2022 IEEE International Conference on Big Data (Big Data)

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020661

PROBER: A System for Real-time Propaganda Behavior Analytics on Social Media and Web Data Streams

Yasas Senarath
George Mason University
Fairfax, VA, USA
ywijesu@gmu.edu

Antonios Anastasopoulos
George Mason University
Fairfax, VA, USA
antonis@gmu.edu

Tonya Thornton
Delta Point Solutions, LLC
Port Republic, VA, USA
tonya.e.thornton@gmail.com

Hemant Purohit
George Mason University
Fairfax, VA, USA
hpurohit@gmu.edu

Abstract—**Social media and online platforms provide a public space for many people to share opinions. Social media has numerous benefits to society; however, previous research has identified that individuals use social media for propagandizing purposes which can be detrimental to society, especially during humanitarian crises. Therefore, communities must look into this content to understand and effectively mitigate propaganda, especially when social media messages contain targeted hate or fake/disinformation. In this paper, we propose a human-centered system called *PROBER*, for *pro*paganda *be*havio*r* analytics in social and web data streams, which the relevant authorities, such as government institutions, could use for operational decision support and informing policy analysts for crisis management.**

Index Terms—**propaganda, social-media, natural language processing**

I. Introduction

With the popularity of various social media platforms, many people use them to share their opinion with the world. Organizations use it for advertising their products or services to the people. Further, social media has been used by governments, non-profits, and volunteer groups to coordinate and help the public in a wide array of natural disasters [1]. However, this extent of popularity and the ability to freely reach a broad audience has made social media an opportune platform for people with malicious intent to utilize social media for spreading information with specific agenda, i.e., propaganda [2]. Formally, Merriam-Webster dictionary defines propaganda as the ideas, facts, or allegations spread deliberately to further one's cause or to damage an opposing cause. The spreaders of such content might craft it to manipulate or influence people on a large scale. For example, for COVID-19 pandemic treatments, World Health Organization (WHO) has highlighted how the misinformation, partly spreading through propaganda tactics, damaged the response to the global pandemic[1].

Although propaganda existed before the popularity of social media, it was indeed limited in regional scope. However, now due to the easy accessibility and anonymity of social media, anyone can spread propaganda from anywhere without concern for consequences. Carley (2020) [3] coined the term "social cybersecurity" to collectively identify cyber-mediated threats to democracy. The author identifies propaganda as a threat since it manipulates people's views and disseminates disinformation. It could be a major issue leading to cascading disasters during a humanitarian crisis.

However, analyzing propaganda on social media can be quite challenging. Recently, we have seen a rise in propaganda on social media infodemic related to COVID-19, which has driven the interest of the research community. Moreover, some studies identified propaganda related to events such as the US Presidential campaign in 2016 and Brexit [2]. Such events have motivated the increasing number of research works on combating propaganda in social media.

Many studies on computational propaganda aim to detect or understand propaganda tactics [2], [4]. Specifically, identifying techniques used for persuasion, such as name-calling, loaded language, and repetition [4]. Many research studies focus on the content of a social media message. While some studies considered textual content [5], others also considered images (e.g., memes) [4]. Further, some studies explore social network features for propaganda diffusion [2]. However, only a few studies have investigated the design of an end-to-end system for identifying, analyzing, and visualizing propaganda behavior online. One such tool is *Prta* (Propaganda Persuasion Techniques Analyzer) [6], which analyzes articles collected from online sources. It provides analytical insights by highlighting spans that use propaganda techniques in text.

In this study, we propose a real-time data analytics system called *PROBER* that stands for ***pro***paganda ***be***havio***r*** analytics on social media and web data streams. For specificity, we illustrate the propaganda content on Twitter as a case study in this paper. There are several challenges when creating a system for propaganda analytics in social media streams such as Twitter. First, unlike news articles, the context represented in social media post is limited due to the limited size of the post. Only 256 characters are allowed in a single tweet; therefore, a user cannot include extensive contextual information. In propaganda, a single tweet may not indicate the whole picture; therefore, the analysis should be performed by aggregating many posts during a certain limited period.

[1] https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time

978-1-6654-8045-1/22/$31.00 ©2022 IEEE

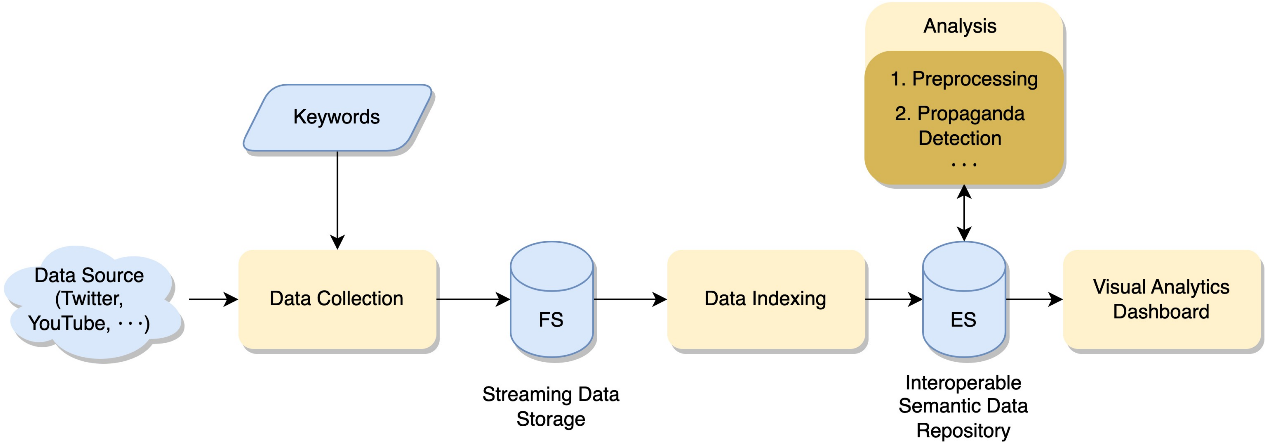


Fig. 1: The high-level architecture of the proposed real-time propaganda analytics system, *PROBER*.

TABLE I: Several example sentences for three propaganda techniques from existing datasets. The acronyms represent the propaganda techniques: RE - Repetition, NC - Name Calling / Labeling, and LL - Loaded Language.

Text	RE	NC	LL
I am tired of sin.	Yes	No	No
We already have to deal with the million unexpected things that come up each day.	No	No	Yes
Disturbing video of young Trump Supporter having his MAGA hat stolen and a drink thrown in his face.	Yes	Yes	Yes
Those, however, who maintain its continued use often see it more as a tool for revenge.	No	Yes	No

We extend the approach proposed in our existing tool, CitizenHelper [7]–[9] that focuses on streaming data analytics of social media data to analyze propaganda behavior. Our contribution in this study is to improve the generic CitizenHelper tool to support the detection and analysis of propaganda in the following ways:

1) We introduce a model capable of detecting propaganda techniques. Table I shows some example sentences for three primary propaganda techniques identified in past research. The Section II-C2 provides more details of the dataset and its labels; and
2) We introduce visualizations that are helpful to study the dynamics of propaganda behavior using the representative set of salient keywords that signify propaganda and a statistical summary of the presence of propaganda tactics in social media.

II. System Design

When designing PROBER, we considered several human-centered design principles as requirements from which the practitioner users (those who utilize the system) can benefit. The core design requirements of the proposed system are:

1) **Simplicity**: The system should provide interfaces that a non-tech savvy user can learn and operate easily [10];
2) **Interactivity**: The system should provide greater human agency and control for adapting AI models behind visual analytics through interactive visualizations [11];
3) **Extensibility**: The system should be interoperable to support the new capabilities of interest for analysis, without much development efforts.

Figure 1 shows the high-level architecture of the proposed real-time system for detecting and analyzing propaganda content on a specific topic. The following subsections will describe the components of this system.

A. Data Collection

Streaming data collection is the first step in our analytics process. We allow the practitioners to set the topic of interest for the data collection manually. We describe the topic of interest by a set of keywords related to the query. For instance, in the case of Twitter, the system uses the *filtered stream* endpoint of Twitter API to collect real-time data on a given topic. We store the collected data as batched JSON [2](JavaScript Object Notation) files (see *FS* in Figure 1).

B. Data Indexing

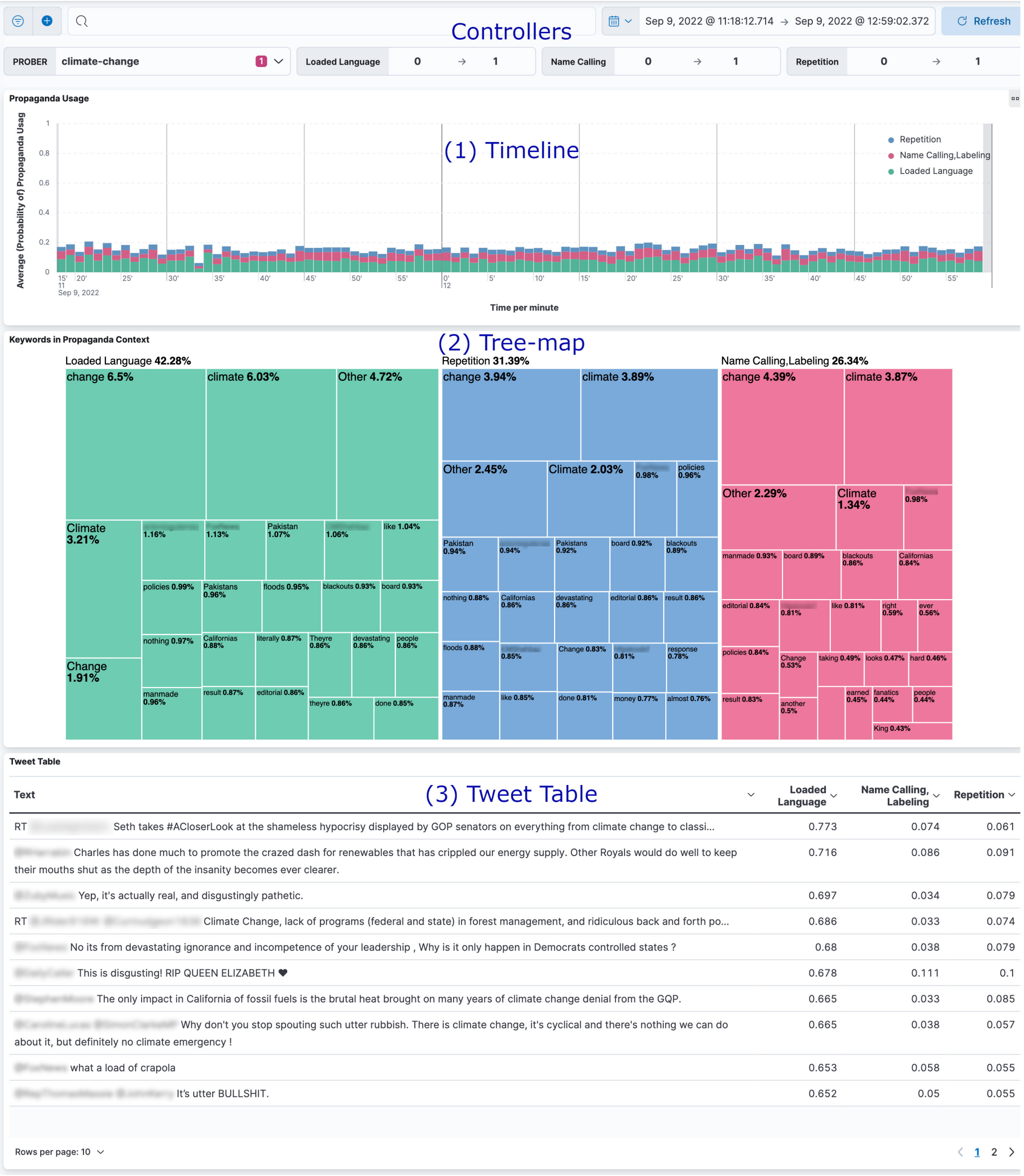
The second process stores the collected data in an index of an Elasticsearch database [3] (see *ES* in Figure 1). Elasticsearch provides faster search capabilities compared to other databases on large number of records, including full-text search support. Additionally, it allows storing JSON objects that can easily extend to incorporate metadata for various analyses as needed. This can enable searching for social media posts with targeted analytics and support decision-making at response agencies.

C. Analysis

1) Pre-processing: We pre-process the collected data to remove the hyperlinks and platform-specific conventions (e.g., RT @USER), and then tokenize the input.

2) Propaganda Detection Model: Here we describe how we created the propaganda detection model.

[2] https://www.json.org/json-en.html
[3] https://www.elastic.co/


Sep 9, 2022 @ 11:18:12.714 → Sep 9, 2022 @ 12:59:02.372

PROBER: climate-change

Loaded Language: 0 → 1

Name Calling: 0 → 1

Repetition: 0 → 1

Text	Loaded Language	Name Calling, Labeling	Repetition
RT [redacted] Seth takes #ACloserLook at the shameless hypocrisy displayed by GOP senators on everything from climate change to classi...	0.773	0.074	0.061
[redacted] Charles has done much to promote the crazed dash for renewables that has crippled our energy supply. Other Royals would do well to keep their mouths shut as the depth of the insanity becomes ever clearer.	0.716	0.086	0.091
[redacted] Yep, it's actually real, and disgustingly pathetic.	0.697	0.034	0.079
RT [redacted] Climate Change, lack of programs (federal and state) in forest management, and ridiculous back and forth po...	0.686	0.033	0.074
[redacted] No its from devastating ignorance and incompetence of your leadership , Why is it only happen in Democrats controlled states ?	0.68	0.038	0.079
[redacted] This is disgusting! RIP QUEEN ELIZABETH ♥	0.678	0.111	0.1
[redacted] The only impact in California of fossil fuels is the brutal heat brought on many years of climate change denial from the GQP.	0.665	0.033	0.085
[redacted] Why don't you stop spouting such utter rubbish. There is climate change, it's cyclical and there's nothing we can do about it, but definitely no climate emergency !	0.665	0.038	0.057
[redacted] what a load of crapola	0.653	0.058	0.055
[redacted] It’s utter BULLSHIT.	0.652	0.05	0.055

Rows per page: 10

Fig. 2: The *PROBER* dashboard for tweets collected with keywords for topic “climate-change”.

TABLE II: Examples where the fine-tuned BERT model detects (✓) or misses (✗) in contrast to the ground truth. The acronyms represent the propaganda techniques: LL - Loaded Language, NC - Name Calling / Labeling, and RE - Repetition.

Text	LL	NC	RE
They have betrayed the British people.	Yes (✗)	No (✓)	No(✗)
Huge victory for the president	No (✗)	No (✓)	Yes (✗)
And what of the training of little jihadis? Nothing!	Yes (✗)	Yes (✗)	Yes (✗)

Dataset: We identified a dataset [4] containing memes with English text from public Facebook accounts and a dataset [5] that has English news articles. The authors of both datasets use similar label schema of propaganda techniques. Since the memes dataset is multimodal and our study primarily focuses on text content, we use the news articles dataset. As news articles are large documents and not representative of short social media posts, we divided the documents in the dataset into separate sentences simulating a setting where a user posts a coherent sequence of tweets (a thread). Next, we identify the presence of each of the following three propaganda techniques: 1) Loaded Language, 2) Name Calling / Labeling, and 3) Repetition. These propaganda techniques had the most number of training examples in the dataset. We split the dataset into training (84%) and testing (16%) sets. The number of training samples for each label is 1837, 924, and 563.

The state-of-the-art techniques for detecting propaganda utilize (pre-trained) language models such as BERT and RoBERTa [12], [13]. Following a similar approach, we fine-tuned BERT [14] model to detect the presence of propaganda techniques in a sentence. We augmented the BERT encoder with a dense output layer of three outputs with sigmoid activation for multi-label classification. We limited the fine-tuning process to five epochs and used binary cross-entropy loss when fine-tuning. Our evaluation revealed that the micro F_1 score of the fine-tuned model is 38%, and the AUC is 61% on a separate test set. Table II provide some examples where the model detects or misses propaganda technique label. We observe that the fine-tuned BERT model cannot capture the repetitions adequately. It could be due to the absence of context at sentence-level prediction.

D. Dashboard

We utilized the open-source tool of Kibana[4] to create a visual analytics dashboard that shows the real-time propaganda analysis. Kibana can directly interface with the Elasticsearch database to obtain and aggregate the data necessary for creating the visualizations in real-time efficiently. We illustrate an analysis of climate change-related tweets in Figure 2. We identify three main visualization methods used to represent the propaganda in the Twitter data stream: 1) *timeline* of the presence of propaganda, 2) a *tree-map* of prominent keywords, 3) *tweet table*. A timeline of the presence of propaganda in tweets provides a general idea about the prevalence of propaganda around the topic of interest for each time period. Secondly, we include a tree-map of frequently used words for each propaganda technique. Finally, the dashboard contains a table of tweets to represent the detailed analysis of the model predictions for each tweet. It helps a practitioner user to use the system to identify specific concerns and try to address those. Moreover, the controllers in the dashboard help the practitioner search for specific tweets by filtering the topic, time, and probability of propaganda techniques.

III. Conclusion

This paper introduced a real-time propaganda analytics system called *PROBER* for social media streams. The proposed system architecture is designed to provide the information proactively on potential, emerging threats to help decision makers at crisis management agencies. Moreover, we train a sentence-level propaganda detection model for detecting propaganda in social media content and report preliminary results. Finally, we created a visual analytics dashboard for summarizing the real-time propaganda on a provided topic.

However, there are several limitations to our approach. We need to consider the context (e.g., historical related posts) in our current modeling process when identifying propaganda in a social media post. Moreover, our system currently does not support detecting propaganda in multimodal content or multilingual settings.

Future Work: We plan to improve the propaganda detection model by in-cooperating a history of related messages and external knowledge (e.g., domain-specific and linguistic knowledge bases). Moreover, in the future, we aim to in-cooperate multimodal and multilingual models to support more modalities and languages as well as test our system for different social media platforms.

Acknowledgment

Authors acknowledge the partial support from the grant H-4Q21-009 from the Commonwealth Cyber Initiative, an investment in the advancement of cyber R&D, innovation, and workforce development (for more information about CCI, visit www.cyberinitiative.org).

References

[1] H. Purohit and S. Peterson, "Social media mining for disaster management and community resilience," in *Big Data in Emergency Management: Exploitation Techniques for Social and Mobile Data*. Springer, 2020, pp. 93–107.

[2] G. Da San Martino, S. Cresci, A. Barrón-Cedeño, S. Yu, R. Di Pietro, and P. Nakov, "A survey on computational propaganda detection," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 4826–4832.

[3] K. M. Carley, "Social cybersecurity: an emerging science," *Computational and mathematical organization theory*, vol. 26, no. 4, pp. 365–381, 2020.

[4] D. Dimitrov, B. B. Ali, S. Shaar, F. Alam, F. Silvestri, H. Firooz, P. Nakov, and G. Da San Martino, "Semeval-2021 task 6: Detection of persuasion techniques in texts and images," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, 2021, pp. 70–98.

[5] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, and P. Nakov, "Semeval-2020 task 11: Detection of propaganda techniques in news articles," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 2020, pp. 1377–1414.

[6] G. Da San Martino, S. Shaar, Y. Zhang, S. Yu, A. Barrón-Cedeno, and P. Nakov, "Prta: A system to support the analysis of propaganda techniques in the news," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020, pp. 287–293.

[7] Y. Senarath, R. Pandey, S. Peterson, and H. Purohit, *Citizen-Helper System for Human-Centered AI Use in Disaster Management*. Singapore: Springer Nature Singapore, 2022.

[8] P. Karuna, M. Rana, and H. Purohit, "Citizenhelper: A streaming analytics system to mine citizen and web data for humanitarian organizations," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[9] R. Pandey and H. Purohit, "Citizenhelper-adaptive: Expert-augmented streaming analytics system for emergency services and humanitarian organizations," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 630–633.

[10] K. Karvonen, "The beauty of simplicity," in *Proceedings on the 2000 conference on Universal Usability*, 2000, pp. 85–90.

[11] B. Shneiderman, "Design lessons from ai's two grand goals: Human emulation and useful applications," *IEEE Transactions on Technology and Society*, vol. 1, no. 2, pp. 73–82, 2020.

[12] J. Tian, M. Gui, C. Li, M. Yan, and W. Xiao, "Mind at semeval-2021 task 6: Propaganda detection using transfer learning and multimodal fusion," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, 2021, pp. 1082–1087.

[13] K. Gupta, D. Gautam, and R. Mamidi, "Volta at SemEval-2021 task 6: Towards detecting persuasive texts and images using textual and multimodal ensemble," in *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 1075–1081. [Online]. Available: https://aclanthology.org/2021.semeval-1.149

[14] J. Devlin, M.-W. Chang, K. Lee, and K. 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, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

[4] https://www.elastic.co/kibana/