Social Cyber-Security

Kathleen M. Carley[1] [0000-0002-6356-0238], Guido Cervone[2][0000-0002-6509-0735], Nitin Agarwal[2][0000-0002-5612-4753] , Huan Liu[2][0000-0002-3254-7904]

[1] Carnegie Mellon University, Pittsburgh PA 15213, USA
[2] Pennsylvania State University, University Park PA 16801, USA
[3] University of Arkansas Little Rock, Little Rock, AR 72204, USA
[4] Arizona State University, Tempe AZ 85281, USA

Abstract. Social Cyber-Security is an emerging scientific discipline. Its methodological and scientific foundation, key challenges, and scientific direction are described. The multi-disciplinary nature of this field and its emphasis on dynamic information strategies is considered.

Keywords: Social Cyber-Security, Network Science, Social Media Analytics.

1 The Social Cyber-Security Perspective

Social Cyber-security is an emerging scientific area focused on the science to characterize, understand, and forecast **cyber-mediated** changes in human behavior, social, cultural and political outcomes, and to build the cyber-infrastructure needed for **society** to persist in its essential character in a **cyber-mediated** information environment under changing conditions, actual or imminent social cyber-threats. An example is the technology and theory needed to assess, predict and mitigate instances of influence and community manipulation through alterations in, or control of, the cyber-mediated information environment via bots, cyborgs (combination of bot and human) and humans.

Fundamental to this area is the perspective that we need to maintain and preserve a free and open information environment in which ideas can be exchanged freely, the information source is known, disinformation and false data are identifiable and minimized, and technology is not used to distort public opinion. This relies on the notion that movement of information should not compromise the infrastructure, and that actors should not be able to compromise the cyber-environment so as to unduly influence or manipulate individuals, groups and communities. Types of events to be prevented include viral retweeting of messages containing images which if downloaded release malware, or the use of bots to manipulate groups into accepting fake news as real.

In cyber-security much of the emphasis has been on attacks on and through the cyber-infrastructure aimed at impacting technology, stealing or destroying information, and stealing money or identities [1]. In contrast, in social cyber-security the emphasis is influencing or manipulating individuals, groups or communities and so affecting their behaviors with an emphasis on socio-political-cultural consequences. An example is Russian interference in US elections and spread of fake news after Black Panther

movie. While some issues overlap both cyber-security and social cyber-security, the emphasis is different. Cyber-security focuses on technology and social cyber-security on social context and policy. The research in social cyber-security is not focused on maintaining individual privacy, but at how groups are manipulated and opinions shaped. While phishing is in both areas, for those interested in privacy the goal is to avoid individual data being compromised, whereas the goal for those in social cyber-security is the use of phishing as part of a group-level social influence campaign.

2 Social Cyber-Security as Computational Social Science

Social cyber-security is an inherently multi-disciplinary multi-methodological multi-level computational social science. Emerging theories blend political science, sociology, communication science, organization science, marketing, linguistics, anthropology, forensics, decision science, and social psychology. Key relevant theories are related to persuasion [2], social influence [3], participatory democracy [4], individualized collective action [5], information diffusion [6], manipulation [7], group formation and dissolution [8], identity creation [9], strategic messaging [10], information warfare [11], digital forensics [12] and power [13]. Researchers in this area employ multi-technology computational social science tool chains [14] employing network analysis and visualization [15], language technologies [16], data-mining and statistics [17], spatial analytics [18], and machine learning [19]. Finally, the theoretical results and analytics are often multi-level focusing simultaneously on change at the community and conversation level, change at the individual and group level, and so forth.

Social cyber-security is a computational social science and as such, the approach is noticeably distinct from a pure computer science approach or a pure social science approach. The methods and theories being developed: a) take the socio-political context into account methodologically and empirically; b) are predicated on issues of influence, persuasion, manipulation, and theories that link human behavior to behavior in the cyber-mediated environment; and c) are focused on operational utility rather than just improving scores for machine learning algorithms or theory testing. To illustrate the difference, we consider the issue of disinformation and fake news in Twitter.

A purely computer science machine learning approach would start with a training set containing a set of tweets which had been labeled whether containing fake news or not. This set might be split in two groups, one used to train new algorithms and one used to assess their efficacy. Algorithms would then be devised to empirically categorize tweets as to whether or not, and with what certainty, they contained fake news. The precision and recall of the algorithm would be measured and compared against older algorithms to determine their utility. The goal is prediction; however, the algorithms would have limited utility in context other than that in which they were trained. Data sets are widely shared and reused; but, few relevant social cyber-security data sets exist.

In contrast, a pure social science approach might take a set of tweets in some context, identify through secondary sources which were fake, and then statistically assess differences in the number, content, users etc., using the analysis to test a theory about fake news that is predicated on human social behavior but ignores the role of the technology.

Data reuse is often confined to the research group and rare for qualitative data. Qualitative or quantitative support for theories determines their utility. The goal is explanation; however, those explanations are often nuanced to specific socio-cultural settings.

A social cyber-security approach considers both how the technology can be employed to impact: 1) messaging – i.e., who gets what messages when, presentation and access; and 2) group formation – i.e., who communicates with whom when, influence, and group and actor identification. Complex network analytics, visualization, statistics and text mining are used to create empirical profiles of messages that do and don't contain fake news, users that do and don't send the messages, and users who are or are not receptive. New methods are often tested on both new and old data. Method and theory are co-developed, reusable, and extensible to new domains. Their utility resides their ability to support explanation, and prediction in the wild.

Table 1. Contrasting Approaches to Fake News.

Characteristics	Computer Science	Social Science	Social Cyber-Security
Operationally Focused	No	No	Yes
Data reuse	High	Low	Medium
Utility based on	Precision and recall	Theory development and validation	Operational value assessment and prediction value
Tests theory about human behavior	No	Yes	Yes
Empirically driven	Yes	Sometimes	Yes
Considers:			
socio-political context	No	Yes	Yes
media's features	Minimally	Minimally	Strongly
adversarial actions	No	Sometimes	Yes
social influence	No	Yes	Yes
individuals & groups	No	Sometimes	Yes
classes of users	Sometimes	Sometimes	Sometimes

3 What are key challenges to doing research in social cyber-security?

The rapid rate of change in cyber-technologies, evolving legal and policy constraints, and rapid global information flow are creating an environment in which technical, policy and economic issues are strongly impacting what science can be done, what science needs to be done, how that science can be done, and what is required for those who can do that science.

A key challenge is data control. Data are held by and controlled by a few providers who restrict who, how, when and what can be accessed, as well as how, or if, the data

are maintained. While data access is always problematic, the degree of external management, volume of data and pervasiveness of controlled data is unprecedented. While Twitter is only a small portion of the digital landscape it is like a canary in a mine, the early indicator of evolving trends in cyber-space. Unlike other platforms, Twitter is more science friendly due to public tweets falling under the creative commons license and therefore being open and free data that can be harvested for automated analysis. Many scientific papers in the social cyber-security area have focused on Twitter.

Twitter data are not, however, as open as it might seem. There are three dominant ways to access this data: 1) use one of the two Twitter APIs, 2) gain access from Twitter to the 10% feed, and 3) buy Tweets from one of the intermediaries who have access to the 100% feed and historical data. The Twitter APIs provide access to only some of the meta-data around the Tweet, focus on more recent Tweets, and the quality of the sample depends on whether bounding boxes or search terms are used [20]. Further, the samples are biased [21]. Gaining access to the 10% feed typically requires getting one of the few Twitter grants or buying data. The 1% API and the 10% feed are not a random sample of all Twitter data given the search criteria; however, the biases are not well known. Buying the data is extremely costly, but can give you some historical data. Intermediates who provide Twitter data are expected to continuously clean the data and remove recalled Tweets and those by suspended users. They also cannot provide the full meta-data which can reduce the ability to link data sets. Further, these companies may "enhance" the data by adding their determination of language, location or whether the Tweeter is a bot – without explaining how this was determined. Consequently, basic research is needed on bias estimation, impact of missing data, and learning from irreproducible results as the data needed for reproduction may have been deleted.

On the policy side, policies and laws are out-of-sync with the new technologies. Importantly, the rate of change in the technology is such that new forms of illicit activity are emerging at an unprecedented rate. Policies designed to impede, punish or otherwise curtail such activities lag behind the technology. Many policy and law makers have minimal understanding of the technology and so design policy and law that are often irrelevant, or unenforceable, or so restrictive that they prevent the science from being done that would inhibit or detect early social cyber-attacks. Illustrative areas are organizational security, privacy versus detection, and global policies.

Organizations are at risk from social cyber-security attacks. Phony Facebook updates, malware embedded in tweeted image, phishing etc., create organizational insecurities ranging from brand manipulation to compromising personnel to get access to intelligence to destruction of data or machines from social media delivered malware.[1] A 2016 report argued that one in five organizations suffers from a malware attack via social media.[2] The cyber-environment creates yet another risk, in that data-mining coupled with massive on-line data opens the door to corporate secrets being discovered simply by assessing corporate activity including purchasing, personnel hiring, changes in board of directors and so on. Organizations are responding by creating various social

---

[1] https://www.infosecurity-magazine.com/blogs/top-10-worst-social-media-cyber/

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

cyber-security policies such as restricting access to the internet from work, using institutional settings on platforms such as email and dropbox, and increased social cyber-security training general cyber-security training. Drawing from the lessons learned in the nuclear industry, effective organizational policies need to be concerned with heedful interaction, and creating a social cyber-security awareness.

4 Summary

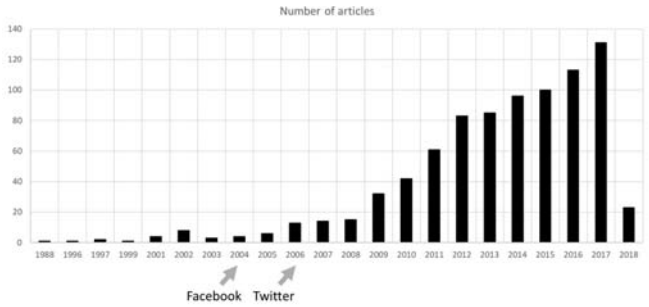

Fig. 1. Number of articles in social cyber-security by year

Social cyber-security is an emerging scientific area concerned with social influence and group manipulation. An estimate of the number of articles based on a snowball from key words in the area and removing those focused exclusively on machine learning algorithms, privacy, or using only a social science approach reveals an exponential growth - see Figure 1. New research is needed in many areas including bias estimation and reduction in data; movement of actors and ideas within and between media; semi-automated identification, assessment of impact of, and effectiveness of counter-messaging for different forms of information strategies; approaches to inoculate individuals and groups against disinformation and effectiveness of those strategies. Future research in this new scientific area is needed to shape the social cyber-environment and promote social cyber-security.

References

1. Reveron, D.S., ed.: Cyberspace and national security: threats, opportunities, and power in a virtual world. Georgetown University Press, Washington D.C. (2012).
2. Gass, R.H., Seiter, J.S.: Persuasion: Social influence and compliance gaining. Routledge, UK (2015).
3. Benigni, M. Joseph, K., Carley, K.M.: Online Extremism and the Communities that Sustain It: Detecting the ISIS Supporting Community on Twitter. PLOS ONE, 12(12), e0181405 (2017).
4. Sunstein, C.R.: # Republic: Divided democracy in the age of social media. Princeton University Press, Princeton NJ (2018).
5. Bennett, W.L.: The personalization of politics: Political identity, social media, and changing patterns of participation. The ANNALS of the American Academy of Political and Social Science 644(1), 20-39 (2012).

6. Wu, L., Liu, H.: Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate. In the Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM2018), ACM, NY NY (2018).
7. Colliander, J., Dahlén, M.: Following the Fashionable Friend: The Power of Social Media: Weighing Publicity Effectiveness of Blogs versus Online Magazines. Journal of advertising research 51(1), 313-320 (2011).
8. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 44-54. ACM, NY NY (2006).
9. Joseph, K., Wei , W., Benigni, M., Carley, K.M.: A Social-event Based Approach to Sentiment Analysis of Identities and Behaviors in Text. Journal of Mathematical Sociology. 40(3), 137-166 (2016).
10. Benigni,M., Joseph, K., Carley, K.M.: Mining Online Communities to Inform Strategic Messaging: practical methods to identify community-level insights. Computational and Mathematical Organization Theory, pp. 1-19 (2017).
11. Cordesman, A.H., Cordesman, J.G.: Cyber-threats, information warfare, and critical infrastructure protection: defending the US homeland. Greenwood Publishing Group, Westport, CT (2002).
12. Al-khateeb, S., Hussain, M.N., Agarwal, N.: Social Cyber Forensics Approach to Study Twitter's and Blogs' Influence on Propaganda Campaigns. In: Lee, D., Lin, Y.R., Osgood, N., Thomson, R. (eds.) Proceedings of the International Conference on Social Computing, Behavioral-Cultural and Prediction and Behavior Representation in Modeling and Simulation, pp. 108-113. Springer, Switzerland (2017).
13. Entman, R.M.: Framing bias: Media in the distribution of power. Journal of communication 57(1), 163-173 (2007).
14. Benigni, M., Carley, K.M.: From Tweets to Intelligence: Understanding the Islamic Jihad Supporting Community on Twitter. In Xu, K.S., Reitter, D., Lee, D., Osgood, N. (eds.) Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 346-355. Springer, Switzerland (2016).
15. Carley, K.M., Wei, W., Joseph, K.: High Dimensional Network Analytics: Mapping Topic Networks in Twitter Data During the Arab Spring. In Shuguan Cui, Alfred Hero, Zhi-Quan Luo and Jose Moura (eds) Big Data Over Networks, Cambridge University Press, Boston MA (2016).
16. Hu, X., Liu, H.: Text analytics in social media. In Mining text data, pp. 385-414. Springer US (2012).
17. Agarwal, N., Kumar, S., Gao, H., Zafarani, R., Liu, H.: Analyzing behavior of the influentials across social media. In Behavior Computing, pp. 3-19. Springer, London, (2012).
18. Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., Waters, N.: Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. International Journal of Remote Sensing 37(1), 100-124 (2016).
19. Wei, W., Joseph, K., Liu, H., Carley, K.M.: Exploring Characteristics of Suspended Users and Network Stability on Twitter. Social network analysis and mining, 6(1), 51. (2016).
20. Carley, K.M., Momin, M., Landwehr, P/M., Pfeffer, J. Kowalchuck, M.: 2016, Crowd Sourcing Disaster Management: The Complex Nature of Twitter Usage in Padang Indonesia. Safety Science, 90, 48-61 (2016).
21. Morstatter, F., Pfeffer, J., Liu, H. Carley, K.M.: "Is the Sample Good Enough? Comparing Data (2013).