basis to build a historical view of the ecosystem which will be used to measure the evolution of propagandist messaging as well as to evaluate the models over time.

Based on the literature search to date, while there are many mentions of the similarities between information spread and other disciplines, such as epidemiology and the social sciences, this is the first attempt to explicitly utilize a multi-disciplinary theoretical approach into the design of the ontology. A high-level sketch of the framework is illustrated in figure 4. Two data sets have been identified as particularly suitable to disinformation and propaganda detection. The *FakeNewsNet* [23–25] data set was chosen for its popularity, so results can be compared to prior work. The *PTC Corpus* [14] was chosen for its recency, and because it is annotated with propaganda devices for fine-grained propaganda detection. In addition, external knowledge graphs such as DBpedia will be leveraged to enrich the knowledge graphs generated from text.

5 EVALUATION PLAN

The framework (see figure 4) will have two models which will need to be evaluated:

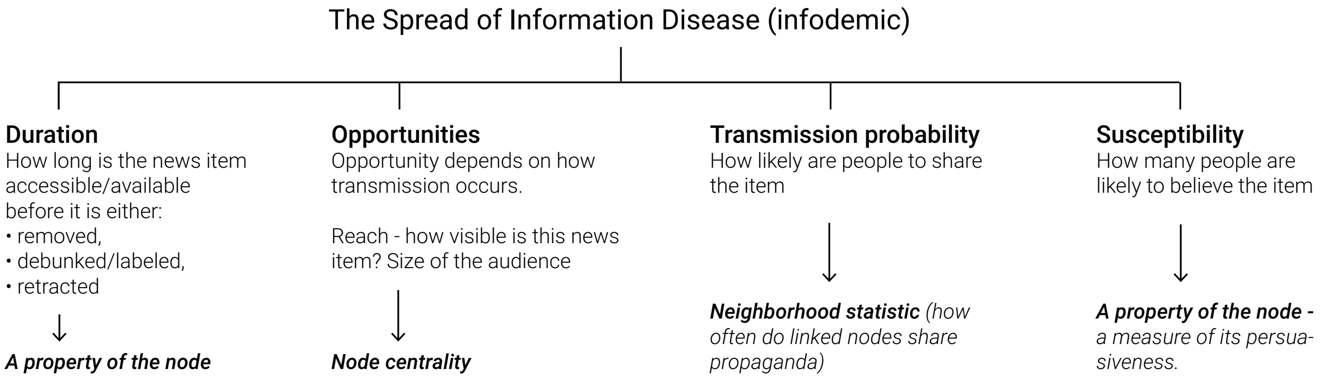

Figure 3: A high-level overview of the *R* number decomposition.

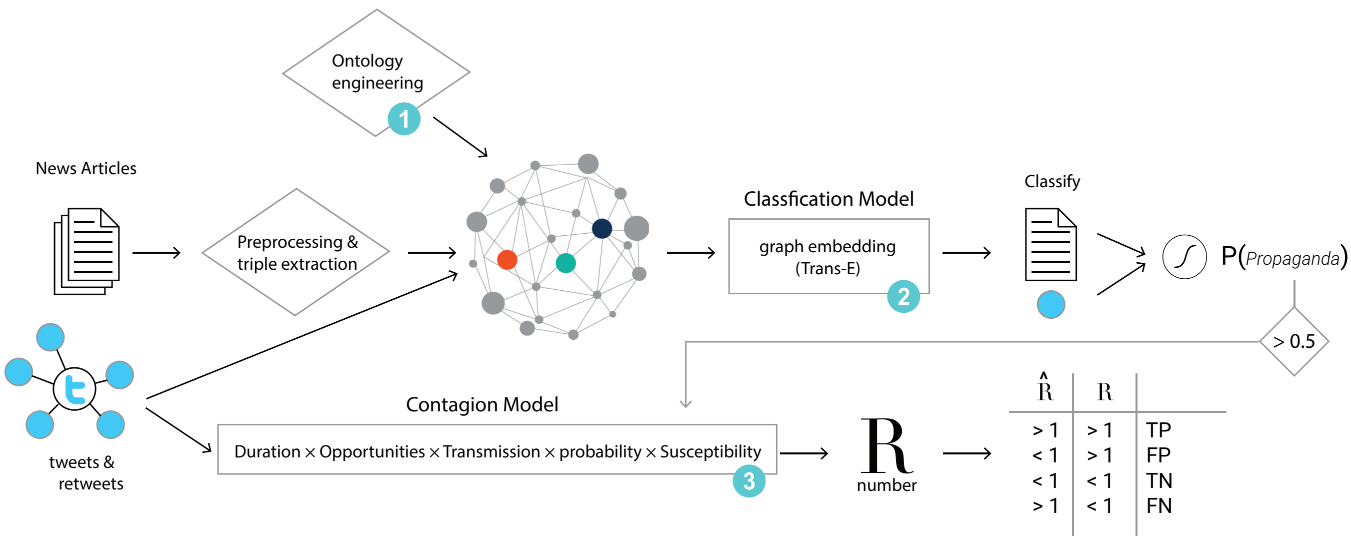

Figure 4: Evaluation stages of the proposed framework

(1) How accurately does the classification model predict propaganda? The ontology will be evaluated using OWL2. The classification model will be evaluated using an F1 score and analysed with a confusion matrix. If the probability of propaganda is greater than a predetermined threshold (0.5 by default), then the contagion model will be executed to determine spread.
(2) How well does the contagion model predict spread? The contagion model evaluation requires more work, but suffice it to say that there is a test set against which the goodness of the model will be measured. An R number below 1 indicates that a piece of propaganda will not spread, while larger R values indicate that it will.

6 CONCLUSIONS AND LIMITATIONS

If this work is successful, there are obvious benefits to society. The ability to model the spread of propaganda and disinformation would help businesses and policy makers implement better long term business models and practices.

There are also some obvious limitations, both exogenous and self imposed. The research is limited to publicly available data. Private social media accounts are not accessible but could potentially hold very valuable information especially in the planning phase of adverse events. Similarly, news articles behind paywalls will not be considered. Furthermore, not all social networks make it possible to obtain their data. While Twitter has at least a limited public API, Facebook does not, as is also the case with other platforms. Currently, this work is focused only on (English) text, while propaganda and disinformation can be spread using images (i.e. *Instagram*), and audio/visual media (i.e. *Youtube*).

7 ACKNOWLEDGEMENTS

Sponsored by Science Foundation Ireland.

REFERENCES

[1] Yochai Benkler, Rob Faris, and Hal Roberts. 2018. *Network propaganda: manipulation, disinformation, and radicalization in American politics.* Oxford University Press.

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

[3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), Vol. 26. Curran Associates, Inc., 2787–2795. https://proceedings.neurips.cc/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf

[4] Elena Cabrio and Serena Villata. 2018. Five Years of Argument Mining: a Data-driven Analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence.* International Joint Conferences on Artificial Intelligence Organization, 5427–5433. https://doi.org/10.24963/ijcai.2018/766

[5] Iván Castillo-Zúñiga, Francisco Javier Luna-Rosas, Laura C. Rodríguez-Martínez, Jaime Muñoz-Arteaga, Jaime Iván López-Veyna, and Mario A. Rodríguez-Díaz. 2020. Internet Data Analysis Methodology for Cyberterrorism Vocabulary Detection, Combining Techniques of Big Data Analytics, NLP and Semantic Web:. *International Journal on Semantic Web and Information Systems* 16, 1 (Jan 2020), 69–86. https://doi.org/10.4018/IJSWIS.2020010104

[6] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M. Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. *PLOS ONE* 10, 6 (Jun 2015), e0128193. https://doi.org/10.1371/journal.pone.0128193

[7] Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-Grained Analysis of Propaganda in News Article. 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).* Association for Computational Linguistics, 5635–5645. https://doi.org/10.18653/v1/D19-1565

[8] Dai Tho Dang, Ngoc Thanh Nguyen, and Dosam Hwang. 2019. Multi-step Consensus: An Effective Approach for Determining Consensus in Large Collectives. *Cybernetics and Systems* 50, 2 (Feb 2019), 208–229. https://doi.org/10.1080/01969722.2019.1565117

[9] F. H. van Eemeren, Bart Garssen, E. C. W. Krabbe, Arnolda Francisca Snoeck Henkemans, Bart Verheij, and Jean Hubert Martin Wagemans. 2014. *Handbook of argumentation theory.* Springer Reference.

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

[11] Lee Howell, World Economic Forum, and Risk Response Network. 2013. *Global risks 2013.* World Economic Forum.

[12] Jooyeon Kim, Behzad Tabibian, Alice Oh, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 2018. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18.* ACM Press, 324–332. https://doi.org/10.1145/3159652.3159734

[13] A. Kucharski. 2020. *The Rules of Contagion: Why Things Spread–And Why They Stop.* Basic Books. https://books.google.ie/books?id=sCDIDwAAQBAJ

[14] Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. [n.d.]. SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles. ([n. d.]), 38.

[15] Jelena Mitrovic, Cliff O'Reilly, Miljana Mladenovic, and Siegfried Handschuh. 2017. Ontological representations of rhetorical figures for argument mining. *Argument & Computation* 8, 3 (2017), 267–287. https://doi.org/10.3233/AAC-170027

[16] Panagiotis Mitzias, Ioannis Kompatsiaris, Efstratios Kontopoulos, James Staite, Tony Day, George Kalpakis, Theodora Tsikrika, Helen Gibson, Stefanos Vrochidis, and Babak Akhgar. 2019. Deploying Semantic Web Technologies for Information Fusion of Terrorism-related Content and Threat Detection on the Web. In *IEEE/WIC/ACM International Conference on Web Intelligence on - WI '19 Companion.* ACM Press, 193–199. https://doi.org/10.1145/3358695.3360896

[17] Garth S. Jowett; Victoria J. O'Donnell. 2018. *Propaganda & Persuasion.* SAGE Publications.

[18] Jeff Z. Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. *Content based fake news detection using Knowledge Graphs.* Vol. 11136. Springer International Publishing, 669–683. https://doi.org/10.1007/978-3-030-00671-6_39

[19] Gordon Pennycook and David G Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (2019), 2521–2526. https://doi.org/10.1073/pnas.1806781116

[20] David Rothkopf. 2003. When the Buzz Bites Back. https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/

[21] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* ACM, 797–806. https://doi.org/10.1145/3132847.3132877

[22] S. Shabani and M. Sokhn. 2018. Hybrid machine-crowd approach for fake news detection. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC).* 299–306. https://doi.org/10.1109/CIC.2018.00048

[23] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. *arXiv preprint arXiv:1809.01286* (2018).

[24] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[25] Kai Shu, Suhang Wang, and Huan Liu. 2017. Exploiting Tri-Relationship for Fake News Detection. *arXiv preprint arXiv:1712.07709* (2017).

[26] Kai Shu, Suhang Wang, and Huan Liu. 2019. 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.* ACM, 312–320. https://doi.org/10.1145/3289600.3290994

[27] Chris Stirewalt. 2021. Op-Ed: I called Arizona for Biden on Fox News. Here's what I learned. https://www.latimes.com/opinion/story/2021-01-28/fox-news-chris-stirewalt-firing-arizona

[28] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake news detection in social networks via crowd signals. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18.* ACM Press, 517–524. https://doi.org/10.1145/3184558.3188722

[29] Nguyen Vo and Kyumin Lee. 2018. The rise of guardians: fact-checking url recommendation to combat fake news. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* ACM, 275–284. https://doi.org/10.1145/3209978.3210037

[30] Douglas Walton. 2016. *Argument Evaluation and Evidence.* Law, Governance and Technology Series, Vol. 23. Springer International Publishing. https://doi.org/10.1007/978-3-319-19626-8

[31] Jin Zhiwei, Juan Cao, Yongdong Zhang, and Jeibo Luo. 2016. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs. In *Thirtieth AAAI Conference on Artificial Intelligence.* AAAI Press. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12128/12049 Technical Papers: NLP and Text Mining.

[32] Xinyi Zhou, Atishay Jain, Vir V. Phoha, and Reza Zafarani. 2020. Fake news early detection: a theory-driven model. *Digital Threats: Research and Practice* 1, 2 (Jul 2020), 1–25. https://doi.org/10.1145/3377478

A PROPAGANDA ECOSYSTEM DIAGRAM - WORK IN PROGRESS

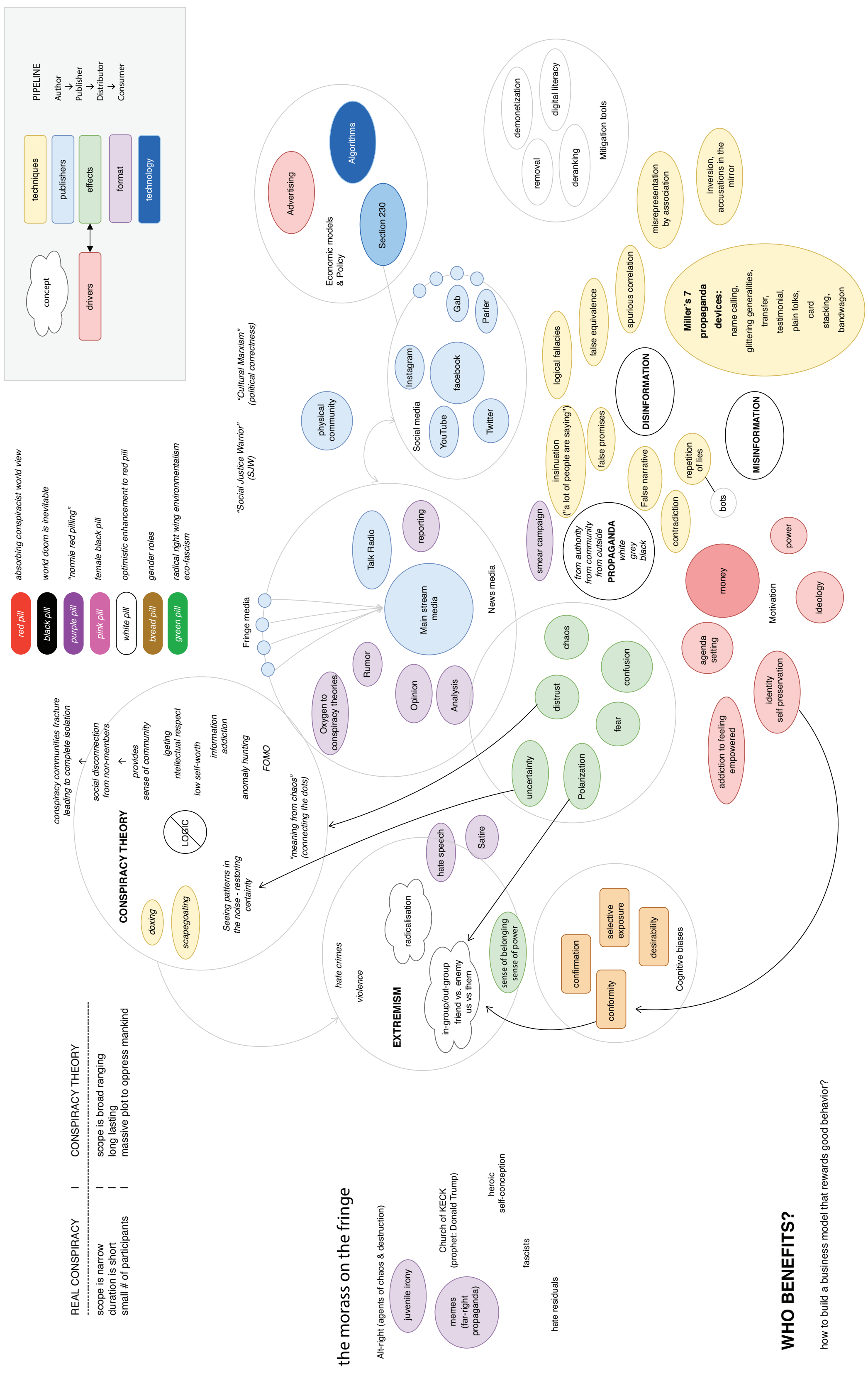

Figure 5: Propaganda and disinformation ecosystem