Big Data & Society
January–June: 1–14
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517211013848
journals.sagepub.com/home/bds

Countering misinformation: A multidisciplinary approach

Kacper T Gradoń[1], Janusz A. Hołyst[2], Wesley R Moy[3], Julian Sienkiewicz[2] and Krzysztof Suchecki[2]

Abstract

The article explores the concept of infodemics during the COVID-19 pandemic, focusing on the propagation of false or inaccurate information proliferating worldwide throughout the SARS-CoV-2 health crisis. We provide an overview of disinformation, misinformation and malinformation and discuss the notion of "fake news", and highlight the threats these phenomena bear for health policies and national and international security. We discuss the mis-/disinformation as a significant challenge to the public health, intelligence, and policymaking communities and highlight the necessity to design measures enabling the prevention, interdiction, and mitigation of such threats. We then present an overview of selected opportunities for applying technology to study and combat disinformation, outlining several approaches currently being used to understand, describe, and model the phenomena of misinformation and disinformation. We focus specifically on complex networks, machine learning, data- and text-mining methods in misinformation detection, sentiment analysis, and agent-based models of misinformation spreading and the detection of misinformation sources in the network. We conclude with the set of recommendations supporting the World Health Organization's initiative on infodemiology. We support the implementation of integrated preventive procedures and internationalization of infodemic management. We also endorse the application of the cross-disciplinary methodology of Crime Science discipline, supplemented by Big Data analysis and related information technologies to prevent, disrupt, and detect mis- and disinformation efficiently.

Keywords

COVID-19 and SARS-CoV-2, misinformation and disinformation, machine learning, complex networks, agent-based models, fake news

This article is a part of special theme on Studying the COVID-19 Infodemic at Scale. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/studyinginfodemicatscale

Introduction

SARS-CoV-2 coronavirus and disease it has caused, COVID-19, was identified in December 2019 in China (Zhu et al., 2020). In January 2020, the World Health Organization (WHO, 2020a) announced a Public Health Emergency of International Concern and on 11 March 2020 declared it a pandemic (WHO, 2020b). Since then, COVID-19 has developed into a global health crisis, disrupting lives worldwide and causing unprecedented disruption (Organisation for Economic Co-operation and Development, 2020). The United Nations (2020) assessed that the pandemic had wiped out decades of development gains. By early April 2021, it had infected over 132 million and resulted in the death of 2.8 million people worldwide (Johns Hopkins University, 2020).

One of unique aspects of the COVID-19 pandemic has been the proliferation of an enormous volume of information, both accurate and incorrect, making it challenging for the general public to find trustworthy

[1] Department of Security and Crime Science, University College London, London, UK
[2] Faculty of Physics, Warsaw University of Technology, Warszawa, Poland
[3] Krieger School of Arts & Sciences, Johns Hopkins University, Washington, DC, USA

Corresponding author:
Kacper T Gradoń, Department of Security and Crime Science, University College London, 35 Tavistock Square, Bloomsbury, London WC1H 9EZ, UK.
Email: k.gradon@ucl.ac.uk

and reliable sources of information (WHO, 2020d). The extraordinary volume of misinformation, often based on conspiracy theories, further amplified the information chaos related to COVID-19 (Gradon, 2020). This "fake news" has been viewed by investigators as one of the greatest threats to democracy, journalism, and freedom of expression (Zhou and Zafarani, 2020). As Gallotti et al. (2020) noted on the verge of the global pandemic emergency, human communication is largely characterized by the production of informational noise and misleading or false information. Waves of unreliable, low-quality information have potentially dangerous impacts on society's capacity to respond and may prevent actions that could contribute to containing the spread of the pandemic. False or misleading information may prevent timely and effective public adoption of appropriate behaviors and of health recommendations (Gallotti et al., 2020). The propagation of false and misleading information is further amplified by the activities of bots or automated social media accounts (Caldarelli et al., 2020).

Information related to specific narratives, especially conspiracy theories and politicized news, generates and sustains polarized communities with similar information consumption patterns (Del Vicario et al., 2016). It has been noted that everyone on the Internet can produce, access, and disseminate content thereby actively participating in creation, diffusion, and reinforcement of different narratives, and that such a large heterogeneity of information fosters the aggregation of people around common interests, worldviews, and narratives (Bessi et al., 2015).

Infodemics

WHO noted the information issues early in the pandemic and, since March 2020, began to use the term of *infodemics*, coined by Rothkopf (2003) during the Severe Acute Respiratory Syndrome (SARS) epidemic, to describe the phenomenon. As Cinelli et al. (2020) observed, the term infodemic emphasizes the perils of the misinformation phenomena during the management of disease outbreaks, since it could even accelerate the epidemic spread by influencing and fragmenting both government and public response. The case of COVID-19 shows the critical impact of this new information environment: the erroneous information can negatively influence public behavior and degrade the effectiveness of public health measures (Cinelli et al., 2020).

According to WHO (2020c), an infodemic is the excessive amount of both accurate and inaccurate health information that can spread misinformation, disinformation, malinformation, and rumors during a health emergency and can hamper an effective public health response. In July 2020, WHO organized the first-ever global Conference on Infodemiology to design methods for managing an infodemic, establishing research agenda to address the issues, and build an international expert community of practice and research (WHO, 2020d).

An issue raised during the WHO (2020d) conference was the need for a lexicon to be used with infodemics. For this purpose, we will use the definitions adapted from Wardle (2018) and Wardle and Derakhshan (2017):

- *Propaganda* is true or false information spread to persuade an audience, but often has a political connotation and is connected to information produced by governments.
- *Disinformation* is false information that is deliberately created or disseminated with the express purpose to cause harm. Producers of disinformation typically have political, financial, psychological or social motivations.
- *Misinformation* is information that is false, but not distributed with intent to cause harm. Individuals who don't know a piece of information is false may spread it on social media in attempt to be helpful.
- *Malinformation* is genuine information that is shared to cause harm. This includes private or revealing information that is spread to harm a person or reputation.

The overlap between Disinformation, Misinformation and Malinformation, highlighting falseness and intention to cause harm is presented in Figure 1 (FirstDraft, 2021).

It is also important to understand the problems related to the 'fake news' expression. As Carmi et al. (2020) noted, although the term *fake news* was coined to capture the use of dis- and mis-information in news reporting it is being used by political actors in attempt to discredit news reporting and reported facts they dislike. Due to the lack of a definition of fake news, some authorities such as the UK Government (Digital, Culture, Media and Sport Committee (DCMSC), 2019) avoid using the term altogether. The UK Government stated explicitly that fake news is a poorly defined and misleading term that conflates a variety of false information, from genuine error through to foreign interference in democratic processes; instead the Government has sought to address disinformation and wider information manipulation as the deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences, either for the purposes of

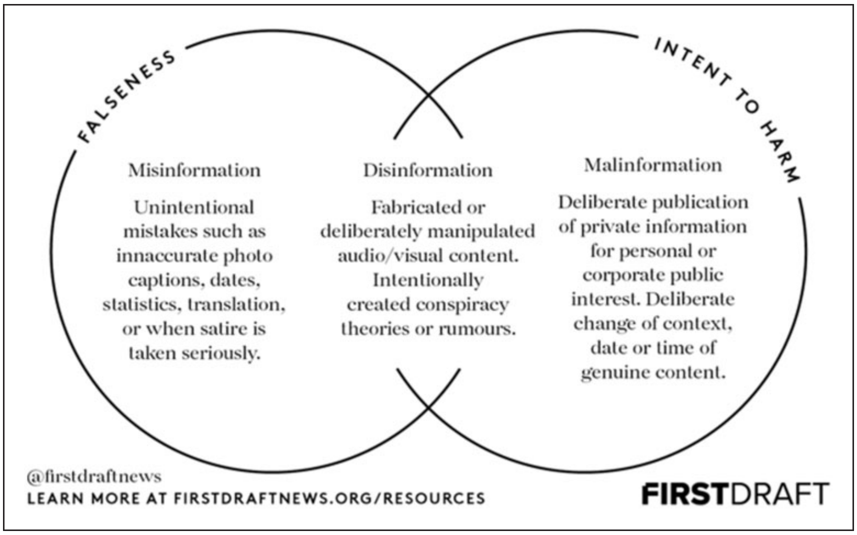
FALSENESS
INTENT TO HARM
Misinformation
Unintentional mistakes such as innaccurate photo captions, dates, statistics, translation, or when satire is taken seriously.
Disinformation
Fabricated or deliberately manipulated audio/visual content. Intentionally created conspiracy theories or rumours.
Malinformation
Deliberate publication of private information for personal or corporate public interest. Deliberate change of context, date or time of genuine content.
@firstdraftnews
LEARN MORE AT FIRSTDRAFTNEWS.ORG/RESOURCES
FIRSTDRAFT

Figure 1. The overlap between Disinformation, Misinformation and Malinformation, highlighting *falseness* and *intention to cause harm*. Adapted from: FirstDraftNews under the CreativeCommons license CC-BY-NC-ND-3.0 (FirstDraft, 2021).

causing harm, or for political, personal or financial gain (DCMSC, 2018).

Security threats

Like the WHO, international law enforcement agencies have observed the emergence of infodemics, noting the potential for new types of cybercrime that leverage both the information overflow and health-related panic (Europol, 2020). During the pandemic, Interpol (2020) observed a shift in cybercrime targets from individuals and small businesses to major corporations, governments, and critical infrastructure. Similarly, the crisis has enabled campaigns bearing the characteristics of hybrid warfare with the intelligence services of both the European Union and United States recognizing the intensified activity of foreign actors. Russia and China especially have been targeting Western democracies during the pandemic (European External Action Service's East StratCom Task Force News and Analysis (EUvsDisInfo), 2020; U.S. Department of State, 2020). As Moy and Gradon (2020) observed, hostile nation states are likely to use effects of the COVID-19 pandemic as part of their ongoing campaigns to undermine Western democratic legitimacy and active measures campaigns will continue to utilize all types of influence activities.

The threat posed by a pandemic to global, regional, and national security systems is not new and had been recognized before COVID-19. The national security and foreign policy communities have increasingly identified global health problems as threats to security (Feldbaum et al., 2006). According to Bouskill and Smith (2019), emergency health and security represent a set of bidirectional relationships that are wedded to achieving stability; they warned that the risk of epidemics and the increased proximity of human, animal, and environmental interaction are consistently underestimated.

National security has been intertwined with public health during the COVID-19 pandemic (Buckley et al., 2020). Disinformation, coupled with the sheer volume of data that needs to be analyzed, further exacerbates these challenges. Information overload makes it more difficult for people to process and understand ultimately leading to misinterpretation and poor decision making (Tylutki, 2018). The complexity of data sources, especially those coming through social media, further contributes to this (Nemr and Gangware, 2019). The problems associated with disinformation aggravating the health crisis are not new. While examining the Ebola epidemic in the Democratic Republic of the Congo, Bouskill and Smith (2019) posited that social media amplified the spread of misinformation and disinformation, induced paranoia and chaos, and complicated recovery efforts. The increasing prevalence of disinformation has become a characteristic of crises and disasters. The U.S. Department of Homeland Security (2018) stated that rumors, misinformation, and false information on social media proliferate before, during, and after disasters and emergencies.

The prevention, combating, interdiction, and mitigation of disinformation during emergencies and disasters are whole-of-society problems (M P et al., 2019). A whole-of-society problem requires all sectors of the population, including government, business, and civil society, to be involved in terms of preparedness and response. During critical health emergency situations, concerted and collaborative effort is needed to sustain essential infrastructure and mitigate impacts on the economy and the functioning of society (WHO, 2009).

The volume of data generated during COVID-19 pandemic not only interferes with the management of the disease, but also jeopardizes the capacity of governments to react (Gjorv, 2020; Pepper and Burton, 2020). Law enforcement agencies realize that data overload cannot be resolved by continuing to hire more analysts, but instead they must leverage technology (Brueggemann, 2008). Shahi and Nadini (2020) noted that detecting misinformation is a cumbersome task requiring a specially trained workforce to distinguish between fake news and real news. The velocity, veracity, and diversity of fake news available on social media platforms, newspapers, and news channels occur in multiple domains. Due to the volume of information and resource constraints, an automated tool for misinformation detection is required (Shahi and Nadini, 2020).

According to the United National Interregional Crime and Justice Research Institute (UNICRI, 2020), information technology experts are developing new solutions to identify the spread of large-scale disinformation using data science, Big Data visualization, and machine learning (ML). These technologies help researchers visualize the spread of disinformation and potentially track down the origin of false narratives. Data-driven techniques allow experts to extract information from millions of human and social bots, find similar texts, and visualize disinformation and misinformation themes. A promising starting point is the design of a repository called ReCOVery that provides multimodal information of news articles on COVID-19, including textual, visual, temporal, and network information, to facilitate a reliability assessment of news (Zhou et al., 2020).

Misinformation detection by data science and complex systems' approaches

It is not the objective of this article to explore all of the opportunities for applying technology to study and combat disinformation, but we find several areas of Big Data analysis especially promising. In the following sections we outline several interdependent theoretical approaches currently being used to understand, describe, and model the phenomenon of misinformation and disinformation. These methods come from different areas of science: sociology (complex networks), computer sciences and linguistics (text mining), mathematics and statistics (ML), and physics (agent-based modeling). When combined they create a powerful tool set to potentially restrain the propagation of dis/misinformation.

Complex networks as universal paradigm for infodemic modeling and restraining

An epidemic of information or *infodemic* has emerged from interactions between constantly evolving information sources, networks, and social groups. The development of the Internet and social media platforms made the information transmission and human-to-human communication fast and cheap. Increasingly, the platforms are misused for propagation of spam, misinformation, and so called *fake-news*. If we are to restrain these misuses, we need to better understand the structure and functions of these technologically enabled social networks. Social groups are complex. It is impossible in practice to fully describe and understand characteristics, motivation, or decisions of those involved, including choice of friends or contacts. The *complex networks* approach solves this by considering all unknown factors to be random, and focusing only on selected features of the whole network, most often expressed in terms of statistical regularities (Albert and Barabási, 2002). If the real system is too complex to understand each connection, focus should be placed on the features or behaviors that matter. For example, the density of network connections determines whether an illness becomes an epidemic or quickly disappears. It does not matter if a specific person has three or five friends, but it does matter if the average is three or five. This approach allows us to find the relations between statistical features of networks and phenomena that occur there.

A *network* (Figure 2(a)) is a system of *nodes*, e.g., Twitter or Facebook users, and *links* connecting these nodes such as followers' relations on Twitter or friendship links at Facebook. Facebook links are *undirected* (*B* and *D* at Figure 2(a)) but links on Twitter are *directed* (*A* follows *B*). It is common that directed connections between two users are not *reciprocal* (as *A* and *B*) but some can be bidirectional links (like *A* and *C*). The number of links coming to a node, followers of the user A, is called its *in-degree* $k^{in}(A)$. If the user A is following k^{out} other users then $k^{out}(A)$ is called the *out-degree* of this node. Usually $k^{in}(A) \neq k^{out}(A)$. For undirected networks such as Facebook or power grids, the node degree k_A means just the number of nearest neighbors of the node A.

Twitter, Facebook, other social media, and networks possess a *complex architecture* that can be statistically described by several measures, most tied to specific effects they produce:

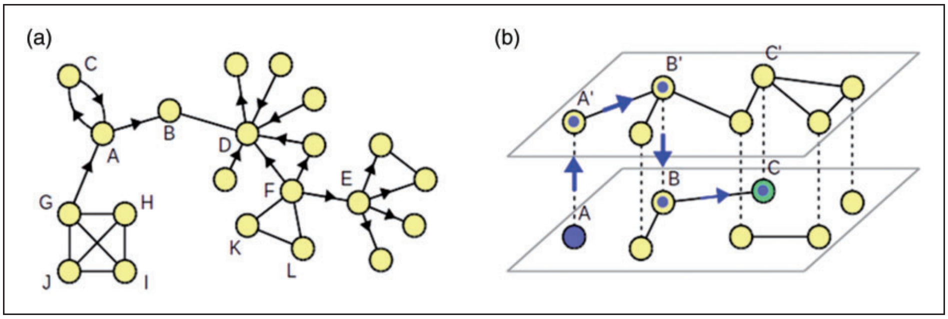

Figure 2. Social contacts or interactions can be represented as models (a) featuring complicated structure that arises from interactions of many persons. As exact quantitative description of human characteristics and decisions escapes grasp of any model, they are often described statistically and called *complex networks*. Among features of real social networks is ability of people to interact on different levels (such as offline and online, or via different online platforms), which can be represented by *multiplex network* structure (b).

- *degree distribution – number of nodes N(k) possessing certain degree k. Social networks often contain hubs, meaning nodes possessing very large degrees, such as nodes D (in-degree hub with kin = 5) and E (out-degree hub with kou t= 4) in Figure 2(a). In Twitter, the kin hubs correspond to users that possess many followers like Barack Obama, who has over 120 million followers, while kout hubs correspond to account holders, such as press agencies that observe many other Twitter users. Presence of hubs may amplify the small world property (see next point) and accelerate information spreading.*
- *distribution of distances between all pairs of nodes, understood as the number of links one needs to pass to come from one node to another. Complex networks demonstrate the small world property, which means the distance between two randomly taken nodes is small even for very large networks (Albert and Barabási, 2002). The mean distance in Twitter between any two accounts is 10, although the total number of monthly active Twitter users is over 300,000,000.*
- *betweenness-centrality of nodes. In Figure 2(a), node B has only k = 2, but is a part of many paths between other nodes (e.g. A–D, C–F, E–G). Betweenness-centrality of a node B is the fraction of shortest paths between any two nodes that pass through B. Nodes with high value (like B) are often responsible for the spreading of information between different parts of the whole network, while those with low (like K or H) do not influence the process significantly even if they participate in it.*
- *clustering that is related to loops in the network topology. The shortest loops of length 3 are connected triangles such as FKL, JGH, or GHI. They correspond to a relationship frequently met in social networks: if K and L are familiar with F then K and L are also familiar with each other. Existence of loops makes networks less vulnerable to failures but makes it more difficult to reverse-engineer the spreading process since it is usually unknown which path misinformation actually took.*
- *community structure. Cliques, groups of nodes that are fully connected with one another (e.g. GHIJ), are examples of network modules or communities, parts of a network with a denser internal connections. The modular structure of many social networks is responsible for the phenomenon of social echo chambers, the situation in which beliefs in a social group drift away from the rest of the network, reinforced by internal communication while being almost isolated from the beliefs and criticism from the outside. This enables misinformation to take hold and persist for a long time, even if a proof of it being false exists.*

The social networks can also display *multilayer* or *multidimensional* structure (Kivelä et al., 2014). The example of such a structure is a *multiplex* network as presented in Figure 2(b). Every user is represented in both network layers, e.g. vertices *A–A'*, *B–B'* and *C–C'* and so on. Each layer may correspond to a different social networking platform, such as one to Facebook with friendship connections and Twitter with follow connections. Spreading of information from one person to another can include paths in both networks, e.g., Twitter user *A* cannot reach Twitter user *C* using only the Twitter platform, but does so via Facebook accounts *A'* and *B'*, and continue via Twitter from *B* to *C* as shown by the path marked with blue arrows in Figure 2(b). Such a multilayer structure means controlling and monitoring of only one platform is likely insufficient to stop proliferation of erroneous and malicious messages. Multidimensional networks are also a challenge for detection of the origin of misinformation.

Data science approach to misinformation

Machine learning

Misinformation detection relies on text analysis accompanied by additional characteristics connected to specific properties of agents in the social network and/or the way the message propagates in the system. In order to describe methods devoted to identifying misinformation we use the fundamentals of ML and text mining. We divide ML into two large branches: *supervised* and *unsupervised* learning with the former being the core of this discussion. In this case, we are in the possession of the so-called training set allowing teaching the algorithm so that it is in position to recognize the *class* (i.e., the *label*) of a new observation based on its properties.

Various statistical techniques have been introduced during last century. These range from Straight forward ones such as Linear Discriminant Analysis (LDA) to complex approaches including the "wisdom of crowds" concept (Breiman, 2001; Hastie et al., 2009). Although many of these methods use sophisticated mathematical approaches, the basic idea is similar, we seek to divide one class from another. Figure 3(a) presents this idea in a simplified form taken from a real-world example. Suppose we have two variables (properties) describing our data, the number of words in a text document and number of users that shared this text. We also know that some of these text messages are misinformation and others not. A supervised ML method will find a line that optimally divides the data into two groups. This decision line is then used to "guess" the class, true or false information, of a new observation, represented by X sign in Figure 3(a). Due to randomness, classes often overlap making it difficult or impossible to create a perfect classifier. One usually aims to achieve high level of accuracy such as the ratio of correctly classified cases to the total number of cases.

Text representation

Misinformation is often conveyed as message text. While a reader usually does not have a significant problem comparing two documents, performance on a massive scale creating a representation of text is necessary. This representation fulfills two primary goals of text mining, first the ability to perform some statistical operation on a set of documents and second, a comparison of two or more documents. The text will need to go through a series of operations such as tokenization (division of the text into primary elements, usually single words) and *lemmatization* or *stemming* – bringing inflected words to their lemma or dictionary form, e.g., "cats" -> "cat".

Using the output of this process gives us the potential to create a bag-of-words model (BOW), in which each text is a group of unique words marked with the number of times they appear in the text. We can then go from BOW to a Vector Space Model, Figure 3(b) in which we treat each word as an axis (a direction) representing the whole document as point in this space, the proximity of two documents being the angle between them (Salton et al., 1975). If documents can be labeled with classes then we may use in a straightforward way the methodology shown in Figure 3(a), trying to find a division among some text. This can be understood, as a prototype method of misinformation detection, provided the wording in documents is a key factor distinguishing these classes.

In the last 30 years, several new methods have been created to support text interpretation using distributional semantics that assumes items with similar

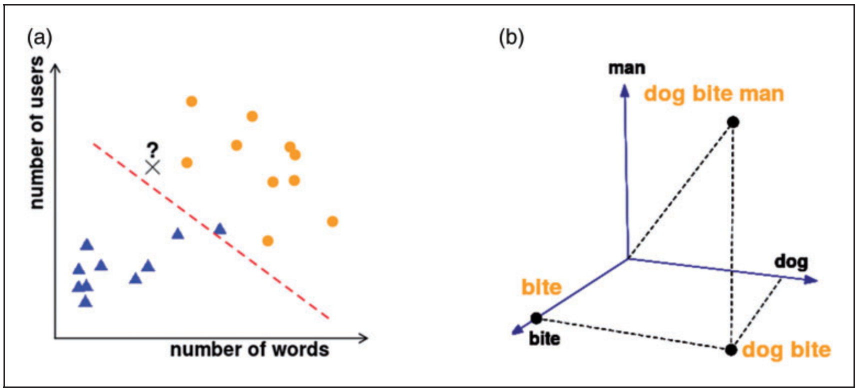

Figure 3. (a) An example of supervised ML approach: an algorithm, here an instance of Linear Discriminant Analysis, is fed with two classes of observations (e.g., triangles – misinformation and circles – true news) that are described by two properties – number of words and number of users. As an output we obtain a line that divides these two sets and allows for classification of a new coming observation. (b) Illustration of the Vector Space Model: each axis (direction) is a single word, here, "man", "bite" and "dog" and a document such as "dog bite man" is a point defined by these directions, i.e., a point in a space constructed from single terms.

distributions tend to have similar meanings. This has led to new methods such as topic modeling such as Latent Dirichlet Allocation and embedding methods like Word2vec (Blei et al., 2003; Mikolov et al., 2013). Although the methodology behind these techniques is sophisticated, they are used like the Vector Space Model and are useful in misinformation recognition. Text properties including length, complexity and sentiment are important to the analysis. In addition to representing text using models, we can also analyze its reception by an audience, assuming reader profiles. A basic text property is length, measured by the number of words or number of characters. Text-length often plays a pivotal yet nuanced role in its reception as has been shown with scientific publications (Sienkiewicz and Altmann, 2016).

Another dimension is text complexity which relates to the level of difficulty the document poses to potential readers. We can determine on reliance on unique words as opposed to more common words in a text (Herdan, 1960). Another tool is the Gunning-Fog or readability index, often used in journalism, that considers the number of complex words, in English, words with three or more syllables, used as a proxy for the number of formal years of education needed to understand the text (Gunning, 1952).

As with all forms of communications, we convey emotions, defined as rapid, unstable reactions provoked by a preceding event. In contrast to face-to-face communications in which humans are taught from early childhood to recognize emotions and use them in social context, detecting emotions in text seems to be difficult. To address this, we assume that emotions are discrete states, such as joy, hate, or sadness, and describe them in two coordinates: valence and arousal (Russell, 1980). The first value is the emotional content of the message, negative to neutral-positive, and the latter connected to the intensity of emotion, low to high. The simplest approach to extract the emotional content in text is to employ lexicon-based methods, i.e., using a dictionary with assigned values of valence and arousal; however, following this path results in low accuracy in contrast to supervised methods (Warriner et al., 2013). In this case, we follow the same scheme represented in Figure 3(a) and (b) – provided that a group of competent referees produced a set of training documents with valence and arousal. Such studies have shown to produce a very high level of accuracy (Paltoglou and Thelwall, 2010). Progress in automated detection of emotions has greatly contributed to findings pointing to phenomena in social sciences including on the collective character of online emotions and their role in sustaining discourse in e-communities or emotional dynamics in the presence of misinformation (Chmiel et al., 2011; Zollo et al., 2015).

Misinformation detection

Automatic misinformation detection, either by examining the text itself or inspecting its environment, requires that we first consider the user that initially spreads the information, the path of the message in the social network and the reaction that it has provoked (Zhou and Zafarani, 2020). When considering the message, we can address it by *knowledge-based methods* that either check fact manually using expert knowledge or automatically by such fact extractors and news aggregators as EventRegistry (Leban et al., 2014). The output of this approach gave rise to a set of online services such as PolitiFact (http://www.politifact.com/), FactCheck (https://www.factcheck.org/, or TruthOrFiction (https://www.truthorfiction.com/), that gather news. The majority of these services are connected to US politics and provide annotations for misinformation which serves as reference for testing supervised models.

Another option is to focus on *style-based features* that reveal the underlying intention behind the news that is contrary to knowledge-based methods, reflecting the authenticity of names, places, and time. These features can belong to the measures of length, complexity, and sentiment but can also express uncertainty. Factors to consider include the number of modal verbs and question marks, informality, typographical errors, profanity, and the use of passive voice. Apart from explicitly observed features it is also possible to obtain latent ones as the output of a topic model or text embedding method. Style-based features can be formed and might serve as an input for supervised methods leading to high prediction values. As an example, a combination of non-latent features classified with Random Forest resulted in over 80% accuracy (Zhou and Zafrani, 2020). Another option is to examine patterns of style-based features separately for accurate news and misinformation, as the latter is associated with higher informality, diversity, and emotional content (Zhou et al., 2020).

The third set of methods consider *propagation-based features*, applying the concept of a cascade, with a tree-like structure depicting how a message propagates in social media, and examining the maximum number of steps the news travelled, how many users were present in the cascade, or the structure of the underlying social network. These attributes may be directly applied in a ML approach or inspected for indicators that distinguish misinformation from accurate news. Misinformation spreaders form much more dense networks in comparison with the disseminators of accurate news (Zhou and Zafarani, 2019). As observed in

Twitter data, misinformation is known to penetrate deeper and wider, creating larger cascades and reaches users much faster in comparison with accurate news. This phenomenon has also been noted in a wider context of misinformation (Del Vicario et al., 2016; Vosoughi et al., 2018).

Finally, the last option is to examine the credibility of the *source of the information*, understood as the first spreader. For this we either need to reconstruct the author's network, indicating which nodes were emitting accurate or inaccurate news, or utilize specialized algorithms that can detect the source of information in the network based on observations (Sitaula et al., 2020).

In conclusion, we must point to Chołoniewski et al. (2020) who suggest that when using the data coming from the Media Bias/Fact Check service (https://mediabiasfactcheck.com), it is possible to quantify the reporting style of news outlets with a single measure based on outlets' activity. They found that news sources with a liberal bias react less in comparison with conservative outlets that react faster and more strongly. Although this approach does not specifically allow detecting misinformation, it opens new perspectives for research on the connections between the content and user activity.

Agent-based models

Agent-based models are mathematical constructs that represent people, institutions, companies, websites, or any other entities commonly called agents that possess some internal variables and are subject to a set of rules dictating what they do. Such models resemble a computer program, with specified variables (agents) and operational code (rules) expressed in a mathematical formulation. Agent-based models are easy to simulate but potentially hard to describe analytically, even if the rules are simple. Agent-based models are different from ML approaches. The focus is on most accurate representation of the variables and rules rather than on predictive power.

The strength of agent-based modeling is in predicting how rules followed by individual agents produce phenomena observed on the scale of the system. An example of that is a Susceptible-Infected-Removed/Recovered (SIR) epidemic model which could tell us whether a disease with given infectivity will spread as epidemic or quickly disappear. The more infectious the disease, the more chance it will develop into an epidemic, but the model can help predict when it might happen and how fast the number of the infected will grow (Figure 3). The relation between individual rules and system phenomena can be also modeled backwards, as agent-based models that reproduce observed phenomena could be used for explaining the unknown rules.

The spread information is similar to the spread of infectious disease (Daley and Kendall, 1965). Information, like a pathogen, multiplies as it spreads from person-to-person, place-to-place, or online. At the base of many models aimed specifically at information spreading are epidemic models such as the Susceptible-Infected (SI) or the SIR models (Hethcote, 2000). Both assume that individuals or agents have single discrete variable representing their state: Susceptible (not infected yet, but can be in the future), Infected (and spreading infection to others), or Removed/Recovered (in the SIR model, the person recovered, became isolated, or died, and no longer spreads infection). A Susceptible agent that interacts with Infected agent has a fixed probability β to become Infected itself, and in SIR model, an Infected agent has fixed probability γ to become Removed per unit of time. The population SIR model, where everyone in a population can interact with every else gave us important results including predicting the increase of infections in time (Figure 4) and, based on reproduction ratio β/γ, whether the infection spreads ($\beta/\gamma > 1$) or declines ($\beta/\gamma < 1$). The critical β/γ point is called epidemic threshold.

In agent-based models with network of interactions, the network can be a critical factor. Reproduction rates now depend on average number of neighbors since an infected agent exposes all neighbors simultaneously. If the network has large hubs, like many actual social networks, then the epidemic threshold vanishes and it becomes statistically likely that infection would result in global epidemic (Pastor-Satorras and Vespignani, 2001). But local structure such as clustering may preserve the existence of threshold even in scale-free networks (Eguıluz and Klemm, 2002). The SI model is a simplified version of SIR and has no concept of

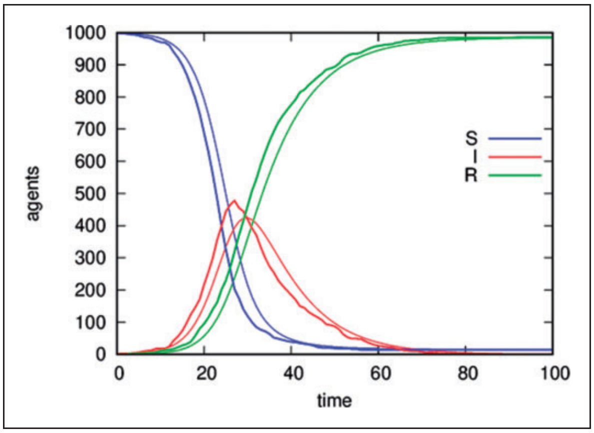

Figure 4. The number of Susceptible, Infected and Removed agents in SIR epidemic model in time. Thin lines show numbers from analytical description, while thick ones show an example numerical simulation of agent-based model.

epidemic threshold since infection of all agents directly or indirectly connected with infected one is guaranteed over time. The spreading rate β impacts the speed at which the infection will progress, just as it does in full SIR model.

The SIR model has been expanded with different variants (Hethcote, 2000) including more states (e.g., MSEIR) or adding extra variables (e.g., SIS, SIRS). Disease-specific models have been introduced, such as the SEIRA model with an asymptomatic stage, for COVID-19 modeling (Contreras et al., 2020). It is possible to adapt models to fit information spreading more precisely by adding an Active super-spreader state, which can describe the behavior of microblogs better (Liu et al., 2016). A noteworthy modification in rumor models is the introduction of stifling that accounts for long-known discrepancies and turn them into additional rules in a SIR-like model. Infected agents can become Recovered if they interact with another Infected or Recovered agent (Daley and Kendall, 1965; Moreno et al., 2004). This recognizes the fact that realization that others already know the rumor, it is old news and not worth spreading further. Similar model fit Twitter hashtag use better than most epidemic models (Skaza and Blais, 2017).

All the models above assume that people want to and will share information that they get. But often that is not exactly what happens. Most people only want to pass information that we agree with. Sharing information then is dependent on wanting to convince others instead of informing them. This proclivity may be modeled using the threshold model in which a person must know some of their friends already supports a given message would spread it themselves (Watts, 2002). This is similar to innovation or opinion spreading than simple messaging, but misinformation or conspiracy theories are often more opinion than fact.

There are also models, such as SAR, which interpolate between regular threshold and typical SIR model (Wang et al., 2015). Of special interest may be various co-evolutionary models in which both the state of agents and the links between them vary. These are termed adaptive networks (Gross and Blasius, 2008). A significant result of co-evolutionary models is fragmentation of a network in an opinion model (Vazquez et al., 2007). The interaction between the opinion-based spread of information and the changing the shape of contact network can lead to polarization of opinions and creation of *echo chambers* (Törnberg, 2018).

The study of how conspiracy theories and scientific news spread online shows that the models, including polarization, can explain the characteristics of real misinformation cascades (Del Vicario et al., 2016). Other works consider the role of similarity of interests in information transmission or sentimental content of the messages and emotions of people involved in transmission of true and false information (Myers and Leskovec, 2014; Zollo et al., 2015). Most studies incorporate these as adjustments to probability based on data, rather than building a dedicated model. The effects of emotions and interests impacting the spread of information from a model standpoint require more research.

Locating sources of (mis)information

Identifying the source of information, whether accurate or not, can be very difficult. The original source of retweet or share of previous post may seem obvious but that is not necessarily the case. People may receive information and pass the content in their own message, the working may be altered, or the content changed entirely (Adamic et al., 2016). It must be questioned if the original source on content can even be found online. There are, however, methods that can locate likely sources provided we can track or identify the information itself.

If the entire history of how specific content has spread, determining the source is a straight forward task. But that's rarely the case. There are common challenges with locating patient zero in epidemic outbreaks, locating the first postings of specific, or the origin of malicious Trojan software spreading. Generally, there is limited knowledge about the actual process by which the contagion is spread and there is a complex network of interactions that is responsible for the spreading process. Tracing misinformation is further complicated because it may be partially shared on private social networks such as Facebook.

There are two scenarios that are most common in the spread of information. Up to a specific point in time, it can be known where it has spread and the state of the spreading process can be assessed. The spread of information to specific agents can be determined through, for example, public pages on Facebook and the exact time information can be identified or there is time information from a set of observers in the network. In both cases, there are two things that create difficulty. First is the complex, deterministic structure of networks and, second, the stochastic nature of the spreading process. The first issue can often be solved by mathematical methods, but the second means that the information is fundamentally uncertain. All algorithms assume we know exact or at least the approximate network responsible for the spreading process. This unfortunately places significant restrictions on when the methods can be applied as the network structure may be partially or completely hidden.

There are several methods developed for estimating the source of the spreading process in complex networks, but most are snapshot- or observer-based (Jiang et al., 2017). Snapshot methods often calculate an estimator for each agent, which shows how likely it is that a given agent is the actual source. This can be a simple measure such as betweenness centrality for an infected subgraph, but more specialized centrality measures such as rumor centrality, and Jordan centrality based on eccentricity perform better and can locate the true source to a certain degree even in incomplete snapshots (Luo et al., 2014; Shah and Zaman, 2011; Zhu and Ying, 2016). It is also possible to calculate more complex estimators based, for example, on dynamic message passing algorithms with better accuracy than centrality-based methods, in most cases (Lokhov et al., 2014).

Observer-based methods also make use of estimators calculated for potential sources. These range from correlations between arrival times and distances from a supposed source to comparisons between actual arrival times and those expected from the spreading model. If the spreading uses independent cascades, then the time to arrive at an agent should correlate with distance from the true source. By observing this relationship, the most likely source can be estimated. These methods offer relatively good accuracy and are resilient to variation in the spreading process including the precise means or variances of the time to spread along each link (Brockmann and Helbing, 2013; Xu et al., 2019).

An additional approach is to calculate the distribution of times of arrival at all observers for each potential source, then compare this with the actual observations (Pinto et al., 2012). This approach has better accuracy than simple distance–time correlation methods because it uses the correlations between arrival times at observers. It is, however, computationally expensive and approximates a tree network which leads to discrepancies for networks with a local structure. These drawbacks have been alleviated by optimized methods that are much faster, or those that take full network structure into account, offering improved accuracy at a higher computational cost (Gajewski et al., 2019; Paluch et al., 2018). Other methods seek to reduce the assumptions required about the spreading process, either by estimating the time distribution from the data or by considering local correlation of arrival times and distances (She et al., 2016; Wang and Sun, 2020).

The methods discussed have been developed from propagation models and tested on either artificial or real networks. The scarcity of time data on real information cascades means that they have not been tested against actual information spreading processes in social networks. Some, however, have been developed for and tested against spreading of infectious diseases (Brockmann and Helbing, 2013; Pinto et al., 2012). The accuracy of these methods varies with both spreading process parameters including randomness, network structure, and observers for observer-based methods. The placement of observers can also play a role although random placement performs quite well (Paluch, 2020).

While locating a source is particularly important, the contact network is particularly important. Often it is the network that identify the source with certainty. The network may be estimated by examining the spreading history. The structure of real contact networks may be determined that might be otherwise unobtainable. This is known as network tomography and relies mainly on a method known as compressed sensing to determine most likely network based on history of spreading processes (Han et al., 2015; Kakkavas et al., 2020; Shen et al., 2014). A single spreading cascade does not have enough information to estimate $N (N - 1)$ potential connections from dynamic information events of at most N agents. A multiple number of information spreading events may be required for reliable estimates.

Conclusions and recommendations

The WHO's initiative on infodemiology is a fundamental advance in handling the volume of information flow that impedes the progress of health policy and affects behavior on a global scale (WHO, 2020d). Disinformation campaigns and the propagation of misinformation are among the major impediments to the effective implementation of public health strategies and, in the case of the current pandemic, to the mitigation of the crisis affecting lives worldwide. According to threat assessments related to disinformation, the prevention, interdiction, and mitigation must be given priority using a full spectrum of measures (Europol, 2020; Interpol, 2020; UNICRI, 2020; U.S. Department of State, 2020).

One of the issues to be addressed is the massive volume of data. The information flow with a broad range of narratives, both true and false, cannot be analyzed quickly, accurately, and efficiently even aided by fact-checking organizations. The enormous volume of data renders traditional approaches impossible.

The opportunities to analyze large data sets, originating in the numerous types of media must rely on the application of properly calibrated Information Technologies. Traditional media, electronic media, and, perhaps most critically, social media need to be included in the analysis. Medical information gathered at national, regional, and local levels must inform this effort. Finally, spatial and temporal data related to demographics and mobility must be considered. Technology tools may enable national and international agencies responsible both for the administration of health policies and policing of disinformation to examine and process real-time information on potential threats.

It is essential to evaluate the opportunities for the use of artificial intelligence, data science, and ML to aid responsible government agencies, health care providers, news media of all types, and civil society organizations to process and analyze information to deliver reliable information to stakeholders and decision makers. Such approaches require collaborative action on a global scale. Crises such as the ongoing COVID-19 pandemic are borderless and they cannot be handled by individual states, regardless of their wealth and power. The implementation of integrated preventive measures necessitates the internationalization of infodemic management.

There are several potential techniques for the application of Big Data analysis to the infodemic problem. There is not a definitive list of the opportunities but these will shape the direction that the Information Technologies applied to countering disinformation should take. The effectiveness of the methods applied to combat and prevent disinformation relies heavily not only on the internationalization of this effort, but also on a multidisciplinary approach. The methodology represented by the Crime Science could be applied to the understanding and mitigation of disinformation. Crime Science is the application of science to crime control: reducing crime by its prevention, disruption, and detection (Laycock, 2008). It focuses on the near or immediate causes and circumstances of crime and is problem driven and evidence based, relying heavily on collecting and analyzing empirical and verifiable data (Gradon, 2013). In essence, it is the use of scientific methods and knowledge from many disciplines to the development of practical and ethical ways to reduce crime and increase security (Wortley et al., 2018). It has already been proven that Crime Science can be successfully applied to the "non-traditional" categories of delinquency such as cybercrime (Hartel et al., 2010).

As an interdisciplinary field, Crime Science offers techniques to the problem of dissemination of false information with intent to cause harm, instigating malign influence operations, or countering organized disinformation campaigns. Utilizing Information Technologies, Intelligence Analysis, Natural Language Processing, and Psychology would provide empirically tested solutions for the prevention, interdiction, and mitigation of disinformation. This strategy would support the early detection of disinformation, the study of its propagation, and the exposure of its original sources. Empirically investigating incidents would explain disinformation rules and patterns as well as the factors influencing the spread of false information through networks. It would also explain the online environment of disinformation creation and propagation and describe what affects the choices to share false information on particular occasions. Finally, Crime Science would highlight potential interventions and measures. Thus, it would allow for the formulation of prevention and mitigation strategies. The creation of an international, interdisciplinary infodemiology research center of excellence under the auspices of the WHO could coordinate the design of prevention and mitigation countermeasures.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially funded (JH, JS, KS) by "IDUB against COVID-19" project granted by Warsaw University of Technology under the program Excellence Initiative: Research University (IDUB).

ORCID iDs

Wesley R Moy https://orcid.org/0000-0001-9747-7638
Julian Sienkiewicz https://orcid.org/0000-0003-2097-1499
Krzysztof Suchecki https://orcid.org/0000-0001-6670-8528

References

Adamic L, Lento T, Adar E, et al. (2016) *Information evolution in social networks. In: WSDM 2016 – Proceedings of the 9th ACM international conference on web search and data mining*, San Francisco, CA, USA, February 2016, pp.473–482. New York NY: Association for Computing Machinery.

Albert R and Barabási AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47–97.

Bessi A, Coletto M, Davidescu GA, et al. (2015) Science vs conspiracy: Collective narratives in the age of misinformation. *PLoS ONE* 10(2): e0118093.

Blei DM, Ng AY and Jordan MI (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

Bouskill KE and Smith E (2019) Global health and security. Threats and opportunities. RAND Perspective, December. Available at: https://www.rand.org/pubs/perspectives/PE332.html (accessed 6 December 2020).

Breiman L (2001) Random forests. *Machine Learning* 45(1): 5–32.

Brockmann D and Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science (New York, N.Y.)* 342(6164): 1337–1342.

Brueggemann CE (2008) *Mitigating information overload: The impact of "context-based approach" to the design of tools for intelligence analysts*. Master's Thesis, Naval Postgraduate School Monterey, USA. Available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a479839.pdf (accessed 3 December 2020).

Buckley C, Clem R and Herron E (2020) National security implications of the COVID-19 crisis: The urgent need to build state capacity. In: *Minerva* r*esearch* i*nitiative*, 16 April. Available at: https://minerva.defense.gov/Owl-In-the-Olive-Tree/Owl_View/Article/2152823/national-security-implications-of-the-covid-19-crisis-the-urgent-need-to-build (accessed 3 December 2020).

Caldarelli G, De Nicola R, Del Vigna F, et al. (2020) The role of bot squads in the political propaganda on Twitter. *Communications Physics* 3: 81.

Carmi E, Yates SJ, Lockley E, et al. (2020) Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review* 9(2): 1–22.

Chmiel A, Sienkiewicz J, Thelwall M, et al. (2011) Collective emotions online and their influence on community life. *PLoS ONE* 6(7): e22207.

Chołoniewski J, Sienkiewicz J, Dretnik N, et al. (2020) A calibrated measure to compare fluctuations of different entities across timescales. *Scientific Reports* 10(1): 20673.

Cinelli M, Quattrociocchi W, Galeazzi A, et al. (2020) The COVID-19 social media infodemic. *Scientific Reports* 10: 16598. https://doi.org/10.1038/s41598-020-73510-5

Contreras S, Villavicencio H, Medina-Ortiz D, et al. (2020) A multi-group seira model for the spread of covid-19 among heterogeneous populations. *Chaos, Solitons and Fractals* 136: 109925.

Daley D and Kendall D (1965) Stochastic rumours. *IMA Journal of Applied Mathematics* 1(1): 42–55.

Del Vicario M, Bessi A, Zollo F, et al. (2016) The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113(3): 554–559.

Digital, Culture, Media and Sport Committee (DCMSC) (2018) *Disinformation and 'fake news'*. Interim Report, Government Response to the Committee's Fifth Report of Session 2017–2019, 23 October. Westminster: House of Commons.

Digital, Culture, Media and Sport Committee (DCMSC) (2019) *Disinformation and 'fake news'*. Final Report – Eighth Report of Session 2017–2019, 14 February. Westminster: House of Commons.

Eguı́luz V and Klemm K (2002) Epidemic threshold in structured scale-free networks. *Physical Review Letters* 89(10): 108701.

European External Action Service's East StratCom Task Force News and Analysis (EUvsDisinfo) (2020) New stories, old plots. How the pro-Kremlin disinformation network capitalises on COVID-19 outbreak. Available at: https://euvsdisinfo.eu/new-stories-old-plots/ (accessed 10 November 2020).

Europol (2020) Catching the virus cybercrime, disinformation and the COVID-19 pandemic. Report of the European Union Agency for Law Enforcement Cooperation, 3April. Hague: Europol.

Feldbaum H, Patel P, Sondorp E, et al. (2006) Global health and national security: The need for critical engagement. *Medicine, Conflict, and Survival* 22: 192.

FirstDraft (2021) First draft news flashcards for use in publications and under the Creative Commons license (CC BY-NC-ND 3.0). Available at: https://firstdraftnews.org/first-draft-flashcards/ (accessed 19 February 2021).

Gajewski LG, Suchecki K and Holyst JA (2019) Multiple propagation paths enhance locating the source of diffusion in complex networks. *Physica A: Statistical Mechanics and Its Applications* 519: 34–41.

Gallotti R, Valle F, Castaldo N, et al. (2020) Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nature Human Behaviour* 4: 1285–1293.

Gjorv GH (2020) Coronavirus, invisible threats and preparing for resilience. In: *NATO review – Opinion,* a*nalysis and* d*ebate on* s*ecurity* i*ssues*, 20 May. Available at: https://www.nato.int/docu/review/articles/2020/05/20/coronavirus-invisible-threats-and-preparing-for-resilience/index.html (accessed 1 December 2020).

Gradon K (2013) Crime science and the internet battlefield: Securing the analog world from digital crime. *IEEE Security & Privacy* 11(5): 93–95.

Gradon K (2020) Crime in the time of the plague: Fake news pandemic and the challenges to law-enforcement and intelligence community. *Society Register* 4(2): 133–148. https://doi.org/10.14746/sr.2020.4.2.10

Gross T and Blasius B (2008) Adaptive coevolutionary networks: A review. *Journal of the Royal Society, Interface* 5(20): 259–271.

Gunning R (1952) Technique of Clear Writing. New York, NY: McGraw-Hill.

Han X, Shen Z, Wang WX, et al. (2015) Robust reconstruction of complex networks from sparse data. *Physical Review Letters* 114(2): 028701.

Hartel PH, Junger M and Wieringa RJ (2010) Cyber-crime science = Crime Science + Information Security. *CTIT Technical Report Series; No. 10-34*. Centre for Telematics and Information Technology. Available at: https://research.utwente.nl/en/publications/cyber-crime-science-crime-science-information-security (accessed 19 February 2021).

Hastie T, Tibshirani R and Friedman J (2009) *Elements of Statistical Learning*. New York, NY: Springer.

Herdan G (1960) *Language as Choice and Chance*. Berlin: Springer.

Hethcote H (2000) Mathematics of infectious diseases. *SIAM Review* 42(4): 599–653.

Interpol (2020) *Cybercrime: COVID-19 impact. Report of the Interpol, 4 August*. Lyon: Interpol.

Jiang J, Wen S, Yu S, et al. (2017) Identifying propagation sources in networks: State-of-the-art and comparative studies. *IEEE Communications Surveys & Tutorials* 19(1): 465–481.

Johns Hopkins University (2020) COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available at: https://coronavirus.jhu.edu/map.html (accessed 06 April 2021).

Kakkavas G, Gkatzioura D, Karyotis V, et al. (2020) A review of advanced algebraic approaches enabling network tomography for future network infrastructures. *Future Internet* 12(2): 20.

Kivelä M, Arenas A, Barthelemy M, et al. (2014) Multilayer networks. *Journal of Complex Networks* 2(3): 203–271.

Laycock G (2008) Special edition on crime science. *Policing: A Journal of Policy and Practice* 2(2): 149–153.

Leban G, Fortuna B, Brank J, et al. (2014) Event registry: Learning about world events from news. In: *Proceedings of the 23rd international conference on World Wide Web, WWW '14 Companion*. New York, NY: Association for Computing Machinery. ISBN 9781450327459, pp.107–110.

Liu Y, Wang B, Wu B, et al. (2016) Characterizing super-spreading in microblog: An epidemic-based information propagation model. *Physica A* 463: 202–218.

Lokhov A, Mézard M, Ohta H, et al. (2014) Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E* 90(1): 012801.

Luo W, Tay WP and Leng M (2014) How to identify an infection source with limited observations. *IEEE Journal of Selected Topics in Signal Processing* 8(4): 586–597.

M P, Alexander S, Cambridge A, et al. (2019) Combatting Targeted Disinformation Campaigns: A whole-of-society issue. Public-Private Analytic Exchange Program. Office of the Director of National Intelligence, Oct. 2019. Available at: https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf (accessed 6 December 2020).

Mikolov T, Sutskever I, Chen K, et al. (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th international conference on neural information processing systems – Volume 2, NIPS'13. Red Hook, NY: Curran Associates Inc., pp.3111–3119.

Moreno Y, Nekovee M and Pacheco A (2004) Dynamics of rumor spreading in complex networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 69(6–2): 066130–066131.

Moy W and Gradon K (2020) COVID-19 effects and Russian disinformation. *Homeland Security Affairs 16:* 8.

Myers S and Leskovec J (2014) The bursty dynamics of the twitter information network. In: *WWW 2014 – Proceedings of the 23rd International Conference on World Wide Web*, Seoul, Korea, April 2014, pp.913–923. New York NY: Association for Computing Machinery.

Nemr C and Gangware W (2019) *Weapons of Mass Distraction. Foreign State Sponsored Disinformation in the Digital Age*. Washington, DC: Park Advisors.

Organisation for Economic Co-operation and Development (2020) The territorial impact of COVID-19: Managing the crisis across levels of government. Available at: http://www.oecd.org/coronavirus/policy-responses/the-territorial-impact-of-covid-19-managing-the-crisis-across-levels-of-government-d3e314e1/#biblio-d1e7012 (accessed 15 November 2020).

Paltoglou G and Thelwall M (2010) A study of information retrieval weighting schemes for sentiment analysis. In: *Proceedings of the 48th annual meeting of the association for computational linguistics*, Uppsala, Sweden, 11–16 July 2010, pp.1386–1395. Uppsala: Association for Computational Linguistics. Available at: https://www.aclweb.org/anthology/P10-1141 (accessed 8 December 2020).

Paluch R, Gajewski L, Hołyst J, et al. (2020) Optimizing sensors placement in complex networks for localization of hidden signal source: a review. *Future Generation Computer Systems* 112: 1070–1092.

Paluch R, Lu X, Suchecki K, et al. (2018) Fast and accurate detection of spread source in large complex networks. *Scientific Reports* 8(1): 1–10.

Pastor-Satorras R and Vespignani A (2001) Epidemic spreading in scale-free networks. *Physical Review Letters* 86(14): 3200–3203.

Pepper MS and Burton SG (2020) Sheer volume of misinformation risks diverting focus from fighting coronavirus. In: *The Conversation,* 29 April. Available at: https://theconversation.com/sheer-volume-of-misinformation-risks-diverting-focus-from-fighting-coronavirus-137408 (accessed 28 November 2020).

Pinto P, Thiran P and Vetterli M (2012) Locating the source of diffusion in large-scale networks. *Physical Review Letters* 109(6): 068702.

Rothkopf DJ (2003) When the buzz bites back. *The Washington Post*, 11 May, B01.

Russell JA (1980) A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6): 1161–1178.

Salton G, Wong A and Yang CS (1975) A vector space model for automatic indexing. *Communications of the ACM* 18(11): 613–620.

Shah D and Zaman T (2011) Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory* 57(8): 5163–5181.

Shahi GK and Nadini D (2020) FakeCovid – A multilingual cross-domain fact check news dataset for COVID-19. In: *CySoc 2020 international workshop on cyber social threats ICWSM 2020*, Atlanta, USA, 8 Jun 2020.

She X, Li X, Liu Y, et al. (2016) A novel source locating strategy without consistent assumptions. In: *12th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)*, Changsha, China, 13–15 August 2016, pp.702–708.

Shen Z, Wang WX, Fan Y, et al. (2014) Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nature Communications* 5: 4323.

Sienkiewicz J and Altmann EG (2016) Impact of lexical and sentiment factors on the popularity of scientific papers. *Royal Society Open Science* 3(6): 160140.

Sitaula N, Mohan CK, Grygiel J, et al. (2020) *Credibility-Based Misinformation Detection*. Cham: Springer International Publishing, pp.163–182.

Skaza J and Blais B (2017) Modeling the infectiousness of twitter hashtags. *Physica A: Statistical Mechanics and Its Applications* 465: 289–296.

Törnberg P (2018) Echo chambers and viral misinformation: Modeling misinformation as complex contagion. *PLoS ONE* 13(9): e0203958.

Tylutki K (2018) The information of a mass destruction range – OSINT in intelligence activities. *Internal Security Review* 19: 384–404.

United Nations (2020) Describing COVID-19 pandemic as wake-up call, dress rehearsal for future challenges, secretary-general opens annual general assembly debate with vision for solidarity. Available at: https://www.un.org/press/en/2020/ga12268.doc.htm (accessed 20 October 2020).

United Nations Interregional Crime and Justice Research Institute (UNICRI) (2020) Stop the virus of disinformation. The risk of malicious use of social media during COVID-19 and the technology options to fight it. Report of UNICRI, November. Torino: UNICRI.

U.S. Department of Homeland Security (2018) Countering false information on social media in disasters and emergencies. Social Media Working Group for Emergency Services and Disaster Management Report. Washington, DC: DHS Science&Technology.

U.S. Department of State (2020) *Special briefing on disinformation and propaganda related to COVID-19*. USDS Report, 27 March. Washington, DC: USDS.

Vazquez F, González-Avella J, Eguíluz V, et al. (2007) Time-scale competition leading to fragmentation and recombination transitions in the coevolution of network and states. *Physical Review E* 76(4): 046120.

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science (New York, N.Y.)* 359(6380): 1146–1151.

Wang HJ and Sun KJ (2020) Locating source of heterogeneous propagation model by universal algorithm. *EPL (Europhysics Letters)* 131(4): 48001.

Wang W, Tang M, Zhang HF, et al. (2015) Dynamics of social contagions with memory of nonredundant information. *Physical Review E* 92(1): 012820.

Wardle C (2018) Information Disorder: The Essential Glossary. Harvard, MA: Harvard Kennedy School.

Wardle C and Derakhshan H (2017) *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe Report DGI (2017)09, 27 September. Brussels: Council of Europe.

Warriner AB, Kuperman V and Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4): 1191–1207.

Watts D (2002) A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America* 99(9): 5766–5771.

World Health Organization (WHO) (2009) Whole-of-society pandemic readiness: WHO guidelines for pandemic preparedness and response in the non-health sector. Available at: https://www.who.int/influenza/preparedness/pandemic/2009-0808_wos_pandemic_readiness_final.pdf (accessed 19 February 2021).

World Health Organization (WHO) (2020a) Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Available at: https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed 10 November 2020).

World Health Organization (WHO) (2020b) WHO Director-General's opening remarks at the Mission briefing on COVID-19, 12 March. Available at: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19—-12-march-2020 (accessed 10 November 2020).

World Health Organization (WHO) (2020c) *Coronavirus disease 2019 (COVID-19)*. Situation report - 45. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf (accessed 20 November 2020).

World Health Organization (WHO) (2020d) *1st WHO infodemiology conference: How infodemics affect the world & how they can be managed*. Conference booklet. Available at: https://www.who.int/docs/default-source/epi-win/infodemic-management/infodemiology-scientific-conference-booklet.pdf?sfvrsn = 179de76a_4 (accessed 21 November 2020).

Wortley R, Sidebottom A, Tilley N, et al. (2018) *Routledge Handbook of Crime Science*. New York, NY: Routledge.

Xu S, Teng C, Zhou Y, et al. (2019) Identifying the diffusion source in complex networks with limited observers. *Physica A: Statistical Mechanics and Its Applications* 527: 121267.

Zhou X, Mulay A, Ferrara E, et al. (2020) ReCOVery: A multimodal repository for COVID-19 news credibility research. In: *Proceedings of the 29th ACM international conference on information and knowledge management (CIKM '20)*, Virtual Event, Ireland, 19–23 October 2020. New York, NY: ACM. Available at: https://doi.org/10.1145/3340531.3412880 (accessed 9 December 2020).

Zhou X and Zafarani R (2019) Network based fake news: A pattern-driven approach. *ACM SIGKDD Explorations Newsletter* 21(2): 48–60.

Zhou X and Zafarani R (2020) A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys* 53(5): 109.

Zhu H, Wei L and Niu P (2020) The novel coronavirus outbreak in Wuhan, China. *Global Health Research and Policy* 5: 6.

Zhu K and Ying L (2016) Information source detection in the SIR model: A sample-path-based approach. *IEEE/ACM Transactions on Networking* 24(1): 408–421.

Zollo F, Novak P, Del VM, et al. (2015) Emotional dynamics in the age of misinformation. *PLOSONE* 10(9): e138740.