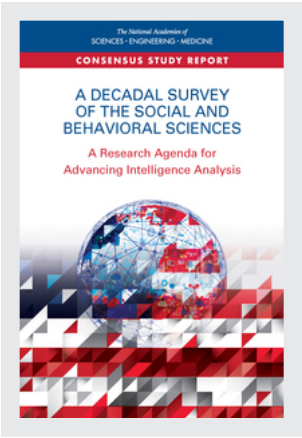

A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis (2019)

DETAILS

340 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-48761-0 | DOI 10.17226/25335

CONTRIBUTORS

Committee on a Decadal Survey of Social and Behavioral Sciences for Applications to National Security; Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2019. *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25335.

6
Integrating Social and Behavioral Sciences (SBS) Research to Enhance Security in Cyberspace

Cyber-related developments have both dramatically altered the nature of security threats and expanded the landscape of potential tools for countering those threats. Experts from multiple disciplines, including electrical engineering, software engineering, computer science, and computer engineering, have a laser focus on cybersecurity, but that focus has been primarily on technical or data challenges, such as identification and prevention of malware, prevention of denial-of-service attacks, self-fixing code, unauthorized data breaches, tools for the cyber analyst, and privacy. Indeed, cybersecurity is often characterized as the set of techniques used to protect the integrity of networks, programs, and data from attack, damage, or unauthorized access.[1] These techniques have undisputed value, but they address only technological challenges, not the human behaviors and motivations that shape those challenges.

The tools of cybersecurity have obvious relevance for national security. Intelligence analysts, however, seek to understand a different but related set of critical problems—those that involve cyber-mediated communication (communication that takes place through computer networks). To understand this phenomenon, it is necessary to integrate insights about constantly evolving technology with understanding of fundamentally human phenomena. The emerging field of social cybersecurity science has developed to fill this need.[2] Researchers in this field build on foundational work in the social and behavioral sciences (SBS) to characterize cyber-mediated changes in individual, group, societal, and political behaviors and outcomes, as well as to support the building of the cyber infrastructure needed to guard against cyber-mediated threats. This chapter describes this emerging discipline, explores the opportunities it offers for the Intelligence Community (IC), illustrates its relevance to intelligence analysis with an example, and describes research needed in the coming decade to fully exploit the field's applications to the work of intelligence analysis.

WHAT IS SOCIAL CYBERSECURITY SCIENCE?

The field of social cybersecurity developed to meet a national need. It was developed by researchers with backgrounds in numerous fields to meet two primary objectives:

[1]See, e.g., https://searchsecurity.techtarget.com/definition/cybersecurity [December 2018] and https://www.paloaltonetworks.com/cyberpedia/what-is-cyber-security [December 2018].

[2]The term "social cybersecurity" is also sometimes used to refer to cyber-mediated security threats themselves, with emphasis on the human, as opposed to the technological, aspects of those threats. Examples of such threats are recruitment of members of covert groups and their training in social media, the spread of fake news and disinformation, attacks on democracy through manipulation of how citizens receive news, the fomenting of crises by creating a perception of the rampant spread of disease or state instability, phishing and spear phishing attacks (i.e., attempts to obtain sensitive or protected information online by posing as a trustworthy entity), recruitment of individuals to act as insider threats (see Chapter 5) through social media, and online brand manipulation and rumors designed to destroy corporations.

- characterize, understand, and forecast cyber-mediated changes in human behavior and in social, cultural, and political outcomes; and
- build a social cyber infrastructure that will allow the essential character of a society to persist in a cyber-mediated information environment that is characterized by changing conditions, actual or imminent social cyberthreats, and cyber-mediated threats.

Scientists in this field seek to develop the technology and theory needed to assess, predict, and mitigate instances of individual influence and community manipulation in which either humans or bots attempt to alter or control the cyber-mediated information environment (Carley et al., 2018). While researchers in the social cybersecurity area come from a large number of disciplines, many identify themselves as computational social scientists. The field is rapidly expanding to meet a growing need; the number of academic papers published in this area has risen exponentially in the past 10 years (Carley et al., 2018). The number of researchers in this area is also growing because of widespread concern about the global consequences of such social cybersecurity attacks as disinformation campaigns, social media manipulation, and phishing to develop insider threats. Many researchers came to this area of study independently, but they are quickly coalescing as a formal discipline through participation in emerging groups, such as the social cybersecurity working group,[3] and domestic and international conferences, such as the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation.[4]

Experts in cybersecurity focus on attacks made on and through the cyber infrastructure that are intended to interfere with technology, steal or destroy information, or steal money or identities (Reveron, 2012). While cybersecurity experts do draw on social science research (see Box 6-1), social cybersecurity researchers have a different approach: they focus on activities aimed at influencing or manipulating individuals, groups, or communities, particularly activities that have large consequences for social groups, organizations, and countries. The solutions to some problems, such as denial-of-service attacks, malware distribution, and insider threats, require both types of expertise, but the emphasis in the two fields is quite different. What links researchers in social cybersecurity is that they

- take the sociopolitical context of cyber activity into account both methodologically and empirically;
- integrate theory and research on influence, persuasion, and manipulation with study of human behavior in the cyber-mediated environment; and
- focus on identifying operationally useful applications of their research.

BOX 6-1
Social and Behavioral Science (SBS) Fields in Cybersecurity

Cybersecurity involves humans who may be attackers, defenders, network administrators, computer users, organizations, and even children surfing the Internet. It is not surprising, then, that SBS research has been applied to many aspects of cybersecurity. In many cases, these

[3]Available: www.social-cybersecurity.org [February 2019].
[4]See http://sbp-brims.org/2019/about [January 2019].

applications have been focused on cybersecurity for the end user. This work has addressed questions about cyberhygiene (practices adopted by system users to maintain the system's health and security), data privacy, passwords, user authentication, identity theft, and end-user beliefs and mindsets and appears regularly in the Proceedings of the Symposium on Usable Privacy and Security.[5]

Other research has applied SBS methods to understand cybersecurity operations. D'Amico and colleagues have used task analyses to understand the work of cyber defenders (D'Amico and Whitley, 2008; D'Amico et al., 2005; Gutzwiller et al., 2016, 2018; Horn and D'Amico, 2011; Vieane et al., 2016). Similarly, cognitive modeling in the form of instance-based learning theory (a type of learning algorithm in which comparison to prior examples is the basis for analysis) has been used to model the analyst and the detection of attacks (Dutt et al., 2011, 2013).

Human factors research has played a prominent role in attempts to improve cybersecurity analysis by understanding the effects of distraction (Gutzwiller et al., 2018), situation awareness (Gutzwiller et al., 2016; Liu et al., 2017), and interruption (Vieane et al., 2017) on the cyber defender's task. In addition, research on teamwork among cyber analysts has shown that while it is generally minimal in practice, it enables analysts to detect a wider array of threats (Rajivan and Cooke, 2018). Armed with a better understanding of the cyber defender's task, human factors researchers have developed visualizations (Goodall, 2009), coordinated displays (Vieane et al., 2016), and other tools designed to facilitate the performance of that task.

Finally, in a recent article, Dawson and Thomson (2018) describe the competencies that will be needed in the future cybersecurity workforce. They emphasize that cyber workers will need competencies that extend beyond technical skills, such as systematic thinking, teamwork skills, communication skills, and social skills.

SOURCE: Generated by the committee.

The boundaries between cybersecurity and social cybersecurity are not altogether sharp, but Table 6-1 lists some key differences between the two. Because cybersecurity focuses primarily on technology, for example, a cyberbreach conducted to steal or compromise data would be in that realm (Carley et al., 2018). In contrast, the manipulation of groups to provide funding for covert actors or extremist groups (Benigni et al., 2017b), sway opinion to win elections (Allcott and Gentzkow, 2017), artificially boost the perceived popularity of actors (Woolley, 2016), or build groups so as to have an audience for recruitment (Benigni et al., 2019) would all best be addressed by social cybersecurity.

TABLE 6-1 Key Differences between Cybersecurity and Social Cybersecurity

Characteristic	Cybersecurity	Social Cybersecurity
Core disciplines	Electrical engineering, software engineering, computer science, computer engineering	Computational social science, societal computing, data science, policy studies

[5] Available: https://www.usenix.org/sites/default/files/soups2018_full_proceedings.pdf [February 2019].

Illustrative problems	Encryption, malware detection. denial-of-service attack protection	Spread of disinformation, spam, altering who appears influential, creating echo chambers
Core methods	Cryptography, software engineering, computer forensics, biometrics	Network science/social networks, language technologies, social media analytics
Illustrative level of data	Packets	Social media posts
Focus on the issue of insider threat	Encryption to prevent ease of reading, software to prevent or detect illicit data sharing, firewalls	Social engineering to seduce insiders to share information, information leakage in social media
Focus on the issue of spreading malware via kitten images on Twitter	How malware is embedded and detected	Use of bots to promote message sharing, what groups are at risk to download
Focus on the issue of denial-of-service attacks	Technology to detect, enable, or prevent denial-of-service attacks	Social media and dark web identification of hackers who perpetrate denial-of-service attacks; analysis of how these hackers are trained
Illustrative tools	SysInternals, Windows GodMode, Microsoft EMET, Secure@Source, Q-Radar, ArcSight	ORA-PRO, Maltego, TalkWalker, Scraawl, Pulse, TweetTracker, BlogTrackers
National infrastructure support	United States Computer Emergency Readiness Team (US-CERT)	Nothing comparable—emergent self-management by social media providers
Illustrative central conferences	RSA, Black Hat, DEFCON, InfoSec World, International Conference on Cybersecurity	World Wide Web, SBP-BRiMS, ASONAM, Social Com, Web and Social Media

SOURCE: Generated by the committee.

Drawing on Other Disciplines

SBS research plays a role in both cybersecurity and social cybersecurity; examples include research on deception and motivations for attacks at the individual and state levels (discussed below) and research on teams (see Chapter 7). Both cybersecurity and social cybersecurity are applied fields in which new technologies are developed and tested. The field of social cybersecurity does not simply supplant the important work of SBS research. Rather, researchers in the field build on some existing work and extend other work to generate new knowledge and in some cases develop new theory and methods that arise from the transdisciplinary approach for studying the cyber environment. Social cybersecurity is a computational social science, one of a growing number of social science fields that are using digital data and developing computational tools and models (Mann, 2016). Computational social science is not the application of computer science techniques to social science problems and data (Wallach, 2018); it is the use of social science theories to drive the development of new computational techniques, combined with further development of those theories using computational techniques for data collection, analysis, and simulation.

In the case of social cybersecurity science, computer scientists and engineers on the one hand and social scientists and policy analysts on the other have not always recognized the

implications of each other's perspectives for their own research. For example, computer scientists' attempts to identify disinformation usually begin with fact checking. However, most disinformation campaigns rely less on blatant falsehood than on other strategies, such as illogic, satire, facts out of context, misuse of statistics, dismissal of topics, intimidation, appeals to ethnic bias, and simple distraction, all topics of SBS research (Babcock et al., 2018). Similarly, when SBS researchers seek to invent or reinvent computer science techniques, the results typically do not scale, are difficult to maintain, and lack generalizability. For example, affect control theory (a valuable computational model of human emotions based on social psychology) cannot be scaled to handle large social groups and populations. Computational social science, in contrast, requires deep engagement in and integration of knowledge, theories, and methods from both computer science and social science. Social cybersecurity science is often viewed as going beyond the interdisciplinary approach of integrating the methods and knowledge of diverse disciplines, having become a truly transdisciplinary science in the sense that it is creating new knowledge, theories, and methods. The objective of social cyber experts is to account for the peculiarities of the cyber environment and the specific opportunities for exploitation available in the communication and entertainment technology used by actors engaged, explicitly or implicitly, in information warfare or marketing.

As the field has matured, "social cybersecurity" has become the recognized term for this work, but the approach has been associated with other terms, including "social cyberforensics," "social cyberattack," "social media analytics," "cyber-physical-social based security," social cyberdefense," "computational propaganda," and "social media information warfare," and a variety of terms are used for key concepts in the field. Table 6-2, although not comprehensive, indicates this variety.

TABLE 6-2 Intersections between Social Cybersecurity and Other Disciplines

Discipline	Key Terms	Key Methods Other Than Network Science/Social Networks	Key Sources of Data Other Than Social Media	Illustrative Question
Sociology	Influence in social media, online influence	Language technology	Demographics	Do online groups and group processes resemble those offline?
Forensics	Social cyberforensics	Forensics	Website scraping, dark web	Who is responsible for a particular social cybersecurity attack?
Political science	Digital democracy, participatory democracy	Forum creation	Forums, legal and policy documents	How can social media be used to support or cripple democracy?

Anthropology	Digital anthropology, online ethnography	Rapid ethnographic assessment, area studies	Interviews, participant observation	How do people in different cultures use social media?
Information science	Cyber-physical-social security, social media analytics	Machine learning	Phone and banking data	How and when does information diffuse in social media?
Psychology	Social engineering	Social media analytics, case studies	Email, laboratory experiments	When do people contribute to conversations in social media?
Marketing	Viral marketing, online marketing	Social media analytics, statistics	Economic indicators, brand diagnostics	How can social media be exploited to market goods and services?
International relations	Social cyberattacks, social cyberdefense, e-government	Case studies, historical and policy assessment	News reports, court cases, dark web	How can state and nonstate actors use social media to gain influence and win battles via nonkinetic activities?
Economics	Digital economy, cybersecurity economics	Economic incentive assessment, econometrics	Money trails, price indices, cryptocurrency rates and usage	How do social media influence the economy?

SOURCE: Generated by the committee.

One constant for researchers in social cybersecurity is the application of network science and social network analysis (see Chapter 5), often in combination with other methods. The field also builds on other computational social science methods, including those used in data science, visual analytics, machine learning, text mining, natural language processing, social media analytics, and spatiotemporal data mining. Key methods include detection of change in networks, assessment and forecasting of diffusion, study of belief formation, influence assessment, identification of network elites, group identification, analysis of mergers and breakups, cyberforensics, actor activity prediction, and topic analysis. As evidence mounts that social media manipulation involves manipulation of both social and knowledge networks, researchers in this area increasingly combine social network analysis and narrative methods (see Chapter 5).

Another constant is reliance on social media data. Social cybersecurity experts are particularly concerned with social influence and group manipulation, the emergence of norms within and between online groups, and the formation or destruction of groups that are either receptive to or proponents of particular ideas and willing to engage in particular actions. Thus, key areas of study include models and methods associated with dynamically evolving data, patterns of life, information and belief diffusion, social influence, narrative construction and manipulation, group inoculation, and group resilience. Increasingly, research in the field is concerned with cultural variations, which often manifest as geographically specific enablers, constraints, and variation. Researchers seek to understand differences across groups by exploring variations in how people in different parts of the world generate, consume, and are affected by social media. They also explore how geospatial constraints, such as the location of ports, the existence of water features, the characteristics of landscape, and the types of natural disasters to which an area is prone may influence how information spreads in cyberspace, and why. Research areas include methods of psychological and social manipulation, cognitive biases in information handling, social biases in accessing information, trust building, and disinformation strategies.

A Social Cybersecurity Approach to Studying a False Information Campaign

The issue of the spread of false information on Twitter illustrates the distinction between the approaches of social cybersecurity and either pure computer science or pure social science.

Analysis of this problem using a purely computer science machine learning approach would begin with a training set containing tweets that had been marked as containing false information, such as a doctored image or a fact that had been checked and found to be inaccurate. Narrative would be assessed in terms of what words, concepts, sentiment, or gist could be extracted computationally (see Chapter 5). These extracted features would become part of the vector of information used in the machine learning model, and as a result, values for these features would become associated with the presence of false news. A desired end-state might be an automated fact checker, similar to spam checkers, which could run on multiple platforms independently of human intervention.

It is not uncommon for a reliable training set to have 2 to 10,000 marked items. This set might be split in half, with some tweets used to train new algorithms and others used to assess their efficacy. Algorithms would then be devised for empirically categorizing tweets according to whether, and with what certainty, they contained false information. The utility of the new algorithms would then be determined by comparing their precision and recall against those features of older algorithms. The new algorithms would have limited utility in any context other than that in which they had been developed. It is common for other researchers to reuse such training data in developing alternative models for comparison, but a mislabeled training set can yield misleading conclusions. Box 6-2 highlights other data challenges for this research, which would affect all three approaches.

BOX 6-2
Challenges in Data Access

The fields of computer science, social science, and social cybersecurity all face several challenges in data access. The first is variation in the policies, laws, and regulations of the corporations that build the data platforms or collect the data, the federal government, states, and

the governments of foreign countries with respect to social media and privacy and data access, storage, and sharing, which change frequently (Anderson, 2017). Communication and entertainment technologies are evolving rapidly, their potentially exploitable features are constantly changing, and new adversarial and marketing technologies for making use of the data are continually appearing (Van Dijck and Poell, 2013). These policy and technology changes alter what researchers can study, what the IC can do, and how easy it is for adversaries to manipulate the information environment (e.g., Stribley, 2018).

Other factors limit data sharing. The process of collecting, cleaning, and validating social media data is extremely time-consuming, and researchers may be reluctant to share their data out of concern that others may not take its nuances into account. It is possible to purchase some kinds of data, but the prices are well beyond the means of most researchers. Policies related to data storage and cleaning, such as the Twitter policy of removing access to tweets from users who are suspended, also inhibit research (Wei et al., 2015, 2016; Thomas et al., 2011).

As a result, there is a paucity of publicly available, sharable data (Baggili and Breitinger, 2015). These issues also make it far more challenging—sometimes impossible—to replicate the results of research (see Chapter 9 and Appendix C). Sharing of data and results is challenging even within the IC because of varying policies and regulations regarding what sorts of data agencies can collect, store, and link (Lawson, 2014; Kris, 2017; Konkel, 2014).

SOURCE: Generated by the committee.

In contrast, a pure social science approach to the same problem might be to begin by defining false information and its nuances in the context of a set of tweets, so that false tweets relative to that context could be identified. Then a quantitative researcher might statistically assess differences between sets of tweets with false information and sets of tweets without false information, using such metrics as the number of tweets, the topic areas addressed, the number of times tweets were retweeted or liked, and so on. This analysis would test a series of hypotheses derived from theories of human behavior (not technology) about, for example, rumor diffusion, attitude formation, persuasion, and social influence. Given the same set of tweets used by the computer scientists for training, the social scientist might assess the characteristics of the tweets and tweeters that affected interrater reliability[6] in determining whether a tweet contained false information.

Social science researchers would likely use multiple qualitative and/or quantitative methods to support the utility of their theoretical model—for example, to understand whether narratives containing false information were different from those without such information, whether different actors used different narratives, what characteristics of actors or groups made them susceptible to believing false information, or what features of narratives containing false information made them persuasive.

In other words, the computer scientist might seek to develop algorithms for identifying false news and deceptive actors in order to eliminate vulnerabilities in social media technologies to prevent the spread of misinformation. In contrast, the social scientist might seek to understand the differences in types of disinformation; the social, economic, and psychological motivations

[6]Interrater reliability refers to the level of agreement between those rating or coding a particular item.

behind deception; and the aspects of human cognition, social cognition, and attitude formation that affect when an individual or group is susceptible to false information.

Drawing on the potential benefits of both of these approaches, a social cybersecurity researcher would take into account

- how social media technology can be manipulated to affect who receives which messages at which times;
- the way the messages are presented and accessed;
- the way humans, individually and in groups, can create, access, be influenced by, and influence others using these features of the technology;
- how the content of a message can be manipulated to affect its persuasiveness, or the tendency of the technology to suspend the sender or recommend the message;
- the features of the content that impact its longevity (e.g., the presence of images);
- the similarities and differences in messages, and so the narratives and counternarratives coming from, going to, and being accessed by different users; and
- how the messages and technology could be manipulated to build up, link, or break down groups, and manipulate both the social network and people's perception of it.

Social cybersecurity researchers engage simultaneously in developing both method and theory, determining whether SBS hypotheses hold up in real-world settings. They would use high-dimensional network analytics[7] to analyze such questions as who is interacting with whom and who shares what narratives with whom. They would use visual analytics, statistics, and text mining to extract narrative features in order to characterize the empirical profile of messages that do and do not contain false information, the dialogues in which those messages are embedded, the narratives and counternarratives under discussion, the users that do and do not send the messages, the types of users and their motivations for sending those messages, and the groups that are or are not receptive to the messages. New methods would likely be tested on a combination of new and old data. As theoretical accounts are modified, social cybersecurity researcher develop new algorithms for collecting data on specific activities or measuring key features of those activities. The utility of these new methods and theories resides in the extent to which they support explanation and prediction in the wild, are reusable, and can be extended to new domains.

OPPORTUNITIES FOR THE IC

The field of social cybersecurity offers two primary benefits for the IC. First, it provides a means of strengthening the capacity of the United States to assess, predict, and mitigate the impact of attacks in the cyber-mediated environment that are aimed at affecting the hearts, minds, and welfare of U.S. citizens, corporations, and institutions. Second, the field provides a means of increasing U.S. capacity to assess, monitor, and forecast changes in behavior in other countries using social cyberintelligence.

[7]High-dimensional network analytics is the use of networks with multiple dimensions, such as a series of time-varying networks, networks with geocoordinates (geonetworks), or a set of networks varying in types of nodes and links (a metanetwork) (Carley, 2002).

The United States is engaged in an ongoing war in cyberspace, which is being conducted to a significant degree in and through social media (Waltzman, 2015; Shallcross, 2017): social cybersecurity threats are pervasive and on the rise because foreign adversaries and criminals exploit features of social media; 50 percent of the 10 worst social media–based cyberattacks occurred in 2017 (Wolfe, 2017). Spear phishing (sending a malicious file or link through an innocuous message) is also on the rise (Frenkel, 2017).

A key role of the intelligence analyst is to understand, explain, assess, and forecast the social threats in cyberspace and to counter those threats, which include the manipulation of information for nefarious purposes. Russia and China both have and use technologies that can manipulate content on social media by altering or disguising what is being said or who appears to be saying it, and influencing who will read or receive what information. Social cyber-mediated interference in elections is common. Bots, trolls, and cyborgs have supported information and disinformation campaigns aimed at influencing elections in the United States, Britain, Germany, and Sweden. Social cybersecurity attacks are prevalent: by one estimate, as many as one in five businesses have been subjected to a social media–based malware attack.[8] Such cyberattacks are conducted by individuals, groups, nonstate actors, state actors, and actors sponsored by states, often supported by the use of bots. Because these actors vary in their capabilities, so, too, does the quality of their information maneuvers (Darczewska, 2014; Snegovaya, 2015; Zheng and Wu, 2005).

Virtually anything that can be represented in digital form can be falsified. Tools for falsifying content include fake actors (personas) (Mansfield-Devine, 2008), fake antivirus software (Stone-Gross et al., 2013), and fake websites (Holz et al., 2009). The spread of such intentionally deceptive material, particularly the spread of false information, has the potential to undermine societies and is a growing concern for governments around the world (van der Linden et al., 2017; Roozenbeek and van der Linden, 2018; Allcott and Gentzkow, 2017). The accuracy of recorded sound and images can no longer be taken for granted. Software can be used to alter digital images, mimic the sounds of human voices, and create simulated videos (e.g., Piotrowski and Gajewski, 2007; Kim et al., 2018). This technology can be used to portray people saying things they did not say and doing things that never actually occurred. The growing ability to fabricate audio and digital information not only complicates the task of societies in distinguishing between reality and false narratives but also complicates the intelligence analyst's task in detecting deception (Joseph, 2017).

The IC must rely on open-source information in addressing a range of issues (Best and Cumming, 2007; Bean, 2011). Intelligence analysts collect, manage, and assess open-source data, seeking to understand the biases contained in the data, recognize when the data have been manipulated by an adversarial party, and recognize when individuals and communities in the United States are under attack in the open-source information environment (Omand et al., 2012). Social cybersecurity science provides many of the tools and methods that can help meet these challenges.

Finally, the analyst has a need to understand which individuals, groups, and communities are at risk of being manipulated through social media and how that risk can be mitigated. This task includes understanding when the analyst and the IC organization are at risk. Meeting this challenge requires effective means of training IC analysts to recognize indicators and warnings

[8]See https://www.pandasecurity.com/mediacenter/social-media/uh-oh-one-out-of-five-businesses-are-infected-by-malware-through-social-media [July 2018].

that social cyberattacks are occurring, to be aware of the kinds of social cyber-mediated attacks that can occur and their consequences, and to operate safely in the social cyberenvironment. The IC needs to recognize quickly when it is under social cyberattack, as well as to identify the ways in which it is susceptible to related risks, such as insider threat, information maneuvers designed to discredit an investigation, or denial-of-service events conducted through social media. The field of social cybersecurity offers important perspectives on how to recognize and respond to such attacks. Other SBS research, particularly in the application of organization theory to high-risk organizations, provides guidance on how to promote heedful interaction in the cyber-mediated realm and how to develop and sustain an effective social cybersafety culture.

EXAMPLE APPLICATION: SOCIAL INFLUENCE ON TWITTER

An example illustrates the contributions of the social cybersecurity approach to intelligence analysis. The Islamic State of Iraq and ash-Sham (ISIS) makes extensive use of social media in its operations (Blaker, 2015; Veilleux-Lepage, 2015). It uses social media for recruitment (Gates and Podder, 2015; Berger and Morgan, 2015); information warfare on local populations (Farwell, 2014); and possibly intelligence gathering and training. Similarly, Russian information operations use social media to influence social opinion and alter behavior. Social cybersecurity theories and methods have been used to identify what tactics are being used for these purposes and to explore their potential impact. Much of this work has been done using data extracted from Twitter, although cyberforensic techniques allow researchers to connect to information in other media (e.g., Facebook and YouTube) as well.

Consider an influence operation using Twitter to benefit ISIS. The high volume of tweets is such that Twitter may not have the resources to send every tweet from a particular user to all of that user's followers, so prioritization schemes are needed to determine what to send and to whom. Twitter is organized organically into a set of topic-groups—dense communities of users that frequently mention each other and share topics, as shown in Figure 6-1 (Benigni et al., 2017a, 2017b). A topic-group is simply the way humans self-organize in many social media systems. Social media platforms often have ways of measuring the size of these topic-groups and use information about the group's size, membership, and topics of discussion to determine what messages, topics, or people to prioritize in various lists.

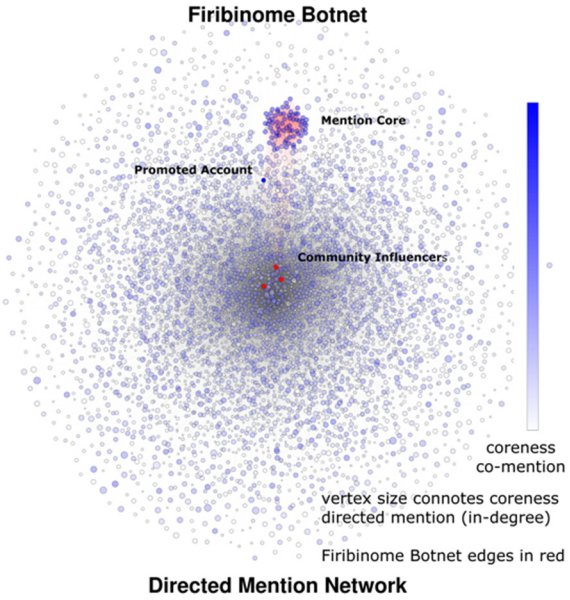

FIGURE 6-1 ISIS and Syrian online extremist community and Firibi Gnome bot on Twitter. NOTE: Each dot (vertex) is a Twitter user who has sent a tweet about a topic of interest. Dots in red are linked to the Firibi Gnome bot. The mention core is the set of Twitter users who are densely connected by mentions. The largest, densest mention core is near the top. The Firibi Gnome bot is in that core. The promoted account is the Twitter site associated with a website that is collecting money for the children of Syria. The influencer is the imam's Twitter account. SOURCE: Benigni et al., 2019.

In our example, one of these topic-groups is focused on issues related to Syria. Initially, it also included many individuals who were, if not members of ISIS, at least sympathizers, and much of the discussion was related to ISIS recruitment and propaganda. Within Twitter topic-groups, users vary in their communicative power, so some individuals have a disproportionate ability to reach others in the topic-group when they tweet. These individuals are often identified using metrics from social network analysis such as page rank, superspreader, or superfriend. Superspreaders in particular have a large number of followers and are central figures in their topic-group. When such individuals tweet, their messages are more likely to be read and/or retweeted than are the messages of others in the topic-group, and tweets that mention such users are more likely to be retweeted, in part because of the algorithms used by Twitter to prioritize messages and users. The analytic theories and methods used in narrative studies, especially regarding what gives certain narratives and messages power (e.g., an underlying emotional message that leads to specific attitudes and beliefs), are relevant to this challenge (see Chapter 5). However, the application of these theories to social media technologies work needs further study, particularly because of the limits in how emotions can be conveyed in or understood from short text statements.

Twitter's algorithms seem to prioritize tweets from superspreaders in the set of tweets received by other users. The term "echo chamber" refers to a set of users who tend to mention one another. The closer a topic-group is to being an echo chamber, the more rapidly information will diffuse within it. In these situations, emotions can escalate rapidly, and contradictory

information is less likely to be broadcast. Messages from echo chambers appear to be prioritized in the set of messages Twitter sends to the topic-groups associated with those echo chambers. Thus if an echo chamber retweets a message from a superspreader, the members of the echo chamber are more likely to appear in the list of messages received by members of the topic-group.

Within the Syrian topic-group in our example, one superspreader is an imam. At this point, it is not known whether he is associated with ISIS. Enter the Firibi Gnome bot (an automated agent that engages in Internet activity and sends tweets). This bot—actually a network of bots—functions as a pure echo chamber. It retweeted messages from the imam, which caused members of the bot network to appear in the feeds of other members of the topic-group, who were often human. Thus, members of the topic-group began to follow these bots. At some point, the bot network started tweeting messages with a link to a charity website that was ostensibly collecting money for the children of Syria, a site that some believe is linked to money laundering for ISIS. Without any active behavior from the imam, members of the topic-group were swayed by this bot to give money. Retweets by those who sympathized with messages from the bot were sufficient to manipulate the topic-group and to change the Twitter algorithm's prioritization of messages and their recipients (see Figure 6-1).

Social cyber researchers have developed methods that make it possible to track and understand such online developments. These methods can be used to identify topic-groups and echo chambers (Benigni et al., 2017b); identify influential users in social media, such as superspreaders and superfriends (Altman et al., 2018); identify core topics (Alvanaki et al., 2012); identify cross-media linkages (Dawson et al., 2018); and measure the potential reach of a message (Hong et al., 2011). Research is also under way on technology that could be used to support the identification or spread of false information. Examples include technology for fact checking (Rubin et al., 2015; Snopes[9]); image modification (Schneider and Chang, 1996); duplication of images (Ke et al., 2004); brandjacking attacks[10] (Youngblood, 2016); sentiment mining (Pang and Lee, 2008); stance[11] detection (Somasundaran and Wiebe, 2010); personality, gender, and age identification (Schwartz et al., 2013); location identification (Huan and Carley, 2017); and event detection (Wei et al., 2015). This work builds on ongoing computer science research that is well funded and in which advances are already being made.

Social cybersecurity research based on this work uses these computational methods in developing new sociotechnical theories and methods focused on the spread of multiple types of information maneuvers that were previously treated as a single phenomenon.

RESEARCH NEEDED IN THE COMING DECADE

Research in the field of social cybersecurity is needed on two parallel fronts: (1) research to establish new scientific methods and techniques capable of processing and analyzing the new types of data and high-dimensional networks made prevalent by social media; and (2) research to translate the resulting findings and techniques to operational tools that can be used by the IC.

[9]Available: https://www.snopes.com/fact-check/category/fake-news [November 2018].

[10]Brandjacking is the practice of mimicking the online identity of a business for the purpose of deceiving or defrauding users.

[11]In this context, "stance" refers to a publicly stated opinion, particularly one that is shared by an online community.

Advances in computer science, such as in the use and application of machine learning, have provided powerful tools for analyzing online activity. However, these advances are not readily transferable to the analysis of online activity in real time, nor are they sufficient to illuminate the broader context in which the activity is taking place. Multidisciplinary computational social science research building on both technological advances in computer science and SBS research has the potential to advance the research infrastructure in the field of social cybersecurity and expand the intelligence analyst's capability to address cybersecurity questions.

Having the tools necessary to predict and prevent attacks in the social cyberspace will require an aggressive research effort to identify, characterize, and understand such attacks. The committee sees opportunities to address a number of issues of concern for intelligence analysis:

- identifying who is conducting social cybersecurity attacks,
- identifying the strategies used to conduct such an attack,
- identifying the perpetrator's motive,
- tracing the attackers and the impact of the attack across multiple social media platforms,
- quantifying the effectiveness of the attack,
- identifying who is most susceptible to such attacks, and
- mitigating these attacks.

For each of these opportunities, we provide an overview of the challenge, summarize the recent work on which future developments would build, and specify the nature of further work that can pay significant dividends in the coming decade.

Social Cyberforensics: Identifying Who Is Conducting Social Cybersecurity Attacks

One of the keys to mitigating and responding to social cybersecurity attacks is being able to identify the perpetrators and impose sanctions against them. However, identification of perpetrators is a difficult problem in the cyber-mediated environment. Overcoming technical issues such as IP spoofing[12] can partly address this problem (Tanase, 2003), but the possibilities for overcoming the problem would be greatly expanded if it were possible to identify behavioral patterns at the individual and group levels, such as those associated with language use, location, credit taking, patterns of verbal communication, and use of images (e.g., Chen, 2015; Krombholz et al., 2015). Yet another need is the capacity to identify two actors appearing in two different media as in fact the same actor (such as when terrorist group members move from Twitter to a site on the dark web) (Maddox et al., 2016).

To illustrate, a writer's identity may be communicated through the linguistic style of a piece of text, and some work has suggested that such clues can be traced. Researchers have developed a method of encoding stylistic attributes to develop so-called "writeprints" (markers akin to fingerprints) (Abbasi and Chen, 2008). This method is based on the premise that aspects of any individual writer's usage (e.g., lexical, syntactic, structural, content-specific, and

[12]IP spoofing has been defined as an attack in which "a hacker uses tools to modify the source address in the packet header to make the receiving computer system think the packet is from a trusted source, such as another computer on a legitimate network, and accept it"; see https://usa.kaspersky.com/resource-center/threats/ip-spoofing [January 2019].

idiosyncratic features) are unconscious and persist from one document to another, so that they can effectively identify an individual author (Pearl and Steyvers, 2012). Topic models have been built using these features. Thus it is now theoretically possible to develop documents that can exactly match the features of a particular author. Most of this work, however, is in early stages and is limited to English (Mbaziira and Jones, 2016).

Other detection tools are possible in the near term. Recent advances in social network/network science (Benigni et al., 2019) and social cyberforensic techniques (Al-Khateeb et al., 2017) offer promising possibilities for identifying perpetrators. The social media reach of perpetrators is often enhanced by the use of bot, cyborg, Sybil, or troll techniques (Johansson et al., 2013; Klausen, 2015).[13] Indeed, as discussed above, many of the actors in social media may be bots; one study suggests that this may be the case for 48 million Twitter accounts (Varol et al., 2017). And according to a recent Pew Research Center report, two-thirds of all links shared on Twitter were shared by suspected bots (Wojcik, 2018). Emerging techniques are making it easier to identify whether perpetrators are humans, bots, or cyborgs, and further research is needed to increase the operational utility of these techniques for the intelligence analyst (e.g., Beskow and Carley, 2018; Morstatter et al., 2016). Bots and cyborgs that are used to influence and manipulate individuals and communities, often by exploiting features of a particular social medium, are evolving in sophistication and form as media platforms and bot-detection techniques evolve. At present, however, understanding of how bots and cyborgs evolve is limited to knowing that they are becoming more sophisticated, and no technology for predicting their evolution exists.

Research Directions

Continue work on developing better theories and methods for identifying perpetrators of cyberattacks.

This research could build on cyberforensic techniques, coupled with social network/network science techniques, machine learning, and deep understanding of sociopolitical contexts and the skills needed by perpetrators to manipulate social media and entertainment technologies. Some promising avenues include

- research to improve the capacity to detect online behavioral patterns of perpetrators;
- translational research to improve the utility for the IC of existing computational social science research techniques for identifying whether apparent attacks stem from humans, bots, or cyborgs;
- interdisciplinary research led by SBS researchers (those with deep understanding of how and why people manipulate media technology)

[13] A cyborg is an actor that is part human and part bot, frequently a human assisted by algorithms. Sybil is another, less widely used name for a bot. A troll is a user who posts inflammatory or off-topic messages in an online community in order to start quarrels or upset people; a troll account may be used by a single person, a group, or cyborgs.

to develop tools for predicting how bots and cyborgs will evolve in the future; and
- the development of sharable and continuously expanding data about known bots and cyborgs.

.

Information Maneuvers: Identifying the Strategies Used to Conduct Such an Attack

Used to manipulate individuals and groups, an information maneuver is any communication strategy intended to exaggerate or mitigate the spread of selected information or opinions, garner information, influence opinion, build or break connections among individuals to enable or prevent the spread of information or opinion, or exaggerate or minimize the influence of key actors (Al-Khateeb and Agarwal, 2016). A typical analytic approach to identifying information maneuvers is to look for something odd in social media posts, such as an increase in messages or the appearance of a new actor, and then collect specific data related to this anomaly. In so doing, an analyst working today would conduct detailed legwork involving tracking and reading messages. This approach is inherently costly, cannot be applied on a large scale, and is difficult to teach. A growing body of multidisciplinary research, however, has laid the foundation for new tools to augment intelligence analysis by detecting information maneuvers in a semiautomated fashion, identifying their intended audience, and classifying them by type. Much of this research has grown out of work on information warfare, marketing studies, and analyses of bot activity.

As discussed in Chapter 5, a central research challenge has been to investigate how fragilities of human social cognition and emotion can be exploited in an online context to shape information access and opinions, as well as how primary influencers exert their influence, and to better understand the nature of groups that are influenced through social media. These questions are important in seeking to understand information maneuvers and social cyberattacks, which typically operate at both the social network level (who is communicating with whom/influenced by whom) and the knowledge network level (who shares what information or opinions with whom). Such attacks typically exploit social cognition, including people's perception of the generalized other (that nebulous entity that represents one's opinion of what is common across the group), generalization strategies, and social influence procedures (Benigni et al., 2017a).

Information maneuvers can take different forms with very subtle nuances, and they require elaborate setups. Examples include maneuvers to manipulate an election (Metaxas and Mustafaraj, 2012), social engineering campaigns (Kandias et al., 2013), and satire campaigns (Babcock et al., 2018).

A social engineering campaign is the psychological manipulation of individuals to get them to perform specific actions, such as divulging confidential information or state secrets. Social engineering is one of the many tactics used in social influence campaigns on social media, such as those aimed at insiders (Kandias et al., 2013). Social engineering attacks, such as phishing and vishing (voice phishing), exploit not only factors well known to drive people's responses (see, e.g., Kumaraguru et al., 2007; Cialdini, 2001) but also how those responses are constrained and amplified by new technology. Traditional social science theories suggest that, whether they are conscious of it or not, people are motivated by

- reciprocity, or a sense of obligation to return favors;
- commitment, or a sense of obligation to do what one says one will;

- authority, or an inclination to obey or follow authority figures;
- social influence, or a tendency to do what others do;
- sociability, or a tendency to do what those one likes suggest; and
- scarcity reduction, or a tendency to desire what is scarce.

In a cyber-mediated environment, these motivations act somewhat differently because of the influence of other factors, such as a preference for easy modes of response, readily available information, and minimization of effort. Further, the features of the communication technologies influence who is motivated by what, and when, by making it possible to alter

- the way information is prioritized;
- constraints on choices;
- the attractiveness of options (e.g., using color, font and images, or repetition); and
- how easy it is to tell whether one is interacting with people, organizations, or bots.

A satire campaign is the use of exaggeration, humor, or irony to expose the inappropriate actions or views of particular people, groups, or organizations. In social media contexts, however, satire often appears out of context and so may not be recognized as such. Satire attacks can go viral and may be mistaken for news and then recharacterized as "fake news." This latter pattern is sometimes referred to as "the Stewart/Colbert effect," referring to the unintended persuasiveness of comedians' personas (Amarasingam, 2011). Satire attacks are among the many tactics used in social influence campaigns on social media, such as those aimed at political groups (Babcock et al., 2018).

The literature on information warfare also sheds some light on new forms of information maneuvers. Research on information warfare typically considers four broad strategies: distort, dissuade, distract, and dismay (Snegovaya, 2015). Classically, these strategies depend on how messages are constructed and communicated; there are well-known rhetorical strategies for persuasion (Ferris, 1994). Although forms of information maneuvers would fit into these four broad strategies, it is not yet known whether there are other strategic purposes to consider, or whether automatic characterization of an information maneuver or social media campaign as representing one of these strategies is possible. In addition, recent research has demonstrated that information maneuvers in social media may not only manipulate what is being said but also foster or undermine online communities or topic-groups associated with a message (Benigni et al., 2017b) or identities and brands (i.e., brand-jacking [e.g., Ramsey, 2010]).

Researchers are currently seeking ways to use features of social media posts and actors and the delivery/response sequence in characterizing information maneuvers. They have identified features that could work, including manipulation of emotions (Stieglitz and Dang-Xuan, 2013), presence (Naylor et al., 2012), group formation (Benigni et al., 2019), image manipulation (Tsikerdekis and Zeadally, 2014a), speed of spread (Vosoughi et al., 2018), and manipulation of the message and the group by bots (Benigni et al., 2017b). A large body of research explores the relationships among emotions, emotion manipulation, and the presentation of emotion in social media (Pang and Lee, 2008; Liu and Zhang, 2012; Gilbert and Hutto, 2014; Asur and Huberman, 2010; Steiglitz and Dang-Xuan, 2013). Additionally, particular emotions have specific triggers and functions, all of which lead to different cognitions and prime different actions/decisions.

Thus, research on how specific discrete underlying emotional states lead to different consequences would be useful.

Much of the work in this area uses highly simplistic measures of emotion focused on the valence and strength of words in general, in context, or across a body of posts (stance). Meanwhile, more sophisticated approaches to emotion, such as affect control theory and discrete emotions theory (Heise, 1987; Robinson et al., 2006), provide the basis for relating emotions to behavior and identity construction empirically. However, these approaches are generally not applied to social media (an exception being Joseph and Carley [2016]). The research on social media and emotions, however, is still not well connected to the research on affect control, emotion management, and group behavior.

Another approach to characterizing information maneuvers—the use of images and videos, including doctored or fake images—has become possible with the advent of new platforms that better support images and videos, as well as increasing bandwidth, the prevalence of smartphones, and growing consumer interest in moving from text to images or videos to communicate. Recent studies in this area have explored the use of images and videos in social media by terrorist groups to recruit, spread messages, distort opinions, sow fear, and spread misleading health information (Farwell, 2014; Syed-Abdul et al., 2013; Huey, 2015; Mangold and Faulds, 2009). Automated image and video analysis, however, is being carried out largely in the field of computer science and has not made its way to the field of social cybersecurity. Although hundreds of social cybersecurity studies have used computational text analysis methods, there appear to be only a few that have used any form of computational image or video processing.[14] An area of research prime for breakthroughs in the near future, then, is understanding how the presentation of emotion-laden messages and images in social media can influence groups, how such presentation varies across messages containing true and false information, and how the impact of such messages and images can be countered within and through information maneuvers.

Research Directions

Conduct interdisciplinary research to develop computational models and theories about information maneuvers in cyberspace and the respective strategies of influence and manipulation.

This research in social cybersecurity can build on foundational work on information warfare in political science, social psychology, and military science; rhetoric and communication theories relevant to marketing and manipulation; theories of social influence, social cognition, and group identification in anthropology, sociology, psychology, and political science; and studies of emotions and affect control in cognitive science and psychology. Moving beyond these theories to account for the technical, global, and temporal nature of the new cyber environment will be a valuable step forward. Promising avenues include

[14]This observation is based on an examination of all papers identified by Carley and colleagues (2018) as being in the area of social cybersecurity.

- research to understand how operational features of specific social media and entertainment technologies are being exploited as part of these information maneuvers;
- research expanding on new work to characterize information maneuvers and to develop a unified list of such maneuvers and associated data;
- research to develop theories for identifying, explaining, predicting, and countering information maneuvers in cyberspace;
- research to further develop tactics, techniques, and procedures, currently in their infancy, for detecting information manipulation as it is happening and identifying the strategies being used, and for reducing the societal and group-level risks of such manipulation; and
- translational research on the operational technology that can allow the IC to identify and characterize information maneuvers and their intended audience rapidly, at scale, and in a semiautomated fashion.

Intent Identification: Identifying the Perpetrator's Motive

Although progress is being made in the development of methods for identifying when information maneuvers have occurred, understanding the intent behind these maneuvers presents its own challenges (e.g., Sydell, 2016). People choose to deceive others for many reasons, including to avoid something negative; to fulfill a desire for fun, economic benefit, or personal advantage; to bolster self-esteem, make others laugh, or act altruistically; or to be polite. They may also, of course, seek to deceive for malicious reasons (Bhattacharjee, 2017). Research on deception by state and nonstate actors in cyberspace has distinguished among three types of cyberattacks: they may be conducted for economic reasons (Lotrionte, 2014) or strategic cyberespionage and military reasons (Geers et al., 2013), or be opportunistic and politically motivated (Kumar et al., 2016).

An intriguing aspect of the motivation for information maneuvers is that much of the activity in social media is not malicious, but is aimed at spreading news or information on new products, sharing information on social activities, and building communities of like interest and concern. At a high level, bots and information maneuvers have been used in similar ways for both illicit and legal gain and with both malicious and nonmalicious intent. Thus information maneuvers useful for spreading false information are also useful for spreading true information. Tactics used to market real products (e.g., Safko, 2010; Scott, 2015) are also used to market illegal products (Benigni et al., 2019). And procedures used to recruit and support followers for sports teams are also used to recruit and support followers for terrorist groups (compare Henderson and Bowley [2010] and Farwell [2014]). Researchers have suggested that differences in metadata, word choice, and timing of messages may provide clues to the intent behind messages (Java et al., 2007; King, 2008), but determining the intent of a particular actor, or at least distinguishing malicious and nonmalicious activity in an automated fashion, remains a challenge.

Assessment of images and videos is frequently used to develop insight into the intent of those who spread deceptive information. In one example, a dismay maneuver used images of a bomb attack in the White House with the intent to spread terror (Weimann, 2014). In another case, a Russian information operation used fake images, some from video games (Luhn, 2017; Murphy, 2017), in tweets and Facebook posts claiming that the United States was supporting

ISIS. Fake images of frightening phenomena, such as sharks in subways or airports flooded with water, are routinely circulated in the immediate aftermath of disasters to contribute to disruption (Gupta et al., 2013). Indeed, compendiums of such images have been developed, so many are reused or doctored and reused whenever disasters occur. Image analysis holds promise for understanding intent in such cases.

Researchers are also exploring other possible indicators that can be used to identify deception, including linguistic markers (Briscoe et al., 2014; Zhou et al., 2003; Zhou and Zhang, 2008); activity indicators (e.g., those used in detecting bots [Subrahamainian et al., 2016]); nonverbal behavior and the use of multiple accounts (Tsikerdekis and Zeadally, 2014a); and social structural behavior (i.e., behaviors that change who is interacting with whom and who is important in the social network) (Pak and Zhou, 2014). However, the ability to engage in deceptive behavior and the types of behaviors possible are dependent on the technology itself (Tsikerdekis and Zeadally, 2014b), language (Levine, 2014), the human social network (Chow and Chan, 2008; Tsikerdekis and Zeadally, 2014b), and human cognition (Spence et al., 2004). Other research has examined the profiles, characteristics, and motivations of hackers or cybercriminals who create fakeries or use deception or deceptive messages (Décary-Hétu et al., 2012; Papadimitriou, 2009; Seigfried-Spellar and Treadway, 2014). Still other work seeks to identify the characteristics of individuals and groups that make them vulnerable to deceptive messaging (Pennycook and Rand, 2018).

Some of this work has led to automated fact checkers that rely on both human- and machine-labeled input (e.g., Snopes[15]; Hassan et al., 2015), software tools for identifying deception based on verbal cues in texts (Zhou et al., 2004), tools for creating and detecting fake personas (even those that create personas with disabilities) (DeMello et al., 2005; Schultz and Fuglerud, 2012), and software for modifying text and auditory and video/image data streams to engender trust in the false information (Stamm et al., 2010; Emam, 2006). While there has been a fair amount of work on detecting in-person deception based on auditory and visual cues, tools for autoidentification based on findings about auditory or visual human "tells" are less well developed (Vrij et al., 2010). Thus, ongoing research in social cybersecurity is seeking ways to uncover intent and deception computationally.

Research Directions

Conduct research to develop techniques and tools with the capabilities to determine automatically and rapidly the intent of those conducting social cybersecurity information maneuvers.

Although such techniques and tools exist, they need to be better linked to theories of motivation and tools for linking motivation to behaviors. Future research in this area would build on work in social psychology, forensics, historical analysis, anthropology, sociology, cognitive psychology, political science, and statistical comparison. Some current work ripe for expansion includes

[15]Available: https://www.snopes.com/fact-check/fake-news-stories [October 12, 2018].

- the development of methods for linking available metadata to actors' intents;
- the development of methods for linking image and video analysis at scale to network science analysis and language technologies;
- identification of nonverbal indicators of veracity and deception in social media encounters and the combinations of linguistic, nonverbal, and audiovisual elements that signal truthfulness versus deception on the part of persons of interest to the IC, such as the leaders of states and nonstate entities, their followers, criminals, money launderers, and other bad actors in both online and offline interactions; and
- determination of differences in early indicators and motivations for types of deception employed at the individual, state, and nonstate actor levels.

Cross-Media Movement and Information Diffusion: Tracing the Attackers and the Impact of the Attack across Multiple Social Media Platforms

Classic theories of information diffusion are largely agnostic with respect to what media are used, and those that consider the media used often focus on social presence (Cheung et al., 2011), speed and network externality effects (Lin and Lu, 2011), and media features (Lee et al., 2015). In social media, however, there is not one medium but many. Studies have shown that movement among media or links from a message in one medium to another can increase the spread and reach of messages (Suh et al., 2010; Agarwal and Bandeli, 2017). Such movement among media can be engineered by bots (Wojcik, 2018), and allows actors to "hide" moving groups and messages they take with them when they move between media (Al-Khateeb and Agarwal, 2016; Liang, 2015), which allows them to create "safe havens."[16] An article on the online news site *Wired* describes the phenomenon this way:

> The Islamic State maximized its reach by exploiting a variety of platforms: social media networks such as Twitter and Facebook, peer-to-peer messaging apps like Telegram and Surespot, and content sharing systems like JustPaste.it. More important, it decentralized its media operations, keeping its feeds flush with content made by autonomous production units from West Africa to the Caucasus—a geographical range that illustrates why it is no longer accurate to refer to the group merely as the Islamic State of Iraq and al-Sham (ISIS), a moniker that undersells its current breadth.[17]

Some social media platforms are more likely to be used to receive rather than to generate messages. Most rumors on Twitter, for example, originate in other media (Liu et al., 2015), most notably in blogs. People in general use different media for different purposes (Haythornthwaite

[16]For example, "terrorists and extremists are increasingly moving their activities online—and areas of the web have become a safe haven for Islamic State to plot its next attacks, according to a report published last week by the London-based Henry Jackson Society" (quoted from http://www.homelandsecuritynewswire.com/dr20180409-stealth-terrorists-use-encryption-the-darknet-and-cryptocurrencies [April 2018]).

[17]See https://www.wired.com/2016/03/isis-winning-social-media-war-heres-beat [April 2018].

and Wellman, 1998). To be sure, diffusion models exist for social media such as Twitter (Xiong et al., 2012) and Flickr (Zhao et al., 2010). However, there are only a few theories of or models for information diffusion when multiple social media are present and in use (an exception being a model called Construct [Carley et al., 2009, 2014]), and even this model needs to be extended to account for the newer social media platforms). Although technologies are available for tracking a message or an individual across media (e.g., Maltego), theories on and the ability to predict such moves do not exist (Al-Khateeb et al., 2017). Work in this area is currently limited by barriers to data collection, diffusion theories that do not account for who uses what media when, and a lack of good digital forensic skills and techniques (Bidgoli, 2006; Huber et al., 2011). Thus, most of this research considers only a single medium, such as blogs (Gruhl et al., 2004), Twitter (Romero et al., 2011), or email (Mezzour and Carley, 2014). In stark contrast, most marketing guidance recommends the use of multiple media (e.g., Hovde, 2017). The technology exists to conduct cross-media assessment, but it is in its infancy and not widely available.

Research Directions

Conduct research to develop multimedia diffusion theories and a better understanding of the co-movement of people and ideas through cyberspace.

As the technology for cross-media assessment becomes more prevalent, SBS research in this area should be highly productive. This research can build on social cyberforensics technologies, social network/network science techniques, and social theories of information diffusion and group formation. New theories of information diffusion that account for multimedia use in cyberspace can then be applied to the development of techniques and tools for tracking, explaining, and predicting the movement of individuals, groups, ideas, and beliefs through and across multiple social media.

Real-Time Measurement of the Effectiveness of Information Campaigns: Quantifying the Effectiveness of the Attack

Real-time measurement of the impacts of information campaigns is a classically difficult problem, as those impacts often are slow to develop. In general, research is sparse on how to assess empirically and in real time the impact or success of an information maneuver (Carrier-Sabourin, 2011). The vast quantity of data and increased speed of communication that characterize social media create an environment in which it may be possible to make progress in this area. A number of metrics for measuring the reach and influence of messages and actors on social media have been suggested (Sterne, 2010; Hoffman and Fodor, 2010). Some of these are predicated on notions of social network influence (Benigni et al., 2019) and still others on rhetoric-based conceptions of reach (Carley and Kaufer, 1993). Nevertheless, there is little consensus among researchers on what to measure, how to use these measures strategically, and whether proposed metrics are valid (Barger and Labrecque, 2013). Furthermore, it is unknown how the data collection strategy affects these metrics and whether, as a consequence, the measurement results could be biased. Another key challenge in this area is the creation and use

of measures that can capture the dynamics of the underlying social and knowledge networks. Although approaches for assessing network dynamics exist (Snijders, 2001; Ahn et al., 2011; Carley, 2017), only those that can be used for incremental assessment scale well (Kas et al., 2013). Existing methods also cannot handle high-dimensional networks and so cannot assess impact in the social and knowledge networks simultaneously.

Research Directions

Develop methods for measuring the impact of an information campaign, in both the short and long terms.

Given the benefits of such methods for intelligence analysis, progress on real-time measurement of the effects of information maneuvers in cyberspace is an important area for future research in social cybersecurity. Such research could build on research on social networks and change detection, communication theories, and studies on group formation and participation in sociology and political science. Promising avenues include research to

- identify, operationalize, and validate these metrics;
- remove bias due to data collection; and
- assess the certainty of the results for large-scale, dynamic, high-dimensional networks.

At-Risk Groups: Identifying Who Is Most Susceptible to Such Attacks

The risk of being susceptible to information maneuvers has traditionally been considered greatest among those who are socially or economically disadvantaged, and risk reduction has been viewed as a function of education, awareness, empowerment, and reduction of disparities. Studies focused on the 2016 elections, however, found that while education was positively associated with accurate recognition of the falsity of news stories, so, too, were age and total media consumption (Alcott and Gentzkow, 2017). Current research suggests several factors that could influence those at risk: lost trust in mainstream media (Ekovich, 2017), overly filtered information through the use of personalized news (Flaxman et al., 2016), being embedded in topic-groups that are echo chambers (Benigni et al., 2019), and the inability to recognize that the information received is from bots (Benigni et al., 2017b). Other research, however, suggests that the majority of people do not trust information on social media (Ekovich, 2017) and that the spare empirical evidence available is not definitive on the impact of filtering (Zuiderveen Borgesius et al., 2016).

Inoculation techniques to reduce the susceptibility of individuals and groups to the spread of disinformation and to being affected by information warfare activities often take the form of media education. These techniques, however, are not based on empirical evidence and a deep understanding of the features of communication and entertainment technologies that can be exploited to spread disinformation. Such features include the short length of tweets, which makes it difficult to tell whether a message is satire (Babcock et al., 2018); marketing services that use bots to send tweets from an individual's account as that person (Benigni et al., 2017b); and the

removal (by Google) of image information that had made it easier to identify the falsity of information (Stribley, 2018).

SBS research shows that people will continue to persist in beliefs even when the evidence for those beliefs is discredited; facts do not change opinions (Kolbert, 2017). Thus knowing that news is manufactured does not keep people from believing it (Lilienfeld, 2014). A variety of mechanisms underlie this phenomenon (Shermer, 2002)—for example, (1) the belief that the untrue is fun, (2) the belief that true information from an untrusted source is not trustworthy, (3) social influence, (4) a reduction in cognitive dissonance, and (5) confirmation bias. Given the high volume of data in social media, it is often argued that trust in the source is used as a way of filtering information and reducing cognitive load, in which case false information from a trusted source is more trusted than true information from an untrusted source (Tang and Liu, 2015). Furthermore, a number of mechanisms have been suggested as supporting the sharing of false information, such as a preference for believing and sharing novel over more familiar information; a preference for stories that generate particular emotional reactions, such as surprise or disgust (see, e.g., Itti and Baldi, 2009; Aral and Van Alstyne, 2011; Berger and Milkman, 2012); a preference for believing what others believe (Friedkin, 2006); and appeals to the generalized other (Benigni et al., 2019).

It has become commonplace for social media providers, such as Facebook and Twitter, to use hidden algorithms to guide users to particular types of content and to other users with similar interests. This algorithmic strategy increases the likelihood that users will experience repeated exposure to particular individuals, groups, messages, and narratives. Bots and cyborgs can exploit these algorithms and create online communities in which alternative messages are suppressed, appeals to the generalized other foster group acceptance (Holdsworth and Morgan, 2007; Mead, 1934), images and humor are used to limit discussion (Meyer, 2000), and users are exposed to artificially enhanced social influence (Benigni et al., 2019). Social influence is critical in affecting one's beliefs and attitudes (Friedkin, 2006), and repeat exposure to these "contained" online communities increases the likelihood that an individual will embrace particular information and messages.[18] Spammed messages to email or social media accounts is another mechanism that has been instrumental in driving people to fake websites and the adoption of malware (Moore et al., 2009).

Research has explored the spread of false information and has begun to document its potency. In a recent large-scale study, for example, Vosoughi and colleagues (2018) found that false information diffused "significantly farther, faster, deeper and more broadly than the truth in all categories of information" (p. 1147), although other studies have found that this is the case only when an offline receptive group exists (Babcock et al., 2018).

Research Directions

Better characterize those groups at risk of social cyberattacks, and identify ways to increase awareness of malicious information maneuvers and strengthen the resistance of at-risk topic-groups to such attacks.

[18]See, e.g., work by Unkelbach (2007); Unkelbach and Stahl (2009); Alter and Oppenheimer (2009); and Fazio et al. (2015).

Such research could build on research in education and social psychology; theories of social influence, marketing, participatory democracy, and cognitive bias reduction; and social and political theories of group formation and dissolution. Promising avenues include

- empirical research on the key factors that put individuals and groups at risk of being targeted by information maneuvers in cyberspace, how these factors and the individuals and groups targeted may vary depending on the specific social media platform, and how that risk can be measured and reduced in specific media;
- research to better understand how recipients are influenced by information maneuvers, and any differences among certain populations; and
- research to develop techniques for measuring the actual and potential impacts of deceptive action or the misplacement of trust at the group or population level.

The Most Effective Responses: Mitigating These Attacks

Direct counterattacks on those conducting information maneuvers are often unsuccessful. Terrorists suspended from Twitter, for example, will recreate new accounts and engage in this activity even more vigorously (Al-Khateeb et al., 2017). Strategies focused on the receivers of the information and countermessages tend to be more effective,[19] but the success of such strategies depends on how messages are constructed and communicated vis-à-vis the group that is to be counterinfluenced. Deep understanding of the sociopolitical context is also necessary to keep the messaging attempt from backfiring. Research has yielded numerous guidelines for the creation of effective countermessages—for example, increasing credibility through the use of visuals (Murakami et al., 2009), not engaging in direct confrontation (Goulston, 2015), including a URL (Suh et al., 2010), being unyielding in stance (Lajeunesse, 2008), creating trust in the source (Tarran, 2017), and providing for sufficient resources and planning (Southwell et al., 2017). Because information maneuvers in social media involve manipulation of both groups and messages, moreover, new approaches to countermessages that include attention to the nature of the group are needed. Examples of such approaches include the application of research on participatory democracy and deliberative democracy techniques (Mutz, 2006), as well as influence maximization (Chen et al., 2010).

Although such research provides some information to guide countermessaging, it does not address a key problem occurring in social media—that, as discussed earlier, those with similar opinions form topic-groups through which they receive constant social support for not listening to counterarguments (the echo chamber effect [Bakshy et al., 2015]). Individuals confined to a topic-group may attend selectively only to certain messages and not even be exposed to any counterarguments (what is known as the filter-bubble effect [Flaxman et al., 2016]). One potential countermessaging approach to address this problem is the use of a context-aware system that directs messages from one actor to another (Conroy et al., 2015; Fischer, 2012).

[19]In operation, policies dictate which kinds of strategies are permitted under the law.

A key limitation of this research, however, is that it tends to focus either on winning the argument or on diffusing the message, and not on winning a diffusion contest against a competing message. The majority of the empirical work on the diffusion of competing ideas has used simulation (e.g., Krackhardt, 2001; Carley, 1990). However, these studies do not address how the type of communication medium affects the spread of ideas. A practical challenge in this area is that even if the perfect countermessaging strategy were known, its use might not be possible under current rules governing the IC.

In the area of cybersecurity relative to such issues as the spread of malware through social media and phishing attacks, research has expanded to look at policies, defenses, and engineering solutions that can mitigate the impact of such attacks (Zargar et al., 2013; Fette et al., 2007; Kumar et al., 2016; Galbally, Marcel, and Fierrez, 2014; Lin et al., 2009; Zahedi et al., 2015; Yin et al., 2007). At the organizational level, much of this work has focused on technical solutions to preventing or minimizing the impact of malware spread by social media (Timm and Perez, 2010) and on training and toolbars to avoid phishing (Wu et al., 2006). Research is increasingly showing that a mitigation strategy needs to employ a three-pronged approach, encompassing corporate policy, social cybersecurity training, and technology (Cross, 2013; Oxley, 2013). Much of this research has been based in the areas of policy and cybersecurity without drawing on the wealth of research in organizational science. The organizational literature suggests that in general, when in a high-risk situation, an organization needs to have a safety culture (Guldenmund, 2000), elements of which include heedful interaction, awareness of the risk, and support for maintaining a safe environment. Although much of the work in this area has focused on health (Pronovost and Sexton, 2005) and nuclear power plants (Pidgeon, 1991), its general claims are equally relevant to social cybersecurity risks. Engaging in heedful social cyber interaction and developing and maintaining a social cybersafety culture can potentially reduce risks associated with social cyberattacks. The IC has itself been a victim of such attacks, and therefore may wish to explore how an IC-specific social cybersecurity safety culture can be instituted.

Research Directions

Support the design of countermessaging strategies in cyberspace.

Such research could build on work on information warfare from social psychology; research on cognitive biases, marketing, and communication; theories of social cognition; and knowledge of participation and group formation gleaned from sociology and political science. Promising avenues include

- research focused specifically on identifying effective countermessaging strategies while taking into account the technical features of the social media;
- research directed at identifying effective countermessaging strategies while taking into account the authorities governing those doing the countermessaging; and
- research on how to implement, measure the prevalence of, and assess the effectiveness of a social cybersecurity safety culture.

CONCLUSIONS

Cyber-mediated threats are a growing area of concern for the IC. Because their use is increasing and their platforms change rapidly, social media serve both as a mechanism for monitoring developments and cyber-mediated threats and as a mechanism that can be manipulated to influence behaviors in ways that may pose threats to national security. We note that current work related to cyberspace issues—including data collection, cybersecurity, and social cybersecurity—is fragmented across a large number of U.S. government agencies and parts of the IC. The tools used by these entities, the authority they have to collect information, and their agreements with third-party vendors to collect data or run assessments all vary. The IC may wish to explore whether a central office to coordinate cyberintelligence efforts is needed. We caution, however, that issues associated with terrorism, social cybersecurity, and cybersecurity each demand distinct sets of skills and authorities.

Designing ways to protect against such threats requires the ability to collect data on and analyze and visualize high-dimensional dynamic networks with both social network and knowledge network components; Twitter networks, for example, generate both social data on who replies, retweets, or mentions or which individuals are quoted, and knowledge data on hashtags or topics that co-occur. However, available machine learning techniques and standard computer science methods are of limited utility for answering nuanced questions about developing situations (Lazer et al., 2014). Nor are traditional social science methods sufficient to address complex issues in today's information environment.

The promising next frontier is the combining of computer science techniques with deep understanding of how the media and entertainment technology used to collect these data operate, the sociocultural phenomena being studied, and relevant social and cognitive science theories (Wang et al., 2007; Carley et al., 2018). Social network/network science methods coupled with

language technologies, geospatial crowdsourced information, or machine learning and applied to large-scale data form the methodological cornerstone on which new advances will be realized. This kind of data is "big" not just because of the quantity involved, but also because of the number of networks in which the messages are embedded over time (National Research Council, 2013).

Empirical assessment of influence and manipulation in social cyberspace is yielding methods capable of processing large volumes of data, often from multiple media, and carrying out high-dimensional network analysis. Such methods have been used for successfully addressing a number of issues, such as the likelihood of retweeting (Suh et al., 2010), information diffusion (Romero et al., 2011), disaster planning (Landwehr et al., 2016), extremist recruiting (Benigni et al., 2019), and political polarization (Conover et al., 2011). Furthermore, geospatial assessments have shown great diversity in the ways in which social media are used by region, time, and political context (Carley et al., 2015).

This work provides a starting point for the development of tools that could be used by the IC for efficiently identifying propaganda, false information, and other social cyberthreats. In addition to building a body of research in this new field, researchers will need to address a number of methodological and data challenges if social cybersecurity research is to make the progress that is needed in the coming decade. These challenges include the development of both policy solutions for improving researchers' access to data and more sophisticated techniques for working with large but often incomplete and biased datasets (Tufekci, 2014).

CONCLUSION 6-1: A comprehensive multidisciplinary research strategy for identifying, monitoring, and countering social cyberattacks, predicated on computational social science, would provide significant support for the IC's efforts to address the social cybersecurity threat in the coming decade. The emerging field of social cybersecurity research can yield insights that would supplement the IC's training and technology acquisition in the area of social cybersecurity threats and foster an effective social cybersafety culture. These insights could support development of the capacity to, for example, detect bots and malicious online actors and track the impact of social cyberattacks.

CONCLUSION 6-2: The IC could strengthen its capacity to safeguard the nation against social cyber-mediated threats by supporting research with the objectives of developing

- generally applicable scientific methods for assessing bias in online data, drawing conclusions based on missing data, and triangulating to interpolate missing or incorrect data using multiple data sources; and
- new computational social science methods that would simultaneously consider change in social networks and narratives within social media–based groups from a geotemporal social-cyber perspective; and operational computational social science theories of influence and manipulation in a cyber-mediated environment that simultaneously take into account the network structure of online communities, the types of actors in those communities, social cognition, emotion, cognitive biases, narratives and counternarratives, and exploitable features of the social media technology.

References

Abbasi, A., and Chen, H. (2008). Cybergate: A design framework and system for text analysis of computer-mediated communication. *MIS Quarterly: Management Information Systems*, *32*(4), 811–837.

Agarwal, N., and Bandeli, K. K. (2017). Blogs, fake news, and information activities. In G. Bertolin (Ed.), *Digital Hydra: Security Implications of False Information Online* (pp. 31–46). Riga, Latvia: NATO Strategic Communications Centre of Excellence.

Ahn, J., Taieb-Maimon, M., Sopan, A., Plaisant, C., and Shneiderman, B. (2011). Temporal visualization of social network dynamics: Prototypes for nation of neighbors. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 309–316). Berlin/Heidelberg, Germany: Springer. Available: http://www.cs.umd.edu/hcil/trs/2010-28/2010-28.pdf [December 2018].

Al-Khateeb, S., and Agarwal, N. (2016). Understanding strategic information maneuvers in network media to advance cyber operations: A case study analysing pro-Russian separatists' cyber information operations in Crimean water crisis. *Journal on Baltic Security*, *2*(1), 6–27.

Al-Khateeb, S., Hussain, M. N., and Agarwal, N. (2017). Analyzing deviant socio-technical behaviors using social network analysis and cyber forensics-based methodologies. In O. Savas and J. Deng (Eds.), *Big Data Analytics in Cybersecurity and IT Management* (Chapter 12). New York: CRC Press, Taylor & Francis.

Allcott, H., and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Alter, A. L., and Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. doi:10.1177/1088868309341564.

Altman, N., Carley, K. C., and Reminga, J. (2018). ORA User's Guide 2018. Technical Report CMU-ISR-18-103. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Institute for Software Research.

Alvanaki, F., Michel, S., Ramamritham, K., and Weikum, G. (2012). See what's enBlogue: Real-time emergent topic identification in social media. In *Proceedings of the 15th International Conference on Extending Database Technology* (pp. 336–347). New York: Association for Computing Machinery.

Amarasingam, A. (Ed.). (2011). *The Stewart/Colbert Effect: Essays on the Real Impacts of Fake News*. Jefferson, NC: McFarland & Company.

Anderson, K. E. (2017). Getting acquainted with social networks and apps: Social media in 2017. *Library Hi Tech News*, *34*(10), 1–6.

Aral, S., and Van Alstyne, M. (2011). The diversity-bandwidth trade-off. *American Journal of Sociology, 117*(1), 90–171.

Asur, S., and Huberman, B. A. (2010). Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1) (pp. 492–499). Washington, DC: IEEE Computer Society.

Babcock, M., Beskow, D., and Carley, K. M. (2018). Beaten up on Twitter? Exploring fake news and satirical responses during the Black Panther movie event. In R. Thomson, C. Dancy, A. Hyder, and H. Bisgin (Eds.), *Proceedings of the 2018 SBP-BRiMS Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior*

Representation in Modeling and Simulation (pp. 97–103). Washington, DC: Springer. Available: https://link.springer.com/chapter/10.1007/978-3-319-93372-6_12 [December 2018].

Baggili, I., and Breitinger, F. (2015). Data sources for advancing cyber forensics: What the social world has to offer. In *Sociotechnical Behavior Mining: From Data to Decisions? Papers from the 2015 AAAI Spring Symposium* (pp. 6–9). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132.

Barger, V. A., and Labrecque, L. (2013). An integrated marketing communications perspective on social media metrics. *International Journal of Integrated Marketing Communications*, 64–76. Available: https://ssrn.com/abstract=2280132 [December 2018].

Bean, H. (2011). *No More Secrets: Open Source Information and the Reshaping of U.S. Intelligence*. Santa Barbara, CA: ABC-CLIO.

Benigni, M., Joseph, K., and Carley, K. M. (2017a). Mining online communities to inform strategic messaging: Practical methods to identify community-level insights. *Computational and Mathematical Organization Theory*, 1–19. doi:10.1007/s10588-017-9255-3.

Benigni, M., Joseph, K., and Carley, K. M. (2017b). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PLOS ONE, 12*(12), e0181405.

Benigni, M., Joseph, K. and Carley, K. M. (2019). Bot-ivistm: Assessing information manipulation in social media using network analytics. In N. Agrawal, N. Dokoohaki, and S. Tokdemir (Eds.), *Emerging Research Challenges and Opportunities in Social Network Analysis and Mining* (pp. 19–42). Cham, Switzerland: Springer.

Berger, J. M., and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205. doi:10.1509/jmr.10.0353.

Berger, J. M., and Morgan J. (2015). Defining and describing the population of ISIS supporters on Twitter. *Brookings*, March 5. Available: http://www.brookings.edu/research/papers/2015/03/isis-twitter-census-berger-morgan [December 2018].

Beskow, D. M., and Carley, K. M. (2018). Bot conversations are different: Leveraging network metrics for bot detection in Twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 825–832). Washington, DC: IEEE Computer Society. doi:10.1109/ASONAM.2018.8508322.

Best Jr., R. A., and Cumming, A. (2007). *Open Source Intelligence (OSINT): Issues for Congress.* Available: https://fas.org/sgp/crs/intel/RL34270.pdf [December 2018].

Bhattacharjee, Y. (2017). *Why We Lie: The Science Behind Our Deceptive Ways.* Available: https://www.nationalgeographic.com/magazine/2017/06/lying-hoax-false-fibs-science [December 2018].

Bidgoli, H. (2006). *Handbook of Information Security, Information Warfare, Social, Legal, and International Issues and Security Foundations*. Hoboken, NJ: John Wiley & Sons.

Blaker, L. (2015). The Islamic State's use of online social media. *Military Cyber Affairs, 1*(1), Article 4. doi:10.5038/2378-0789.1.1.1004.

Briscoe, E. J., Appling, D. S., and Hayes, H. (2014). Cues to deception in social media communications. In *47th Hawaii International Conference on System Sciences* (HICSS)

(pp. 1435–1443). Washington, DC: IEEE Computer Society. doi:10.1109/HICSS.2014.186.

Carley, K. M. (1990). Group stability: A socio-cognitive approach. In E. Lawler, B. Markovsky, C. Ridgeway, and H. Walker (Eds.), *Advances in Group Processes: Theory and Research* (Vol. VII) (pp. 1–44). Greenwich, CN: JAI Press.

Carley, K. M. (2002). Smart agents and organizations of the future. In L. Lievrouw and S. Livingstone (Eds.), *The Handbook of New Media* (pp. 206–220). Thousand Oaks, CA, SAGE Publications.

Carley, K. M. (2017). ORA: A toolkit for dynamic network analysis and visualization. In R. Alhajj and J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining*. Washington, DC: Springer. doi:10.1007/978-1-4614-7163-9_309-1.

Carley, K. M., and Kaufer, D. (1993). Semantic connectivity: An approach for analyzing semantic networks. *Communication Theory, 3*(3), 183–213.

Carley, K. M., Martin, M. K., and Hirshman, B. (2009). The etiology of social change. *Topics in Cognitive Science*, *1*(4), 621–650.

Carley, K. M., Lanham, M. J., Joseph, K., Kowalchuck, M., and Morgan, G. P. (2014). *Construct User's Guide*. Report CMU-ISR-14-105R. Pittsburgh, PA: School of Computer Science, Institute for Software Research. Available: http://reports-archive.adm.cs.cmu.edu/anon/isr2014/CMU-ISR-14-105R.pdf [December 2018].

Carley, K. M., Wei, W., and Joseph, K. (2015). High dimensional network analytics: Mapping topic networks in Twitter data during the Arab Spring. In S. Cui, A. Hero, Z.-Q. Luo, and J. Moura (Eds.), *Big Data Over Networks* (pp. 278–300). Cambridge, MA: Cambridge University Press.

Carley, K. M., Cervone, G., Agarwal, N., and Liu, H. (2018). Social cyber-security. In R. Thomson, C. Dancy, A. Hyder, and H. Bisgin (Eds.), *Proceedings of the 2018 SBP-BRiMS Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation* (pp. 389–394). Washington, DC: Springer. doi:10.1007/978-3-319-93372-6_42.

Carrier-Sabourin, K. (2011). *Measuring Effects and Success in Influence Operations: Challenges, Limitations and Opportunities.* Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a568381.pdf [April 2018].

Chen, A. (2015). The agency. *The New York Times Magazine*, June 2. Available: https://www.nytimes.com/2015/06/07/magazine/the-agency.html [December 2018].

Chen, W., Wang, C., and Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1029–1038). New York: Association for Computing Machinery.

Cheung, C. M. K., Chiu, P.-Y., and Lee, M. K. O. (2011). Online social networks: Why do students use Facebook?. *Computers in Human Behavior, 27*(4), 1337–1343.

Chow, W. S., and Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. *Information & Management, 45*(7), 458–465. doi:10.1016/j.im.2008.06.007.

Cialdini, R. B. (1987). *Influence* (Vol. 3). Port Harcourt, Nigeria: Albin Michel.

Cialdini, R. B. (2001). The science of persuasion. *Scientific American, 284*(2), 76–81.

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on Twitter. In *Proceedings of the Fifth International AAAI*

Conference on Weblogs and Social Media (Vol. 133) (pp. 89–96). Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275 [December 2018].

Conroy, N. J., Rubin, V. L., and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (p. 82). St. Louis, MO: American Society for Information Science. Available: https://pdfs.semanticscholar.org/939f/eec48ae1abb222cf9881932680b7ec3c68a7.pdf [December 2018].

Cross, M. (2013). Social media security: Leveraging social networking while mitigating risk. *Newnes*, November 1.

D'Amico, A. D., and Whitley, K. (2008). The real work of computer network defense analysts: The analysis roles and processes that transform network data into security situation awareness. In J. Goodall, G. Conti, and K. Ma (Eds.), *Proceedings of the Workshop on Visualization for Computer Security* (pp. 19–37). Berlin/Heidelberg, Germany: Springer. doi:10.1007/978-3-540-78243-8_2.

D'Amico, A., Whitley, K., Tesone, D., O'Brien, B., and Roth, E. (2005). Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *49*(3), 229–233. doi:10.1177/154193120504900304.

Darczewska, J. (2014). *The Anatomy of Russian Information Warfare. The Crimean Operation, a Case Study*. Available: https://www.osw.waw.pl/en/publikacje/point-view/2014-05-22/anatomy-russian-information-warfare-crimean-operation-a-case-study [December 2018].

Dawson, J., and Thomson, R. (2018). The future cybersecurity workforce: Going beyond technical skills for successful cyber performance. *Frontiers in Psychology*, *9*, 744. doi:10.3389/fpsyg.2018.00744.

Dawson, M., Lieble, M., and Adeboje, A. (2018). Open source intelligence: Performing data mining and link analysis to track terrorist activities. In *Information Technology-New Generations* (pp. 159–163). Cham, Switzerland: Springer.

Décary-Hétu, D., Morselli, C., and Leman-Langlois, S. (2012). Welcome to the scene: A study of social organization and recognition among Warez hackers. *Journal of Research in Crime and Delinquency,* 49(3), 359–382. doi:10.1177/0022427811420876.

DeMello, M. A., Keely, L. B., Byrum, F. D., Yaacovi, Y., and Hughes, K. E. (2005). *Method and System for Binding Enhanced Software Features to a Persona*. U.S. Patent 6,891,953, issued May 10, 2005. Available: https://patents.google.com/patent/WO2002001330A3/en [February 2019].

Dutt, V., Ahn, Y., and Gonzalez, C. (2011). Cyber situation awareness: Modeling the security analyst in a cyber-attack scenario through Instance-Based Learning. In Y. Li (Ed.), *Lecture Notes on Computer Science: Data and Applications Security and Privacy XXV* (pp. 280–292). International Federation for Information Processing. Available: https://pdfs.semanticscholar.org/f09a/9c917a376fe937ce8dc168a53cdb0c8cf040.pdf [December 2018].

Dutt, V., Ahn, Y.-S., and Gonzalez, C. (2013). Cyber situation awareness: Modeling detection of cyber attacks with instance-based learning theory. *Human Factors, 55*(3), 605–618.

Ekovich, S. R. (2017). Listening to Donald Trump. *Contemporary French and Francophone Studies, 21*(5), 498–506.

Emam, O. (2006). *System and Method for Creating Artificial TV News Programs*. U.S. Patent Application 11/236,457, filed June 22, 2006.

Farwell, J. P. (2014). The media strategy of ISIS. *Survival, 56*(6), 49–55.

Fazio, L. K., Brashier, N. M., Payne, B. K., and Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. doi.org/10.1037/xge0000098.

Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of English*, *28*(1), 45–65. Available: https://www.jstor.org/stable/40171324 [December 2018].

Fette, I., Sadeh, N., and Tomasic. A. (2007). Learning to detect phishing emails. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 649–656). New York: Association for Computing Machinery.

Fischer, G. (2012). Context-aware systems: The "right" information, at the "right" time, in the "right" place, in the "right" way, to the "right" person. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (p. 287–294). New York: Association for Computing Machinery. doi:10.1145/2254556.2254611.

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1), 298–320.

Frenkel, S. (2017). Hackers hide cyberattacks in social media posts. *The New York Times*, May 28. https://www.nytimes.com/2017/05/28/technology/hackers-hide-cyberattacks-in-social-media-posts.html [April 2018].

Friedkin, N. E. (2006). *A Structural Theory of Social Influence* (Vol. 13). Cambridge, MA: Cambridge University Press.

Galbally, J., Marcel, S., and Fierrez, J. (2014). Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing, 23*(2), 710–724.

Gates, S., and Podder, S. (2015). Social media, recruitment, allegiance and the Islamic State. *Perspectives on Terrorism, 9*(4), 107–116.

Geers, K., Kindlund, D., Moran, N., and Rachwald, R. (2013). *World War C: Understanding Nation–State Motives Behind Today's Advanced Cyber Attacks*. Available: https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/fireeye-wwc-report.pdf [February 2019].

Gilbert, C. J., and Hutto, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Available: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf [December 2018].

Goodall, J. R. (2009). Visualization is better! A comparative evaluation. In *International Workshop on Visualization for Cyber Security* (pp. 57–68). Available: https://web.ornl.gov/~jgoodall/goodall-vizsec09.pdf [December 2018].

Goulston, M. J. (2015). *Talking to "Crazy": How to Deal with the Irrational and Impossible People in Your Life*. New York: American Management Association.

Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. (2004). Information diffusion through blogspace. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 491–501). New York: Association for Computing Machinery.

Guldenmund, F. W. (2000). The nature of safety culture: A review of theory and research. *Safety Science*, *34*(1-3), 215–257. doi:10.1016/S0925-7535(00)00014-X.

Gupta, A., Lamba, H., Kumaraguru, P., and Joshi, A. (2013). Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 729–736). New York: Association for Computing Machinery.

Gutzwiller, R. S., Hunt, S. M., and Lange, D. S. (2016). A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2016* (pp. 14–20). doi:10.1109/COGSIMA.2016.7497780.

Gutzwiller, R. S., Ferguson-Walter, K., Fugate, S., and Rogers, A. (2018). "Oh, look, a butterfly!" A framework for distracting attackers to improve cyber defense. *Proceedings of the Human Factors and Ergonomics Society*, *62*(1), 272–276. doi:10.1177/1541931218621063.

Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J., and Yu, C. (2015). *The Quest to Automate Fact-Checking*. Available: http://cj2015.brown.columbia.edu/papers/automate-fact-checking.pdf [December 2018].

Haythornthwaite, C., and Wellman, B. (1998). Work, friendship, and media use for information exchange in a networked organization. *Journal of the American Society for Information Science, 49*(12), 1101–1114.

Heise, D. R. (1987). Affect control theory: Concepts and model. *Journal of Mathematical Sociology, 13*(1-2), 1–33.

Henderson, A., and Bowley, R. (2010). Authentic dialogue? The role of "friendship" in a social media recruitment campaign. *Journal of Communication Management, 14*(3), 237–257.

Hoffman, D. L., and Fodor, M. (2010). Can you measure the ROI of your social media marketing?. *MIT Sloan Management Review, 52*(1), 41. Available: https://sloanreview.mit.edu/article/can-you-measure-the-roi-of-your-social-media-marketing [December 2018].

Holdsworth, C., and Morgan, D. (2007). Revisiting the generalized other: An exploration. *Sociology, 41*(3), 401–417.

Holz, T., Engelberth, M., and Freiling, F. (2009). Learning more about the underground economy: A case-study of keyloggers and dropzones. In *European Symposium on Research in Computer Security* (pp. 1–18). Berlin/Heidelberg, Germany: Springer.

Hong, L. G., Dan, O., and Davison, B. D. (2011). Predicting popular messages in Twitter. *Proceedings of the 20th International Conference Companion on World Wide Web* (pp. 57–58). New York: Association for Computing Machinery. Available: http://www.cse.lehigh.edu/~brian/pubs/2011/WWW/predicting-popular-messages-twitter.pdf [February 2019].

Horn, C., and D'Amico, A. (2011). Visual analysis of goal-directed network defense decisions. In *Proceedings of the 8th International Symposium on Visualization for Cyber Security. VizSec '11* (Article No. 5). New York: Association for Computing Machinery. doi:10.1145/2016904.2016909.

Hovde, K. (2017). Why every business should be using multiple social media accounts. *Business.com*, February 22. Available: https://www.business.com/articles/why-every-business-should-be-using-multiple-social-media-accounts [April 2018].

Huan, B., and Carley, K. M. (2017). On predicting geolocation of tweets using convolutional neural network. In D. Lee, Y. Lin, R. Thompson, and N. Osgood (Eds.), *Proceedings of the International Conference Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2017)* (pp. 281–291). Washington, DC: Springer.

Huber, M., Mulazzani, M., Leithner, M., Schrittwieser, S., Wondracek, G., and Weippl, E. (2011). Social snapshots: Digital forensics for online social networks. In *Proceedings of the 27th Annual Computer Security Applications Conference* (pp. 113–122). New York: Association for Computing Machinery. doi:10.1145/2076732.2076748.

Huey, L. (2015). This is not your mother's terrorism: Social media, online radicalization and the practice of political jamming. *Journal of Terrorism Research, 6*(2). doi:10.15664/jtr.1159.

Itti, L., and Baldi, P. (2009). Bayesian surprise attacks human attention. *Vision Research, 49*(10), 1295–1306. doi:10.1016/j.visres.2008.09.007.

Java, A., Song, X., Finin, T., and Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (pp. 56–65). New York: Association for Computing Machinery.

Johansson, F., Kaati, L., and Shrestha, A. (2013). Detecting multiple aliases in social media. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1004–1011). New York: Association for Computing Machinery.

Joseph, R. (2017). Supersynthizers: Confronting the coming analytical crisis in an age of influence. Presentation at the National Academies of Sciences, Engineering, and Medicine's Workshop on Leveraging Advances in Social Network Thinking for National Security. October 11, 2017.

Joseph, K., and Carley, K. M. (2016). Relating semantic similarity and semantic association to how humans label other people. In *Proceedings of 2016 EMNLP Workshop on Natural Language Processing and Computational Social Science* (pp. 1–10). Austin, TX: Association for Computational Linguistics. Available: http://www.aclweb.org/anthology/W16-5601 [December 2018].

Kandias, M., Galbogini, K., Mitrou, L., and Gritzalis, D. (2013). Insiders trapped in the mirror reveal themselves in social media. In *International Conference on Network and System Security* (pp. 220–235). Berlin/Heidelberg, Germany: Springer.

Kas, M., Wachs, M., Carley, K. M., and Carley, L. R. (2013). Incremental algorithm for updating betweenness centrality in dynamically growing networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference* (pp. 33–40). Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi:10.1109/ASONAM.2013.6785684.

Ke, Y., Sukthankar, R., and Houston, L. (2004). Efficient near-duplicate detection and sub-image retrieval. In *ACM International Conference on Multimedia* (pp. 869–876). New York: Association for Computing Machinery. Available: http://www.cs.cmu.edu/~rahuls/pub/mm2004-pcasift-rahuls.pdf [February 2019].

Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Nießner, M., Pérez, P., Richardt, C., Zollhöfer, M., and Theobalt, C. (2018). *Deep Video Portraits.* Available:

https://web.stanford.edu/~zollhoef/papers/SG2018_DeepVideo/paper.pdf [December 2018].

King, R. (2008). How companies use Twitter to bolster their brands. *Bloomberg*, September 6. Available: https://www.bloomberg.com/news/articles/2008-09-06/how-companies-use-twitter-to-bolster-their-brandsbusinessweek-business-news-stock-market-and-financial-advice [December 2018].

Klausen, J. (2015). Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism, 38*(1), 1–22.

Kolbert, E. (2017). Why facts don't change our minds. *The New Yorker*, February 27. Available: https://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds [December 2017].

Konkel, F. (2014). The intelligence community's big-data problem. *FCW*, March 13. Available: https://fcw.com/articles/2014/03/13/ic-big-data.aspx [December 2018].

Krackhardt, D. (2001). Viscosity models and the diffusion of controversial innovation. In A. Lomi and E. R. Larsen (Eds.), *Dynamics of Organizations: Computational Modeling and Organization Theories* (pp. 243–268). Cambridge, MA: MIT Press.

Kris, D. (2017). The CIA's new guidelines governing publicly available information. *Lawfare*, March 21. Available: https://www.lawfareblog.com/cias-new-guidelines-governing-publicly-available-information [December 2018].

Krombholz, K., Hobel, H., Huber, M., and Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and Applications, 22*, 113–122. doi:10.1016/j.jisa.2014.09.005.

Kumar, S., Benigni, M., and Carley, K. M. (2016). The impact of U.S. cyber policies on cyber-attacks trend. In *Proceedings of the 2016 IEEE Conference on Intelligence and Security Informatics (ISI)* (pp. 181–186). doi:10.1109/ISI.2016.7745464.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., and Nunge, E. (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)* (pp. 905–914). doi:10.1145/1240624.1240760.

Lajeunesse, G. C. (2008). Winning the war of ideas. *Small Wars Journal*. Available: http://smallwarsjournal.com/blog/journal/docs-temp/110-lajeunesse.pdf [December 2018].

Landwehr, P. M., Wei, W., Kowalchuck, M., and Carley, K. M. (2016). Using Tweets to support disaster planning, warning and response. *Safety Science*, *90*, 33–47. doi:10.1016/j.ssci.2016.04.012.

Lawson, S. (2014). The U.S. military's social media civil war: Technology as antagonism in discourses of information-age conflict. *Cambridge Review of International Affairs, 27*(2), 226–245.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.

Lee, E., Lee, J.-A., Moon, J. H., and Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking, 18*(9), 552–556.

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology, 33*(4), 378–392.

Liang, C. S. (2015). *Cyber Jihad: Understanding and Countering Islamic State Propaganda*. GSCP Policy Paper 2015/2. Available: https://www.gcsp.ch/News-Knowledge/Publications/Cyber-Jihad-Understanding-and-Countering-Islamic-State-Propaganda [December 2018].

Lilienfeld, S. (2014). Evidence-based practice: The misunderstandings continue. *Psychology Today*, January 27. Available: https://www.psychologytoday.com/us/blog/the-skeptical-psychologist/201401/evidence-based-practice-the-misunderstandings-continue [February 2019]

Lin, K.-Y., and Lu, H.-P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior, 27*(3), 1152–1161.

Lin, Z., He, J., Tang, X., and Tang, C.-K. (2009). Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recognition, 42*(11), 2492–2501.

Liu, B., and Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. Aggarwal and C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Boston, MA: Springer. doi:10.1007/978-1-4614-3223-4_13.

Liu, X., Nourbakhsh, A., Li, Q., Fang, R., and Shah, S. (2015). Real-time rumor debunking on Twitter. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management* (pp. 1867–1870). New York: Association for Computing Machinery.

Liu, P., Jajodia, S., and Wang, C. (Eds.). (2017). *Recent Advances in Cyber Situation Awareness*. Boston, MA: Springer.

Lotrionte, C. (2014). Countering state-sponsored cyber economic espionage under international law. *NCJ Int'l L. & Com. Reg.,* 40, 443.

Luhn, A. (2017). Russia uses video game pictures to claim U.S. helped ISIL. The Telegraph, November 9. Available: https://www.telegraph.co.uk/news/2017/11/14/russia-use-video-game-picture-claim-us-helped-isil [February 2019].

Maddox, A., Barratt, M. J., Allen, M., and Lenton, S. (2016). Constructive activism in the dark web: Cryptomarkets and illicit drugs in the digital "demimonde". *Information, Communication & Society, 19*(1), 111–126.

Mangold, W. G., and Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons, 52*(4), 357–365.

Mann, A. (2016). Core concept: Computational social science. *Proceedings of the National Academy of Sciences of the United States of America,* 113(3), 468–470.

Mansfield-Devine, S. (2008). Anti-social networking: Exploiting the trusting environment of Web 2.0. *Network Security*, *2008*(11), 4–7. doi:10.1016/S1353-4858(08)70127-2.

Mbaziira, A., and Jones, J. (2016). A text-based deception detection model for cybercrime. In *International Conference on Technology and Management*. Available: https://www.researchgate.net/publication/307594168_A_Text-based_Deception_Detection_Model_for_Cybercrime [December 2018].

Mead, G. H. (1934). *Mind, Self and Society* (Vol. 111). Chicago, IL: University of Chicago Press.

Metaxas, P. T., and Mustafaraj, E. (2012). Social media and the elections. *Science, 338*(6106), 472–473. doi:10.1126/science.1230456.

Meyer, J. C. (2000). Humor as a double-edged sword: Four functions of humor in communication. *Communication Theory,* 10(3), 310–331.

Mezzour, G., and Carley, K. M. (2014). Spam diffusion in a social network initiated by hacked e-mail accounts. *International Journal of Security and Networks, 9*(3), 144–153.

Moore, T., Clayton, R., and Stern, H. (2009). Temporal correlations between spam and phishing websites. In *Proceedings (LEET'09) of the 2nd USENIX Conference on Large-scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More* (p. 5). Berkeley, CA: USENIX Association. Available: https://www.cl.cam.ac.uk/~rnc1/leet09.pdf [February 2019].

Morstatter, F., Wu, L., Nazer, T. H., Carley, K. M., and Liu, H., (2016). A new approach to bot detection: The importance of recall. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 533–540). Piscataway, NJ: IEEE Press.

Murakami, K., Nichols, E., Matsuyoshi, S., Sumida, A., Masuda, S., Inui, K., and Matumoto, Y. (2009). Statement map: Assisting information credibility analysis by visualizing arguments. In *Proceedings of the 3rd Workshop on Information Credibility on the Web* (pp. 43–50). New York: Association for Computing Machinery.

Murphy, P. (2017). Russia uses fake photos to accuse U.S. of supporting ISIS. *CNN*, November 14. Available: https://www.cnn.com/2017/11/14/us/russia-fake-photos-accusation-trnd/index.html [February 2019].

Mutz, D. C. (2006). *Hearing the Other Side: Deliberative versus Participatory Democracy*. New York: Cambridge University Press.

National Research Council. (2013). *Frontiers in Massive Data Analysis*. Committee on the Analysis of Massive Data, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences. Washington DC: The National Academies Press. doi:10.17226/18374.

Naylor, R. W., Lamberton, C. P., and West, P. M. (2012). Beyond the "like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing*, *76*(6), 105–120.

Omand, D., Bartlett, J., and Miller, C. (2012). Introducing social media intelligence (SOCMINT). *Intelligence and National Security*, *27*(6), 801–823. doi:10.1080/02684527.2012.716965.

Oxley, A. (2013). *Security Risks in Social Media Technologies: Safe Practices in Public Service Applications*. Oxford, UK: Chandos Publishing.

Pak, J., and Zhou, L. (2014). Social structural behavior of deception in computer-mediated communication. *Decision Support Systems, 63*, 95–103. doi:10.1016/j.dss.2013.08.010.

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Papadimitriou, F. (2009). A nexus of cyber-geography and cyber-psychology: Topos/"notopia" and identity in hacking. *Computers in Human Behavior*, *25*(6), 1331–1334. doi:10.1016/j.chb.2009.05.009.

Pearl, L., and Steyvers, M. (2012). Detecting authorship deception: A supervised Machine learning approach using author writeprints. *Literary and Linguistic Computing*, 27(2), 183–196.

Pennycook, G., and Rand, D. G. (2018). *Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking*. doi:10.2139/ssrn.3023545.

Pidgeon, N. F. (1991). Safety culture and risk management in organizations. *Journal of Cross-Cultural Psychology*, *22*(1), 129–140. doi:10.1177/0022022191221009.

Piotrowski, Z., and Gajewski, P. (2007). Voice spoofing as an impersonation attack and the way of protection. *Journal of Information Assurance and Security, 2*(3), 223–225.

Pronovost, P., and Sexton, B. (2005). Assessing safety culture: Guidelines and recommendations. *Quality and Safety in Health Care, 14*(4), 231–233. doi:10.1136/qshc.2005.015180.

Rajivan, P., and Cooke, N. J. (2018). Information pooling bias in collaborative security incident correlation analysis. *Human Factors, 60*(5), 626–639. doi:10.1177/0018720818769249.

Ramsey, L. P. (2010). Brandjacking on social networks: Trademark infringement by impersonation of markholders. *Buffalo Law Review, 58*, 851–929. Available: http://www.buffalolawreview.org/past_issues/58_4/Ramsey.pdf [December 2018].

Reveron, D. S. (2012). An introduction to national security and cyberspace. In D. S. Reveron (Ed.), *Cyberspace and National Security: Threats, Opportunities, and Power in a Virtual World*. Washington, DC: Georgetown University Press.

Robinson, D. T., Smith-Lovin, L., and Wisecup, A. K. (2006). Affect control theory. In *Handbook of the Sociology of Emotions* (pp. 179–202). Boston, MA: Springer.

Romero, D. M., Meeder, B., and Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 695–704). New York: Association for Computing Machinery. doi:10.1145/1963405.1963503.

Roozenbeek, J., and van der Linden, S. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*. doi:10.1080/13669877.2018.1443491.

Rubin, V. L., Chen, Y., and Conroy, N. J. (2015). Deception detection for news: Three types of fakes. *Proceedings of the Association for Information Science and Technology, 52*(1), 1–4.

Safko, L. (2010). *The Social Media Bible: Tactics, Tools, and Strategies for Business Success*. Hoboken, NJ: John Wiley & Sons.

Schneider, M., and Chang, S.-F. (1996). A robust content based digital signature for image authentication. In *Proceedings of 3rd IEEE International Conference on Image Processing* (Vol. 3) (pp. 227–230). doi:10.1109/ICIP.1996.560425.

Schultz, T., and Fuglerud, K. S. (2012). Creating personas with disabilities. In *International Conference on Computers for Handicapped Persons* (pp. 145–152). Berlin/Heidelberg, Germany: Springer.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., and Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One, 8*(9), e73791. doi:10.1371/journal.pone.0073791.

Scott, D. M. (2015). *The New Rules of Marketing and PR: How to Use Social Media, Online Video, Mobile Applications, Blogs, News Releases, and Viral Marketing to Reach Buyers Directly*. Hoboken, NJ: John Wiley & Sons.

Seigfried-Spellar, K. C., and Treadway, K. N (2014). Differentiating hackers, identity thieves, cyberbullies, and virus writers by college major and individual differences. *Deviant Behavior*, *35*(10), 782–803. doi:10.1080/01639625.2014.884333.

Shallcross, N. (2017). Social media and information operations in the 21st century. *Journal of Information Warfare, 16*(1), 1–12.

Shermer, M. (2002). *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time*. New York: Henry Holt and Company.

Snegovaya, M. (2015). *Putin's Information Warfare in Ukraine: Soviet Origins of Russia's Hybrid Warfare*. Washington, DC: Institute for the Study of War. Available: https://www.stratcomcoe.org/msnegovaya-putins-information-warfare-ukraine [December 2018].

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology, 31*(1), 361–395.

Somasundaran, S., and Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (pp. 116–124). Los Angeles, CA: Association for Computational Linguistics. Available: http://anthology.aclweb.org/W/W10/W10-0214.pdf [December 2018].

Southwell, B. G., Thorson, E. A., and Sheble, L. (2017). The persistence and peril of misinformation. *American Scientist, 105*(6), 372–375.

Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., and Ganesan, V. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society B: Biological Sciences, 359*(1451), 1755–1762. doi:10.1098/rstb.2004.1555.

Stamm, M. C., Tjoa, S. K., Lin, W. S., and Liu, K. J. R. (2010). Undetectable image tampering through JPEG compression anti-forensics. In *Image Processing (ICIP), 2010 17th IEEE International Conference* (pp. 2109–2112). doi:10.1109/ICIP.2010.5652553.

Sterne, J. (2010). *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. Hoboken, NJ: John Wiley & Sons.

Stieglitz, S., and Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217–248.

Stone-Gross, B., Abman, R., Kemmerer, R. A., Kruegel, C., Steigerwald, D. G., and Vigna, G. (2013). The underground economy of fake antivirus software. In *Economics of Information Security and Privacy III* (pp. 55–78). New York: Springer. doi:10.1007/978-1-4614-1981-5_4.

Stribley, R. A. (2018). Google just made it harder to spot fake news. *Medium,* February 22. Available: https://medium.com/s/story/google-just-made-it-harder-to-spot-fake-news-39a1ecff4c40 [April 20, 2018].

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., and Menczer, F. (2016). The DARPA Twitter bot challenge. *Computer, 49*(6), 38–46. doi:10.1109/MC.2016.183.

Suh, B., Hong, L., Pirolli, P., and Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *SOCIALCOM '10 Proceedings of the 2010 IEEE Second International Conference on Social Computing* (pp. 177–184). Washington, DC: IEEE Computer Society. doi:10.1109/SocialCom.2010.33.

Sydell, L. (2016). We tracked down a fake-news creator in the suburbs. Here's what we learned. *National Public Radio*, November 23. Available:

https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs [December 2018].

Syed-Abdul, S., Fernandez-Luque, L., Jian, W. S., Li, Y. C., Crain, S., Hsu, M.-H., Wang, Y. C., Khandregzen, D., Chuluunbaatar, E., Nguyen, P. A., and Liou, D. M. (2013). Misleading health-related information promoted through video-based social media: Anorexia on YouTube. *Journal of Medical Internet Research, 15*(2):e30. doi:10.2196/jmir.2237.

Tanase, M. (2003). IP spoofing: An introduction. *Symantec*, March 11. Available: https://www.symantec.com/connect/articles/ip-spoofing-introduction [December 2018].

Tang, J., and Liu, H. (2015). Trust in social media. *Synthesis Lectures on Information Security, Privacy, & Trust,* 10(1), 1–129.

Tarran, B. (2017). Why facts are not enough in the fight against fake news. *Significance, 14*(5), 6–7.

Thomas, K., Grier, C., Song, D., and Paxson, V. (2011). Suspended accounts in retrospect: An analysis of Twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (p. 243–258). New York: Association for Computing Machinery. Available: http://www.icir.org/vern/papers/twitter-susp-accounts.imc2011.pdf [February 2019].

Timm, C., and Perez, R. (2010). *Seven Deadliest Social Network Attacks (Syngress Seven Deadliest Attacks)*. Burlington, MA: Elsevier. doi:10.1016/C2009-0-61910-3.

Tsikerdekis, M., and Zeadally, S. (2014a). Multiple account identity deception detection in social media using nonverbal behavior. *IEEE Transactions on Information Forensics and Security, 9*(8), 1311–1321.

Tsikerdekis, M., and Zeadally, S. (2014b). Online deception in social media. *Communications of the ACM, 57*(9), 72–80.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 505–514). Available: https://arxiv.org/ftp/arxiv/papers/1403/1403.7400.pdf [December 2018].

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230. doi:10.1037/0278-7393.33.1.219.

Unkelbach, C., and Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, *18*(1), 22–38. doi:10.1016/j.concog.2008.09.006.

van der Linden, S., Maibach, E., Cook, J., Leiserowitz, A., and Lewandowsky, S. (2017). Inoculating against misinformation. *Science, 358*(6367), 1141–1142.

Van Dijck, J., and Poell, T. (2013). Understanding social media logic. *Media and Communication*, *1*(1), 2–14. doi:10.17645/mac.v1i1.70.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)* (pp. 280–289). Available: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817 [December 2018].

Veilleux-Lepage, Y. (2015). Paradigmatic shifts in Jihadism in cyberspace: The emerging role of unaffiliated sympathizers in the Islamic state's social media strategy. *Journal of Terrorism Research, 7*(1), 36–51. doi:10.15664/jtr.1183.

Vieane, A., Funke, G., Mancuso, V., Greenlee, E., Dye, G., Borghetti, B., Miller, B., Menke, L., and Brown, R. (2016). Coordinated displays to assist cyber defenders. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 344–348. doi:10.1177/1541931213601078.

Vieane, A., Funke, G., Greenlee, E., Mancuso, V., Borghetti, B., Miller, B., Menke, L., Brown, R., Foroughi, C. K., and Boehm-Davis, D. (2017). Task interruptions undermine cyber defense. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 375–379. doi:10.1177/1541931213601576.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. doi:10.1126/science.aap9559.

Vrij, A., Granhag, P. A., and Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest, 11*(3), 89–121.

Wallach, H. (2018). Computational social science ≠ computer science + social data. *Communications of the ACM*, *61*(3), 42–44. doi:10.1145/3132698. Available: https://cacm.acm.org/magazines/2018/3/225484-computational-social-science-computer-science-social-data/fulltext [April 16, 2018].

Waltzman, R. (2015). The U.S. is losing the social media war. *TIME*, October 12. http://time.com/4064698/social-media-propaganda [March 1, 2018].

Wang, F. Y., Carley, K. M., Zeng, D. and Mao, W. (2007). Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems, 22*(2), 79–83. Available: https://pdfs.semanticscholar.org/f430/9d8913cc9f0d72ec08a4bfb9829866d321d1.pdf [December 2018].

Weeks, B. E., Ardèvol-Abreu, A., and de Zúñiga, H. G. (2017). Online influence? Social media use, opinion leadership, and political persuasion. *International Journal of Public Opinion Research, 29*(2), 214–239.

Wei, W., Joseph, K., Wei, L., and Carley, K. M. (2015). A Bayesian graphical model to discover latent events from Twitter. In *Proceedings of the 9th The International AAAI Conference on Web and Social Media (ICWSM'2015)*. Available: https://www.cs.cmu.edu/~kjoseph/papers/wei_icwsm_15.pdf [December 2018].

Wei, W., Joseph, K., Liu, H., and Carley, K. M. (2016). Exploring characteristics of suspended users and network stability on Twitter. *Social Network Analysis and Mining,* 6(1), 51.

Weimann, G. (2014). *New Terrorism and New Media.* Washington, DC: Commons Lab of the Woodrow Wilson International Center for Scholars.

Wojcik, S. (2018). 5 things to know about bots on Twitter. *Pew Research Center*, April 9. Available: http://www.pewresearch.org/fact-tank/2018/04/09/5-things-to-know-about-bots-on-twitter [April 2018].

Wolfe, S. (2017). The top 10 worst social media cyber-attacks. *Infosecurity,* October 20. https://www.infosecurity-magazine.com/blogs/top-10-worst-social-media-cyber [December 2018].

Woolley, S. C. (2016). Automating power: Social bot interference in global politics. *First Monday, 21*(4). Available: https://firstmonday.org/article/view/6161/5300 [December 2018].

Wu, M., Miller, R. C., and Garfinkel, S. L. (2006). Do security toolbars actually prevent phishing attacks?. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 601–610). New York: Association for Computing Machinery.

Xiong, F., Liu, Y., Zhang, Z.-J., Zhu, J., and Zhang, Y. (2012). An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A, 376*(30-31), 2103–2108.

Yin, H., Song, D., Egele, M., Kruegel, C., and Kirda, E. (2007). Panorama: Capturing system-wide information flow for malware detection and analysis. In *Proceedings of the 14th ACM Conference on Computer and Communications Security* (pp. 116–127). New York: Association for Computing Machinery.

Youngblood, J. R. (2016). *Business Theft and Fraud: Detection and Prevention*. Boca Raton, FL: CRC Press.

Zahedi, F. M., Abbasi, A., and Chen, Y. (2015). Fake-website detection tools: Identifying elements that promote individuals' use and enhance their performance. *Journal of the Association for Information Systems, 16*(6), Article 2. Available: https://aisel.aisnet.org/jais/vol16/iss6/2 [December 2018].

Zargar, S. T., Joshi, J., and Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials, 15*(4), 2046–2069.

Zhao, J., Wu, J., and Xu, K. (2010). Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E, 82*(1), 016105.

Zhou, L., and Zhang, D. S. (2008). Following linguistic footprints: Automatic deception detection in online communications. *Communications of the ACM*, *5*(9), 119–112. doi:10.1145/1378727.1389972.

Zhou, L., Burgoon, J., and Twitchell, D. P. (2003). A longitudinal analysis of language behavior of deception in e-mail. In H. Chen, R. Miranda, D. Zeng, T. Madhusudan, C. Demchak, and J. Schroeder (Eds.), *Proceedings of the First NSF/NIJ Symposium on Intelligence and Security Informatics (ISI 2003), Lecture Notes in Computer Science (LNCS 2665)* (pp. 102–110). Berlin/Heidelberg, Germany: Springer-Verlag.

Zheng, Y., and Wu, G. (2005). Information technology, public space, and collective action in China. *Comparative Political Studies, 38*(5), 507–536.

Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., and Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). doi:10.14763/2016.1.401.