Received April 11, 2020, accepted May 1, 2020, date of publication May 14, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994658

Socio-Technical Mitigation Effort to Combat Cyber Propaganda: A Systematic Literature Mapping

AIMI NADRAH MASERI[1], AZAH ANIR NORMAN[1], CHRISTOPHER IFEANYI EKE[1,2], ATIF AHMAD[3], AND NURUL NUHA ABDUL MOLOK[4]

[1]Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia
[2]Department of Computer Science, Faculty of Science, Federal University, Lafia 234830, Nigeria
[3]School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC 3010, Australia
[4]Department of Information Systems, Faculty of Information and Communication Technology, International Islamic University Malaysia, Selangor 53100, Malaysia

Corresponding author: Azah Anir Norman (azahnorman@um.edu.my)

This work was supported by the University of Malaya Postgraduate Research Grant (PPP), under Project PG124-2015B.

ABSTRACT This systematic mapping literature aims to identify current research and directions for future studies in terms of combating cyber propaganda in the social media, which is used by both human effort and technological approaches (socio-technical) for mitigation. Out of 5176 retrieved articles, only 98 of them were selected for primary studies; classified based on research artifacts, mitigation effort, and the social media platforms involved in the research. The search was conducted using selected databases and applying selection criteria set for this research. Through the analysis, important research trends were identified based on human effort and technological approaches in mitigating and combating the cyber-propaganda issues. The authors also identified various mitigation socio-technical approaches such as identification, detection, image recognition, prediction, truth discovery and comprehension of rumours flow. The study also highlights areas for further improvements, to complement the performances of existing techniques. Besides, the study provides a brief review of cyber propaganda detection using classification techniques. Hence, it has set forth applicable research focus on the areas dealing with the mitigation of risk borne by cyber propaganda in the social media.

INDEX TERMS Systematic mapping review, social media, cyber propaganda, mitigation, human effort, socio-technical approaches.

I. INTRODUCTION

Social media sites such as Facebook and Twitter are known as powerful communication tools that have fundamentally changed the way information is shared and as well as how human relationships are developed. By leveraging internet-working technologies, social media has enabled much of the world's population to remain connected, hence overcoming the traditional communication barriers of time and distance. The advancement in social media technology such as the newsfeed algorithm has helped in the propagation of messages in a quicker engagement. At its worse, this cycle can transform the social media into a kind of confirmation bias machine, one perfectly tailored for the spread of propaganda. Therefore, existence of social media has sparked concerns about the adverse consequences of propaganda or other forms of false information in global societies. This is because, the social media is increasingly being held responsible for the content on their sites, as the world tries to grapple events in real-time as they unfold [1].

There are many terms used by scholars, politicians, and journalists to represent the accuracy of the information and the intentions of spreading information in the media, some of them include propaganda, misinformation, fake news, and rumours. However, there is no generally agreed definition for this information manipulation [2], [3] and these terms are also considered as synonyms [4]. However, in this study, we define the term 'cyber propaganda' as information that is created in a controversial way to shape public opinion, by suggesting something negative that will manipulate public opinion in the social media [3], [5]. Propaganda has gained significant attention in the US presidential election of 2016. During the election, most of the fake news stories mentioned tended to favour Donald Trump over Hilary Clinton, and many people

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

thought that Donald Trump would not have been elected as the president [6]. This propaganda did not only happen in the US that year but it also affected politics in nine other countries within the year 2016 [7]. To date, propaganda has begun to rise exponentially and the global effect has risen by 150% for the past two years according to the Computational Propaganda Research Project (COMPROP) [6]–[8]. The social bots that reside in social media are also believed to be one of the major factors that contributed to the propaganda dispersion in 2017, where approximately 23 million social bots were found in Twitter accounts [9]. The spread of propaganda has not only brought about disruption in global politics but has also caused cyber-hate, riots, threats, and even numerous instances of violence. For instance, the rise of disturbing online anti-Muslim propaganda has caused offline incidents such as Muslim women being targeted for wearing the headscarf, and Muslims in general, facing an attack on the streets, and various mosques being vandalized [10], [11]. The root of this matter is believed to have been started by the behaviour of humans and the intricate technology found in the social media. To give an instance, humans are more likely to respond to contents that tap into existing grievances and beliefs, thus inflammatory messages will generate fast engagement. Meanwhile, the intricate technology in the social media plays an important role in dispersing propaganda; considering the rise of social bots', newsfeed algorithms in the social media, and the platform functionality that makes social engagement more rapid.

Human behaviour and the technology of the social media factors (also known as socio-technical) are inseparable. Thus, there is a need to understand how they influence each other and co-evolve [12]. It is crucial to incorporate a holistic view of factors in mitigating propaganda, this includes socio-technical approaches. Currently, there are different ways of combating propaganda. For instance, a project conducted by the University of Oxford developed various methods in mitigating propaganda through algorithms, tools' data memos, methodologies and others as far back as 2016 [5]. Not only has this caught researchers' attention in providing solutions to this problem, Facebook and Google have also taken their initiative in combating this issue. After the issue of fake news propagation during the 2016 US presidential elections, Facebook administrators began the labelling and warning of inaccurate news by implementing a flag button. To further strengthen it, Google administrators in March 2018 launched Google News Initiative (GNI), to fight against the spread of false information [13]. Cyber information propagation behaviour and advanced social media power in world history and public memory, has come to the forefront in the recent years; especially with the increase in multi-functional social networking tools like Facebook's Likealyzer and Twitter's Tweetdeck. The question of how cyber information propagation behaviour and social media power interacts and associates, however, has been poorly addressed. Some scholars on social media governance argued that research in the domain should focus on the dynamics of behaviours such as free speech [14]. Yet, scholars on Internet technology have noted that internet technology narratives often praise the intricate technology and its method to be the core influence [15].

To date, however, little work has thoroughly examined the above contributions as the core attribute to the aggressive cyber-propaganda in the current social media. As a result of ttthis, scholars endorse a narrative that characterizes behaviour and technology as separate elements, making both conditions separable and non-associated [16], while in reality, it is not [17]. Mitigating propaganda in socio-technical approaches, however, has been proposed and investigated by many researchers and experts to address the spread of false information that has led to numerous global threats. The major contributions of this paper are as follows:

- A critical review of the current state of literature on the mitigation of cyber propaganda in the social media.
- A detailed review of the classification approach carried out on some of the selected studies for cyber propaganda detection (section III)
- A literature mapping review on how to combat cyber propaganda (Figure 8).

The remainder of this paper is arranged as follows: In section II, the research motivation, questions, and objectives are presented. Section III provides a review of cyber propaganda detection using classification techniques. In section IV, systematic mapping is carried out. Section V presents the data extraction for the study. In section VI, the results of the mapping are presented. Section VII provides the discussion while section VIII finally brings the paper into its conclusion.

II. MOTIVATION, RESEARCH QUESTIONS, AND OBJECTIVES

There has been an increasing research interest on methods for combating propaganda, and this interest persists despite the reality that these two areas are not yet well established and there are no systematic reviews available for further research. Systematic Literature Reviews (SLRs), are employed under circumstances when there is insufficient empirical evidence available, or when the topic area is too vast for an SLR to be conducted. [18]. Thus, systematic mapping was conducted because systematic mapping would give an extensive, wide and detailed overview of the research area where it would map out and categorize existing literature; which would fit the objectives of this study. For conducting the review, the guidelines followed were described by Petersen [18] in his paper on conducting systematic mapping. This paper aims at presenting not only an overview but also an understanding of how mitigation effort contributes to different areas.

A. RESEARCH QUESTIONS AND OBJECTIVES

Table 1 below shows the research objectives and questions for this research.

TABLE 1. Research questions and objectives.

S/N	Research Question	Research Objectives
1	What is the most frequent social media platform, mentioned in cyber propaganda studies?	To identify the most frequent social media platforms used in cyber propaganda activities
2	What are the existing socio-technical approaches that contribute to mitigating cyber propaganda?	To identify the existing socio-technical approach that contributes to mitigating cyber propaganda
3	What are the socio-technical compositions, especially on the effort that has been developed to mitigate propaganda?	To study the socio-technical compositions, especially on the effort that has been developed to mitigate propaganda.

III. REVIEW OF CYBER PROPAGANDA DETECTION USING CLASSIFICATION TECHNIQUES

This section provides a review of the classification techniques for cyber propaganda detection on the selected studies under the aspect of datasets, feature sets, modelling, and performance metrics. The review process in this section adopted a similar approach used in [19]. Out of the 98 selected studies, 30 of them employed classification techniques for detection. This section is divided into four different sub-sections. In subsection A, a review of the datasets is provided. Subsection B gives the review of the feature used for cyber propaganda classification whereas subsection C provides a review of the modelling aspect. Finally, subsection D reviews the performance metric employed for cyber-propaganda classification. The summary of the review is presented in Table 2.

A. REVIEW OF DATASET FOR CYBER PROPAGANDA DETECTION

In any data mining task, datasets and its collection are very crucial. The dataset serves as the starting point requirement because feature extraction is carried out on the dataset. Through the dataset, some discriminative features are extracted, which serve as an input to the machine learning algorithm. A literature survey on this domain shows that most datasets are sourced by the authors. The analysis of the selected studies shows that the datasets for cyber propaganda detection can be broadly classified into homogeneous and heterogeneous data. In homogeneous data, the study uses only a single type of dataset. Most of the studies that utilized this type of dataset have obtained the data from Twitter. For example, Conti *et al.* [20] in their study on rumour detection using different window time, employed Twitter datasets. The authors made use of longitudinal and near-complete data that have content of all the Twitter public user communication that has been collected over $3^1/_2$ years. With the dataset, the past rumour propagation instance investigation was made possible. Similarly, Yang *et al.* [21], on the attempt on automatic detection of rumour, employed Sina Weibo datasets. In their research, they shifted their attention from Twitter data and investigated the credibility of information on Sina Weibo. According to literature, Sina Weibo has over 374.1 million registered users [21] that generates over 100 million microblogs daily.

On the other hand, heterogeneous data is said to have been utilized when two or more datasets for cyber propaganda detection are used to enhance the predictive performance of the model. For instance, Jin *et al.* [22] in their study on rumour detection on microblog collected dataset from Weibo and Twitter. The Weibo dataset was crawled on all the posts for false rumour between May 2012 and January 2016 from the official debunking system for a rumour. For Twitter data, the study adopted the MediaEval data Boididou *et al.* [23] which is meant for the detection of multimedia content on the social media. The dataset consists of the Training set (that comprised 6,000 non-rumour and 9,000 rumour tweet obtained from 17 rumour connected events), and a test set that comprised 200 tweets obtained from 35 rumour connected events. The review of the selected studies revealed that most studies employed homogeneous data for detection experiments. The summary of the dataset used is shown in Table 2.

B. REVIEW OF THE FEATURE EXTRACTION FOR CYBER PROPAGANDA DETECTION

Features serve as a key factor in constructing the classification algorithm to obtain the performance of the upper bound cyber propaganda detection. Thus, feature extraction is a crucial step in any classification task. A varying range of features for automatic identification of rumour has been presented in the selected studies. The feature analysis used in the studies can be broadly categorized into three different classes of features - social content, image, textual and propagation-based feature [22], [24]. The textual feature denotes the semantic and statistical properties found in a tweet text. The statistical feature helps in capturing the important tweets statistics such as frequency of the punctuation marks, the word count and the count of the capitalized characters in the tweets [25]. On the other hand, the semantic feature denotes the abstract representation of the semantics of text such as opinionated words and sentiment scores. In some of the selected studies [26], [27], the bag-of-words feature extraction technique has been used to capture the textual features that show the relation existing between tweets. Besides, topic model related feature extraction techniques such as LDA (Latent Dirichlet allocation) are also useful for the representation of abstract semantic features for rumour detection tasks [28], [29]. Textual features are extracted manually and as a result, have some limitations such as context extraction and not being able to take account of word order during the extraction. To address the limitation, Ma *et al.* [30] employed a deep neural network (RNN) to effectively represent tweets.

Social context features are obtained from the social network features found in the microblog. This set of features is most useful in rumour detection in the social media [29]. The features are constructed to obtain the interaction that exists in social media such as user replies, re-tweets, user mention, hashtag (#) and URLs. For instance, Ratkiewicz *et al.* [31], in their study, employed social context features such as a

TABLE 2. Summary of cyber propaganda classification techniques.

Study	Features	Modelling	Performance measure
[30]	Content-based, user-based and propagation-based features	SVM, DT, RF	Acc, REC, PR, F-M
[39]	Stylometric, time-based, sentiment-based feature	SVM, NB, Adaboost	ACC
[38]	Data independent feature, Data-dependent feature	Adaboost	ACC, REC, PR
[40]	Time (hour of the day), links, message source, message language, hashtags, mentions	Not mentioned	ACC, REC, PR, F-M
[41]	Number of followers and followees, URLs, spam word, replies, Hashtags	NB, DT, Clustering	ACC, True positive, False positive
[42]	Content feature, propagation feature, user feature	LR	ACC
[43]	Source credibility, source identity, source diversity, source location and witness, message believe, event propagation	SVM, DT, RF	ACC
[44]	Topic feature, sentiment feature	LDA, Taxamura's semantic orientation dictionary	ACC, REC, PR, F-M
[45]	The author based, registration age, number of followers, number of positive tweets, friend follower ratio, content-based, number of friends, subjectivity score, user mentions, first pronoun, sentiment score, URLs	Linear SVM, DT, RF, and RF-ext	ACC, REC, PR, F-M
[46]	Number of followers, number of friends, number of followers/ number of friends, number of tweets, number of favourites, number of listed, number of active days of the account.	SVM	ACC, REC, PR, F-M
[47]	Content-based, network-based, hashtag, tokens from hashtags, expected hitting time, harmonic closeness	Adaboost	PR, sensitivity, specificity
[48]	Message-based, user-based, propagation-based	SVM	REC, AUC, PR, F-Rate
[49]	Parts of speech, feature phrase	NB, DT, RF, and K-NN	REC, PR, F-M
[50]	Author retweet, URL content, Media content, average number of retweets replies and favourite account, replies to rumour tweet or other tweets, number of favourite counts, followers count and status count of the author, existence of hashtag on rumour tweet replies	CNN	ACC, F-M
[51]	Content related features	K-means	Similarity measure
[52]	Factive verb, assertive verb, mitigating words, report verbs, discourse makers, subjective/bias	RF	ACC, PR, REC
[53]	Textual feature, temporal feature, user feature	SVM, DT, LSTM-1, GRU-2	ACC, F-M
[54]	User embedding features	GRU-based RNN	ACC, PR, REC, F-M
[22]	Number of exclamation/question mark, Number of words/ characters, Number of positive/negative word, Number of first/ second/ third order of pronoun, Number of URLs, @, #, sentiment score, contains happyemo/Sademo, Number of uppercase characters, Number of retweets	RNN	ACC, PR, REC, F-M
[34]	Text content feature, user content, propagation	SVM, LR, Kstar, RF	ACC, PR, REC, F-M
[55]	Periodic feature, tweet features, user features	K-means, LR, MNB, GBT	ACC, PR, REC, F-M
[56]	Retweeter related feature, URLs, Hashtag, mention based feature	XGBoost	ACC, PR, REC, F-M, AUC
[35]	User-related feature, linguistic feature, Network feature, temporal feature	Not mentioned	F-M
[57]	Text feature	K-means, ME, SVM, NB, BP neural network	ACC, AUC
[20]	High-level properties of propagation feature, topological properties, evolution properties	LD, RF, MLP	ACC, PR, REC, F-M, AUC
[58]	N-gram, by using n=1,2,3,4,5	NB	PR, REC, F-M
[59]	Textual feature, visual feature	Att-RNN	ACC, PR, REC, F-M
[60]	Textual feature, image feature	CNN, RNN	ACC, PR, REC, F-M
[61]	Editorial articles features	KNN, NB, SVM	ACC
[62]	Sentence char, parts of speech, emotion, extreme words, sentiment, novelty, pseudo-feedback	SVM	ACC

** SVM= support vector machine, RF= random forest, NB= Naïve Bayes, KNN= K- nearest neighbours, DT= Decision tree, LR= logistic regression, GBT= gradient boosted trees, MNB= Multinomial Naïve Bayes, LD= Linear discriminant, MLP= multi-layer perception, CNN= convolutional neural network, RNN= recurrent neural networks, att-RNN= attention recurrent network, GRU= Gated Recurrent Unit, RBF= radial basis function, LSTM= long short term memory, BP= back propagation, ACC= accuracy, **PR**=Precision, **REC**=recall, **F-M**=F-measure, **AUC**=Area under the curve, TP= true positive, FP= false positive.

mention, hashtag, and links to create a "Truthy system" for the identification of deceptive political memes using Twitter data. Image feature has also been employed as a feature for rumour detection. The verification of the reliability of multimedia content, besides the textual content, has been carried out by a few studies. For instance, Morris *et al.* [32], published a survey outcome that indicated the reliability of the user profile image on the user's post. Similarly, Boididou *et al.* [33], in a study on rumour detection, carried out an automatic prediction on the reality of tweet that shares multimedia contents. In the verification, the study employed the "Multimedia Use task", which was a part of the MediaEval benchmark between 2015 and 2016. However, the combination of image and text forensic features were captured as a feature baseline for the task. In a related study, Jin *et al.* [34], investigated various image features of the tweet based on the image statistics and visual appearance. However, the fusion of these novel images and text features indicated effectiveness in rumour detection. The summary of the feature extraction is depicted in Table 2.

C. REVIEW OF THE CLASSIFICATION TECHNIQUES FOR CYBER PROPAGANDA DETECTION

In any classification task, features extracted from the dataset are employed as an input to the classifiers. In such cases, the constructed model can classify the labelled data as rumour or not rumour. According to the findings of selected studies, various classification algorithms have been applied for rumour detection in social media. Moreover, some studies employed more than one classifier for performance comparison of individual classifiers on their proposed technique. The review of the studies shows that different researchers on rumour detection employed different datasets. As a result, the comparison of different classifier performance during the classification phase in such a case becomes problematic. For instance, Ma *et al.* [30], in their study on a novel approach for automatic identification of rumour in Microblog website, utilized Twitter and Sina Weibo dataset. They created a model called DSTs (Dynamic Series-Time Structure) that investigated different social context features, which include the content-based feature, propagation-based, and user-based feature. They trained the model with various classification algorithms found in Weka such as SVM, DT, RF, RF-ext, and SVM-RBF. The experimental results show enhanced performance in both Twitter and Weibo data on rumour detection concerning the early period of diffusion and full events lifecycle. However, the comparison of the results becomes a problem due to differences in the datasets.

Kwon *et al.* [35] in their study on ''Rumour detection over varying Times window'', utilized Random Forest classifier to test and compare the predictive performance of the selected features (which include user, network, linguistic and temporal domain) to obtain the power of individual features. However, a three-fold cross-validation experiment was performed on each selected feature and the iteration was carried out 10 times due to the insufficiency of the feature (a total of 111 humour and non-humour) in drawing a conclusion on the performance results. Thus, only the feature that occurred above 8 times was selected as a discriminating feature for humour classification. In another study, Jin *et al.* [22] experimented with a deep neural network with a multimodal feature to investigate the presence of rumour in the tweet data. The study employed RNN (Recurrent neural network) to obtain the intrinsic correlation that exists in textual, visual, and social context feature on the instance of a tweet. The experiment was carried out on two separate datasets (Twitter and Weibo). However, the predictive performance shows that using Weibo data with an att-RNN model, enhanced the predictive performance of rumour detection from 65% to 78.8% by using a single modality approach. Hence, it performed better than the feature fusion approach by 12%. By using Twitter dataset, on the other hand, the model boosted from 59.6% to 68.2% when compared with the best result obtained in the single modality method (visual feature). In addition, a better performance was also obtained than the feature fusion modelling by over 6%. The summary of the modelling approaches is shown in Table 2.

D. REVIEW OF THE PERFORMANCE MEASURE FOR CYBER PROPAGANDA DETECTION

In cyber propaganda classification, the evaluation of predictive performance can be determined by employing different performance metrics such as similarity measure, accuracy (ACC), Sensitivity, Specificity, recall (REC), F1-Score (F-S), Precision (PR), and Area under the curve (AUC). The computation of these metrics can be carried out by employing values of false positive (FP), true positive (TP), false negative (FN), and true negative (TN), which are the contents of the confusion matrix. However, the choice of choosing the performance metrics depends on the purpose of detecting cyber propaganda. Even though the review studies show that the most useful performance metrics are F-measure, accuracy, precision, and recall, yet these metrics may not be enough to accurately measure the predictive performance of the model in all cases. This is due to the imbalance class that exists on different datasets in most selected studies. Thus, AUC might be the right option in such an instance because it is suitable for individual class performance evaluation [36], [37].

For instance, Kaati *et al.* [38] in their study ''Detecting Multipliers of Jihadism on Twitter '' collected two different sets of data that involved in media Mujahideen and Jihadist propaganda. The Mujahideen consists of 835 English tweeps and 337 Arabic tweeps. On the other hand, the jihadist data consists of 87,753 English and 61,013 Arabic sets of jihadist propaganda tweets. The set of these data has been employed in classification. The study has utilized precision, recall, and accuracy as the performance metrics to measure the predictive performance of the classification. It is obvious that the two sets of data are naturally imbalanced and in such an instance, there may be a bias in relying only on those performance metrics. Therefore, the correct measure to accurately measure the predictive performance of the Multiplier of jihadism detection is AUC. The reason is that AUC has a strong resistance to the skewness in datasets. Table 2 shows a summary of the performance measures employed.

IV. SYSTEMATIC MAPPING

Following the guideline by Petersen *et al.* [18], this study adopted Systematic Mapping to categorize and summarize the existing information technology platforms that contribute to disseminating propaganda in an unbiased manner. The objective of this paper is to structure the existing literature of the field of socio-technical mitigation efforts on cyber propaganda. The systematic mapping process includes defining research questions, searching for relevant articles, paper screening, keywording using abstracts, keywords, and titles, and lastly data extraction and mapping out the information Figure 1. Based on Figure 1, each step has an outcome. The outcome of the process is the systematic map.

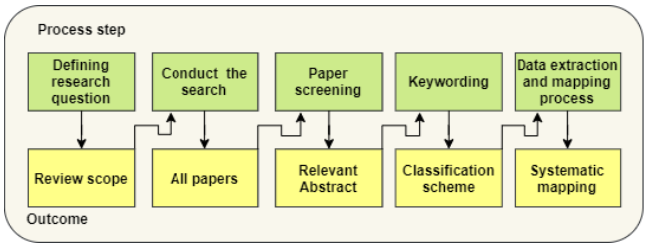


FIGURE 1. Systematic mapping process.

A. ARTICLE SCREENING/SEARCH STRATEGY

The search strategy on publications to answer the above research questions employed four standard digital libraries namely IEEE Explore, ACM Digital Library, Elsevier Science Direct and ISI Web of Knowledge. These databases were selected, based on experience as suggested by Petersen [18] and they were known for either empirical studies or literature surveys for systematic reviews [63].

B. KEYWORD RANGE

For each digital library, query strings were created for the search tool by employing Population, Intervention, Comparison, Outcomes (PICO) keyword search method proposed by Petersen *et al.* [18] and Kitchenham *et al.* [63]; to identify keywords and formulate search strings from the research questions. This method has been chosen due to its appropriate method in attaining the specified research objectives of this study.

- Population: It refers to a particular group that the research intends to investigate. In our context, the population is the number of social media users that are influenced by the propaganda. The used keywords are ''social media'', ''social networking'', ''social network'' and ''social site''.
- Intervention: Following the guideline by Petersen *et al.* [18], intervention refers to a methodology, tool, technology, or procedure. This research does not have a specific intervention to be investigated.
- Comparison: Our systematic mapping study compares several mitigation efforts of cyber propaganda.

Outcomes: The collected publications must represent a wide coverage of areas within the field of cyber propaganda. This will ensure the validity and objectivity of this systematic mapping study. The identified keywords were grouped into sets and their synonyms were considered to formulate different strings. We also used a Boolean operator ''OR'' to join alternative phrases and synonyms in each component (i.e. population comparison and outcomes) and also ''AND'' operator to join the terms respectively from the three components.

C. STUDY SELECTION AND CRITERIA

We have excluded articles that have been retrieved from the above databases, based on inclusion and exclusion criteria, as well as full-text reading and quality assessment.

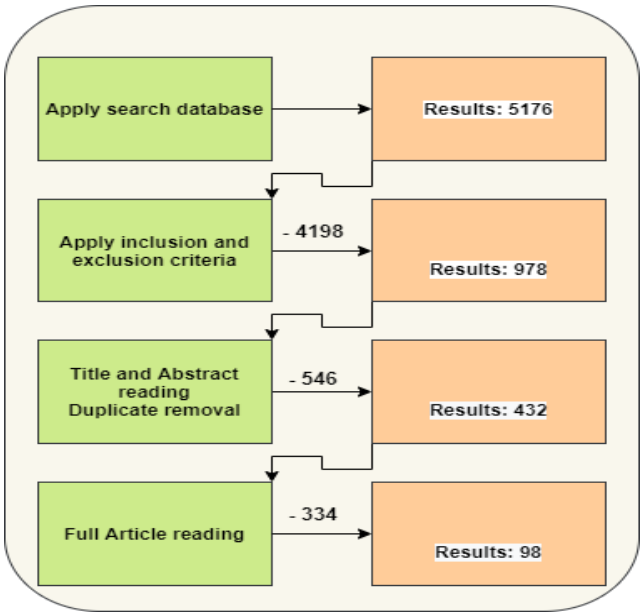


FIGURE 2. Process of article selection.

1) INCLUSION CRITERIA

- Field of study: Social media and propaganda.
- Articles access: Articles that are open and full-text access only.
- Year of Publication: The papers must have been published between 2014 and 2018. In particular, these five years was selected because it focuses only on the present problems and trends in social media for the context of propaganda.
- The relevance of the study: This study will contribute to the body of knowledge on mitigating cyber-propaganda. The contribution can be in the form of practices, techniques, algorithms, framework, or development of models.

2) EXCLUSION CRITERIA

- Technical reports and documents in the form of either abstracts or presentations (i.e., elements of ''grey'' literature) are not considered
- Primary studies that are not available in an electronic format are not included.
- Papers that are not written in the English language are excluded.

V. DATA EXTRACTION

To extract the data required to answer the study questions, all selected papers were read. From each selected research, the data items are described in Table 3, and the extracted data were then tabulated in a spreadsheet. Before data extraction, the definitions of the data items have also been extracted and discussed to clarify the meanings of the data items. Once the data from the studies were recorded, an analysis was carried out. Thus, the contributions, mitigation efforts, and social

TABLE 3. The first stage of the retrieved article.

Database	Search	Search result
IEEE Explore	(“social media” OR “social networking”	225
ACM Digital Library	OR “social network” OR “social site”)	217
Elsevier Science Direct	AND (“propaganda” OR "fake news"	4106
ISI Web of Knowledge	OR "rumours" OR "misinformation")	628

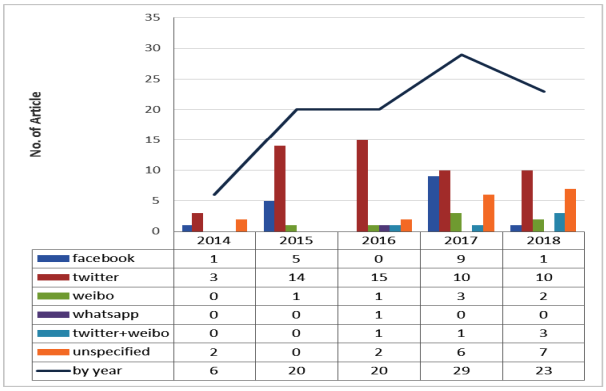
media platforms were identified according to the formulated objectives and research questions.

VI. RESULTS OF THE MAPPING

The articles published until 2018, were searched and a total number of 5176 articles were found from the selected databases. Out of these articles, 978 articles were selected after applying the inclusion and exclusion criteria. 432 articles were chosen after the reading of each article title and abstract, and the removal of duplicate copies. Finally, a total of 98 articles were selected after the full article reading process. Thus, a total of 98 articles in the areas of communication, politics, computer science, health, and marketing were selected (with a predominance of periodicals from the computing area). However, it is noteworthy to state that most of the articles were from developed nations such as America, China, and Italy. Besides, other words found that hold the same meaning as a rumour, fake news, misinformation and propaganda are ‘data incest’ and ‘online water army’. At this point, it is interesting to note that most of the articles discussed politics, disaster, war, attacks, customer reviews, and health as a topic in misinformation studies. Additionally, there are quite a notable amount of papers that mentioned the technology of social bots that has become a major tool of diffusion of propaganda in the social media. Most of the mitigation efforts use data mining, machine learning and sentiment analysis as an approach to mitigate cyber propaganda. When this study started, the researcher discovered as many articles up to 5176 dealing with social media and false information were identified in the first stage of the search. However, not all of those articles had a misinformed main topic. We found out 432 articles talked about misinformation as a problem on the social media. Nevertheless, most of these articles did not focus on mitigating misinformation in the first place. For instance, some papers improved on previous rumour models and used it to improve the vaccination issue, customer review, language barrier and so on. When the final selection filter was applied, this resulted in a high number of papers, which indeed, is a serious topic to be mitigated.

VII. DISCUSSION

The extensive systematic mapping review of the research articles on mitigation efforts on combating cyber propaganda in social media published between 2014 and 2018 has been investigated in this study. The study is associated with propaganda, misinformation, fake news, and rumours. This study



	2014	2015	2016	2017	2018
facebook	1	5	0	9	1
twitter	3	14	15	10	10
weibo	0	1	1	3	2
whatsapp	0	0	1	0	0
twitter+weibo	0	0	1	1	3
unspecified	2	0	2	6	7
by year	6	20	20	29	23

FIGURE 3. Annual trend of cyber-propaganda and social media.

aims at identifying current research and directions for future studies in terms of combating cyber propaganda in the social media by using both human and technological efforts for mitigation.

The first research question *“What is the most frequent social media platform mentioned in cyber propaganda studies?”*

Figure 3 above shows the annual trend of cyber-propaganda research papers from 2014 to 2018. This figure illustrates the recent but explosive emergence of cyber-propaganda research and the social media platform frequently used in investigating the dispersion of false news. Based on Figure 3, it is evident that most of the researchers developed or initiated mitigation efforts from the Twitter dataset analysis and used it as a medium to mitigate cyber propaganda. It is also worth mentioning that some of the articles did not mention what social media platforms they focused in helping the mitigation of the spread of false information. These articles used words like ‘social media’ and ‘social network’, instead of mentioning which of the social media platforms are suitable to use the mitigation effort that the paper contributed to.

In research question two, “*What are the existing socio-technical approach that contributes to mitigating cyber propaganda?”,* the contribution facet from our mapping diagram is adapted from the studies found in [64], [65], which initially focused on software development. It was then altered, according to the most current research that contributes to mitigating propaganda. The contribution types in this study are described as the kind of contribution a study provides. Table 4, shows our contribution facet on mitigating false information in social media. Based on the findings of the study, it is important to analyze contribution facet, to gauge the current contributions, while identifying the gap in the research. Once we identify the gap, we then could determine the most appropriate segment that should be emphasized in future research.

The research contribution focused on design science research such as model, framework, method, technique, approach, system, and mechanism. Based on Figure 4 three main contributions produced by the research on cyber-propaganda are frameworks/methods (33%), models

TABLE 4. Data item extracted from each study.

S/N	Data Names	Description	Relevancy to QRs
D1	Title	Title of the article	None
D2	Year	The publication year of the study	None
D3	Publication Type	The publication type of the paper	None
D4	Country	The Country published paper	None
D5	Topic discussed	Topics that have been discussed as a fake news	None
D6	Social media	Social media that has been discussed or mentioned in the research	RQ1
D7	Contribution Facet	Where the quality of the studies were measured	RQ2
D8	Mitigation Effort	The socio-technical mitigation effort proposed or presented to combat cyber-propaganda	RQ3

TABLE 5. Contribution facet (adapted from [64], [65]).

Contribution	Description
Model	Representation of observed reality, by concepts or related concepts after a conceptualization process
Framework/Methods	Models related to constructing software or managing development processes
Technique	A new or better way of performing certain tasks, such as design, implementation, maintenance, measurement, evaluation, selection of alternatives; involves techniques for execution, representation, management and analysis; a technique should be operational not advice or guidelines, but a procedure
Approach	It includes a set of logical assumptions that could be created to better understand problems. It may also be viewed as a word that will give birth to your systematic plans and the strategies you will use to attain specific goals.
Mechanism/System	It encompasses a set of things working together, as parts of a mechanism.

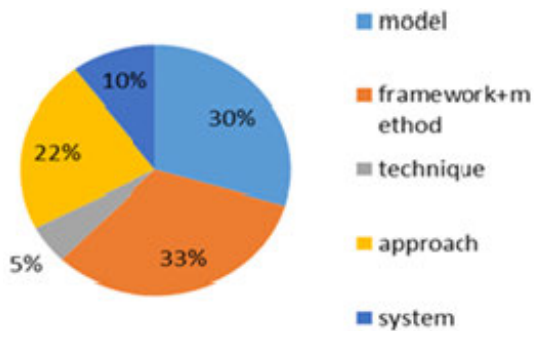


FIGURE 4. Contributions in mitigating propaganda from 2014 to 2018.

(30%), and approaches (22%). Figure 4 shows the model displaying significant attention among researchers. Techniques (5%), seem not to be the main concern in the context of a mitigating effort in combating cyber-propaganda.

This review provides the answer to research question three: *"What are the socio-technical compositions, especially on the effort that has been developed to mitigate propaganda?"*. Although the research has not provided a comprehensive method(s) of combating cyber propaganda yet, there are several contributions towards this end goal. The contributions listed from the findings are factors that have contributed to dispersing propaganda, identification of propaganda, and the spreaders' detection of propaganda messages' decision making support in preventing the cyber propaganda; truth discovery or information credibility, comprehension of propaganda spreading in the network, blocking and propaganda prediction. The overview of the findings is tabulated in Table 6. In Figure 6: Composition of Effort in Mitigating Cyber-Propaganda from 2014-2018, it is obvious that most of the papers focused on propaganda mitigation through detection (40%). Figure 4 shows that the detection of propaganda messages mostly focuses on developing a method or framework and has the least contribution to the technique that detects propaganda messages.

The 'comprehension' of rumour spreading (21%), discusses how the propagation of rumours happen i.e.: the propagation characteristics of online social networks rumours, propagation behaviour, the pattern and the trend of propagation [66]–[69]. It is worth mentioning that this type of effort mostly contributes to a model, as a way of mitigating cyber propaganda. From the bubble plot diagram shown in Figure 7, it can be concluded that these models mostly mention how rumours spread in the network, but

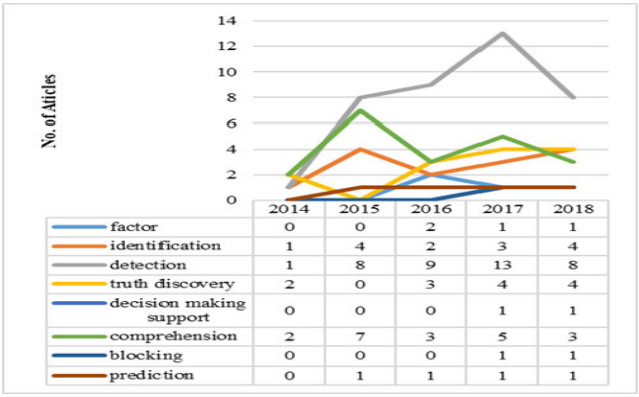


FIGURE 5. Propaganda mitigation effort from 2014 to 2018.

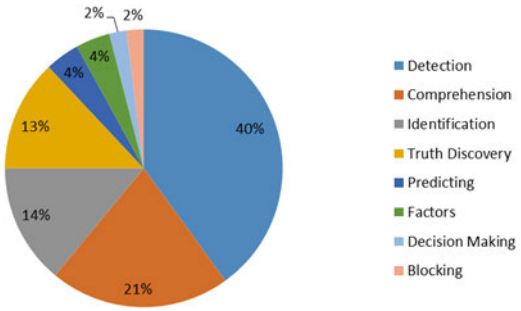


FIGURE 6. Composition of Effort in mitigating cyber-propaganda from 2014-2018.

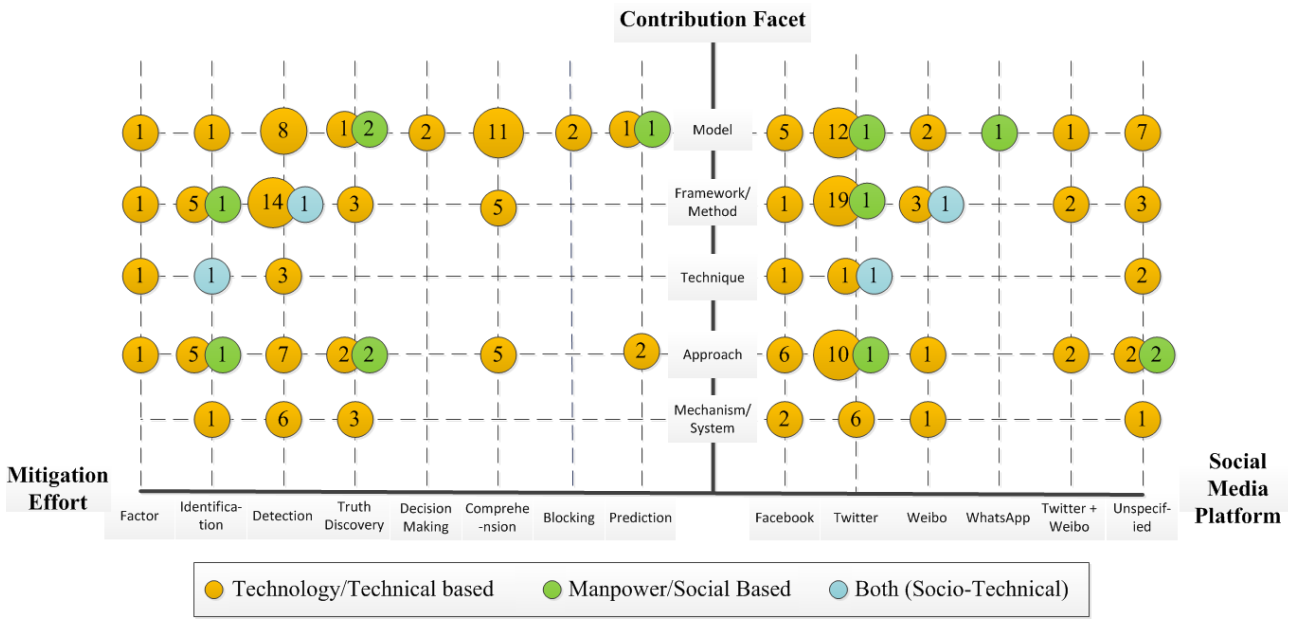


FIGURE 7. Socio-technical propaganda mitigation effort articles, produced from 2014 until 2018.

TABLE 6. Composition of effort in mitigating propaganda.

Composition of Effort	Description
Detection	Focus on propaganda messages detection in social media [82], [83] .
Comprehension	Discuss and explain the process by which false information, propaganda, rumour and misinformation occur [84].
Identification	Explain how to identify rumoured messages/ fake news/ propaganda/ misinformation and spreaders in social media [85], [86].
Truth Discovery	Seeks evaluation of information credibility by seeking the truth of information using manpower/social networking and technology [87], [88].
Predicting	Predict whether the message will become rumours in the future.[45], [72]
Factors	Factors on how social media is used to spread and re-transmit cyber propaganda [28][29].
Decision Making	Guide users in making their own decision on correcting the rumours [33].
Blocking	Block these rumoured messages/fake news/ propaganda/ misinformation messages from appearing in the social media [34][35]

never state the social media platform it focuses on. This type of propaganda mitigation effort has helped in providing comprehensive information on rumour spreading and thus can guide on how to control the same.

The 'Identification' (14%) mitigation effort, discusses how the socio-technical approach can identify rumoured messages and spreaders through the linguistic pattern; behaviour of the user; identification of the spreaders such as journalist and media streams organization; and lots of other [70], [71]. 'Truth discovery' (13%), often seeks evaluation of information credibility, by seeking the truth of information using manpower/social networking and technology itself. In the bubble plot diagram shown in Figure 7, it is notable to mention that there is no contribution to the technique that mitigates propaganda through truth discovery. There are several efforts discovered after extracting the articles on mitigating propaganda, misinformation, fake news and rumours; through predicting the (4%) rumour, whether the message will become rumour in the future. Several approaches have been introduced, to help in identifying future rumours and refuting them before they cause harm through model, method, and approaches [62], [45], [72]. Though, only one article has been focusing on the socio aspect in mitigating propaganda; which claims to have a strong prediction power, for the pro-social rumour combating behaviour [73].

The 'factors' (4%) effort of mitigating propaganda, contributes to the discovering factors on how social media is being used to spread propaganda and what factors have made the re-transmission of propaganda [74], [75]. Some researchers have also discovered that journalists; media; or organization official accounts; have become a major factor in the propagation of online rumours [76], [77]. As for the 'decision making' (2%) mitigation effort, it has given theoretical support to help the user with the decision making process [78]. The model of correcting behaviour by the user demonstrates how the user can make their decision on correcting the rumours [79]. Last but not least is the 'blocking' (2%) mitigation effort, which holds the least number of articles to produce. Only 2 contributions are found in blocking effort, and both of these articles introduced a model that can help in blocking of the propaganda messages [80], [81]. Table 6 below applies the classification schema on the primary studies that provide an overview of the field of cyber-propaganda research.

Figure 7, shows the visualization of overall data gathered as suggested by [18], which is the bubble plot diagram

TABLE 7. Systematic map overview.

#	Paper Title	Year	Social media platform	Contribution type	Mitigation Effort	Socio-technical pertinence	Ref.
1	Alethiometer: A Framework for Assessing Trustworthiness and Content Validity in Social Media	2014	Twitter	Framework	Truth Discovery	Technical	[89]
2	Fake Tweet Buster: A Webtool to Identify Users Promoting Fake News On Twitter	2014	Twitter	Technique	Identification	Socio-technical	[90]
3	Misinformation propagation in the age of Twitter	2014	Twitter	Model	Comprehension	Technical	[84]
4	Rumour Source Detection with Multiple Observations: Fundamental Limits and Algorithms	2014	unspecified	Framework	Detection	Technical	[91]
5	Social network rumours spread model based on cellular automata	2014	unspecified	Model	Comprehension	Technical	[92]
6	To shut them up or to clarify: Restraining the spread of rumours in online social networks	2014	Facebook	Method	Truth Discovery	Technical	[93]
7	A Novel Agent-Based Rumour Spreading Model in Twitter	2015	Twitter	Model	Comprehension	Technical	[94]
8	A Part-of-Speech Based Sentiment Classification Method Considering Subject-Predicate Relation	2015	Twitter	Method	Detection	Technical	[95]
9	Characterizing Online Rumouring Behavior Using Multi-Dimensional Signatures	2015	Twitter	Method	Comprehension	Technical	[96]
10	Crowdsourced Rumour Identification During Emergencies	2015	Twitter	Method	Identification	Social	[85]
11	Crowdsourcing the Annotation of Rumourous Conversations in Social Media	2015	Twitter	Method	Identification	Technical	[97]
12	Detect Rumours Using Time Series of Social Context Information on Microblogging Websites	2015	Twitter	Model	Detection	Technical	[83]
13	Evaluating Algorithms for Detection of Compromised Social Media User Accounts	2015	Twitter	Method	Detection	Technical	[40]
14	Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks	2015	Facebook	Model	Comprehension	Technical	[98]
15	Improving spam detection in Online Social Networks	2015	Twitter	Approach	Detection	Technical	[41]
16	Mining Streaming Tweets for Real-Time Event Credibility Prediction in Twitter	2015	Twitter	Model	Prediction	Technical	[45]
17	Real-Time News Cer Tification System on Sina Weibo	2015	Weibo	System	Detection	Technical	[42]
18	Real-time Rumour Debunking on Twitter	2015	Twitter	Method	Identification	Technical	[43]
19	Rumour Spreading Maximization and Source Identification in a Social Network	2015	Facebook	Approach	Identification	Technical	[99]
20	Science vs Conspiracy: Collective Narratives in the Age of Misinformation	2015	Facebook	Approach	Comprehension	Technical	[100]
21	Trend of Narratives in the Age of Misinformation	2015	Facebook	Approach	Comprehension	Technical	[101]
22	Tweet credibility analysis evaluation by improving sentiment dictionary	2015	Twitter	Method	Detection	Technical	[44]
23	Tweeting Propaganda, Radicalization and Recruitment: Islamic State Supporters Multi-sided Twitter Networks	2015	Twitter	Framework	Comprehension	Technical	[102]
24	Viral Misinformation: The Role of Homophily and Polarization	2015	Facebook	Method	Comprehension	Technical	[103]
25	Detecting Jihadist Messages on Twitter	2016	Twitter	Approach	Detection	Technical	[39]
26	Detecting Multipliers of Jihadism on Twitter	2016	Twitter	Approach	Detection	Technical	[38]
27	An exploration of rumour combating behavior on social media in the context of social crises	2016	unspecified	Model	Prediction	Technical	[73]
28	Applying a Tendency to Be Well Retweeted to False Information Detection	2016	Twitter	Method	Detection	Technical	[46]
29	Automated classification of extremist Twitter accounts using content-based and network-based features	2016	Twitter	Technique	Detection	Technical	[47]
30	Could This Be True?: I Think So! Expressed Uncertainty in Online Rumouring	2016	Twitter	Framework	Identification	Technical	[70]
31	Detecting Rumours Through Modelling Information Propagation Networks in a Social Media Environment	2016	Weibo	Model	Detection	Technical	[48]
32	ECRModel: An Elastic Collision-Based Rumour-Propagation Model in Online Social Networks	2016	Twitter	Model	Comprehension	Technical	[104]
33	Generalization of Information Spreading Forensics via Sequential Dependent Snapshots	2016	Twitter	Framework	Comprehension	Technical	[105]
34	Geoparsing and Geosemantics for Social Media: Spatiotemporal Grounding of Content Propagating Rumours to Support Trust and Veracity Analysis During Breaking News	2016	Twitter	Approach	Truth Discovery	Technical	[49]
35	How Information Snowballs: Exploring the Role of Exposure in Online Rumour Propagation	2016	Twitter	Approach	Comprehension	Technical	[69]
36	Keeping Up with the Tweet-dashians: The Impact of 'Official' Accounts on Online Rumouring	2016	Twitter	Approach	Factors	Technical	[76]
37	Kidnapping WhatsApp – Rumours during the search and rescue operation of three kidnapped youth	2016	WhatsApp	Model	Truth Discovery	social	[106]
38	Leveraging the Implicit Structure Within Social Media for Emergent Rumour Detection	2016	Twitter	Method	Detection	Technical	[107]
39	Misinformation in Online Social Networks: Detect Them All with a Limited Budget	2016	Twitter	Model	Detection	Technical	[108]
40	Rumour Source Obfuscation on Irregular Trees	2016	unspecified	Technique	Detection	Technical	[109]
41	Social media analytics to identify and counter Islamist extremism: Systematic detection, evaluation, and challenging of extremist narratives online	2016	Twitter	System	Detection	Technical	[110]
42	Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumours	2016	Twitter, Weibo	Approach	Identification	Technical	[111]
43	The Retransmission of Rumour-related Tweets: Characteristics of Source and Message	2016	Twitter	Model	Factors	Technical	[74]
44	Twitter Truths: Authenticating Analysis of Information Credibility	2016	Twitter	Method	Truth Discovery	Technical	[88]

TABLE 7. ***(Continued.)*** **Systematic map overview.**

45	Visualization of the social bot's fingerprints	2016	Twitter	System	Detection	Technical	[112]
46	Credibility investigation of newsworthy tweets using a visualising Petri net model	2017	Twitter	Model	Detection	Technical	[113]
47	A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumours	2017	Twitter	Model	Decision Making	Technical	[79]
48	A Temporal Attentional Model for Rumour Stance Classification	2017	Twitter	Approach	Truth Discovery	Technical	[50]
49	Behaviour Profiling of Reactions in Facebook Posts for Anomaly Detection	2017	Facebook	Approach	Detection	Technical	[51]
50	Centralized, Parallel, and Distributed Information Processing During Collective Sensemaking	2017	Twitter	Model	Comprehension	Technical	[114]
51	ClaimVerif: A Real-time Claim Verification System Using the Web and Fact Databases	2017	unspecified	System	Truth Discovery	Technical	[52]
52	Crowdsourcing the Verification of Fake News and Alternative Facts	2017	Twitter	System	Detection	Technical	[115]
53	CSI: A Hybrid Deep Model for Fake News Detection	2017	Twitter	Model	Detection	Technical	[53]
54	Distress and rumour exposure on social media during a campus lockdown	2017	Twitter	Approach	Comprehension	Technical	[116]
55	DRIMUX: Dynamic rumour influence minimization with user experience in social networks	2017	unspecified	Model	Blocking	Technical	[81]
56	Filipino and English Clickbait Detection Using a Long Short Term Memory Recurrent Neural Network	2017	unspecified	Model	Detection	Technical	[117]
57	From Retweet to Believability: Utilizing Trust to Identify Rumour Spreaders on Twitter	2017	Twitter	Approach	Identification	Technical	[54]
58	Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election	2017	Facebook	Model	Comprehension	Technical	[118]
59	Identifying Extremism in Social Media with Multi-view Context-Aware Subset Optimization	2017	unspecified	Framework	Detection	Technical	[119]
60	Initial Model of Social Media Islamic Information Credibility	2017	Facebook	Model	Truth Discovery	Technical	[120]
61	Modelling information spread in polarized communities: Transitioning from legacy media to a Facebook world	2017	Facebook	Model	Comprehension	Technical	[6]
62	Multimodal Fusion with Recurrent Neural Networks for Rumour Detection on Microblogs	2017	Twitter, Weibo	Model	Detection	Technical	[22]
63	Novel Visual and Statistical Image Features for Microblogs News Verification	2017	Weibo	Method	Detection	Technical	[34]
64	Preprocessing framework for Twitter bot detection	2017	Twitter	Framework	Detection	Technical	[55]
65	Revealing and Detecting Malicious Retweeter Groups	2017	Twitter	Approach	Detection	Technical	[56]
66	Role of Individual Activity in Rumour Spreading in Scale-free Networks	2017	Facebook	Model	Comprehension	Technical	[121]
67	Rumour Detection over Varying Time Windows	2017	Twitter	System	Detection	Technical	[35]
68	Rumour Identification with Maximum Entropy in MicroNet	2017	Weibo	Method	Identification	Technical	[57]
69	Rumour Restraining Based on Propagation Prediction with Limited Observations in Large-scale Social Networks	2017	Facebook	Method	Prediction	Technical	[72]
70	Rumour Source Detection in Finite Graphs with Boundary Effects by Message-passing Algorithms	2017	unspecified	Approach	Detection	Technical	[122]
71	SWIM: Stepped Weighted Shell Decomposition Influence Maximization for Large-Scale Networks	2017	unspecified	Method	Identification	Technical	[123]
72	The Fake News Spreading Plague: Was It Preventable?	2017	Facebook	Technique	Factors	Technical	[75]
73	Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda	2017	Weibo	Method	Detection	Socio-technical	[124]
74	CrowdsouRS: A crowdsourced reputation system for identifying deceptive online contents	2018	Facebook	System	Detection	Technical	[125]
75	It's always April fools' day!: On the difficulty of social network misinformation classification via propagation features	2018	Facebook	System	Truth Discovery	Technical	[20]
76	A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis	2018	Twitter	Framework	Comprehension	Technical	[67]
77	An Online Water Army Detection Method Based on Network Hot Events	2018	Weibo	Method	Detection	Technical	[126]
78	Anatomy of an online misinformation network	2018	Twitter	System	Truth Discovery	Technical	[127]
79	Approach to automatic identification of terrorist and radical content in social networks messages	2018	unspecified	Approach	Detection	Technical	[58]
80	Detecting pathogenic social media accounts without content or network structure	2018	Twitter	Framework	Detection	Technical	[128]
81	Distributed Rumour Blocking with Multiple Positive Cascades	2018	Facebook	Model	Blocking	Technical	[81]
82	EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection	2018	Twitter, Weibo	Framework	Detection	Technical	[59]
83	Effacing the Dilemma of the Rumouring Subject: A Value-oriented Approach towards Studying Misinformation on Social Media	2018	unspecified	Approach	Truth Discovery	Social	[129]
84	Engage Early, Correct More: How Journalists Participate in False Rumours Online During Crisis Events	2018	Twitter	Approach	Identification	Social	[130]
85	Fake News and its Credibility Evaluation by Dynamic Relational Networks: A Bottom up Approach	2018	unspecified	Approach	Truth Discovery	Social	[87]
86	Fake News Identification on Twitter with Hybrid CNN and RNN Models	2018	Twitter	Framework	Detection	Technical	[60]
87	Identifying Fake News and Fake Users on Twitter	2018	Twitter	System	Identification	Technical	[86]
88	"IRA Propaganda on Twitter: Stoking Antagonism and Tweeting Local News	2018	Twitter	Framework	Factors	Technical	[77]
89	Ising Model of User Behavior Decision in Network Rumour Propagation,	2018	unspecified	Model	Decision Making	Technical	[78]
90	Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation	2018	Twitter, Weibo	Framework	Detection	Technical	[131]
91	Mining Significant Microblogs for Misinformation Identification: An Attention-Based Approach	2018	Twitter, Weibo	Approach	Identification	Technical	[132]

TABLE 7. *(Continued.)* **Systematic map overview.**

92	Polarity Analysis of Editorial Articles Towards Fake News Detection	2018	unspecified	Model	Detection	Technical	[61]
93	Predicting future rumours	2018	Weibo	Approach	Prediction	Technical	[62]
94	The diffusion of misinformation on social media: Temporal pattern, message, and source	2018	Twitter	Framework	Comprehension	Technical	[68]
95	The future of deception: machine-generated and manipulated images, video, and audio?	2018	unspecified	Technique	Detection	Technical	[133]
96	The Rise of Guardians: Fact-checking URL Recommendation to Combat Fake News	2018	Twitter	Model	Truth Discovery	Social	[134]
97	The rumour spectrum	2018	Twitter	Model	Comprehension	Technical	[66]
98	Third person effects of fake news: Fake news regulation and media literacy interventions	2018	unspecified	Model	Identification	Technical	[135]

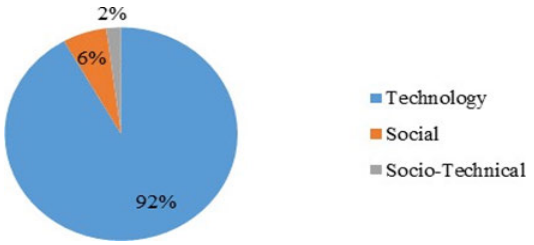


FIGURE 8. Contributions in mitigating propaganda from 2014 to 2018.

of the articles that focused on propaganda mitigation as the main topic. Based on II, it is worth-mentioning that most of the studies have only concentrated on a technical-based/technology-based approach (92%), for propaganda mitigation. For instance, the technical approach of combating propaganda includes a study that contributes to the areas of developing an algorithm, an equation or a system that combats cyber propaganda. Only a very few studies have revealed that the effort in mitigation of propaganda use social or human effort (6%), which, in this case, includes studies that provide ways on how social media users can combat propaganda through social activities in the social media platform. Lastly, the mitigation of cyber propaganda uses both socio-technical approaches that hold the least number with only 2%.

VIII. CONCLUSION

This study presents a systematic mapping survey on the mitigation effort in combating cyber propaganda in the social media. The study covered the published articles between 2014 and 2018. Out of 5176 retrieved articles, only 89 of them meet up with the selection criteria, they were selected for the primary study from four major academic databases. The purpose of the selected articles is to explore an overview of the socio-technical approaches that have been providing ways of mitigating cyber propaganda in the social media. From the inception, we noted that the theme of the social media associated with propaganda, misinformation, fake news and rumours has been widely studied. Evidence from the reviewed primary studies indicates that many studies have been conducted on combating cyber-propaganda. The review result shows that 92% of the studies contribute to the technical approach, 6% to the social effort and only 2% contribute to socio-technical approaches. Thus, there are insufficient studies involving how the humans, the combination of social behaviour and the technical context can govern the side effects of cyber propaganda. The major contributions of this study are the literature review of the classification technique (see section III) and literature mapping created (see Figure 8). Based on the systematic mapping, it is possible to identify how the context of this study has been explored; therefore, determining research gaps and future research opportunities. In conclusion, this systematic mapping study has provided a vital insight into the research area by identifying the current research and directions for future researches in terms of combating cyber propaganda in social media by using both human and technological efforts for mitigation. In the future, the authors will perform a comparative analysis of cyber propaganda detection methods.

REFERENCES

[1] S. Hamidian and M. Diab, "Rumor identification and belief investigation on Twitter," in *Proc. 7th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, 2016, pp. 3–8.
[2] C. Jack, "Lexicon of lies: Terms for problematic information," *Data Soc.*, vol. 3, 2017.
[3] E. C. Tandoc, Z. W. Lim, and R. Ling, "Defining 'Fake News': A typology of scholarly definitions," *Digit. Journalism*, vol. 6, no. 2, pp. 137–153, 2018.
[4] N. Pandey, "Fake news–a manufactured deception, distortion and disinformation is the new challenge to digital literacy," *J. Content, Community Commun.*, vol. 4, no. 8, pp. 15–21, Dec. 2018.
[5] S. C. Woolley and P. Howard, "Computational propaganda worldwide: Executive summary," Tech. Rep., 2017, vol. 36.
[6] I. Patel, H. Nguyen, E. Belyi, Y. Getahun, S. Abdulkareem, P. J. Giabbanelli, and V. Mago, "Modeling information spread in polarized communities: Transitioning from legacy media to a Facebook world," in *Proc. SoutheastCon*, Mar. 2017, pp. 1–8.
[7] S. C. Woolley and P. N. Howard, "Political communication, computational propaganda, and autonomous agents-Introduction," *Int. J. Commun.*, vol. 10, pp. 4882–4890, Oct. 2016.
[8] S. Bradshaw and P. N. Howard, "The global disinformation order: 2019 global inventory of organised social media manipulation," Project on Computational Propaganda, Tech. Rep., 2019.
[9] C. A. de Lima Salge and N. Berente, "Is that social bot behaving unethically?" *Commun. ACM*, vol. 60, no. 9, pp. 29–31, Aug. 2017.
[10] I. Awan, "Islamophobia and Twitter: A typology of online hate against muslims on social media," *Policy Internet*, vol. 6, no. 2, pp. 133–150, Jun. 2014.
[11] I. Awan, "Cyber-extremism: Isis and the power of social media," *Society*, vol. 54, no. 2, pp. 138–149, Apr. 2017.
[12] M. Greene, "Socio-technical transitions and dynamics in everyday consumption practice," *Global Environ. Change*, vol. 52, pp. 1–9, Sep. 2018.
[13] C. Ireton and J. Posetti, *Journalism, Fake News & Disinformation: Handbook for Journalism Education and Training*. Paris, France: UNESCO Publishing, 2018.

[14] T. Gillespie, "Governance of and by platforms," in *SAGE Handbook of Social Media*, Newbury Park, CA, USA: Sage, 2017.

[15] Y. Luo and J. Bu, "How valuable is information and communication technology? A study of emerging economy enterprises," *J. World Bus.*, vol. 51, no. 2, pp. 200–211, Feb. 2016.

[16] D. S. Bogatinov, M. Bogdanoski, and S. Angelevski, "AI-based cyber defense for more secure cyberspace," *Nature-Inspired Comput. Concepts, Methodol. Tools, Appl.*, vol. 3, pp. 1471–1489, 2016.

[17] J. Peterson and J. Densley, "Cyber violence: What do we know and where do we go from here?" *Aggression Violent Behav.*, vol. 34, pp. 193–200, May 2017.

[18] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, Aug. 2015.

[19] C. I. Eke, A. A. Norman, L. Shuib, and H. F. Nweke, "Sarcasm identification in textual data: Systematic review, research challenges and open directions," *Artif. Intell. Rev.*, pp. 1–44, Nov. 2019.

[20] M. Conti, D. Lain, R. Lazzeretti, G. Lovisotto, and W. Quattrociocchi, "It's always april fools' day!: On the difficulty of social network misinformation classification via propagation features," in *Proc. IEEE Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2017, pp. 1–6.

[21] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proc. ACM SIGKDD Workshop Mining Data Semantics (MDS)*, 2012, pp. 1–7.

[22] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on microblogs," in *Proc. ACM Multimedia Conf. (MM)*, 2017, pp. 795–816.

[23] C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schifferes, and N. Newman, "Challenges of computational verification in social multimedia," in *Proc. 23rd Int. Conf. World Wide Web (WWW Companion)*, 2014, pp. 743–748.

[24] Q. Zhang, S. Zhang, J. Dong, J. Xiong, and X. Cheng, "Automatic detection of rumor on social network," in *Natural Language Processing and Chinese Computing*. Cham, Switzerland: Springer, 2015, pp. 113–122.

[25] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 675–684.

[26] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on Twitter," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 153–164.

[27] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 230–239.

[28] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 2972–2978.

[29] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on sina weibo by propagation structures," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Apr. 2015, pp. 651–662.

[30] J. Ma, "Detecting rumors from microblogs with recurrent neural networks," Tech. Rep. 3818, 2016.

[31] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. M. Menczer, "Detecting and tracking political abuse in social media," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 1–8.

[32] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "Tweeting is believing? Understanding microblog credibility perceptions," in *Proc. ACM Conf. Comput. Supported Cooperat. Work*, 2012, pp. 441–450.

[33] C. Boididou, "Verifying multimedia use at MediaEval 2015," *MediaEval*, vol. 3, no. 3, p. 7, 2015.

[34] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 598–608, Mar. 2017.

[35] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PLoS ONE*, vol. 12, no. 1, 2017, Art. no. e0168344.

[36] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions," in *Proc. 3rd Int. Conf. Knowl. Discovery Data Mining*," 1997, pp. 43–48.

[37] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation while comparing induction algorithms," in *Proc. ICML Conf.*, 1998.

[38] L. Kaati, E. Omer, N. Prucha, and A. Shrestha, "Detecting multipliers of jihadism on Twitter," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 954–960.

[39] M. Ashcroft, A. Fisher, L. Kaati, E. Omer, and N. Prucha, "Detecting jihadist messages on Twitter," in *Proc. Eur. Intell. Secur. Informat. Conf.*, Sep. 2015, pp. 161–164.

[40] D. Trang, F. Johansson, and M. Rosell, "Evaluating algorithms for detection of compromised social media user accounts," in *Proc. 2nd Eur. Netw. Intell. Conf.*, Sep. 2015, pp. 75–82.

[41] A. Gupta and R. Kaushal, "Improving spam detection in online social networks," in *Proc. Int. Conf. Cognit. Comput. Inf. Process. (CCIP)*, Mar. 2015, pp. 1–6.

[42] X. Zhou, J. Cao, Z. Jin, F. Xie, Y. Su, D. Chu, X. Cao, and J. Zhang, "Real-time news cer tification system on Sina Weibo," in *Proc. 24th Int. Conf. World Wide Web (WWW Companion)*, 2015, pp. 983–988.

[43] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, "Real-time rumor debunking on Twitter," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2015, pp. 1867–1870.

[44] T. Kawabe, Y. Namihira, K. Suzuki, M. Nara, Y. Sakurai, S. Tsuruta, and R. Knauf, "Tweet credibility analysis evaluation by improving sentiment dictionary," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, May 2015, pp. 2354–2361.

[45] J. Zou, F. Fekri, and S. W. McLaughlin, "Mining streaming tweets for real-time event credibility prediction in Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, 2015, pp. 1586–1589.

[46] Z. Yoshida and M. Aritsugi, "Applying a tendency to be well retweeted to false information detection," in *Proc. 18th Int. Conf. Inf. Integr. Web-Based Appl. Services (iiWAS)*, 2016, pp. 154–159.

[47] U. Xie, J. Xu, and T.-C. Lu, "Automated classification of extremist Twitter accounts using content-based and network-based features," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2545–2549.

[48] Y. Liu and S. Xu, "Detecting rumors through modeling information propagation networks in a social media environment," *IEEE Trans. Comput. Social Syst.*, vol. 3, no. 2, pp. 46–62, Jun. 2016.

[49] S. E. Middleton and V. Krivcovs, "Geoparsing and geosemantics for social media: Spatiotemporal grounding of content propagating rumors to support trust and veracity analysis during breaking news," *ACM Trans. Inf. Syst.*, vol. 34, no. 3, p. 16, Apr. 2016.

[50] A. P. B. Veyseh, J. Ebrahimi, D. Dou, and D. Lowd, "A temporal attentional model for rumor stance classification," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 2335–2338.

[51] P. V. Savyan and S. M. S. Bhanu, "Behaviour profiling of reactions in facebook posts for anomaly detection," in *Proc. 9th Int. Conf. Adv. Comput. (ICoAC)*, Dec. 2017, pp. 220–226.

[52] S. Zhi, Y. Sun, J. Liu, C. Zhang, and J. Han, "ClaimVerif: A real-time claim verification system using the Web and fact databases," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 2555–2558.

[53] N. Ruchansky, S. Seo, and Y. Liu, "CSI: A hybrid deep model for fake news detection," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 797–806.

[54] B. Rath, W. Gao, J. Ma, and J. Srivastava, "From retweet to believability: Utilizing trust to identify rumor spreaders on Twitter," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 179–186.

[55] M. Kantepe and M. C. Ganiz, "Preprocessing framework for Twitter bot detection," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, Oct. 2017, pp. 630–634.

[56] N. Vo, K. Lee, C. Cao, T. Tran, and H. Choi, "Revealing and detecting malicious retweeter groups," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 363–368.

[57] S. Yu, M. Li, and F. Liu, "Rumor identification with maximum entropy in MicroNet," *Complexity*, vol. 2017, Sep. 2017, Art. no. 1703870.

[58] A. I. Kapitanov, I. I. Kapitanova, V. M. Troyanovskiy, V. F. Shangin, and N. O. Krylikov, "Approach to automatic identification of terrorist and radical content in social networks messages," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2018, pp. 1517–1520.

[59] Y. Wang, "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2018, pp. 849–857.

[60] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," in *Proc. 9th Int. Conf. Social Media Soc.*, Jul. 2018, pp. 226–230.

[61] M. J. C. Samonte, "Polarity analysis of editorial articles towards fake news detection," in *Proc. Int. Conf. Internet e-Business (ICIEB)*, Apr. 2018, pp. 108–112.

[62] Y. Qin, W. Dominik, and C. Tang, "Predicting future rumours," *Chin. J. Electron.*, vol. 27, no. 3, pp. 514–520, May 2018.

[63] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.

[64] M. Shaw, "Writing good software engineering research papers," in *Proc. 25th Int. Conf. Softw. Eng.*, 2003, pp. 726–736.

[65] N. Paternoster, C. Giardino, M. Unterkalmsteiner, T. Gorschek, and P. Abrahamsson, "Software development in startup companies: A systematic mapping study," *Inf. Softw. Technol.*, vol. 56, no. 10, pp. 1200–1218, Oct. 2014.

[66] N. Turenne, "The rumour spectrum," *PLoS ONE*, vol. 13, no. 1, 2018, Art. no. e0189080.

[67] S. M. Jang, T. Geng, J.-Y. Queenie Li, R. Xia, C.-T. Huang, H. Kim, and J. Tang, "A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis," *Comput. Hum. Behav.*, vol. 84, pp. 103–113, Jul. 2018.

[68] J. Shin, L. Jian, K. Driscoll, and F. Bar, "The diffusion of misinformation on social media: Temporal pattern, message, and source," *Comput. Hum. Behav.*, vol. 83, pp. 278–287, Jun. 2018.

[69] A. Arif, K. Shanahan, F.-J. Chou, Y. Dosouto, K. Starbird, and E. Spiro, "How information snowballs: Exploring the role of exposure in online rumor propagation," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput. (CSCW)*, 2016, pp. 466–477.

[70] K. Starbird, E. Spiro, I. Edwards, K. Zhou, J. Maddock, and S. Narasimhan, "Could this be true?: I think So! Expressed uncertainty in online rumoring," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 360–371.

[71] K. Starbird, D. Dailey, O. Mohamed, G. Lee, and E. S. Spiro, "Engage early, correct more," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2018, pp. 1–12.

[72] Q. Wu, T. Wang, Y. Cai, H. Tian, and Y. Chen, "Rumor restraining based on propagation prediction with limited observations in large-scale social networks," in *Proc. Australas. Comput. Sci. Week Multiconf.*, 2017, pp. 1–8.

[73] L. Zhao, J. Yin, and Y. Song, "An exploration of rumor combating behavior on social media in the context of social crises," *Comput. Hum. Behav.*, vol. 58, pp. 25–36, May 2016.

[74] A. Y. K. Chua, C.-Y. Tee, A. Pang, and E.-P. Lim, "The retransmission of rumor-related Tweets: Characteristics of source and message," in *Proc. 7th Int. Conf. Social Media Soc.*, 2016, p. 22.

[75] E. Mustafaraj and P. T. Metaxas, "The fake news spreading plague: Was it preventable?" in *Proc. ACM Web Sci. Conf.*, 2017, pp. 235–239.

[76] C. Andrews, E. Fichet, Y. Ding, E. S. Spiro, and K. Starbird, "Keeping up with the Tweet-dashians: The impact of 'Official' accounts on online rumoring," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, 2016, pp. 452–465.

[77] J. Farkas and M. Bastos, "IRA propaganda on Twitter: Stoking antagonism and tweeting local news," in *Proc. 9th Int. Conf. Social Media Soc.*, 2018, pp. 281–285.

[78] C. Li, F. Liu, and P. Li, "Ising model of user behavior decision in network rumor propagation," *Discrete Dyn. Nature Soc.*, vol. 2018, pp. 1–10, Aug. 2018.

[79] A. Arif, "A closer look at the self-correcting crowd: Examining corrections in online rumors," in *Proc. ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, 2017, pp. 155–168.

[80] G. Tong, W. Wu, and D.-Z. Du, "Distributed rumor blocking with multiple positive cascades," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 2, pp. 468–480, Jun. 2018.

[81] B. Wang, G. Chen, L. Fu, L. Song, and X. Wang, "DRIMUX: Dynamic rumor influence minimization with user experience in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2168–2181, Oct. 2017.

[82] Z. Wang, W. Dong, W. Zhang, and C. W. Tan, "Rumor source detection with multiple observations: Fundamental limits and algorithms," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 1, pp. 1–13, Jun. 2014.

[83] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2015, pp. 1751–1754.

[84] F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of Twitter," *Computer*, vol. 47, no. 12, pp. 90–94, Dec. 2014.

[85] R. McCreadie, C. Macdonald, and I. Ounis, "Crowdsourced rumour identification during emergencies," in *Proc. 24th Int. Conf. World Wide Web (WWW Companion)*, 2015, pp. 965–970.

[86] C.-S. Atodiresei, A. Tănăselea, and A. Iftene, "Identifying fake news and fake users on Twitter," *Procedia Comput. Sci.*, vol. 126, pp. 451–461, 2018.

[87] Y. Ishida and S. Kuraya, "Fake news and its credibility evaluation by dynamic relational networks: A bottom up approach," *Procedia Comput. Sci.*, vol. 126, pp. 2228–2237, 2018.

[88] R. Chatterjee and S. Agarwal, "Twitter truths: Authenticating analysis of information credibility," *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 2352–2357.

[89] E. Jaho, E. Tzoannos, A. Papadopoulos, and N. Sarris, "Alethiometer: A framework for assessing trustworthiness and content validity in social media," in *Proc. 23rd Int. Conf. World Wide Web*, 2014, pp. 749–752.

[90] D. Saez-Trumper, "Fake tweet buster: A webtool to identify users promoting fake news OnTwitter," in *Proc. 25th ACM Conf. Hypertext Social Media*, 2014, pp. 316–317.

[91] Z. Wang, W. Dong, W. Zhang, and C. W. Tan, "Rumor source detection with multiple observations: Fundamental limits and algorithms," in *Proc. ACM Int. Conf. Meas. Modeling Comput. Syst.*, 2014, pp. 1–13.

[92] A. Wang, W. Wu, and J. Chen, "Social network rumors spread model based on cellular automata," in *Proc. 10th Int. Conf. Mobile Ad-Hoc Sensor Netw.*, vol. 1, Dec. 2014, pp. 236–242.

[93] S. Wen, J. Jiang, Y. Xiang, S. Yu, W. Zhou, and W. Jia, "To shut them up or to clarify: Restraining the spread of rumors in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3306–3316, Dec. 2014.

[94] E. Serrano, C. Á. Iglesias, and M. Garijo, "A novel agent-based rumor spreading model in Twitter," in *Proc. 24th Int. Conf. World Wide Web (WWW Companion)*, 2015, pp. 811–814.

[95] T. Kawabe, Y. Namihira, K. Suzuki, M. Nara, Y. Yamamoto, S. Tsuruta, and R. Knauf, "A Part-of-Speech based sentiment classification method considering subject-predicate relation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 999–1004.

[96] J. Maddock, K. Starbird, H. J. Al-Hassani, D. E. Sandoval, M. Orand, and R. M. Mason, "Characterizing online rumoring behavior using multidimensional signatures," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput. (CSCW)*, 2015, pp. 228–241.

[97] A. Zubiaga, M. Liakata, R. Procter, K. Bontcheva, and P. Tolmie, "Crowdsourcing the annotation of rumourous conversations in social media," in *Proc. 24th Int. Conf. World Wide Web (WWW Companion)*, 2015, pp. 347–353.

[98] M. Tambuscio, G. Ruffo, A. Flammini, and F. Menczer, "Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 977–982.

[99] W. Luo, W. P. Tay, and M. Leng, "Rumor spreading maximization and source identification in a social network," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 186–193.

[100] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: Collective narratives in the age of misinformation," *PLoS ONE*, vol. 10, no. 2, 2015, Art. no. e0118093.

[101] A. Bessi, F. Zollo, M. Del Vicario, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Trend of narratives in the age of misinformation," *PLoS ONE*, vol. 10, no. 8, 2015, Art. no. e0134641.

[102] A. T. Chatfield, C. G. Reddick, and U. Brajawidagda, "Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided Twitter networks," in *Proc. 16th Annu. Int. Conf. Digit. Government Res.*, 2015, pp. 239–249.

[103] A. Anagnostopoulos, A. Bessi, G. Caldarelli, M. Del Vicario, F. Petroni, A. Scala, F. Zollo, and W. Quattrociocchi, "Viral misinformation: The role of homophily and polarization," 2014, *arXiv:1411.2893*. [Online]. Available: http://arxiv.org/abs/1411.2893

[104] Z. Tan, J. Ning, Y. Liu, X. Wang, G. Yang, and W. Yang, "ECRModel: An elastic collision-based rumor-propagation model in online social networks," *IEEE Access*, vol. 4, pp. 6105–6120, 2016.

[105] K. Cai, H. Xie, and J. C. S. Lui, "Generalization of information spreading forensics via sequential dependent snapshots," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 44, no. 2, pp. 12–14, Sep. 2016.

[106] T. Simon, A. Goldberg, D. Leykin, and B. Adini, "Kidnapping Whatsapp–Rumors during the search and rescue operation of three kidnapped youth," *Comput. Hum. Behav.*, vol. 64, pp. 183–190, Nov. 2016.

[107] J. Sampson, F. Morstatter, L. Wu, and H. Liu, "Leveraging the implicit structure within social media for emergent rumor detection," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 2377–2382.

[108] H. Zhang, M. A. Alim, X. Li, M. T. Thai, and H. T. Nguyen, "Misinformation in online social networks: Detect them all with a limited budget," *ACM Trans. Inf. Syst.*, vol. 34, no. 3, p. 18, Apr. 2016.

[109] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, and P. Viswanath, "Rumor source obfuscation on irregular trees," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 44, no. 1, pp. 153–164, Jun. 2016.

[110] I. B. Arpinar, U. Kursuncu, and D. Achilov, "Social media analytics to identify and counter islamist extremism: Systematic detection, evaluation, and challenging of extremist narratives online," in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, Oct. 2016, pp. 611–612.

[111] I. C.-H. Fung, "Social media's initial reaction to information and misinformation on Ebola, August 2014: Facts and rumors," *Public Health Rep.*, vol. 131, no. 3, pp. 461–473, 2016.

[112] M. Kaya, S. Conley, and A. Varol, "Visualization of the social bot's fingerprints," in *Proc. 4th Int. Symp. Digit. Forensic Secur. (ISDFS)*, Apr. 2016, pp. 161–166.

[113] M. Torky, R. Baberse, R. Ibrahim, A. E. Hassanien, G. Schaefer, I. Korovin, and S. Ying Zhu, "Credibility investigation of newsworthy tweets using a visualising Petri net model," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 3894–3898.

[114] P. Krafft, K. Zhou, I. Edwards, K. Starbird, and E. S. Spiro, "Centralized, parallel, and distributed information processing during collective sensemaking," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 2976–2987.

[115] R. J. Sethi, "Crowdsourcing the verification of fake news and alternative facts," in *Proc. 28th ACM Conf. Hypertext Social Media (HT)*, 2017, pp. 315–316.

[116] N. M. Jones, R. R. Thompson, C. Dunkel Schetter, and R. C. Silver, "Distress and rumor exposure on social media during a campus lockdown," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 44, pp. 11663–11668, Oct. 2017.

[117] P. K. Dimpas, R. V. Po, and M. J. Sabellano, "Filipino and english clickbait detection using a long short term memory recurrent neural network," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Dec. 2017, pp. 276–280.

[118] A. Fourney, M. Z. Racz, G. Ranade, M. Mobius, and E. Horvitz, "Geographic and temporal trends in fake news consumption during the 2016 US presidential election," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 2071–2074.

[119] S. Das Bhattacharjee, B. V. Balantrapu, W. Tolone, and A. Talukder, "Identifying extremism in social media with multi-view context-aware subset optimization," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1–10.

[120] K. A. Kadir, N. S. Ashaari, and J. Salim, "Initial model of social media islamic information credibility," in *Proc. 6th Int. Conf. Electr. Eng. Informat. (ICEEI)*, Nov. 2017, pp. 1–6.

[121] Y. Zhang, M. Xiong, Y. Xu, J. Guan, and S. Zhou, "Role of individual activity in rumor spreading in scale-free networks," in *Proc. Companion 10th Int. Conf. Utility Cloud Comput. (UCC Companion)*, 2017, pp. 125–129.

[122] P.-D. Yu, C. W. Tan, and H.-L. Fu, "Rumor source detection in finite graphs with boundary effects by message-passing algorithms," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 86–90.

[123] A. Vardasbi, H. Faili, and M. Asadpour, "SWIM: Stepped weighted shell decomposition influence maximization for large-scale networks," *ACM Trans. Inf. Syst.*, vol. 36, no. 1, p. 6, Aug. 2017.

[124] B. Qin, D. Strömberg, and Y. Wu, "Why does China allow freer social media? Protests versus surveillance and propaganda," *J. Econ. Perspect.*, vol. 31, no. 1, pp. 117–140, Feb. 2017.

[125] M. R. Siddiki, M. A. Talha, F. Chowdhury, and M. S. Ferdous, "CrowdsouRS: A crowdsourced reputation system for identifying deceptive online contents," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–6.

[126] W. Zhang and J. Lu, "An online water army detection method based on network hot events," in *Proc. 10th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA)*, Feb. 2018, pp. 191–193.

[127] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia, "Anatomy of an online misinformation network," *PLoS ONE*, vol. 13, no. 4, 2018, Art. no. e0196087.

[128] E. Shaabani, R. Guo, and P. Shakarian, "Detecting pathogenic social media accounts without content or network structure," in *Proc. 1st Int. Conf. Data Intell. Secur. (ICDIS)*, Apr. 2018, pp. 57–64.

[129] R. Aricat, "Effacing the dilemma of the rumouring subject: A value-oriented approach towards studying misinformation on social media," *J. Hum. Values*, vol. 24, no. 1, pp. 56–65, Jan. 2018.

[130] K. Starbird, D. Dailey, O. Mohamed, G. Lee, and E. S. Spiro, "Engage early, correct more: How journalists participate in false rumors online during crisis events," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, p. 105.

[131] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez-Rodriguez, "Leveraging the crowd to detect and reduce the spread of fake news and misinformation," in *Proc. 11th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2018, pp. 324–332.

[132] Q. Liu, F. Yu, S. Wu, and L. Wang, "Mining significant microblogs for misinformation identification: An attention-based approach," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 5, p. 50, Apr. 2018.

[133] J. Bakdash, "The future of deception: Machine-generated and manipulated images, video, and audio?" in *Proc. 3rd Int. Workshop Social Sens. (SocialSens)*, Apr. 2018, p. 2.

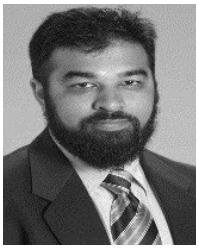
[134] N. Vo and K. Lee, "The rise of guardians: Fact-checking url recommendation to combat fake news," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2018, pp. 275–284.

[135] S. M. Jang and J. K. Kim, "Third person effects of fake news: Fake news regulation and media literacy interventions," *Comput. Hum. Behav.*, vol. 80, pp. 295–302, Mar. 2018.

AIMI NADRAH MASERI received the bachelor's degree in management information system from the University of Malaya, Kuala Lumpur, in 2018, where she is currently pursuing the master's degree in computer science. Her research area is focused on social computing. She has joined a symposium conference held at Kuala Lumpur in 2018 and international.

AZAH ANIR NORMAN received the Ph.D. degree in information systems security. She is currently a Lecturer with the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur. She had previously worked as a Security Consultant with the MSC Trustgate.com (subsidiary body of MDEC Malaysia) a certification authority in Malaysia for more than four years. Her research interests include e-commerce security, information systems security management, security policies, and standards. She was awarded a few research grants that focused on social media security and cybersecurity practices. Her articles in selected research areas have been printed in local and international conferences and ISI/Scopus WOS journals. She actively supervises many students at all levels of study from undergraduate (i.e., bachelor's degree) up to post-graduate (i.e. master's and Ph.D.) supervisions.

CHRISTOPHER IFEANYI EKE received the B.Sc. degree in computer science from Ebonyi State University, Nigeria, and the M.Sc. degree in mobile computing from the University of Bedfordshire, Luton, U.K. He is currently pursuing the Ph.D. degree with the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. His research interests are data science, NLP, information system and security, cloud computing, machine learning, big data, and social media analytics.



ATIF AHMAD is a Senior Academic with the School of Computing and Information Systems, The University of Melbourne. He leads Business Information Security Research and serves as the Deputy Director for the Academic Centre of Cyber Security Excellence. His areas of expertise include strategy, risk and incident response in information security management (ISM). He has authored over 70 scholarly articles in ISM and received over 3M USD in grant funding. He has previously served as a Cybersecurity Consultant for WorleyParsons, Pinkerton, and SinclairKnightMerz. He is a Certified Protection Professional with the American Society for Industrial Security.

NURUL NUHA ABDUL MOLOK received the B.Sc. degree in computer science (artificial intelligence) and the M.Sc. degree in computer science (information systems) from Universiti Malaya, Malaysia, and the Ph.D. degree in IS security from the University of Melbourne, Australia. She is currently an Assistant Professor with the Department of Information Systems, Faculty of Information and Communication Technology, International Islamic University Malaysia. She is a certified ISO 27001 Information Security Management Systems (ISMS) Lead Auditor and the Cyber Defender Associate (CCDA). She is also a working group member of the Information Security Technical Committee, Department of Standards Malaysia. To date, she has more than 30 publications in indexed and non-indexed, international and local, journals, and proceedings.