Social Cybersecurity: Observations and Directions

Kathleen M. Carley

Social cybersecurity 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 social influence manipulation and the spread of disinformation by inauthentic actors such as bots, cyborgs, and trolls. Social cybersecurity is a computational social science in which the socio-political context is taken in to account, advanced smart technologies operate along side humans, and operational utility of theory and methods is prized. Given the massive and ongoing changes in human communication and the changing affordances of available technologies, this is area where it is critical to think beyond the boundaries of disciplines and to move to transdisciplinary theories and empirical research.

The information environment around COVID-19 lays bare the issues in this area. Compared to prior disasters, or elections, the flow of disinformation is higher (more than 20 times higher than most disasters), more varied (over 9 types of stories), with more global participation in creating and consuming the stories, and with higher consequences for individual health (e.g. death due to drinking bleach), institutional stability and trust in institutions (e.g., disinformation about WHO), and the stability of countries (e.g., due disinformation designed to create polarization). Compared to prior events more individuals and organizations are calling out posts that contain disinformation, and media platforms are removing bot certain types of disinformation and sometimes those who spread it. However, the same disinformation continues to resurface. Disinformation evolves, storylines change, and groups of users, some of whom are bots, collaborate to spread certain disinformation stories; meanwhile, others unknowingly spread the disinformation. AI technologies for identifying disinformation, bots, trolls and hate speech have to be continuously rebuilt due to this evolution. Disinformation story lines don't need to contain inaccurate facts or altered images, but can rely on innuendo, illogic, and implication. This presents an edge case where it is difficult to build AI tools that are accurate, and where the implications of incorrectly identifying a story, image, or user as inauthentic has implications for freedom of speech, government control, and privacy.

These points are illustrated using disinformation storylines from COVID-19 and the elections. A pipeline for characterizing and identifying effective counter manipulation in cyber-space is described. Central to this is the assessment of intent using BEND maneuvers. These maneuvers, and how they shed more light on the influence process in social media vis-à-vis disinformation are described. Key research needs in this area are highlighted.

Social cybersecurity is not just a national challenge, but a global challenge. Like other complex systems, any decision made for this area today will have ramifications for what science is needed, what science can be done, and who will do it, in the future.