2019 11th International Conference on Cyber Conflict:
Silent Battle
T. Minárik, S. Alatalu, S. Biondi,
M. Signoretti, I. Tolga, G. Visky (Eds.)
2019 © NATO CCD COE Publications, Tallinn

Challenges and Opportunities to Counter Information Operations Through Social Network Analysis and Theory

Alicia Bargar, MS
Data Scientist
Asymmetric Operations
Johns Hopkins Applied Physics Laboratory
Laurel, Maryland, United States
alicia.bargar@jhuapl.edu

Janis Butkevics, MS
Data Scientist
Asymmetric Operations
Johns Hopkins Applied Physics Laboratory
Laurel, Maryland, United States
janis.butkevics@jhuapl.edu

Stephanie Pitts, PhD
Social Scientist
Asymmetric Operations
Johns Hopkins Applied Physics Laboratory
Laurel, Maryland, United States
stephanie.pitts@jhuapl.edu

Ian McCulloh, PhD
Chief Data Scientist
Accenture Federal Services
Accenture
Arlington, Virginia, United States
ian.mcculloh@accenturefederal.com

Abstract: Information operations on social media have recently attracted the attention of media outlets, research organizations and governments, given the proliferation of high-profile cases such as the alleged foreign interference in the 2016 US presidential election. Nation-states and multilateral organizations continue to face challenges while attempting to counter false narratives, due to lack of familiarity and experience with online environments, limited knowledge and theory of human interaction with and within these spaces, and the limitations imposed by those who own and maintain social media platforms. In particular, these attributes present unique difficulties for the identification and attribution of campaigns, tracing information flows at scale, and

* This work was funded by the Department of the Navy, Office of Naval Research under ONR Grant No. N00014-18-1-2128. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research.

identifying spheres of influence. Complications include the anonymity and competing motivations of online actors, poorly understood platform dynamics, and the sparsity of information regarding message transferal across communication platforms.

We propose that the use of social network analysis (SNA) can aid in addressing some of these challenges. We begin by providing a brief explanation of the field and its utility in understanding online communications. We discuss how theories drawn from SNA, which seek to make statistical inferences about relationships and information transfer, can be applied to the information operations domain. Specifically, we will focus on how current research in social influence, information diffusion, and cluster analysis can be immediately applied and identify opportunities for future research. We then demonstrate how these analytic techniques can work in practice, utilizing multiple online communication datasets. Finally, we conclude by discussing how the use of these methods can lead to the development of tactical approaches countering misinformation campaigns.

Keywords: *information operations, social network analysis, influence operations, information diffusion*

1. INTRODUCTION

A. Information Operations

Recent events require us to reconsider the role of information operations in modern conflict. The online infrastructure that facilitates civilian communication and organization also provides adversaries with new-found capabilities for exerting influence and disrupting democratic processes. Despite familiarity with information operations (IO) at a strategic level, adversaries' presence in the online environment and the intermingling between different actors complicates the development of countermeasures. How do we disrupt information campaigns without impacting civilian rights?

This paper does not promise universal solutions. Instead, we address the space between policy and practice, drawing upon current social network analysis and theory (SNA/T) research to propose alternative methodologies that can be used when detecting, analysing, and countering IO. The field of social network analysis (SNA) has developed theories and methods for understanding how humans relate, communicate, and spread information. Its relevance for understanding online social phenomena has

cast the field into the spotlight. Although the application to IO is novel, SNA's study of the communication channels upon which IO relies makes it a natural fit.

Definitions of IO vary widely. Military descriptions, like those of NATO and the US Department of Defense, figure most prominently. In JP-313, the US military describes how using information-based systems can "… influence, disrupt, corrupt, or usurp the decision making of adversaries [1]". However, United Nations peacekeeping operations, like the 1999 operation in Kosovo [2], similarly invoke information campaigns to spread awareness and influence in "struggles for control over information identifiable in situations of conflict" [3]. These operations differ due to their alternative objectives and potential lack of adversary. Alternatively, the Canadian Forces' nation-state policy focuses less on assertive actions and more on peacetime strategies to: "deter conflict, protect… information and information systems, and [shape] the information environment" [4, 5].

In this report, we focus specifically on the deterrence of adversarial information campaigns. For clarity, we follow NATO's definition, which describes IO as "military information activities [that] create desired effects on the will, understanding, and capability of adversaries, potential adversaries, and other [North Atlantic Council] approved parties" [6].

We also avoid the term 'information warfare.' Offensive activities with national or international significance can be conducted by non-state groups, criminal organizations, or individuals for personal or economic benefit [7]. We thus choose the term 'operations,' which reflects the complexity of the online environment without implying a nation-state origin.

B. Challenges of the Online Environment

Online domains and social media have become platforms for advancing state-sponsored information campaigns. Most famously, Russian-backed accounts posed as US citizens to spread information prior to the 2016 presidential election [8]. The transferral of IO to the online domain introduces new complexities for developing countermeasures. The attributes below illustrate the unique challenges of the Internet.

1) Anonymity

Identifying information sources online remains difficult. People and organizations obscure their identity for purposes like fun, whistleblowing, trolling, criminal activity, and astroturfing [9, 7].

2) Ease of Coordination

The Internet enables people with similar interests, desires, or beliefs to coordinate more

easily. This heightened capacity for ordinary civilians, communities, or organizations to mobilize at a national or international level creates a new social dynamic that is still not fully understood.

3) Virality

The online environment enables the rapid spread and evolution of information. The fast-paced, global spread of information online makes rumour containment challenging.

4) Multi-stakeholder Governance

The Internet's governance structure reduces state power online. Privacy laws [10], private domain limitations, and individuals' rights to counter government statements online illustrate some considerations that states must take when attempting to gain situational awareness or exert influence.

C. Why Use Social Network Analysis?

Social network analysis studies the underlying patterns of relationships and communications using models known as 'networks.' Network models enable us to address questions such as: 'What communities exist? How does information spread? What is a group's organizational structure?' To answer these questions and others, we combine an understanding of 'relational statistics' [11] with methodologies that ground research in social theories on the variables that influence behaviour.

The advent of the Internet and accompanying datasets inspired new computational techniques that apply SNA to large-scale social systems. As a result, there exists an expansive body of work that utilizes SNA approaches to map out communities, information flows and key actors in online environments. Relevant studies for countering information campaigns include network-based interventions for behaviour change [12], methods for identifying influential information sources [13], and approaches for identifying organizational structures of covert groups [14, 15]. In the next section, we will explain these aspects of SNA to show how they can enhance analytic processes for identifying and countering IO.

Note that SNA alone is not sufficient to develop counter-IO tactics. SNA characterizes content dissemination but not the content itself. Regardless, social networks can help illuminate the social influences and forces present that may spread or contain an IO.

2. COUNTERING ONLINE IO WITH SOCIAL NETWORK ANALYSIS

Scholars, policymakers, and members of the private and public sectors have debated varying measures to counter online disinformation (as defined in [12]). Using these resources, we propose the following linkages between identified needs and SNA/T contributions (Table 1). For the remainder of this section, we discuss each contribution alongside illustrations from relevant work.

TABLE 1: LINKAGES BETWEEN THE EUROPEAN COMMISSION'S HEG REPORT ON COUNTERING DISINFORMATION, NATO IO DOCTRINE, AND SNA/T RESEARCH.

Need [12]	Action [6]	SNA/T Contribution
Identify campaigns	Detect	Anomaly Detection
Monitor scale, techniques, tools, nature, impact of disinformation	Probe	Network Metrics
Identify and map sources and mechanisms	Expose, Deter	Attribution Strategies
Safeguard diversity and sustainability	Protect, Safeguard, Support	Influence Analysis
Counter IO efforts	Disrupt/Diminish/Negate/Prevent	Network Intervention

A. Anomaly Detection

Ruses, stratagems, deceits, camouflages and tricks are as old as war itself and their use… is written in the mists of time. – Paul Villatoux, translated [5]

Identifying where information campaigns exist is a critical first step in the countering process. The frequency and velocity of online discussion makes this a non-trivial task considering the variety of ongoing 'influence' campaigns including product marketing, legitimate political efforts and various organic viral content. Malicious IO campaign detection must both identify the various campaigns and determine which are hostile.

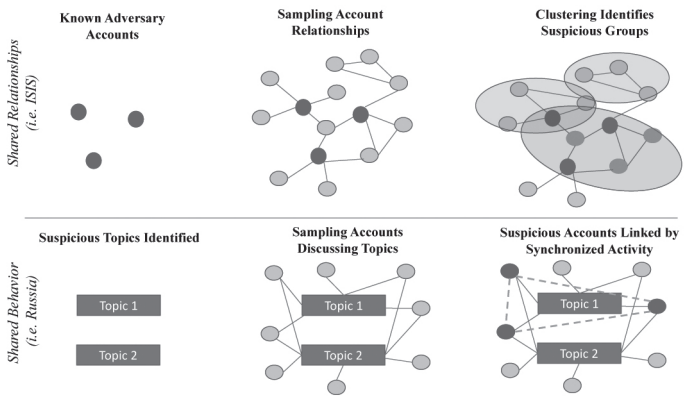
The ability to characterize the interactions between online actors and their intended audiences makes SNA a common tool for information campaign detection [13, 14, 15]. Two popular examples are the Islamic State of Syria (ISIS) online recruitment campaigns and the Russian Federation interference in United States (US) elections. Note that Russian interference is not limited to only US elections, but US elections are a common topic of research and data.

The ISIS online recruitment campaign was a novel approach to manpower sourcing by a terrorist organization [16]. ISIS strongly relied on Twitter to spread propaganda and initialize recruitment across the world. Given ISIS's relatively unique messaging

and tactics, SNA was heavily leveraged [13] for identifying ISIS users on Twitter. ISIS recruiters and propagandists were identified as seed actors, and users who interacted with their accounts were collected. While many users collected in such a manner had no relation to ISIS, the groups of ISIS supporters and non-ISIS Twitter users could be separated into communities through clustering, which is an SNA approach of grouping users into communities based on the attributes of their interactions such as frequency, similarity of connections and other metrics.

Russian influence campaigns opportunistically leverage world events to promote a diversity of objectives. Their use of both human and bot activity allows influence campaigns to scale with large 'astroturf' bot campaigns or targeted posts by humans [8]. Identifying this opportunistic targeting requires different approaches, such as the detection of synchronized actions [17] that appear to focus on a single topic, set of keywords or hashtags, or users. Prominent topic(s) or individual(s) in online discussion can be identified through various SNA metrics such as degree centrality measures, density, or clustering algorithms [14]. Figure 1 illustrates the differences between this approach and the one to identify ISIS accounts.

FIGURE 1. PROCEDURES TO IDENTIFY SUSPICIOUS ACCOUNTS IN ISIS AND RUSSIAN CAMPAIGNS.



SNA techniques can effectively detect change and time of change in networks based on stable relationships between accounts or group-level connections [18, 19]. Another common method for both ISIS and Russian-like campaigns is to pair SNA with machine learning (ML) methods to build systems for automated and possibly near-real-time campaign detection. SNA metrics are coupled with other features like post timing, content analysis and user-specific measures that are then fed into ML models to mine interactions and unique characteristics of the specific campaigns [13, 14, 15, 20].

While joint SNA and ML methods have shown capability to rapidly detect malicious information campaigns, the adversaries continue to adjust tactics, techniques, and procedures (TTPs) to elude them. Along with changing adversary tactics, the limits of available data often curtail effective analysis. Many papers demonstrate detection capabilities on Twitter data, which is relatively easy to obtain. However, data from more secure and private platforms such as Facebook and Instagram are scarce. Furthermore, capabilities to map content and actors across online platforms are in early stages and, therefore, detecting cross-platform information campaigns is currently limited.

B. Network Metrics

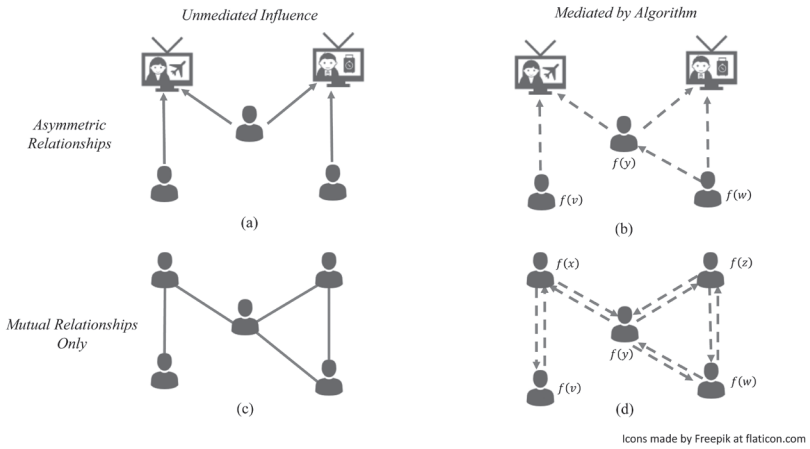
Probe: to examine closely in order to evaluate a system or entity to gain an understanding of its general layout and/or perception. - Allied Joint Doctrine for Information Operations

By finding ways to compare online campaigns, we can begin to build a strategic framework. SNA measures have been applied to study covert network organization [21] and may serve a similar purpose for comparing IO campaign structures. For example, centralization can tell us whether a campaign's communications rely on a few pivotal actors or if its propagation structure is dispersed. Cohesion describes how tightly interconnected people are, whereas modularity measures the extent to which they cluster into groups that infrequently mix. Finally, heterophily represents how often actors with different characteristics interact. In the context of electoral processes, this could measure how often people from differing political backgrounds communicate, thus indicating the likelihood that an IO narrative is shared across political party lines. Figure 2 illustrates how these measures can aid our understanding of a given campaign.

FIGURE 2. THE WHITE HELMETS, A SYRIAN VOLUNTEER RESCUE SQUAD, WERE EVACUATED TO SEVERAL COUNTRIES IN LATE JULY. AS PART OF A BROAD SAMPLING OF TWITTER MESSAGES REGARDING THE SYRIAN CONFLICT, THE AUTHORS COLLECTED ACCOUNTS WARNING THE RECEIVING COUNTRIES OF THE HELMETS' SUPPOSED TERRORIST TIES. THIS IS A TWITTER-MENTION NETWORK FOR THE 'CANADA/WHITE HELMETS/TERRORISM' NARRATIVE IN EARLY AUGUST. THE NETWORK HAS LOW CENTRALIZATION (0.044 USING DEGREE) AND LOW COHESION (0.003 USING EDGE DENSITY), REFLECTING THE LACK OF A DOMINANT ACTOR OR FREQUENT CROSS-NETWORK COMMUNICATION. BECAUSE GROUPS ARE HIGHLY SEPARATE, IT IS HIGHLY MODULAR (0.781 USING UNDIRECTED LOUVAIN). ASSORTATIVITY (0.219) IS ALSO LOW, WHICH REFLECTS THAT PEOPLE TEND TO MENTION OTHERS WITH SIMILAR VIEWS, THOUGH THE VIEWS OF MANY MENTIONED ACCOUNTS ARE UNAVAILABLE.

To create these measures, one must decide on a modelling approach. Network constructions differ by media platform. Figure 3 defines network models based on two dimensions. First, is it possible to have a relationship that is not reciprocated (i.e. to favourite or follow)? Asymmetric relationships are better represented by directed networks (top row) whereas mutual relationships (i.e. to friend) map to undirected networks (bottom row). The second dimension reflects whether algorithms impact the information an actor sees. Unaffected communications are dictated by personal choice and/or timing. When algorithmic influence is present, actors will see different message orderings based on their individual parameterizations.

FIGURE 3. A TAXONOMY OF NETWORK MODELS FOR MEDIA PLATFORMS ALONG TWO DIMENSIONS: ASYMMETRIC VS. MUTUAL RELATIONSHIPS, AND UNMEDIATED VS. ALGORITHMICALLY-MEDIATED COMMUNICATIONS. AS OF 2018, CATEGORICAL EXAMPLES ARE: (A) TRADITIONAL MEDIA OUTLETS (TV, NEWSPAPERS) TO USERS, BLOG LINKAGES, EARLY INSTAGRAM AND TWITTER IMPLEMENTATIONS; (B) CURRENT INSTAGRAM AND TWITTER, YOUTUBE; (C) CHATROOM-LIKE PLATFORMS INCLUDING WHATSAPP, FACEBOOK MESSENGER, SNAPCHAT, AND DISCORD; (D) THE FACEBOOK TIMELINE.



These categories exclude forums like Reddit and 4chan due to the difficulty of distinguishing users with a relationship from users with similar preferences. When this distinction is unnecessary, the mutual relationships/unmediated model can be applied.

Beyond measuring a campaign's organization, we may wish to evaluate its ability to engage and convert users. The innovation-decision process from diffusion of innovation theory maps five stages from awareness to adoption that can frame engagement levels and measures [22]. Characterizing users by stage in a network diagram may help gauge an information operation's impact on a target audience.

'Silent' intermediate objectives intend to shape the network environment and may precede message delivery. For example, researchers studying Russian influence on the US's white supremacy movement found themselves targeted by bot attacks: previously dormant bots followed the researchers *en masse* before flooding their Twitter notifications with messages [23]. This mass-following would be reflected as sudden changes in network measures, including cohesion and centralization.

C. Attribution Strategies

Doing attribution well is at the core of virtually all forms of coercion and deterrence. – Ben Buchanan and Thomas Rid

Attribution is not a common aim for the SNA community [24], but SNA research may be practically applied to address the challenge of online anonymity. For example, methods that emphasize finding consistent patterns and inferring relationships could disambiguate groups across campaigns. Matching accounts across networks can also identify additional data sources for attributional clues.

1) Affiliation Networks

Affiliation networks construct possible relationships between people based on shared event attendance, group membership, or other commonalities [25]. We can use these networks to infer coordination among actors in otherwise potentially unrelated events. For example, Campana reconstructed a human trafficking network's structure using co-event data drawn from court files [26]. By comparing perpetrators' roles with their network positions, the author derived evidence that the trafficking ring was driven by specialized and independent actors rather than a unified organization.

Technical artefacts, including code similarities, media, or metadata, can also define affiliation networks. Saxe and Sanders built a network between malware samples based on shared icons, and found a cluster of linked Trojans. Through additional analysis, they proved that the clustered samples originated from the same source [27]. In IO, shared forums, slogans, or information sources could similarly be employed.

2) Structural Equivalence

If two actors are structurally equivalent, this means that their relationships are identical [28, 29]. This 'structural redundancy' can provide valuable clues. For example, in September 2014, Twitter began aggressively suspending ISIS accounts. That same month, there was a sudden surge of new ISIS-supporting accounts [30]. Preventing banned users from creating new accounts is difficult, but looking for structural equivalence over topics can help identify these 'rebound accounts.'

Analyzing actors across campaigns could be aided by ongoing research into how to compare roles between networks. Jeffrey Johnson's ethnographic approach of operationalizing social roles through detailed case studies is inspired by anthropology, but may be applicable for those actively participating in covert networks like dark web forums [31]. Some computational approaches include block modelling [32] and regular equivalence [33]. The practical need to identify TTPs, track operational consistency, or profile actor types may incentivize extending this theoretical work.

3) Network Deanonymization

Network deanonymization attempts to reconstruct actor identities in a network by matching them to the population of another network containing additional information. Ji et al. survey deanonymization techniques in [34] and provide a table (Table III) with information on their scalability, practicality, and computational efficiency. Most approaches require or are made significantly more effective with the presence of 'seeds', or successfully matched actors. Another challenge is identifying how well the known network's population matches that of the hidden network. [35] explores methods to determine which auxiliary networks are most promising using the nodes' network properties. Despite favourable results, follow-on efforts to explore and define its feasibility are still lacking. Due to ethical and regulatory concerns surrounding privacy [10], it is advised to fully understand one's rights and limitations before attempting this approach. Regardless, it may prove useful for determining whether a known group has instigated a particular campaign.

D. Influence Analysis

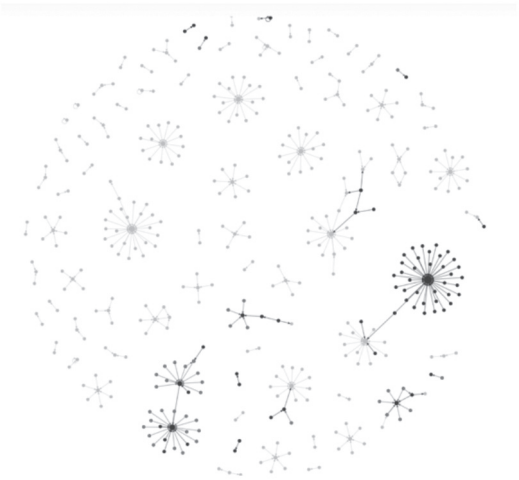
Power is unthinkable outside matrices of force relations; it emerges out of the very way in which figurations of relationships… are patterned and operate. – Mustafa Emirbayer

The European Commission Report states the need to 'safeguard diversity and sustainability' online [12]. The online environment is not a static system. Understanding how the rise and fall of influence is facilitated by social structures, dynamics, and platform design may guide the development of principles for future moderation efforts.

1) Identifying Key Actors

Many centrality measures exist for identifying important actors. Degree centrality helps identify particularly popular individuals. In an asymmetric network, the highest-degree actors are those most followed: examples include media outlets, influential bloggers, or maintainers of popular channels or podcasts. Figure 4 shows how one can track a message's dominance by how frequently its proponents are quoted. Other roles within a network provide different types of influence. For example, one person may serve as a frequent mediator between two groups, such as a translator who interprets messages across linguistically bound communities. This may be captured using betweenness centrality, which measures how frequently an actor is present in communications across the network. Refer to [36] for further discussion of other centrality measures and their applicability and interpretability.

FIGURE 4. IN OUR ANALYSIS ON THE SYRIAN CONVERSATION, WE CREATED THIS NETWORK OF QUOTES AND RETWEETS FROM THE CONVERSATION ABOUT CANADA AND THE WHITE HELMETS. DARK NODES ARE USERS PUSHING THE WHITE HELMETS/TERRORIST NARRATIVE. THE MEDIUM NODES HAVE QUOTED THEM ON OTHER TOPICS, INDICATING THAT THEY WERE LIKELY EXPOSED TO THE NARRATIVE. ACTORS WITH THE HIGHEST DEGREE IN THIS NETWORK ARE THOSE WHO MOST SUCCESSFULLY HAD THEIR MESSAGE AMPLIFIED.



2) Creating Online Influence

Algorithms have an unseen effect on online communications. By altering communications between users, this mechanism changes what impact influencers can have on their connections. For example, YouTube's recommendation algorithm has been accused of promoting extreme content [37]. By recommending certain channels over others, this algorithm influences a user's choice of information sources. Algorithmic newsfeeds curate content based on a user's past preferences and actions [38, 39], thus shifting a user's likelihood of exposure to certain sources or posts. A content provider's ability to utilize these algorithms can determine their own influencing capabilities.

3) Influence Campaigns and Moderation

Online groups constantly seek to better promote their own personal or political beliefs, including state-based operations and extremist groups like ISIS. 4chan's famous trolling forum /pol/ attempted to influence the Google search algorithm to correlate racist terms with innocuous words [40]. Civilians have also used online platforms to increase their political influence, as seen in such high-profile cases as the Arab Spring, the 2017 Women's March, and the Gilets Jaunes.

Furthermore, maintaining the online influence space as a free and balanced marketplace of ideas is as much an economic challenge as it is a technical or political

one. The monetization capability of influence has led to strategic product placement in influencers' posts, via allocated ad spaces, and even using false accounts [9] to promote word-of-mouth recommendations. Social interactions have financial value in the online world, and it is unclear to what extent this complexity has been considered in our current models of online communications.

Platforms and internet providers also have the ability to impact influencers' capabilities. Moderation efforts span from top-down driven administration, like Twitter's efforts to combat ISIS accounts [30], to Wikipedia's decentralized organization [41]. Censorship shocks on the Mandarin Wikipedia demonstrate the possibilities of Internet provider effects [42].

Some of the most recent developments in SNA are dedicated to better understanding these phenomena. Refer to [43] for a variety of techniques designed to tease out the source of a diffused message. Exponential random graph models (ERGMs) and stochastic actor-oriented models (SAOMs) are applied to test theories of how micro-behaviours lead to differences in network structure [44, 45, 46]. Relational event models (REMs) and Dynamic Network Actor Models (DyNAMs) consider the likelihood of an actor's actions based on their relationships and environmental factors [47, 48]. As the computational cost of dynamic modelling is reduced, ongoing work in this area holds promise for further illuminating the causes and influences of online dynamics.

E. Network Interventions

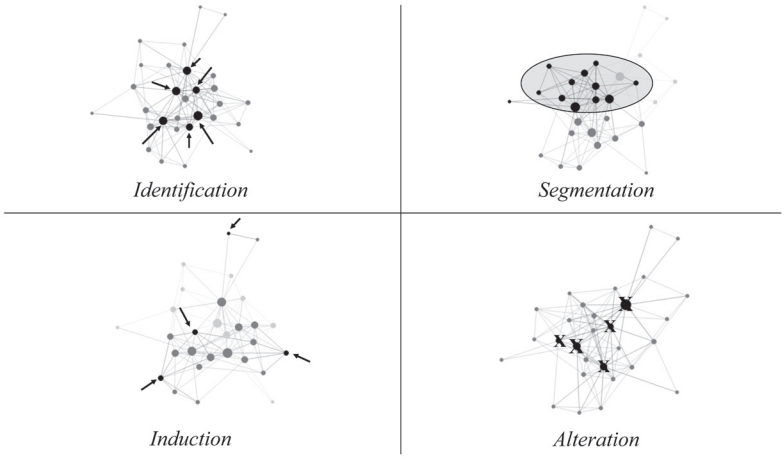
Example is not the main thing in influencing others. It is the only thing. - Albert Schweitzer

Network interventions use social influence forces to promote behaviour change [49, 22]. These interventions are based on the concept that exposure to a behaviour increases one's likelihood of adoption and that the influence of one's peers can be harnessed to spread desired behaviours. The authors have not identified conscious applications of network interventions to counter IO efforts; however, these interventions offer a comprehensive framework to classify suggested counter-IO tactics and inspire new approaches. Table 2 describes each intervention strategy and their unique capabilities, and Figure 5 demonstrates their potential for operationalization.

TABLE 2: NETWORK INTERVENTION STRATEGIES AND EXAMPLES

Strategy	Description	Example	Risks
Identification	Use network structure to identify actors to train to spread desired messaging or behaviour.	Opinion leaders on a social media site are identified and encouraged to spread counter-IO content.	Opinion leaders may not want to share desired content, or may share content that inadvertently increases belief in misinformation.
Segmentation	Simultaneously target actors that are in a well-connected group or in shared positions.	Clusters of highly connected accounts are located and targeted with counter-IO messaging.	Targeted accounts may view content as an attack on their community. Existing connections among members may reinforce current behaviour if members are not accepting of counter-IO information.
Induction	Promote communication across existing relationships in the network to disseminate desired messaging.	Civilians in affected area are encouraged to provide their accounts of on-ground activities.	Campaign hashtags or keywords may provide vehicles for continued misinformation spread. Civilian accounts may be framed to support an undesirable narrative.
Alteration	Change network structure to alter exposure and message spread.	Bridging actors on social media platforms are identified and trained in misinformation detection and handling to reduce the likelihood of their transmission of misinformation.	Actors may misunderstand intervention materials and increase the spread of misinformation through well-intended efforts to correct misinformation.

FIGURE 5. GIVEN THE CANADA/WHITE HELMETS/TERRORISM NARRATIVE, HOW COULD WE DESIGN AN INTERVENTION? IDENTIFICATION COULD TARGET HIGH-DEGREE NODES, WHILE INDUCTION WOULD BE MORE RANDOMLY DISPERSED. SEGMENTATION WOULD LOOK FOR CLUSTERS, AND ALTERATION COULD ADDRESS NODES THAT CONNECT DIVERSE POPULATIONS USING BETWEENNESS CENTRALITY.



Identification techniques engage actors in key positions in a network for training or messaging, with the expectation that their actions will impact the overall network. For example, rumour blocking simulations model the spread of misinformation and credible information simultaneously. Some researchers use identification techniques within these models to identify an optimized subset of users to spread credible information more effectively [50].

A tightly-knit group with few external relationships and frequent sharing of homogenous content can become an echo chamber. Segmentation methods can intervene with an echo chamber as a collective set so all actors receive content simultaneously.

Induction techniques reframe a narrative by actively encouraging people to communicate with one another. An example of this approach would be a word-of-mouth campaign that asks civilians to share their views on a topic with photos or other user-created media. Sharing user experiences from those close to an on-the-ground situation may aid in combating false information pertaining to that situation.

Finally, alteration methods modify network structure by adding or deleting links and/or nodes. Note that removing malicious bots or accounts from online platforms does not necessarily eliminate them: bot masters may make new accounts that are harder to detect or migrate to other platforms. However, not all forms of node removal require explicit removal. Analogous to how vaccinations prevent disease transmission, we can focus on techniques that reduce accounts' transmission of misinformation [49]. Training and messaging actors who play central roles in spreading information may effectively reduce an IO's diffusion through a network. Finally, link-based alteration strategies include encouraging people to connect or disconnect from particular accounts. A recent report suggested that actors that connect more with people that have differing opinions may reduce their belief in misinformation [51].

Node addition may be an overlooked tactic for network alteration. Self-identified bots could serve as assistive devices to provide just-in-time content to counter or distract from disinformation. For example, a monitoring account could analyse tweets and reply with an automated analysis of potentially coercive or emotionally evocative content, though the risk of false positives should be considered. Simulations have been conducted to inform optimal monitor placement within a network for misinformation detection for early containment [52].

3. CONCLUSION AND FUTURE DIRECTIONS

Through SNA, we gain a theoretical lens and applicable methodologies for examining and countering IO. Some of SNA's capabilities, like centrality measures and clustering, have been frequently applied to the online environment while others remain underutilized. Here we seek to broaden the audience's perspective of ongoing research in the field.

We note that social network analysis is not a cure-all for addressing IO. Because it is message-agnostic, theories related to the shaping and framing of a narrative are absent from this work. Furthermore, no tool replaces the need for collaboration among stakeholders.

Regardless, SNA has strong potential when combined with other technical and political techniques. As demonstrated above, SNA-combined approaches lead to more effective ways to identify information campaigns and extremist organizations than machine learning alone. Network-based measures and attributional information can help guide the decision-making process regarding whether to address potential campaigns. Finally, network intervention techniques provide potential strategies for implementing campaign countermeasures. We encourage policymakers and researchers alike to consider how SNA methodologies can further the development of countermeasures against online IO.

Acknowledgments

We thank the anonymous reviewers, Brad Ward, Dr David Silberberg, Dr Anthony Johnson, John Renda, Kristine Henry, and Daniel Kapellmann for their guidance on this work.

BIBLIOGRAPHY

[1] US Joint Staff, "Joint Publication 3-13 Information Operations," Government Printing Office, Washington DC, 2014.

[2] D. Lindley, "Untapped power? The status of UN information operations," *International Peacekeeping*, vol. 11, no. 4, pp. 608-624, 2004.

[3] K. Avruch, J. L. Narel and P. C. Siegel, Information Campaigns for Peace Operations, Washington DC: Office of the Assistant Secretary of Defense, Washington DC Command and Control Research Program (CCRP), 2000.

[4] Canadian Forces, "Canadian Forces Information Operations," Canadian Forces, Ottawa, 1998.

[5] R. Vandomme, "From Intelligence to Influence: The Role of Information Operations," Canadian Forces College, Toronto, 2010.

[6] NATO, "Allied Joint Doctrine for Information Operations," NATO, Talinn, Estonia, 2009.

[7] D. Denning, *Information Warfare and Security*, Reading, MA: Addison Wesley, 1999.

[8] S. Shane and M. Mazzetti, "The Plot to Subvert an Election," *The New York Times*, 20 9 2018.

[9] M. Kovic, A. Rauchfleish, M. Sele and C. Caspar, "Digital astroturfing in politics: Definition, typology, and countermeasures," *Studies in Communication Sciences*, vol. 18, no. 1, pp. 69-85, 2018.
[10] J. Soetbeer, "European Data Protection Regulation – Information Sheet," 1 3 2016. [Online]. Available: https://www.privacy-europe.com/blog/european-data-protection-regulation-information-sheet/. [Accessed 9 12 2018].
[11] U. Brandes, G. Robins, A. McCranie and S. Wasserman, "What is network science?" *Network Science*, vol. 1, no. 1, pp. 1-15, 2013.
[12] European Commission High Level Group, "A multi-dimensional approach to disinformation - Report of the independent High Level Group on fake news and online disinformation," Publications Office of the European Union, Belgium, 2018.
[13] M. C. Benigni, K. Joseph and K. M. Carley, "Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter," *PloS one*, vol. 12, no. 12, p. e0181405, 2017.
[14] O. Varol, E. Ferrara, F. Menczer and A. Flammini, "Early detection of promoted campaigns on social media," *EPJ Data Science*, vol. 6, no. 1, p. 13, 2017.
[15] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Goncalves, A. Flammini and F. Menczer, "Detecting and tracking political abuse in social media," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)*, Barcelona, 2011.
[16] J. M. Berger, "Tailored online interventions: The Islamic State's recruitment strategy," *CTC Sentinel*, vol. 8, no. 10, pp. 19-23, 2015.
[17] Q. Cao, X. Yang, J. Yu and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, Scottsdale, AZ, 2014.
[18] I. McCulloh and K. M. Carley, "Detecting Change in Longitudinal Social Networks," *Journal of Social Structure*, vol. 12, no. 3, pp. 1-37, 2011.
[19] I. McCulloh, M. Webb and K. M. Carley, "Social Network Monitoring of Al-Qaeda," *Network Science*, vol. 1, no. 11, pp. 25-30, 2007.
[20] O. Varol, E. Ferrara, C. A. Davis, F. Menczer and A. Flammini, "Online Human-Bot Interactions: Detection, Estimation, and Characterization," *arXiv*, vol. 1703, no. 03107, 2017.
[21] S. F. Everton, *Disrupting dark networks* Vol. 34, New York: Cambridge University Press, 2012.
[22] E. Rogers, *Diffusion of Innovations*, 5th edition, New York: Free Press, 2003.
[23] J. Cox, "The Day an Army of Bots Turned on Bot Researchers," *The Daily Beast*, pp. https://www.thedailybeast.com/the-day-an-army-of-bots-turned-on-bot-researchers?ref=scroll, 29 8 2017.
[24] N. Hummon and K. M. Carley, "Social networks as normal science," *Social Networks*, vol. 15, no. 1, pp. 71-106, 1993.
[25] S. Borgatti, M. Everett and J. Johnson, "Analyzing Two-Mode Data," in *Analyzing Social Networks*, Los Angeles, CA, SAGE, 2013, pp. 267-286.
[26] P. Campana, "The Structure of Human Trafficking: Lifting the Bonnet on a Nigerian Transnational Network," *The British Journal of Criminology*, vol. 56, no. 1, pp. 68-86, 2016.
[27] J. Saxe and H. Sanders, "Identifying Attack Campaigns with Malware Analysis," in *Malware Data Science: Attack Detection and Attribution*, San Francisco, No Starch Press, Inc., 2018, pp. 54-58.
[28] H. C. White and F. Lorrain, "Structural equivalence of individuals in social networks," *The Journal of Mathematical Sociology*, vol. 1, no. 1, pp. 49-80, 1971.
[29] R. Burt, "Social contagion and innovation: Cohesion versus structural equivalence," *American Journal of Sociology*, vol. 92, no. 6, pp. 1287-1335, 1987.
[30] J. M. Berger and J. Morgan, "The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter," The Brookings Project on U.S. Relations with the Islamic World, Washington D.C., 2015.
[31] J. C. Johnson, C. Avenarius and J. Weatherford, "The Active Participant-Observer: Applying Social Role Analysis to Participant Observation," *Field Methods*, vol. 18, no. 2, pp. 111-134, 2006.
[32] H. C. White, S. A. Boorman and R. L. Breiger, "Social Structure from Multiple Networks," *American Journal of Sociology*, vol. 81, no. 4, pp. 730-780, 1976.
[33] D. R. White and K. P. Reitz, "Graph and semigroup homomorphisms on networks of relations," *Social Networks*, vol. 5, no. 2, pp. 193-234, 1983.
[34] S. Ji, P. Mittal and R. Beyah, "Graph Data Anonymization, De-Anonymization Attacks, and De-Anonymizability Quantification: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1305-1326, 2017.
[35] P. Govindan, S. Soundarajan and T. Eliassi-Rad, "Finding the most appropriate auxiliary data for social graph deanonymization," in *1st KDD Workshop on Data Ethics*, New York, New York, 2014.

[36] S. P. Borgatti, M. G. Everett and J. C. Johnson, “Centrality,” in *Analyzing Social Networks*, Los Angeles, SAGE Publishing, 2013, pp. 189-208.
[37] Z. Tufekci, “YouTube, the Great Radicalizer,” *The New York Times*, p. SR6, 10 3 2018.
[38] M. A. DeVito, “From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed,” *Digital Journalism*, vol. 5, no. 6, pp. 753-773, 2017.
[39] N. Koumchatzky and A. Andryeyev, “Using Deep Learning at Scale in Twitter’s Timelines,” 9 5 2017. [Online]. Available: https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html. [Accessed 9 12 2018].
[40] G. E. Hine, J. Onaolapo, E. D. Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini and J. Blackburn, “Kek, cucks, and god emperor Trump: A measurement study of 4chan’s politically incorrect forum and its effects on the web,” in *Proceedings of the 11th International AAAI Conference on Web and Social Media, ICWSM*, Montreal, 2016.
[41] A. Forte, V. Larco and A. Bruckman, “Decentralization in Wikipedia Governance,” *Journal of Management Information Systems*, vol. 26, no. 1, pp. 49-72, 2009.
[42] A. F. Zhang, D. Livneh, C. Budak, L. Robert and D. Romero, “Shocking the Crowd: The Effect of Censorship Shocks on Chinese Wikipedia,” in *The 11th International Conference on Web and Social Media*, Montreal, Canada, 2017.
[43] Jiaojiao, S. Wen, S. Yu, Y. Xiang and W. Zhou, “Identiying Propagation Sources in Networks: State-of-the-Art and Comparative Studies,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 465-481, 2017.
[44] G. Robins, T. A. B. Snijders, P. Wang and M. Handcock, “Recent developments in exponential random graph (p*) models for social networks,” *Social Networks*, vol. 29, pp. 192-215, 2007.
[45] D. A. McFarland, J. Moody, D. Diehl, J. A. Smith and R. J. Thomas, “Network Ecology and Adolescent Social Structure,” *American Sociological Review*, vol. 79, no. 6, pp. 1088-1121, 2014.
[46] T. A. B. Snijders, “Stochastic actor-oriented models for network change,” *Journal of Mathematical Sociology*, vol. 21, no. 1-2, pp. 149-172, 1996.
[47] C. T. Butts, “A Relational Event Framework for Social Action,” *Sociological Methodology*, vol. 38, no. 1, pp. 155-200, 2008.
[48] C. Stadtfeld, J. Hollway and P. Block, “Dynamic Network Actor Models: Investigating Coordination Ties through Time,” *Sociological Methodology*, vol. 47, no. 1, pp. 1-40, 2017.
[49] T. W. Valente, “Network Interventions,” *Science*, vol. 337, no. 6090, pp. 49-53, 2012.
[50] I. Litou, V. Kalogeraki, I. Katakis and D. Gunopulos, “Real-Time and Cost-Effective Limitation of Misinformation Propagation,” in *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, Porto, 2016.
[51] D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs and C. Mattison, “Combating fake news: An agenda for research and action,” in *Combating fake news: An agenda for research and action,* Cambridge, MA, 2017.
[52] H. Zhang, M. A. Alim, M. T. Thai and H. T. Nguyen, “Monitor placement to timely detect misinformation in Online Social Networks,” in *2015 IEEE International Conference on Communications (ICC)*, London, 2015.