Understanding Social Media Influence, Semantic Network Analysis, and Thematic Campaign Classification Using Machine Learning.

by

Nathan Johnson

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2022 by the
Graduate Supervisory Committee:

Martin Reisslein, Chair
Benjamin Turnbull
Ming Zhao
Yanchao Zhang

ARIZONA STATE UNIVERSITY

May 2022

ABSTRACT

Individuals and organizations have greater access to the world's population than ever before. The effects of Social Media Influence have already impacted the behaviour and actions of the world's population. This research employed mixed methods to investigate the mechanisms to further the understand of how Social Media Influence Campaigns (SMIC) impact the global community as well as develop tools and frameworks to conduct analysis. The research has qualitatively examined the perceptions of Social Media, specifically how leadership believe it will change and it's role within future conflict. This research has developed and tested semantic ontological modelling to provide insights into the nature of network related behaviour of SMICs. This research also developed exemplar data sets of SMICs. The insights gained from initial research were used to train Machine Learning classifiers to identify thematically related campaigns. This work has been conducted in close collaboration with Alliance Plus Network partner, University of New South Wales and the Australian Defence Force.

ACKNOWLEDGMENTS

To my wife and son. My favourite people in the world, whom I love with all my heart. Thank you for accepting my general absence and reclusive behaviour over the past four years. I look forward to reallocating the hours to quality time as a family and finding ways to repay the support. You have and always will, be my steady rock in the stormy ocean of life. To Dr Benjamin Turnbull, your tireless efforts, constant supervision and selfless nature has been integral to this research. I thank you for directing my academic and professional growth over the past 4 years. To Professor Martin Reisslein, thank you for your efforts as my supervisor and enduring the long nights of review and COVID safe interaction with me. I would like to recognise the support of the Australian Defence Force, specifically General Angus Campbell. Without your leadership, belief in the Soldier Scholar and intense sense of greater good, this research would not have been possible. I sincerely thank you for providing the support conditions for my family and I during my academic pursuits. Finally, I recognise the General Sir John Monash Foundation, the American Australian Association and Arizona State University. Their dedication to their scholars is beyond reproach and I will be eternally grateful to have been accepted as a scholar of these institutions. Without their support, this research would also not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

0.1 INTRODUCTION

Daily life is now a hybrid of Social Media, Social Networking, digital communications, as well as physical communications and interactions. Facebook, Twitter, and Instagram boast over two billion monthly active users Dhiraj (2019), and as such, their ability to directly and indirectly connect the world's population has never been easier or more far reaching. Online content delivery, and the algorithms that govern it, have changed both the method and the speed at which our world communicates, consumes, decides, and progresses. There is now sufficient data that has been voluntarily shared on social media by users, that their beliefs and behavioral responses can be predicted and manipulated Cadwalladr (2018). An important open question remains: are the underlying principles and mechanics of influence between the physical and digital (cyber) realms understood? In order to progress the understanding of influence flow between the physical and cyber realms, there is a requirement to not only contextualize but attribute actions to reactions across both realms. To begin this research, the established mechanisms and rational of influence transfer via social media are as follows.

0.1.1 Social Media Influence

Unregulated Publishing

In the United States of America, public broadcasters and publishers are mandated to abide by federal Public Broadcasting Act of 1967. These guidelines have been developed over years to avoid situations that insight mass hysteria or public panic and ensure transparency of what is being reported or published Grainger (1999). In contrast, Social Networking Services (SNS) provide a mechanism for publication of user generated content. The concept of user generated material side skits the issue of fact checking and source referencing the material, as it's opinion based, be it accurate or not. Moreover, whilst FaceBook, YouTube and Twitter are moderated, the volume and ease of creating profiles to publish out paces any attempts to capture all material that breaches their code YouTube (2005). YouTube also sells advertising and shares the profit with the creators, therefore, regulation becomes the responsibility of the creators or they forfeit revenue. Other social media services that do not have payment mechanisms like YouTube, i.e twitter or 4chan, who's economy is attention and influence, offending material must rise above the detection threshold of moderators before content is removed to minimise it's impact.

Reach

At the end of 2019, it was estimated that the number of SNS users was in excess of 3.5 Billion Ema (2001). As of 2021, that number is estimated to be between 3.8 Billion and 4.4 Billion depending on what you define as social media, which is more than 5% growth globally dat (2021). These services have the ability to reach more than half the worlds population immediately. This figure is constantly growing with another billion users online and another billion using mobile devices daily Ema (2001). Regardless of their nature, SMICs have the ability to influence the world on a scale never before seen. To put this in perspective, famous actor, Sacha Baron Cohen gave a speech to the Anti-Defamation League (ADL) and made the comment, "Just Think What Goebbels Could Have Done with Facebook" Cohen (2019)

Information Source

The average user will spend between 1 to 2 hours a day on SNS which is rapidly increasing and inversely proportional to the decline in consumption of traditional media such as television news and printed mediaMcLennan and Miles (2018). Therefore, for the majority of users, this is their sole source of daily news and events. Combined with with the earlier mentioned absence of fact checking and the lack of moderation, both acute and enduring misinformation campaigns can run freely without debate or conjecture.

Content Algorithms

Selection of what information and content is delivered to a user's device is governed by a content delivery algorithm based on patterns from historical metadata collected from the users device or network Ray (2019). Moreover, the user guided experience, i.e. selection of who to follow and who to ignore results in an extremely filtered and monochromatic exposure to information and narrativesdev (2017). The rise of the anti-vaccination movement is an example of distributing incorrect information to undermine proven historical research and fact. As a result, deadly diseases once eliminated in various countries are on the risedev (2017). The power to manipulate what demographics, communities and individuals see and consume is without precedence, and one that is obviously beyond the control of the platform managers. All of these elements suggest that research must continue into in order helo support countries and populations defend themselves against misuse.

Cost of Entry

A significant portion of the challenge of moderating SMICs is that there is little to no cost to enter the conversation. Different to that of a News Paper, Radio program, Television broadcast which requires access to infrastructure. The only requirements are a profile and internet connection which is ubiquitous in the modern world. With the emergence of modern fifth generation (5G) wireless systems, the connection to the internet is becoming even more ubiquitous via various gateway technologies Alalewi *et al.* (2021); Hoeschele *et al.* (2021); Thyagaturu *et al.* (2016) reaching people in crowded settings, e.g., during protests Ali *et al.* (2017); Tyagi *et al.* (2013, 2015); Weitzen *et al.* (2016), as well as people on the move Akkari and Dimitriou (2020). Also, the multi-access Taleb *et al.* (2017); Doan *et al.* (2021) and accelerated computing paradigms Benzaïd and Taleb (2020); Linguaglossa *et al.* (2019); Shantharama *et al.* (2020) that are emerging with 5G systems lower the cost of entry for the manipulation of cyber messaging.

0.1.2 Cyber Influence

Whilst these mechanisms discussed are not exhaustive, they account for the last decade of exponential growth of Social Media. Social Networking Services (SNS) remain largely unregulated by government Tench and Jones (2013), and have been weaponized to become *Cyber Influence* Singer and Brooking (2018a). Social influence being "the change in one's beliefs, behavior or attitudes due to external pressures that may be real or imagined" Guadagno and Cialdini (2010). Cyber influence, is social influence via digital means, which research shows is commensurate and in some cases even more powerful than physical influence Chen *et al.* (2014). Cyber influence has already been employed to mobilise oppressed populations, win elections, fight wars, and undermine drug cartels Singer and Brooking (2018a). Given this unprecedented access and undeniable force, there has been significant research into the diverse nature of Social Media based influence and potential effects.

Examples of Cyber Influence or Social Media influence on commerce from review mechanisms, Lee *et al.* (2021), developmental concerns from Social Media and it's use Boer *et al.* (2021) and increasing Social Media's integration with our lives Shumanov and Johnson (2021). That said, Much of this research does not critically analyse how Social Media is being perceived, and more importantly how this perception impacts elements of human behaviour and human interactions such as conflict. Numerous studies have been conducted into distrust, manipulation and how the influence of Social Media is being abused but are these warnings being heeded by the community? What element of Social Media make people behave this way or approach important issues like conflict? What precautions or preparations need to be taken now to protect the population and their interests in the future?

It cannot be denied that social media has ushered in a new era of connection with communities. However, it is also playing a key role in dividing it. Research has shown that social media is not only impacting behaviour but guiding it Romeo *et al.* (2021), leading to its increased role in connection to conflict Singer and Brooking (2018b). Vast amounts of time, money and research has been devoted to the identification and discouragement of Fake News but what is the perception of influence and trust regarding the information on Social Media and what does this perception tell us about now and the future?

These open questions suggest that there is a requirement to understand how Social Media is perceived and how it is effecting the population with regards to key elements of trust, influence and conflict across the community. Moreover, there are engineering requirements to build better databases that are semantically based with the ability to combine both physical event with online events. It is also paramount to investigate the impacts of Social Media Influence Campaigns waged to help understand the mechanics of the influence exchange which then becomes the basis of tools that take a holistic approach to influence analysis. This approach can then be adapted to help develop Artificial Intelligence and machine learning tools to further protect the community both now and in the future against negative actors. This research aimed to address the concerns raised by first understanding where experts in both social media and influence believed we had the most to learn. Second, this research took these insights then built semantic databases to help with analysis and understanding the link between physical and online events. Cross domain or multi-realm behavior has yet to be fully understood. Finally, this research has also build Artificial Intelligence and machine learning methods to classify and detect thematic campaigns in an attempt to extract them from the massive social media data stream.

0.1.3 Data sets

For this research a number of data sets where purchased and used. Data sets can be generated using the Twitter API, or purchased off a third party like TweetBinder. Due to our time restraints, we elected to purchase the data sets. Each data set is discussed in detail in the thesis in their respective sections that employed them. However, the following is a general over view of each set.

- #Euromaidan. This data set has approximately 1.6 million tweets ranging over the period between 01 November 2013 - 31 March 2014. It includes 3 other hashtags,#, #, #ukraine It covered the Ukrainian crisis and the Russian occupation of Kyiv. The retrospective conducted by TweetBinder captured the global tweets published with the hashtag 'Euromaidan' regardless of their nature, content or intent.

- #BalakotAirstrike. This was the collective data set with approximately 1 million tweets ranging over the period 14 February 2019 - 11 April 2019. This covered 8 hashtags, #BalakotAirstrike, #indiastrikesback, #terroristanpakistan, #pakapologist, #AvengePulwana, #surgicalstrike2, #pakistanstrikesback and #IndianFailedStrike. All of which are related to the Indian and Pakistan conflict of 2019. Again, the TweetBinder search captured global tweets on the 6 hashtags over this period, regardless of their nature, content or intent.

- NN modelling data sets. We needed different data sets for the NN model experiments. The main difference was the requirement for similiar sized data sets that captured various thematic campaings. This time, smaller data sets where purchased from TweetBinder on the following hashtags, #FIFAWWC, #MTVEMA #zagrebearthquake and #ARRESTTRUMPNOW. The methodology did not demand a full campaign nor over a speific time frame of activity, hence, we purchase 40,000 tweets on each hashtag of the most recent capture.

Each data set has been published on the IEEE Data Portal for public access and reproduction of results if required.

0.1.4 Original Contributions

This section summarizes the original contributions made by this research. The majority has been captured in three individual publication within IEEE's Access, Elsevier's Technology in Society and IEEE Transaction Special Issues - Generating Human Readable Interpretations in Natural Language Processing. The following items highlight the original contributions made by each section:-

- Quantitative Survey of Leaders in Social Media and conflict. This research presented the results of a quantitative survey of Civilian and Military leadership within a Social Media and conflict context. Focusing on warlike future conflict issues, it presented the predictions and future trends of how Social Media will influence warlike conflict and potentially evolve. As a result of the research, we were able to suggest key themes and ideas, such as the importance of Networks, to not only be aware of Social Media changes, but prepare for their impacts.

- Social Media Influence Campaigns. In response to the first research article's identification of network importance in Social Media, this article presented a novel 'semantic model and analytical framework' for analysing Social Media Influence Campaigns. It did so by proposing a semantic (relationship) based database and using semantic based search algorithms to generate novel knowledge on campaigns. Moreover, this article confirmed it's claims by conducting two case studies and comparing the outcomes to each other which had previously not been archived.

- Thematic Campaign Classification Framework. Finally, the third article was an extension of the first and second articles analytical concepts. This research investigated employing Artificial Intelligence and Machine Learning methods to classify Social Media campaigns. A novel Neural Network Model was developed to classify campaigns based on their thematic basis (sports, political, conflict or natural disaster) based on a novel network based feature set we named 'Campaign Network Attribute'.

0.2 BACKGROUND AND RELATED RESEARCH

0.2.1 Definition of Social Media

In the available literature, definitions of Social Media vary significantly. By far the most comprehensive study in the field to date is Carr and Hayes (2015).This research refutes the assertion that because there is an 'inherent understanding of Social Media, based on the extent of the technology' there is also a generic definition of Social Media. Carr and Hayes (2015) explores historical definitions and provides a framework that the authors anticipate to remain applicable through 2035. In this paper, the authors build a definition of Social Media and then go on to discuss each element, their definition a distillation of theories available at the time condensed into the following statement, *'Social media are Internet-based channels that allow* *users to opportunistically interact and selectively self-present, either in real-time or asynchronously, with both broad and narrow audiences who derive value from user-generated content'.* In a more recent study Dollarhide (2019), Dollar et al site Carr and Hayes (2015), however, state that Boyd and Ellison's older definition, *"web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system*Boyd and Ellison (2007) remains valid. This research focuses on investigating the context of a definition, specifically, how a lay definition will impact the use and perception of a Social Network Service (SNS). The premise being, 'the definition of Social Media varies depending on how it's used', ie. A definition for categorisation within a 'application store' will be different to that of a definition if being used to describe the brand essencesDollarhide (2019). Key to Dollarhide (2019) is the value in lay definitions and what this conveys. Hence, investigating how leadership thinks about or defines Social Media can provide insight into how their field views Social Media as a whole. When discussing Social Media in terms of definition, Use and Gratification (U&G) theory is also of value. The original research was published in 2010Quan-Haase and Young (2010). Their start point coming from 1993 McQuail and Windahl (2015), that mass media is anything *'having a uniform and immediate influence on individuals, whom they perceived as easily susceptible to influence and unable to form their own opinions.* The outcome of the study being, *When an audience member has a need for escape, there are specific media available to gratify this need in a satisfactory manner. Moreover, the concurrent use of various tools suggests that each fulfills a distinct need making an analysis of U&G essential.* Their research showing that not only are definitions individual but each experience on Social Media differs based on user need, the study finding differences in the gratifications obtained from each type of social media. U&G theory is also employed to analyse the definitions and use of the big four platforms in Alhabash and Ma (2017).

0.2.2 Definition of Social Media Influence

Social Influence already has an established psychological definition which is *'any change in an individual's thoughts, feelings, or behaviors caused by other people, who may be actually present or whose presence is imagined, expected, or only implied.'*American Psychological Association (1985). That said, in a Social Media context, researchers and authors often use the word influence to summarize a collection of actions that can either happen immediately or as a chain of events. In the coarse of this research, when influence is discussed in a social media context, there is both a 'method of influence' and a 'effect of influence'. Bond *et al.* (2012) defines influence in social media as the behavioural change or social cognition. I.e. That behaviour change comes from a willingness to adhere to social norms within a group. Bond *et al.* (2012) specifically investigates the concept of online social media being able to transmit influence (contagious) akin to face to face transmission. Other definition is 'one agent's ability to dominate another'Jackson (2010) however, that study was investigating social and economic networks using graph theory. Chung and Zeng (2020) specifically defines 'Influentials' in a social media context *'users who change the opinions or emotions of others in a large scale'.* Given the variations in definition and interpretation of what Social Media Influence is, the first research question is:

RQ1) Define Social Media Influence

0.2.3 Trusting Social Media

Research has explored the role of trust in influencing behaviour of a Social Network long before the advent of Social Media Eden (1988) and produced cognitive models to explain the connections. Trust defined in 1995 as *a willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party'*. Furthermore, there are five established types of trust in general terms which are, knowledge based trust, institution based trust, calculative based trust, cognition based trust, and personality based trust **?**Kim *et al.* (2008). Trust has also been well defined in relation to e-commerce and online shopping Gefen *et al.* (2003). More recently, in their comprehensive research from 2017, Cheng *et al.* (2017) Cheng et al not only developed a new model for trust within Social Media that introduced the three modes of communication (Interpersonal, Group and Mass), but also found three main conclusions with respects to the antecedents of trust in Social Media. Summarised, Cheng *et al.* (2017) discovered that time-saving and information quality was paramount for mass communications. Common topic and convenience were more important to group communications and finally, competence is salient in interpersonal communications. Trust in Social Media also played a significant role in the hesitancy of the COVID-19 vaccination. The term 'infodemic' was coined to describe the influx of *'complex and dynamic information – both factual and incorrect'*Jennings *et al.* (2021). Trust in Social Media has also been researched in it's growing threat to democracy as well as investigating methods to handle it from a leadership perspective Dubois *et al.* (2020). Therefore, research question two is focused on trust and influence in Social Media, or simply put:

RQ3) Can Social Media be trusted?

0.2.4 Role of Social Media in Future Conflict

The authors propose that the role of Social Media in future conflict can be extrapolated from further understanding two key sub elements: 1) How Social Media influences the physical world 2) How Social Media will evolve in the future. And then layering this with established theories of the role of Social Media in conflict.

Social Media's influence on the physical world has already been thoroughly investigated. The field of crisis management is heavily invested in understanding the linkages between Social Media and Physical events. Using Social Media sentiment and geographical data can help speed up response times, but also help identify the element most in need or danger, or which supportive efforts are of highest value Dou *et al.* (2020); Jamali *et al.* (2020). The ability to predict physical events based on Social Media activity prior to official reporting enhances crisis management's ability to deploy first responders faster and with greater impactBorden *et al.* (2020); Sakaki *et al.* (2010). From an epidemic perspective, identifying outbreaks in early stages can stop the spreed of the disease, which was extremely important in 2020 Li and Cardie (2013); Li *et al.* (2020); Qin *et al.* (2020); Depoux *et al.* (2020); Vyas *et al.* (2021). Whilst features in Social Media can provide insights to these applications, the fundamentals about 'why' they are connected or related are still poorly defined.

Finding accurate studies on how Social Media will evolve is difficult, however, this is to be expected. Given the corporate value of these ideas, the speed of change in Social Media, publication overhead and public uptake making it almost impossible to predict how Social Media will develop from established research. That said, there are researchers attempting to understand how Social Media will evolve their respective fields or establishments. For example, Sarwar *et al.* (2021) looked at how Social Media evolution in University Rankings correlated with world renowned university rankings systems such as Quacquarelli Symonds World University Ranking and Times Higher Education World University Ranking. Kent and Li (2020) explores the evolution of Social Media and how this new paradigm of world communications will impact Public Policy. Feldkamp (2021) reviews the rise of the Social Media Platfrom TikTok and the unique circumstances of COVID-19's impact on it's gain of market share. Interestingly, Feldkamp (2021) identified three pillars of the platforms success being, the hyper-personalised algorithm, an anti-social approach and 'influencers' Hutchinson (2020); Novak (2020); Schwär (2020). As such, given the hypothetical nature of predictions, the responses from our participants are purely speculative. Although, their responses do provides insights into how leadership is thinking about Social Media, identifying fears or hopes for the next generation and Social Media platforms.

The role that Social Media plays within the spectrum of conflict has gained significant interest over the last 10 years. Beginning with low level interpersonal conflict, through to state on state actions. For example, researchers have found positive association with inter-parental conflict and Problematic Social Media Use (PSMU) but also that PSMU mediated other developmental issues such as, self-esteem and maladaptive cognition Wang *et al.* (2021); Horwood and Anglim (2021); Boer *et al.* (2021). Oluchukwu (2021) demonstrated that intermatrial conflict was often resolved using Social Media, advocating Social Media's use to strengthen relationships. The interconnection of Social Media with state on state actions or 'war' has been established by authors and researchers alike. PJ Singer's book, 'Like War' Singer and Brooking (2018b), sights numerous examples from Arab Springs to the ISIS campaigns to Mexican cartels. Research have also been investigating state on state conflict, Johnson *et al.* (2020) exploring case studies such as the Crimean Russian invasion and

Pakistan India boarder clash. These examples confirm that Social Media plays a role within human behaviour and modern conflict, but will this continue in future? Understanding Social Media's role and and specifically, how to prepare for it, will define the difference between victory and defeat for the individual, business and nation. From a positive perspective, techno-optimistic researchers site the Arab Springs movement as the introduction of Social Media into democracy Eltantawy and Wiest (2011). However, comprehensive research from Zeitzoff (2017) in 2017 has shown the increasingly negative impact of Social Media in democracy and subsequent conflict. These insights compounded by the scholastic contribution to the book 'Social Media Impacts on Conflict and Democracy: The Techtonic Shift' Schirch (2021). This collaborative work of over 40 scholars, adapted an accounting ledger approach to weight the positives and negatives of Social Media in conflict and democracy, and unfortunately the result was strongly negative. Therefore, in order to provide insights into questions like: What is our leadership doing to prepare for Social Media's growing impact? Is leadership considering Social Media in their development plans? Does leadership agree with the current research and concur with the negative effects? Hence, the fourth and final research question is:

RQ3) What is Social Media's role in future conflict?

0.2.5 Semantic Modelling of Social Media Influence

The vast data associated with SNS makes isolating influence difficult, however, there is evidence of not only state sponsored influence, but individuals determined on shaping the nature of SNS and the populations that engage with SNS Zannettou *et al.* (2019). Extensive research has already surveyed and defined the hierarchical schema of SNS Razis *et al.* (2020) which have been used in various research studies, for example, text mining research to explore the trending or popular actions Karami *et al.* (2020). Studies have also defined areas for potential future research Abu-Salih *et al.* (2019), while others have applied semantic analysis of SNS to track and assess the influence of content shared across these platforms Bayrakdar *et al.* (2020). Whilst some of this research has touched on the physical realm associated with cyber influence, most of this prior research is limited to only the cyber realm.

The role of Social Media influence in physical conflict is highly important, and is commensurate to that of conflict propaganda Stengel (2019) first popularized in WWII by Joseph Goebbles Zannettou *et al.* (2019). Whilst Goebbles was limited by the media available at the time, his propaganda still had significant influence on the population and final outcomes of the conflict. Propaganda and the role of influence in conflict has evolved continuously from the 'hearts and minds' campaigns from the Vietnam War Miller (2019) to modern information warfare Schaner (2020). With social media fast becoming the sole source of information for individuals, it is very likely that social media will be the future front line of conflict. Shaping what a user is exposed to, consumes, and believes is as effective as any kinetic effect could be Stengel (2019). The race to build tools that absorb and analyse big data will result in a competitive advantage Rauta (2020).

The fundamentals of Social Media influence are mainly derived from the following four areas of related research. The first area of related research is computer platforms used to propagate influence. The second area is the analysis of social networking and their place within modern society. The third area is the theory of influence and how influence is modeled. The fourth area is the role of influence within conflict. These four related areas are discussed in the following four subsections.

0.2.6 Computer Propagated Influence

The founder of the concept of persuasive computers was Fogg Fogg (1998) who theorized that computers have the ability to persuade individuals and coined the term "Computer as Persuasive Technology (CAPT)" Oinas-Kukkonen and Harjumaa (2008); Sara and Mostafa (2018). Further work, termed Mass Interpersonal Persuasion (MIP)Fogg (2008), applied CAPT at scale. When understanding the outcomes and effects of influence, captology only considered compliance of the agent as the result. Xie et al. Xie *et al.* (2016) expanded on the outcomes of social influence by including **obedience** and **conformity**. These terms were drawn from the work of Cialdini in 2004 Cialdini and Goldstein (2004). In light of Cialdini and Goldstein (2004), Xie et al. recognized three effects of cyber influence: *compliance*, *conformity*, and *obedience*. Compliance refers to a particular type of response known as acquiescence, i.e., a request. The request may be explicit or implicit, but in all cases the target recognises being urged by the source to respond in a desired way Cialdini and Goldstein (2004). Conformity is the act of changing one's behaviour to match the responses of others Cialdini and Goldstein (2004); Deutsch and Gerard (1955); Thomson (2005).

0.2.7 Analysis of Social Media

The second area that relates to cyber influence is the analysis of online communities, social media, social networking, and their role within our modern society. Over 4.5 billion people are now estimated to be online Kemp (2019); Cisco Inc. (2020); Edosomwan *et al.* (2011). Facebook reports a monthly active user group of over 2.45 billion and over 1.62 billion daily active users Fac (2019). Combined with smart phones, an individual is easily identified within cyber meet spaces, which allows for categorizing individuals into age, gender, and ethnic groups for the unique and targeted delivery of content and marketing.

The power and implications of social media are well established and have become a popular research field, resulting in a spectrum of studies from the psychological, socio-economic, academic, and industrial fields Altuwairiqi *et al.* (2019); Shibuya and Tanaka (2019); Xu *et al.* (2020); Ning *et al.* (2017); Maistrelli (2019). When analyzing Twitter specifically, the size and scale of the tweet stream is problematic; hence, the development of techniques and methods to categorise tweets is important Belhadi *et al.* (2020); Harakawa *et al.* (2019); He *et al.* (2017); Resende de Mendonça *et al.* (2020); Rosyiq *et al.* (2019); Sharma and Jain (2019); Yu *et al.* (2019); Arora *et al.* (2019a). Many studies assist commercial endeavours for businesses interested in sentiment assessment or optimizing product exposure and promotion Arora *et al.* (2019a); Penas *et al.* (2013); Deparis *et al.* (2011). Subsequently, the ability to fabricate highly technical results by Social-media Data Providers (SDP) can impact corporation models of new and developing businesses. Hence, Zou *et al.* (2019); Sediyono *et al.* (2014) investigated methods of verifying correctness and completeness of their data and the results generated from social media analysis. Also, Lee *et al.* (2017) uses an ontological approach to build trust and transparency into social media data. These related studies show the depth of research already provided and we leverage the knowledge gained by them in building our cross-realm model.

0.2.8 Influence Models and Modeling

The third related area of research draws upon influence modeling. The theory of quantitative models of influence began with the "two step" flow model Lazarsfeld and Merton (1954); Weimann (1991). The "two step" model represented a person with an established reputation within the community (celebrity), who would then pass on their influence to their network. Watts and Dodds critically reviewed the "two step" model and quantified the effect of so-called *influentials* Watts and Dodds. (2007). This research negated the theory that specific individuals have significantly more influence over others within a network. Whilst these individuals do exist, they were considered only one factor, with the state of the network as a whole being a larger factor. This research was completed before the existence of modern SNS, but is applicable to any form of communication that can be represented as a network. The Watts and Dobbs model manipulated the thresholds to activate neighbouring cell to initiate information cascades. By changing the nature of the network, Watts and Dobbs Watts and Dodds. (2007) were able to demonstrate the ability to create information cascades regardless of who within the network was activated first.

The *Sick, Injured, Recovering* (SIR) model for pathogen modeling was theorized by Coleman in 1957 Coleman *et al.* (1957). The SIR model differs from the influence flow model Watts and Dodds. (2007), in that the SIR model has no memory. Rather, the SIR model treats each interaction as "pure", as opposed to observations over time. This relates to social media because, each interaction and information exchanged is typically accepted or considered pure. As such, consumers of information and content from these networks are in a highly vulnerable position. Recently, the AAS study Lazer *et al.* (2018) investigated the cat-and-mouse type game of detecting and countering detection of fake news on social media. The AAS study outlined the importance of intervention measures to protect the public, such as education and personal fact checking, in addition to platform structural changes to prevent exposure to such material. The study Lazer *et al.* (2018) specifically outlines that "There are no comprehensive data-collection system to provide a dynamic understanding of how pervasive systems of fake news provision are evolving".

As mentioned in Section 0.2.7, commercial applications are popular, which has resulted in the emergence of the field of influence maximization Chen *et al.* (2020). The study Chen *et al.* (2020) improves computational efficiency by using cloud computing and specifically designed algorithms. Other influence modeling examines security threats in Social Gaming Networks (SGN) Alturki *et al.* (2020), using influence modeling to identify why certain players are targeted by scams or cyber attacks. Influence modeling has also been employed to improve the detection of subtle and long-running radicalization of individuals Kursuncu *et al.* (2019). These examples show that there are many forms of influence modeling and methods of defining a campaign. For the purpose of our CIC model, we define influence as an outcome of an action, contributing to or resolving conflict.

0.2.9 Social Media Influence Campaigns In Conflict

The power of cyber influence is increasingly being recognized by governments and militaries around the world. In 2016, NATO published the study Brangetto and Veenendaal (2016) which focused on Influence Cyber Operations (ICO) as a subset of Influence Operations. The NATO study Brangetto and Veenendaal (2016) identifies operations that are conducted in the logical layer of cyberspace with ICOs targeting attitudes, behaviours, and decisions, and specifically, “hacking minds by shaping the environment in which political debate takes place”. A key prediction from Brangetto and Veenendaal (2016) is the increased employment of ICOs due to the promise of “victory through non-kinetic means to erode the adversary’s willpower, confuse him, constrain his decision making and undermine public support” with little to no attribution.

Singer and Brooking Singer and Brooking (2018a) draw a parallel between cyber influence and Clausewitz’s concept of war being ’an extension of policy by other means’. War is used to enforce one nations narrative or policy on another. Cyber influence does the same without the physical violence or destruction. Cyber influence can be achieved through communication directly or indirectly with the population itself, thus bypassing or reinforcing diplomatic channels. The US Army Cyber School recognized that adversaries are weaponizing social media to attack the American social and political environment Hart and Klink (2017). The study Hart and Klink (2017) highlights the malicious nature of "foreign governments employing a combination of state sponsored media and personas who support their positions on social media and disrupt free discourse" and America's requirement to advance their cyber and information operations to counter this threat. In order to do so, Hart and Klink (2017) proposed the development of the *1st Troll Battalion* to conduct both offensive and defensive "trolling" operations.

Assessing a state's cyber power by measuring cyber influence was investigated by Herrick (2016). The study Herrick (2016) found that the "logic relies on the assumption that the same skills that allow actors to be successful at social media operations also enable them to be successful at offensive and defensive cyber operations". The study concluded that cyber operations require orders of magnitude of greater skill and technology compared to cyber influence. The study Herrick (2016) also identifies that there is very little cross-over of skill between the two disciplines, cyber skill sets being technology driven, whilst cyber influence skill sets being psychology based. The framework Zeitzoff (2017) posed four research questions: 1) How do groups use social media to recruit and shape the ideology of potential followers? 2) How do elites and world leaders use social media? 3) How do technology advances influence the strategic interactions of actors in highly dynamic settings? 4) Does the reduced entry cost of communication increase partisan and ethic polarization, as well as erode the trust in mainstream media?

There remains significant research to be conducted that investigates the interplay between both the physical and cyber realms in cyber influence. In order to identify and quantify information operations or propaganda, there is a requirement to build organized and flexible data structures and datasets capable of representing influence flow across realms. There is also a pressing need to develop logical and flexible analytical tools that leverage these next-generation data structures to identify influentials, regardless of their genesis, as well as misinformation and automated activities.

0.2.10 Cyber Influence Campaign Examples

The previous section discussed research fields that contributed to the concept of CICs, the role of CICs within conflict, and the requirement to evolve cyber influence concepts within the technical, policy, and academic fields. This section explores real examples of CICs to demonstrate their scale, outcomes, and time frames. This section shows that CICs can transfer influence in both the physical and cyber realms with a spectrum of methods and techniques.

Scale

The concept of scale of a CIC is target dependent, which could be as small as a single agent, or as large as the entire network. Chicago Gangs often use low-level or individual-orientated CICs daily. Commonly refereed to as "Cyber Banging" Patton *et al.* (2017), these individual CICs are typically initiated by single agents to support gang violence Anderson (1999). The CIC techniques include tagging oneself in rival gang territory or posting inflammatory comments on rival gang members posts Decker and Pyrooz (2013); Storrod and Densley (2017); Papachristos *et al.* (2013). More complicated techniques involve gang members increasing status by promoting ones persona to their digital audience Storrod and Densley (2017).

At the other end of scale, there is state-on-state conflict. The use of international

CICs is now becoming commonplace Singer and Brooking (2018a). A recent example was the February 2019 India vs. Pakistan conflict which was started by a terrorist attack against an Indian convoy Maheshwari (2019). After an Indian retaliatory strike on a terrorist camp Philip (2020), a small CIC quickly became a large CIC that leveraged popular celebrities to increases it's impact IAF (2019). The #IndiaStrikesBack and #BalakotAirstrike networks were prominent and quickly became politicized and led the narrative of the conflict as well as the upcoming government elections Kaura (2020).

Outcomes

The desired outcomes or influence effects of a CIC will determine the target or targets, the techniques to be used, and the required scale. Outcomes achievable with a CIC are also on a spectrum from personal or local, all the way to political and international. Small-scale CICs typically target individuals with personal or commercial outcomes. Larger CICs can have far greater and longer lasting outcomes. For example, the Al Hayat Media Center is an Islamic State of Iraq and Syria (ISIS) media branch Durr (2016). Al Hayat are funded to operate CICs for various outcomes, such as recruiting to ensure the survival of the group Anderson (2016), re-branding the group as a legitimate government alternative Durr (2016), influencing potential candidates, and inciting violence using coordinated tactics Alkaff and Mahzam (2018); Awan (2017); Perešin (2015); Speckhard (2015). Al Hayat were also the first group to use a large-scale CIC concurrently with an application called 'Dawn of Glad Tidings' Awan (2017). Whilst eventually shut down, 'The Dawn of Glad Tidings' served to reinforce ideals and opinions, creating what is commonly known as an "Echo Chamber", restricting nuance and only allowing strict ideological messaging Barberá *et al.* (2015).

There are also acute examples of large-scale CICs that can be hijacked or repurposed for other outcomes. In late March 2014, Russian forces were lawfully invited into the Crimean Peninsula to help settle a social unrest Tracker (2014). A legitimate request on the surface, however, it was the result of a long-running large-scale CIC. It began with a domestic unrest due the Ukraine President ceasing discussion on a EU trade agreement. Domestic protests were initiated by individuals using the #euromaidan network. The hashtag gained popularity as a single point of coordination and voice of the people. At the same time, Russia saw this as an opportunity to reclaim Crimea. Russia used what is now known as the Dulles Doctrine MacFarquhar (2018) to dominate narratives within social media. Russia defines operating within social media as an evolutionary Information Warfare, "a permanently operating front through the entire territory of an enemy state", which can asymmetrically lower an adversary's combat potential Gerasimov (2013). Russia used state level resources to push pro-Russian messaging on the #euromaidan network and influence support for Russian intervention in Crimea.

Time Frames

CIC time frames are closely linked to the used techniques and the desired outcomes. Intuitively, there is a linear relationship between the scale of a CIC versus the time frame and investment. Whilst individual small-scale CICs can be launched almost instantly from a single agent account, large-scale influence requires CICs to closely coordinate a critical mass of accounts. In the #euromaidan example, the hijacking was possible because the accounts posting to the network looked legitimate. Russia's Internet Research Agency (IRA) Strudwicke and Grant (2020) or "Troll Factory" Mejias and Vokuev (2017) accounts look legitimate because hundreds of bloggers were paid to build false identities, push pro-Russian messaging, praise Putin, and denounce opposition in forums, social networking, and comments boards; thus, achieving a coordinated effect Garmazhapova (2013); Gregory (2014); Shane (2017).

For purposes of demonstrating our model and ontology we selected two of the reviewed CIC examples for case studies: The #euromaidan campaign is a particularly interesting case due to the corruption of the network as well as the CIC evolving into a state-on-state conflict. The second case study is India vs. Pakistan and the Balakotstrike. This CIC will be valuable due to the highly correlated physical events. The #AlleyesonISIS campaign would also be very interesting given the significant influence and intimidation achieved, however, much of the graphic content posted has been scrubbed from Twitter and thus makes detailed analysis infeasible.

0.2.11 Semantic Modeling, RDF/RDFS, and Ontologies

Semantic Modeling

Ontologies have been well researched with some modern examples found in El Asikri *et al.* (2016); Bonacchi and Krzyzanska (2019); Kuang and Du (2010); Han *et al.* (2012). At its core, an ontology is a graph, using graph theory to collate and organise information. Essentially, an ontology is a number of definitions, relationships, and inference rules. For example, heterogeneous data provided by various devices and sources can all be integrated and applied with commonality and uniformity Bischof *et al.* (2014). An advantage of semantic modeling is the ability to link established ontologies. This means that terminologies of objects with their inherent properties for common concepts have to be defined only once and remain the same in various ontologies. This reduces replication and maintains consistency once an ontology is stable, but also means they can be leveraged by other models. For example, Halawi *et al.* (2018) generates separate ontologies within an evaluation model to categorise tweets and detect spam.

Cyberthreat researchers employ semantic modeling to categorise large and unstructured datasets collected from cyber attacks. This approach allows to "provide a flexible framework for representing and structuring the large variety of data with which security analysts are confronted", the framework can then be used for implementing cyber security analytic tools Bromander *et al.* (2016). One of the key benefits of semantic modeling is that a single query will result in all the information about a particular instance or object, thus improving search and time efficiencies within large datasets.

Semantic modeling and information structures are applicable beyond computer science, cyber security, and engineering. Ontologies are able to logically and conceptually map information, making them versatile and valuable to numerous research fields Breuker *et al.* (2004); Hoekstra *et al.* (2007); Campbell (2020); Brodaric *et al.* (2019). Masolo, Benevides, and Porello Masolo *et al.* (2018) proposed a formal framework to examine the relationship between (scientific) models and empirical observations. The study Masolo *et al.* (2018) uses an ontological approach to address the problem of observational conclusions and the potential for inconsistencies that underline the knowledge gained from the observations.

RDF/RDFS

The Resource Descriptive Framework (RDF) is a standard for data interchange on the web W3C (2017). The Web Ontology Language (OWL) is built using RDF Schema (RDFS) which extend link data via Unique Resource Identifiers (URI) resulting in only one instance of data being allowed to exist. RDF/RDFS allows for linking even if the underlying data schemata are different.

Ontologies

This section outlines a selection of related existing ontologies. The advantage of exploring these existing ontologies is that they can either be leveraged by our ontology or tailored to support our requirements. Generally, as we explore these ontologies, it is important to remember the question, "why should we use an ontology?" The simple answer is because an ontology is well suited to artificial intelligence (AI) applications. As we will discuss in both this section and in Section 0.4.4, the ontology is one method to enable AI to bring meaning to an environment. An ontology achieves this by building causal links of "related" data to enable the discovery of new information. As shown in Razis *et al.* (2020), an ontology can "identify the order of relationship among the entities" which can then be processed by an AI algorithm. This same principle is employed in Bindu and Thomas (2021); Angele *et al.* (2020). Hence, the development of an ontology is a building block of AI research. The focus of this study is on developing and demonstrating a functioning practical ontology for CIC modeling. Future research on AI algorithms can then build on the ontology developed in this study to discover and infer new information and make better, more accurate decisions about CICs based on the developed CIC ontology model.

- Good Ontologies A good ontology as defined by the World Wide Web Consortium (W3C), means that it is well documented, differentiable, used by independent data providers, and possibly supported by existing tools w3c (2017). Many of the ontologies used in research and academia, or published in the public domain, use good ontologies as a baseline. A well-known ontology is Friend of A Friend (FOAF) foa (2011), which represents relational networks of online social media and was one of the first ontologies to highlight the potential of semantic modeling. In foa (2011), an individual person can be linked to others using the `foaf:knows` relationship as well as online artifacts, such as documents and URLs, building an understanding of social media. The Socially Interconnected Online Communities (SIOC) ontology sio (2011) is commonly used to represent communities. The Dublin Core (DC) ontology Weibel and Koch (2000) is a lightweight generalist ontology used to describe metadata. Many of the following ontologies extend these good ontologies for a specific purpose or requirement.

- Consent Ontology In response to the introduction of personal data laws in Turkey in 2016, researchers at Ege University developed an extension of the FOAF ontology to track the consent of a person to process personal medical data. The semantic solution Olca and Can (2018) allowed Turkey to comply with international laws but also to manage this data. The extended ontology Olca and Can (2018) imported the FOAF ontology, leveraging the `Person` class. Secondly, due to the legal age requirements, FOAF was further extended with additional classes, such as '`foaf:HasMinAge` which is Boolean and either above or below 18 years old. To allow for consent to be granted by a parent or legal guardian, additional classes from the Relationship ontology are imported, namely `foaf:MotherOf foaf:FatherOf foaf:RepresentativeOf`. This consent ontology now tracks if consent is provided (another Boolean class of :permission or :prohibition) and who provided that for legal history. The extended ontology Olca and Can (2018) demonstrates the ability to import established ontologies and extend them for other purposes.

- FOAF Academic Kalemi and Martiri Kalemi and Martiri (2011) developed an extension to the existing FOAF vocabulary to include professional achievements and bring people closer to others with similar interests, topics, and research.

Kalemi and Martiri Kalemi and Martiri (2011) focused on the academic community, extending it FOAF to cover academic specific terms and relationships. An example of this is the 'afoaf:university' class. A main class of the ontology, narrowing down the academic community to a geographic location. Foaf academic also defines axioms which allow for richer information, but also assurance of the information. For example, Rule 1: If person A and person B are at University C, they are colleagues. Rule 2: if person X and person Y work at a department D, then they are in the same department and Rule 1 is inferred. This ontology shows the power of axioms and ability to enrich information in meaningful ways.

- OSN extension to FOAF El Kassiri and Belouadha El Kassiri and Belouadha (2017) extended FOAF to address the evolution of Online Social Networks (OSNs) through a Unified Semantic Model (USM). The USM leverages three good ontologies, FOAF, Semantically Interlinked Online Communities (SIOC), and Simple Knowledge Organization System (SKOS). USM extends FOAF by using membership, association, and organization to imply ideals and potential persuasions. The study El Kassiri and Belouadha (2017) demonstrates that unique extensions (including classes not traditionally associated with FOAF) can provide specific deep insights.

- SNS analysis Nie et al. Nie *et al.* (2020) used text based analysis to identify bursty hot events within twitter. They clustered key words using a domain ontology. Leveraging the graph structure of an ontology, enabled measurements of the distance between words through the graph. Hence, key words could be used in different contexts, but their syntactic and semantic definitions remained the same. Fang et al. Fang *et al.* (2019) proposed a unified ontological model for cross-media events, allowing combinations of SNS platform data for SNS analysis. Dhiman and Toshniwal Dhiman and Toshniwal (2020) used an ontological model for specific event detection. Focusing again on text analysis and then the generation of graphs around related textual content to form relationships and linkages Agarwal and Toshniwal (2019)

Whilst important research and applicable to SNS analysis, specifically to the automated detection of malicious events, these studies differ significantly from our work. Both our model and interests are positioned at a higher level of abstraction. Our ontology takes into account both physical and cyber events, enabling our ontology to determine the causal influence between physical and cyber events, as well as to quantify the influence of a physical versus cyber action.

- InfluenceTracker Ontology To the best of our knowledge, the closest study to ours is the InfluenceTracker Ontology developed in 2014 by Razis and Anagnostopoulos Razis and Anagnostopoulos (2014) to specifically represent the influence of twitter accounts on each other. A very specific instance of cyber influence, the study limits influence strictly to the cyber domain and only uses the twitter platform. Important for future studies are the metrics developed to quantify influence of one account over another. For example, the Followers to Following ratio (FtF ratio) and the Tweets Creation Rate (TCR) provides quantifiable and measurable metrics to determine if one account influences the network more than another. The InfluenceTracker leverages the FOAF ontology for representing an agent and uses a similar hierarchy of classes.

0.2.12 Social Media Influence Campaigns

The genesis of Social Media Influence Campaigns originates from data science and event detection in Social Media. This involved the discovery of events within high volume, largely homogeneous, data Wang *et al.* (2021). As popularity and integration of Social Media grew, so did the concept of influence transfer using Social Media Castillo *et al.* (2011), initial studies relating back to fundamental concepts from Fogg and Tseng Fogg and Tseng (1999) regarding computer credibility and persuasive technology. Since then, it has been accepted that influence can be transferred via digital means akin to that of physical interactionsVer Steeg and Galstyan (2012). As such the concept of a Social Media campaign has been popularised for almost a decadeKetter and Avraham (2012); Van Noort *et al.* (2012). The ability to influence or sway a network of users on themes such as political issues or commercial pursuits has increased in interest partially due to the highly public exposure of Facebook and Cambridge AnalyticaWylie (2020).

Investigation into how to detect and track Social Media campaigns has significantly evolved in the last decade. Initial methods focusing on free-text based features, discovering coordinated campaigns and commonality within large scale data setsLee *et al.* (2011, 2014). The integration of graph theory has allowed researchers to explore critical problems with Social Media such as organizational problems, data structure and connectivityChakraborty *et al.* (2018); Pitas (2016). Recently, attention has now turned to detecting and tracking bot and autonomous influence with significant successHeidari and Jones (2020); Cresci *et al.* (2019). The 2016 US presidential election highlighted both autonomous activity as well as the potential of misinformation in Social Media, or 'Fake News' Carlson (2020). The influence and detection of misinformation campaigns has also been thoroughly investigatedWu *et al.* (2019, 2016), addressing the profile rise that these systems can have on users or issues. The use of semantic analysis has also opened new possibilities from detection to prediction and new data discoverySaif *et al.* (2017); Johnson *et al.* (2020). Finally, time series analysis has been used specifically to track changes within campaigns Nusratullah *et al.* (2015); Kafeza *et al.* (2017). As a result, there is a well established base of understanding Social Media Influence Campaigns.

The culmination of much of this research can be seen in the commercial deployment of the 'BotSlayer' systemHui *et al.* (2020). A web-based system that not only allows for real time tracking of campaigns, but provides qualitative insights into the nature and influence of SMICs. BotSlayer provides the entire ecosystem from platform scraping for data collection, through to the quantitative metrics such as 'BotSlayer' or 'BS' level. This metric indicating the level of suspicion associated with a hashtag or campaign. Moreover, BotSlayer also provides integrated tools to conduct further analysis using graph based visualisation. An impressive tool, but not without limitations. Botslayer along with the majority of research into SMICs assumes the homogeneous nature of a Social Media Platform. I.e, the potential for a campaign is the same regardless of the network. Given recent studies into the nature of Social Media networks **??**, the concept of Networks of Networks opens up new concepts for research.

0.2.13 Social Media Network of Networks

Social media is a continuous loop of action-network-reaction. More specifically, the loop begins with the initiation of an **action**, i.e., an original post or social media content. The content is applied to a **network** with followers or subscribers. The followers then **react** via more posts or metadata in relation to the action. Hence, a continuous loop of action-network-reaction is traversed. This continuous loop results in a highly chaotic environment, where the networks are constantly being realigned and restructured **?**. Thus, the agents (users)—both organic and artificial—are influenced via exposure to content and sources that are constantly changing **?**. The related study Johnson *et al.* (2020) identified the classes and predicates required for semantic modelling in order to allow for relationship search and knowledge generation from social media data.

0.2.14 Networks of Networks

The state-of-the-art of the social media analysis **?** considers the SMP itself as homogeneous. As a result of this assumption, when evaluating the influence of a social media campaign, the *potential for influence* is consistent across the entire platform Karlsen and Enjolras (2016). However, that is *not* consistent with how social media is implemented. Rather, interest groups, algorithms, and user preferences create networks within the wider SMP network. Hence, the impact of these networks of networks should be considered when analysing the *potential for influence* of an SMIC

The majority of quantitative SMIC research considers the action (social media content or post) to be the independent variable, responsible for the influence that is transferred, with all other variables being dependent variables **?**. In contrast to **??**, we adopt the concept that SMPs are *not homogeneous*, i.e., that SMPs are communities within communities similar to Alduaiji *et al.* (2018), i.e. networks of networks. We posit that the networks of networks concept more closely reflects the concept of groups and communities within real-life social networks and that these networks impact the potential influence of an action.

0.2.15 Machine Learning

Recent developments regarding the application of Machine Learning (ML) techniques to social media analysis employ a wide spectrum of techniques, including unsupervised learning, supervised learning, and reinforcement learning. Large and detailed social media data sets lend themselves well to ML methods. Generally, the feature extraction in existing studies varies based on the goals. Specifically, studies focused on content analysis commonly extract content features Alizadeh *et al.* (2020), such as text, images, or links within a social media action. Due to the high dimensionality of social media data, unsupervised learning methods, such as clustering, face the curse of dimensionality Köppen (2000), where Euclidean distances approach zero. Whilst co-clustering **?** and Sequential Cluster Estimation **?** were investigated for multidimensional big data, they are still unsuitable for the high number of dimensions in social media network data. Filter Bank Common Spatial Pattern methods work well with electroencephalography (EEG) signals **?**, whilst Natural Language Processing (NLP) techniques build features out of the structure of the language used to discover sentiment, sarcasm, and intent. These are combined to detect physical events or mine for opinions Shi *et al.* (2020). Other studies have focused on metadata regarding the content, such as time and location Redondo *et al.* (2020). The detection of communities, bot and autonomous activities, as well as misinformation in social media uses a combination of feature extraction and definition of new features to support specific research and discovery Heidari *et al.* (2021); Heidari and Jones (2020).

0.2.16 Current ML Techniques and Social Media

A comprehensive review of current established ML techniques and their suitability for the different applications of social media analysis is provided in Balaji *et al.*

(2021). Following on from established ML techniques, research is being conducted into improving the performance of the ML techniques and the algorithms themselves. For example, Partial Least Squares-Structural Equation Modeling (PLS-SEM) Salloum *et al.* (2021) enables researchers to "estimate complex models with many constructs, indicator variables and structural paths without imposing distributional assumptions on the data" Hair *et al.* (2019). PLS-SEM was demonstrated to be highly effective in a study of social science preferences and decision making, for example, predicting student adoption of social media websites based on perceived influence, playfulness, and ease of use of the interface Salloum *et al.* (2021).

The study Mullah and Zainon (2021) reviews advanced ML techniques and investigates the use of ensemble techniques (combining of methods) to achieve greater performance for classification problems. Also, Mullah and Zainon (2021) reviews specific deep learning techniques, such as Long Short Term Memory-Convolutional Neural Networks (LSTM-CNN), suggesting that for the specific purpose of hate speech detection, the BiLSTM-CNN generates the highest F1 score. The study **?** uses efficient Deep CNN for video analysis. Extreme Deep Learning Trees have shown to outperform other neural network based models **?** using MNIST, LeadSnap, and ORL face recognition data sets. Natural Language Processing is important in social media analysis as well, due to the majority of features being content or text based. Research has shown the ability to predict electricity demand based on text mining of social media platforms **?**. Recently developed by Google in 2019, Bidirectional Encoder Representations from Transformers (BERT) is a "pre-trained language model for context embedding and attracted attention due to its deep analyzing capability, valuable linguistic knowledge in the intermediate layer, trained with larger corpus, and fine-tuned for any NLP task" Deepa *et al.* (2021). Essentially, BERT is a pre-trained model for NLP which can be adapted to specific ML tasks or problems.

The problem of sarcasm in social media and NLP is of growing interest, as much of the language is contextual and difficult for traditional ML methods. For ML models that are highly complex, this raises the black box problem, where the details of the model are uninterruptible by humans. The eXplainable Artificial Intelligence (XAI) framework has been developed to alleviate this problem Gunning *et al.* (2019). For NLP, the XAI framework offers two main advantages. The first advantage is transferability as the model has been trained in a controlled environment and has truly discovered the underlying patterns. The second advantage is the ability to aid in determination of the contributing factors, akin to importance coefficients for linear modelling. LIME is an XAI method implementable with both CNNs and Multi Layered Perception (machine vision) modes, allowing for interpretation of the model Ribeiro *et al.* (2016).

The SHapley Additive exPlanations (SHAP) process is a unified process, leveraging the advantages of six different processes including LIME, DeepLIFT, Layer Wise Relevance Propagation, and Classic Shapley Value Estimation Lundberg and Lee (2017). SHAP's output is a model of the model, known as the explanation model for interpreting ML modelsLundberg and Lee (2017). The main benefit of SHAP is that it can be applied to Deep Learning models and ensemble methods. NLP processes the data at the content (Tweet) level, which does not suit the requirements of our research aims. As our approach is different, the features and attributes that are analyzed by our ML are the aggregation of network features of tweets. This allows us to aggregate the importance of network features of a campaign. We note that the implementation of LIME and SHAP using our CNA-TCC framework is an interesting direction for future research. The present study is focused on designing the CNA-TCC framework and establishing a baseline performance based on elementary tuning of the hashtag network size, the employed linear regression model, and the hyperparameters of the neural network model, so as to provide a basis for such future LIME and SHAP implementations.

0.2.17 Social Media in the Future? — Need for Research

Detecting SMICs is typically a challenging task since social media data sets are large and important feature information is often missing due to scraping or collection methods Assenmacher *et al.* (2020). A key aspect of the detection of campaigns is the use of identifiers or key features that relate to the topic of interest. Typically, this acts as a starting point or heuristic and can be anything, such as an account, hashtag, or geo-location. However, there are a number of use cases where these are unavailable. For example, political campaigns can be born of any issue and are not restricted to a particular country or location. The issue itself may not have political origins initially, but after specific events, it may become very important to a political party, government, or security force. Therefore, the search space, when looking for campaigns, is often extremely broad and unmanageable.

Therefore, this research aimed to develop a classification method agnostic to content identifiers. This research has developed a framework to classify campaigns of similar themes using the network behaviors and attributes. For example, historic campaigns with damaging or physical impacts are available, e.g., the #StormTheCapitol campaign which was associated with the U.S. Capitol Hill riots of 2021. This campaign was subsequently blocked by the platforms and hence, cannot be used for future detection. However, by extracting the network behaviors from historic data, the network attributes associated with the #StormTheCapitol campaign can be encoded. These encodings can then be searched for in generic social media data to find campaigns with similar profiles.

0.2.18 Social Media in the Future? — Need for Research

Detecting SMICs is typically a challenging task since social media data sets are large and important feature information is often missing due to scraping or collection methods Assenmacher *et al.* (2020). A key aspect of detection campaigns is the use identifiers or key features that relate to the topic of interest. Typically, this acts as a start point or heuristic and can be anything, such as an account, hashtag, or geo-location. However, there are a number of use cases where these are unavailable. For example, political campaigns can be born of any issue and not restricted to a particular country or location. The issue itself may not have political origins initially, but after specific events, it may become very important to a political party, government, or security force. Therefore, the search space, when looking for campaigns, is often extremely broad and unmanageable.

Therefore, this research aimed to develop a detection method agnostic to content identifiers. This research has developed both a framework and method to detect campaigns of similar themes using the network behavior and attributes. For example, historic campaigns with damaging or physical impacts are available, e.g., the #StormTheCapitol campaign which was associated with the Capitol Hill riots of 2021. This campaign was subsequently blocked by the platforms and hence, cannot be used for future detection. However, by extracting the network behavior from historic data, the network attributes associated with the #StormTheCapitol campaign can be encoded. These encodings can then be searched for in homogeneous social media data to find campaigns with similar profiles.

0.3 QUANTITATIVE SURVEY OF LEADERS IN SOCIAL MEDIA AND CONFLICT

This section is the first point of novelty for this research being *Expert Prediction*. Considering that SNS operators and content creators are constantly evolving to better capture the attention of their audiences Johnson *et al.* (2022a). The nature of SMICs will also evolve with the medium. By investigating what experts believe the future holds for SMICs means the model will be able to account for this evolution and continue to identify SMICs in the future. The have delivered a robust and extensive survey of expert analysis of the future of SMICs and use of SNS. This research leveraged the authors network of high ranking military and government experts as well as established academics and researchers on the subject. This survey used construct validity and established codification methods to produce the following research.

0.3.1 Methods

Participants

Prior to commencing this research, ethical approval was gained though Arizona State University's Institutional Review Board. Each participant was informed in writing of both recruitment and consent to ensure they were aware of their rights and the use of the gathered information. No participant younger than the age of consent was engaged; moreover, during each of the interviews the participant was reminded that participation was voluntary, and all responses were anonymous, unclassified, and non-attributed. Importantly, each participant's responses were solely their own and in no way represented their government's nor their employer's point of view.

Participants for this research were selected based on several key aspects. We applied two collective criteria to all candidates. The first collective criterion was either holding or had recently held a leadership position within their respective fields and careers. This could be defined as either managing a team or being in a position of decision making authority. The second collective criterion for candidate selection was the provision of a unique perspective. Each candidate must provide either a unique experience, education, or perspective. This avoided repetition and ensured representation of leadership from a wide cross-section of the community.

Candidates for the study were categorised into four domains: Military, government, private industry, and academia. To ensure suitable candidates with relevant and applicable experience were selected to participate, each domain had additional specific criteria. The specific criteria for military participants included the following: Must be a serving member of the military forces of Australia, the United Kingdom, or the United States of America. To ensure current perspectives, individual service must be within a Cyber Command, Information Warfare Division, any operational command, or capability development branch. Finally, each candidate must hold the rank of O5 or above, which commonly corresponds to the rank of Lieutenant Colonel or Commander or above. This ensured they had leadership exposure and at least 20 years of military experience. The specific criteria for government candidates included: Currently employed by the government of Australia, the United Kingdom, or the United States of America. Government candidates were selected based on job description. These must be one of the following: Policy advisor, technical advisor, national security representative, or officer of the state department. Each government candidate must have a job descriptions that specifically mentioned working with cyber security, conflict, or Social Media. The specific selection criterion for candidates from private industry included: Being currently employed in a technical role, such as cyber security, telecommunications, or Social Media. Due to the vast variation of jobs in the private sector, each civilian job description was individually reviewed to make sure they could adequately contribute to the study. The civilian candidates had varied backgrounds including marketing, political campaigning, news agencies, technology companies, and Social Media influencing. Finally, the specific selection criteria for academic candidates was that they must either be a Ph.D. candidate or hold a lecturing position at a university and be working in either a social science field or a STEM field. An invitation to join the study was sent to 40 candidates in total, with the final number of actual participants that could be captured within the study time frame being 33 ($N = 33$).

The characteristics of the participants are summarized in Table 13. For this study, a standard definition of Generation X (born 1960-1979), Millennial (born 1980-1996), and Generational Z (born 1996-onwards) was used Dimock (????); the average participant age was 34.5 years old. One member of the study represented both academia and private industry, as he was a Ph.D. student but also working for a commercial engineering laboratory.

Table 1: Characteristics of $N = 33$ participants

Field	Gen X	Millennial	Gen Z
Age Group	10	16	7
Government Defense	6	2	1
Government Civilian	2	3	0
Civilian / Private Industry	1	10	3
Academia	1	2	3
Masters or Higher	10	10	0
Undergraduate Degree	1	4	5
Highschool	0	1	2
Male	9	11	7
Female	1	5	0

As a result of this diverse and expansive collection of leadership, we have captured a wide cross-section of leadership within the global community.

Interview design

The interview was designed to allow the participants freedom in answering, whilst ensuring a consistent interpretation. The process of construct validity Aiken (2009); Brown (2000) was used to build a coherent logical set of interview questions, which are provided in the Appendix. Four independent experts in the Social Sciences from the academic communities of Australia and the United States of America reviewed and refined the question list until it was suitable for pilot testing. The questions were tested using pilot interviews, ensuring that the interview responses addressed the research questions.

Analytical approach

The coding methodology used during the analysis and data extraction plays a significant role in the eventual outcomes and result generated by research Glesne (2016). Taking Saldana (2013) as an authoritative source of codifying guidance, the first cycle coding was Invivo coding, which generated codes based on key words used by the participants. In line with Saldana (2013), these key words are then grouped into categories and then compounded further into themes that form insights into the data. During this process, the researcher also has latitude to use mixed methods, depending on the targeted outcome Saldana (2013). "Structural" coding was used for the second cycle codification. Structural coding allows the researcher to align coding and qualitative data with a set of predetermined lines of inquiry, such as the questions of an interview.

0.3.2 Results

This section provides the results of the interviews. The tabulated data is the raw quantitative breakdown of codes, generated using the methods described in Section 0.3.1. Specifically, each table shows the number of similar responses based on the Invivo coding methodology. Each number n mentioned in the remainder of the paper reflects the number of participants that referred to a specific code, category, or concept. These numbers n are calculated by the Nvivo software and are non-repetitive to avoid misrepresentation. This is important to note, as the total number of participants for a particular category of codes cannot be calculated by simply summing the individual numbers for the individual codes (coding references) within the considered category. For ease of interpretation we associate an actual number of participants, e.g., ($n = 21$), with a percentage that gives the actual number n relative to the number $N = 33$ of interviewed participants, e.g., $21/33 = 64\%$.

In particular, the tabulated data is presented as follows. The first digit in the results column gives the number of individual participants that responded with that code. The second digit is the total number of references to a specific code across all the interviews and questions. This is important since participants can give multiple examples which are coded individually and can come from other questions. Finally, the Total shows the aggregated non-repetitive number of participants that provided a response. The Total metric is automatically generated by the Nvivo software.

Definition of Social Media

The interviews began with questions regarding definitions, not only did this allow the participants to start to think critically about their ideas of what Social Media was, but it also grounded their responses for later questions. After the first and second

Table 2: Codes for 'Definition of Social Media' for $N = 33$ participants.

Category	Code	Result
Mass Media	Online systems and tools	6/7
	Information exchanges	4/4
	Platform specific	2/2
	Total	12/15
Online Presence	Sharing between people	17/19
	Communication	12/12
	Community of people	9/9
	Digital extension	4/4
	Networks of people	2/3
	Total	29/54
Commercial Application	Total	2/2
Influence	Engineered engagement	1/2
	Influence accepted	3/4
	Total	3/6
Social Application	Non professional	3/3
	Anti-social behavior	3/4
	Total	5/7

codifying cycle, the results in Table 16 were obtained.

The results in Table 16 highlight five categories for participants outlining their thoughts on Social Media definitions. The dominant theme being that Social Media is perceived to be more than just digital mass media, as the most frequent response was the concept that Social Media is the projection of one's self in a digital manner or one's digital online presence. By accumulating the codes regarding online presence in Table 16 the total shows this as the dominant proportion of responses with 87% of participants ($n = 29$), remembering that this number is the aggregated non-repetitive total. This percentage suggests that Social Media is perceived to be fundamentally different from traditional media. The themes suggest that Social Media is different by providing mechanisms for getting one's own presences out into the world as a digital projection of one's personality. A participant described it as *'being present without being present'*, which is richer and indicates a new form of media distinct from traditional mass media. Discovering that a large proportion of leadership shared this opinion was unexpected.

The idea of Social Media being different from regular Mass Media is not novel, nor is the idea that Social Media can provide a replacement for community functions. However, this study has discovered that leaders perceive that Social Media can replace these functions, such as a sense of belonging, emotional support, and shared interests in the community. The authors interpret this 'sense of community' being generated by Social Media as potentially derived from communication and engagement with people of common beliefs and ideas. For example, personal validation, education, and guidance can be achieved by online communities as effectively as by physical communities. These community effects demonstrate the importance of the network to the user, which provides further evidence as to why individuals are so acutely influenced through Social Media.

Influence of Social Media

Table 3: Totals of 'Is Social Media more influential than other media?' for $N = 33$ participants.

Category	Code	Total
More Influential	Yes	32/32
	No	1/1

The psychological definition of social influence is a *'non-physical application to the mentality of the individual exposed to it'*. By seeking a definition of Social Media influence, the responses provide insights into how leadership is approaching Social Media influence. The first result in Table 15 indicates that almost unanimously, Social Media is considered to be more influential than other forms of media. First and second cycle codification of participant responses from further questions generated three categories: *effects*, *vectors*, and *factors*, see Table 14. These categories reflect the various interpretations of Social Media influence. If a participant described the "impact" that could be achieved by Social Media influence, this was categorised as "Effects". Participant responses that specifically discussed "how" influence could be achieved were categorised as "Vectors". Responses that talked about what "factors" determine the scale of Social Media influence were categorised as "Factors". The breakdown and codes of each category are shown in Table 14.

As per Table 15, 97% ($n = 32$) of the participants supported the idea that Social Media is more influential than any other forms of mass media, digital or not. This result is justified by the number of novel and unique services provided by Social Media that participants identified (see Table 19), such as interactions, tailored content, speed of communication, exposure times, and a sense of belonging. This result reflects the importance and general understanding of the power of Social Media influence in the community. Only one participant, or 3% ($n = 1$), suggested that Social Media was

Table 4: Codes of Influence categories for $N = 33$ participants.

Category	Code	Result
Influence Vectors	Targeted advertising	7/7
	Communication	2/2
	Sharing ideas	1/1
	Sharing of lives	1/1
	Issue raising	1/1
	Group thematic	1/1
	Total	13/13
Influence Factors	Size of followership	3/4
	Lead narrative	2/3
	Credibility	2/2
	Agenda	2/2
	Authority	1/1
	Platform theme	1/1
	Trust	1/1
	Total	9/15
Influence Effects	Changing a point of view	11/12
	Determining behavior	5/5
	Shaping network	1/1
	Total	15/21

Table 5: Codes for 'Why is Social Media influential?' for $N = 33$ interview participants.

Category	Code	Result
Greater Influence		
	Interaction	12/13
	Tailored	9/11
	Speed of communications	6/7
	Monetized / incentive	4/5
	Exposure time	3/4
	Leverages brain reward	2/2
	Integration and reach	2/2
	No fact checking	2/2
	Sense of belonging	2/2
	Sole information source	2/2
	Human need to connect	1/2
	Relateable	1/1
	Mass effects	1/1
	More access	1/1
	Entertainment	1/1

less influential for older generations, as they are not as integrated with Social Media and mobile technology as deeply as the younger generations. It is the opinion of the authors that this individual is accurate with their logic, suggesting that there is still a segment of the community not reachable or influenced via Social Media.

As shown in Table 14, 47% (n = 15) of the participants responded consistent with the psychological definitions, which have codified influence as an "effect"; be it changing minds, determining future behavior, or shaping public opinion. Vectors of influence was also a frequent response, being discussed by 39% (n = 13) of the participants. Whilst slightly less frequently than the other two, but with still by a large proportion of responses, factors were mentioned by over 28% (n = 9) of the participants.

Trust of Social Media

Table 6: Totals of 'Is Social Media trustworthy?' for $N = 33$ participants.

Category	Code	Total
Trustworthy	No	17/33
	Yes	12/33
	Depends	14/33
	Aggregated	5/12

Table 17 contains the aggregated responses to the line of inquiry about trusting Social Media. Table 18 shows the results of Tables 15 and 17 broken down by generation, which will be further discussed in Section 0.3.4.

With regards to the trustworthy nature of Social Media, Table 17 shows that 53% ($n = 17$) of the participants felt that Social Media can not be trusted. 36% ($n = 12$) felt that it could be trusted. Out of the 36% of participants who trust Social Media,

Table 7: Totals of 'Is Social Media more influential than other media?' and 'Is Social Media trustworthy?' by generation for $N = 33$ participants.

Category	Code	Gen X	Mill	Gen Z
More Influential	Yes	10	16	6
	No	0	0	1
Trust	No	5	7	5
	Yes	4	6	2
	Depends	3	9	2

42% ($n = 5/12$) of these individuals only trusted aggregated data from Social Media. Out of the overall $N = 33$ participants, 39% ($n = 14$) believe that trust depended on the situation and that all factors pertaining to an issue needed to be considered.

Of the 5 participants in Table 17 that believe Social Media is trustworthy through the use of data aggregation, 2 of these participants cited chaos theory as their rationale Williams (1997). The idea being that a deterministic systems may appear random when elements are viewed in isolation. However, when viewed on an appropriate scale, the systems have underlying patterns and order that can be perceived. This results in behavior that is not random at all Williams (1997). The point being, whilst individuals might appear to act randomly or independently, when scaled en masse, behaviour and responses form patterns and adopt a normal distribution which is deterministic and not random. This also supports an idea raised by two participants, suggesting that most modern big businesses use and analyse big data everyday. If this data were untrustworthy, then the outcome would be the businesses going bankrupt, and the reality is that they are not He *et al.* (2017).

Some participants directly referred to trusting Social Media because it was unfiltered. Many other comments from participants regarding Social Media's trustworthiness noted that content, posts, and opinions on Social Media are unfiltered and evocative in nature. From Table 8, 9% ($n = 3$) of participants implied that Social Media is a raw, potentially highly authentic reflection of what the individuals actually think about issues. Also, from Table 8, 30% ($n = 10$) of the participants stated that some Social Media platforms are perceived to be more trustworthy than others. Key elements for platform trust being monitoring and regulated content.

The application of the information or decision being made was also a dependant factor: 18% ($n = 6$, Table 8) of the participants suggested that the information from Social Media is fine for low impact applications or decisions, (e.g., selecting a restaurant) but not for high impact applications (e.g., selecting a doctor or lawyer). This indicates the 'risk versus reward' quotient playing a significant role in whether Social Media can be trusted.

0.3.3 Social Media and future conflict

As mentioned in Section **??**, Social Media's role in future conflict is the combination of the evolution of social media and how it will impact the physical world.

Evolution of Social Media

The results provided in Table 9 show that 64% ($n = 21$) of the participants believed that further regulation of Social Media is inevitable in the future. The forms of regulation varied significantly, but included government control, internal curation, account verification, and censorship. The platforms' incentives to change come from public pressure to be more accountable and to focus on trust and truth. A small proportion of these 63%, namely $n = 2$ responses went so far as to suggest that support for regulation would be accompanied by a more robust legal framework, similar to that of TV broadcasting and news paper publication. Over 48% ($n = 16$) of the participants believed that Social Media would remain in our lives, becoming further

Table 8: Codes for 'Why is Social Media trustworthy?' for $N = 33$ interview participants.

Category	Code	Result
No Trust	Lacks editor	5/5
	Incentive to misrepresent	4/4
	Skepticism	3/3
	Opinion based	2/3
	Fake news	2/2
	Data manipulation	2/2
	Nothing is 100% trustworthy	2/2
	Misinformation	1/1
	Peak distrust	1/1
Trust	Data aggregation	5/5
	Unfiltered	3/5
	Meta data only	2/2
	Research support	2/2
	Emotional responses	1/1
	No reason to lie	1/1
Depends	Source	10/10
	Application (low vs. high impact dec.)	6/6
	Platform	1/1
	Decision	1/1

Table 9: Codes for 'How will Social Media Evolve?' for $N = 33$ interview participants.

Category	Code	Result
Evolution of Social Media		
	Regulation	21/42
	More integrated	16/25
	More interactivity	7/12
	National security issues	7/10
	Free speech concerns	7/9
	Further polarization	6/6
	Less delineation	2/4
	Develop legal framework	2/3
	More AI presence	2/2
	I'm scared	1/1

integrated in the future and usage of Social Media by the population approaching 100%.

Additional analyses for which we could not include tables due to space constraints indicated that participants also believed that online behavior will mature in the future. They mentioned concepts, such as people sharing less sensitive personal information, establishment of online etiquette, group orientation of information, less antisocial content, and perhaps most interestingly, the advent of circuit breaker technology. Circuit breaker technology imposes a time delay between successive content posts combined with sentiment analysis in order to restrict repeat posts on a subject in rapid succession. Hence, circuit breaker technology may allow time for individuals to reflect and understand the implications of their posts.

In general, 18% ($n = 6$, Table 9) of the participants thought that platforms will become even further polarized and that there will be more Artificial Intelligence presence ($n = 2$) in years to come. Also, 21% ($n = 7$) projected that the nature of Social Media will become a matter of national security, lending to support for more government intervention. Finally, 21% ($n = 7$) of the participants raised concerns about free speech, e.g., who determines what free speech is. The concern about free speech poses the challenge of finding a balance between the interdiction of violence and hate speech versus public freedom of expression.

Social Media and the physical world

Asking leaders about Social Media's interrelationship with the physical world yielded three main categories: Psychological factors, Technological factors, and Abstract concepts, as shown in Table 10.

The combined aspects, as shown in Table 10 indicate that more than 78% ($n = 26$) of participants connected the digital domain and physical domain via psychological and technological aspects. The remainder of the responses were more abstract concepts, such as inter-domain effects, diminishing separation, asymmetric effects, and information overload. Examples of inter-domain operational effects include Arab Springs, ISIS, Myanmar, and Crimea Singer and Brooking (2018b). Diminishing separation, suggests that on an individual level, there is no longer a separation of what happens in one's physical life and one's digital life; rather, physical life and digital life for the individual impact each other immediately and directly.

Social Media in future conflict

From Table 11, over 57% ($n = 19$) of the participants were of the opinion that Social Media will continue to shape and influence conflict affected populations: 42% ($n = 14$) believed it will be used as a tool of war similar to propaganda, as well as targeting

Table 10: Codes for ’Why are the digital and physical realms connected?’ for $N = 33$ interview participants. Recall from Section 0.3.2 that second number is the total number of references to a specific code from the entire interview process; numbers above 33 are possible if references are mentioned multiple times by participants.

Category	Code	Result
Technological Aspects	Total	17/30
	Speed of communications	10/12
	Ease of organization	7/10
	Ubiquitous nature	3/3
	Mobile technology	2/2
	Inter connectivity	2/2
	Faster news cycles	2/2
	Amplification	2/2
	Citizen reporters	1/1
Psychological Aspects	Total	21/48
	Mass population effects	9/11
	Sense of belonging	8/10
	Emotional responses	7/8
	Encouragement	4/4
	Echo camber	4/4
	Bystander effect	2/3
	Satisfying as physical world	1/1
	Boredom	1/1
	Challenges values	1/1
Techn. + Psych. Aspects	Total	26/78
Abstract Aspects	Total	11/17
	Inter-domain effect	8/9
	Diminishing separation	5/6
	Asymmetric effects	2/3
	Information overload	2/2

Table 11: Codes for 'What is Social Media's role in future conflict?' for $N = 33$ interview participants

Category	Code	Result
Role of Social Media	Shaping the affected population	19/22
	Weapon or Tool	14/16
	Passage of information	10/13
	Support DIME effects	7/7
	Employed during Phase Zero	5/6
	Should not be in conflict	5/5
	Amplify all effects in conflict	5/5
	Unregulated action space	1/2
	Game changer	1/1
	Low cost of entry	1/1
	Agnostic platforms	1/1
	Platforms foster controversy	1/1
	It will be bias	1/1
	Total	29/81

leadership or insurgents, akin to intelligence lead operations in counterinsurgency warfare Kilcullen (2006). 30% ($n = 10$) of the participants thought that Social Media will be agnostic in it's role, simply as a passage of information for both sides of conflict. However, 15% ($n = 5$) of the participants believed that Social Media will amplify any and all effects in a conflict. Support to Diplomatic, Information, Military, and Economy (DIME) effects on governments was suggested by 21% ($n = 7$) of the participants. 15% ($n = 5$) of the participants believed that Social Media should not be in conflicts at all. Moreover, if Social Media does have a role, it should only be as part of conflict resolution. Finally, 15% ($n = 5$) of participants believed that Social Media's role in conflict was prior to kinetic action, or the use of physical military force. This suggests that Social Media's role is below the threshold of physical conflict. These responses cited Clausewitz's concept of undermining the adversary's will to fight in order to win the conflict.

0.3.4 Discussion

This section discusses the results and provides interpretations where appropriate. In brief, from a theoretical perspective, the present study challenges the existing definition of Social Media as merely a new form of digital mass media; rather, our findings point to a pronounced emphasis on the personal nature of Social Media. Our findings on Social Media influence, trust, and role in conflict are consistent with the existing literature.

Definition of Social Media

The existing theoretical definitions of Social Media specifically focus on the user generated nature of Social Media Carr and Hayes (2015), simply users publishing content. Our study of leadership advances the existing theoretical definition by indicating that the vast majority of participants associate Social Media with a new method of human interaction and projection of personality. The authors offer two pieces of evidence for the advance in theoretical definition of Social Media.

1) Based on the results shown in Table 16 that 87% ($n = 29$) of leadership define Social Media as an online presence, which means that the majority of leadership are adopting and embodying this concept. The importance of this finding is that current leadership are actually adopting Social Media and will find new and creative ways to make Social Media part of their leadership strategy. This progressive position signals a future of dynamic and potentially more intuitive and connected leadership. To date,

there are examples of this leadership strategy both succeeding and failing. We theorise that this combination of success and failure is because leadership is exposed to the power of misinformation and manipulation of Social Media. A leadership position that embraces Social Media must do so using the pillars of Social Media success, which are being "consistent and genuine" Kane (2018).

2) Also based on Table 16, which shows that 45% ($n = 15$) of leaders maintain the importance of traditional mass media. This is interpreted as these leaders adopting a more risk-adverse position with regards to their perceptions of Social Media. The main theory being that these leaders are potentially insulating themselves from the negative aspects of Social Media, which can be interpreted as self-consciousness or a potential weakness. However, this does not necessarily mean missing out on the benefits of Social Media, if leveraged correctly. Further analysis of the interview data, which we cannot include in the tables due to space constraints revealed additional insights: Government employees constitute the majority (66%, $n = 10$) of the 15 participants that perceive Social Media as simply online mass media. We interpret this result to indicate that, in general, government leadership is more conservative compared to the rest of the population and less transparent in their public relations. The implication of conservative government leadership is a dislocation or disconnection with the younger generation who have embodied the Social Media technology as a ubiquitous personal method of communication.

Influence of Social Media

40% ($n = 13$, Table 14) of the participants defined Social Media influence using context consistent with the psychological or agreed definition of Social Media influence. This suggests that approximately one in two leaders understands the psychological impacts of Social Media influence. Of these 40% of the participants, there was a 60%

($n = 8$) / 40% ($n = 5$) split between government and private entity leaders. We interpret this to indicate that more government leadership appreciates Social Media's ability to change the mind of a consumer, be it the consumer of a conceptual or physical product. On the other hand, this means that over half of the leaders do not fully understand the influence of Social Media.

We have identified Influence Factors (see Table 14) that characterize how the participants perceived influence. These influence factors can be further examined in future research in order to provide a quantitative assessment of influence effects. The quantification of influence factors can be useful when selecting and constructing features for Artificial Intelligence (AI) analysis or generally, when conducting analysis of Social Media influence. By designing AI based databases via ontological representations (relationship linkages), these influence factors may enable novel insights into the behaviors of Social Media and the influence campaigns that can be conducted within Social Media.

Trust of Social Media

In response to the question on the trustworthiness of Social Media, a conclusive majority of the participants expressed a sentiment of distrust. The "No Trust" categories in Table 8 provide unique insights as to why Social Media can or cannot be trusted. The main reasons for the lack of trust are derived from the lack of editorial or fact checking as well as the existence of fake news and misinformation on Social Media, contributing to the cautious approach of one participant suggesting *'Nothing is ever 100% trustworthy'*. Interestingly, several codes in the No Trust category in Table 8 relate to the antecedents of trust of Social Media that have recently been identified in Khan *et al.* (2021). For instance, the codes lacks editor ($n = 5$), incentive to misrepresent ($n = 4$), and fake news ($n = 2$) in Table 8 relate to the information quality that has been identified as an antecedent of trust in Khan *et al.* (2021); specifically, these codes in Table 8 indicate that absence, i.e., lack, of information quality is a reason for *not* trusting Social Media. Similarly, the codes data manipulation ($n = 2$) and nothing is 100% trustworthy ($n = 2$) in Table 8 indicate a lack of the antecedent perceived security Khan *et al.* (2021).

From Table 19, 13% ($n = 4$) of participants believe that there are substantial incentives to exaggerate or manipulate content due to the monetization strategies of the Social Media platforms. We believe that these monetization incentives are one of the key reasons why leadership in general does not trust Social Media. The same idea can also be extended to creators, suggesting that monetary incentives to produce content may lead to reduced proportions of authentic content on Social Media, especially if authentic content is more costly in terms of time and resources to produce than fake content.

Only 3% ($n = 1$, Table 8) of leadership believes Social Media encourages emotional responses which contradicts the perception of raw and unfiltered responses. The interview responses seem to suggest that leadership has the perception that everything posted on Social Media has an agenda and does not provide authentic insights into the emotional states of the posters

A frequent response to trustworthiness, at over 42% ($n = 14$, Table 18), was, "it depends". This is interpreted as leadership needing to conduct further research into the actual content, e.g., the author, artifact, report, or comment, before making judgement. Interestingly, the trust antecedents from Section 0.3.2, see in particular the terms (codes) in the Depends category in the bottom section of Table 8, use similar codes as the influence factors from Section 0.3.2, see in particular the Influence Factors in the middle section of Table 14. Based on the semantic meanings of these codes, we postulate the following mappings between the antecedents of trust in the bottom section of Table 8 and the influence factors in the middle section of Table 14: Source =¿ Credibility, Authority, and Size of followership; Application =¿ Lead narrative; Platform =¿ Platform theme; and Decision =¿ Agenda. Future research should quantitatively examine these mappings. Nevertheless, the relationships between the semantic meanings of these respective trust and influence codes appear to indicate that leadership believe there are mappings between trust and influence as they relate to Social Media.

In addition to this outcome, it was noted that 30% ($n = 10$) of the "it depends" responses mentioned the requirement to consider the source, i.e., to consider the layers of information about a source and the information provided by the source, in order to verify the information. We interpret this finding to highlight a common misconception. The confusion regarding the definition of trust and accuracy of information. In these ten responses, the participants are actually describing the accuracy of the information, not the trust in the source, e.g., the individual, that provided the information. Given this response, there is likely little differentiation in the interviewed leadership's minds between someone that believes what they are posting versus it actually being true. One can have all the antecedents of trust Khan *et al.* (2021), and still get the information wrong. This appears to indicate that leadership value accuracy over trust and have a desire to verify the source of information to marginalise the requirement of trust. Regardless, antecedents of trust do not undermine the correlation between trust and influence, i.e., if one trusts the information it will have the same influence regardless of the accuracy of the information.

Social Media and future conflict

More than 63% ($n = 21$, Table 9) of the participants believe that the evolution of Social Media will also include regulation. At the same time Social Media that are more integrated into our lives and provide more interactivity may provide opportunities for more extensive manipulations of the discourse in a community and ultimately the shaping of the ideology of a community. Possibly, a new "arms race" will emerge between efforts to regulate Social Media and efforts to make Social Media a yet more powerful manipulator.

As per Table 10, 78% ($n = 26$) of the participants agreed that there are connections between Social Media actions and the physical world outcomes. The leadership responses regarding Social Media and future conflict in this study relate to and reaffirm the New Communication Technology and Conflict theory by Zeitzoff Zeitzoff (2017). Specifically, the ease of organization noted by $n = 7$ participants and mass population effects noted by $n = 9$ participants (see Table 10) and the explicit low cost of entry response ($n = 1$, Table 11) relate to the element of reduced cost of communication in Zeitzoff (2017). Speed of communication ($n = 10$) was the most common response code in Table 10, reaffirming the increase in speed of communication postulated by Zeitzoff Zeitzoff (2017). The ubiquitous nature of information ($n = 3$) and mobile technology ($n = 2$) response codes in Table 10 as well as the popular ($n = 10$) passage of information code in Table 11 relate to the technological integration and evolution as well as the abundance of information availability that are part of the Zeitzoff theory.

Similarly, the popular ($n = 19$) response code for shaping the affected population and the employed during Phase Zero code ($n = 5$) in Table 11 (whereby Phase Zero refers to the shaping the conflict phase **??**) relate to the wider set of conflict support activities that can be supported by Social Media according to the Proliferating Hybrids theory by Gray and Gordo Gray and Gordo (2014). Thus, the analysis of our leadership interviews revealed response codes in Tables 10 and 11 that reaffirm recent theoretical expositions of the relationships between Social Media and armed conflict.

As mentioned in Section **??**, the connections between Social Media and conflict that have been documented in this study can serve as a basis for future research that seeks to quantify and apply these connections to learning algorithms so as to extract novel information and insights from Social Media data. As to the evolution of Social Media, participants responded that the idea that change is inevitable and potentially welcomed with the advent of more technology to protect the consumer. However, this is only true if the positive initiatives outweigh the potential for negative initiatives. Unfortunately, considering the over-incentivized ecosystem of Social Media, slow adoption of regulation, and the younger consumer age, this appears unlikely.

Generally, the participant responses suggest that the future of Social Media will provide more opportunities for influence and manipulation of populations in conflict. As such, the majority of the participants supported the concept that the role of Social Media in conflict will continue to rise. Greater influence can be achieved via Social Media through aspects such as reach, speed, legal ambiguities, and impact than possible with traditional media. Also, a very important point being the time and method of its use, which can range from recruitment through to targeting operations. This means that the use of Social Media in conflict starts out and remains critical all the way from under the threshold of conflict to the resolution and rebuilding phase. Essentially, Social Media was considered a critical enabler and force multiplier in all phases of conflict. Moreover, many participants lamented the lack of legal framework to act in support of national security in this space, citing the the concept *"who adopts technology first will have an advantage"* Singer and Brooking (2018b). This suggests that the lack of a legal framework in modern countries limits government operations in Social Media. Essentially, the lack of a legal framework has a paralysing effect and allows adversaries to build and gain experience in the meantime.

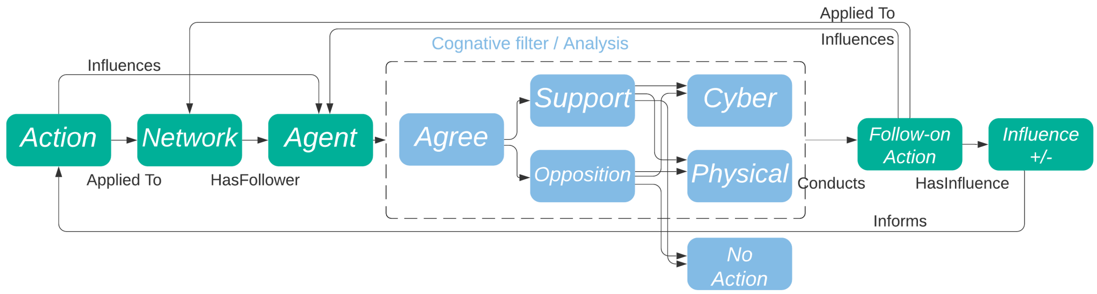

Figure 1: CIC Model with Flow of Influence

0.4 SOCIAL MEDIA INFLUENCE CAMPAIGN MODEL

This section is the second point of novelty for this research. The basis is formed from an article that was published in IEEE access on the 15th of January 2021. Based on the results and discussion from the research above, a significant consideration of the problem with analysing SMICs is the consideration of the network. When looking into SMICs, they are taken as homoegenously applied to a social media platform and each account has the same exposure. This is not the case and the concept of network impacting influence needed to be further explored. Secondly, the term Social Media Influence Campaign (SMIC) has been adopted since this article's publication, but it's definition is the same as Cyber Influence Campaign (CIC). Figure 1 depicts the cyclical CIC model, showing the flow of influence from action, to network to agent, through the cognitive filter and back to action in a cycle. The green boxes are classes, the black links are predicates, and the blue boxes are states of the cognitive process. This model is then integrated into an ontological representation using the Terse Triple Language (TTL) and abbreviated to `cicmod`. Whilst researchers have been able to observe that nefarious injection of content can steer the climate and discourse of an issue Zannettou *et al.* (2019); Ferrara *et al.* (2016); Kumar *et al.* (2017), their outcomes are based on data analysis and pattern recognition. Our model and ontology formalises the underlying relationships, identifying the foundational causal

Table 12: Novel Classes Defined for CIC Model as Well as Predicates Defined to Connect the Classes.

Classes: `cicmod:Action`, `cicmod:Network`, `cicmod:Realm`, `cicmod:Analysis`, `cicmod:Influence`
Predicates: `cicmod:Initiates`, `cicmod:Informs`, `cicmod:AppliedTo`, `cicmod:Influences`, `cicmod:HasFollower`, `cicmod:IsFollowing`, `cicmod:HasAgent`, `cicmod:BelongsTo`, `cicmod:Conducts`, `cicmod:ConductedBy`, `cicmod:HasInfluence`

links between the physical and cyber realms in terms of influence flow. By creating this framework, we can understand how an influence campaign starts with a cyber action, flows through agents and networks, and results in real-world physical actions. The model observes actions taken by the agents, applied to networks of agents who then take further action. The cycle stops only when all agents within a network take no action. Section 0.4 explains the design decisions behind the classes and level of abstraction. We have developed the novel classes and predicates to connect the classes in Table 12. The agent's cognitive filter is represented by the `cicmod:Analysis` class, a process that captures disposition, motivation, and realm in boolean sub classes. The `cicmod:AppliedTo` and `cicmod:Influences` predicates are shown linking back to the next network and agent, respectively, in the cycle. Finally, the `cicmod:Influence` class is the result of observing any action taken by an agent. The `cicmod:Influence` class also provides feedback to the CIC via the `cicmod:Informs` predicate in either positive or negative states.

Our ontology allows for abstract concepts, such as Influence and Follows, to be represented simply with predicates. Whereas, these abstract relationships between objects would be very difficult to capture with statistical models or conventional database structures. Moreover, our ontology allows for objects to be related across domains. A predicate can relate a physical event, such as a protest, to an object in another domain, such as an online message or CIC. As a result, through our CIC ontological model, we can translate abstract cross-domain concepts into a format that can be machine interpreted. Thus, allowing for the application of logic rules to discover information within large and potentially unrelated data.

Key Model Components

The following sections describe each class in detail.

- Action The action class is designed for any type of action that could have an influence effect. Any online publication or communication, e.g., post, tweet, vlog, blog, or opinion, is considered a cyber action or an action taken in the cyber realm. By exclusion, all other activities not taken in cyber space are deemed to be physical actions taken in the physical realm, e.g., talking, voting, and violence. Actions are applied to both the physical and cyber networks. This means that an agent applies an action to a network, and another agent consumes the action by being connected to the network. In the physical realm this would be attending a lecture, presentation, or address. In the cyber realm, this is subscribing, following, liking, or searching for any point of reference of the agent's action. Whilst the agent conducts an action, the influence is the result of the action. Hence, the action connects influence to agents.

  The `cicmod:HasInfluence` predicate represents the resultant influence of the action in either the positive or negative state. Many of the platforms have their own metrics for this already, such as upvote, downvote, like, dislike, thumbs up, and favourite. These metrics align with one of the three established influence categories, see Sec. 0.2.6. The predicate `cicmod:Influences` is the action's effect on the agent and assumes the same metrics as `cicmod:HasInfluence` (like, dislike, and voting). This design allows both simple and complex actions to be represented, from traditional support campaigns to false flag campaigns.

- Network The network class represents the first degree contacts of an agent or thing. An agent or thing can have multiple networks, in both the cyber and physical realms. For example, a thing may have multiple Twitter, Instagram, and YouTube accounts, (#musicfestival #lollapalooza) representing multiple cyber networks. In the physical realm, an agent may have multiple networks, such as, friends, colleagues, and family. In either realm, the networks can be accessed by other agents within the network. For example, when #election2016 was hijacked and used against a running candidate. Or, a town hall meeting can be used to voice the opinion of anyone that attends. As such, the metrics for networks must be native to the network itself. Data properties, such as **Followers** and **Following**, belong with the network class. The following-to-followers ratio is an established influence metric Razis and Anagnostopoulos (2014), however, our model determines scale by considering the `HasFollowers` metadata.

- Agent The ontology employs the `foaf:Agent` foa (2011) with its established definitions. Our cicmod ontology establishes additional predicates and linkages, but the definition of the agent class remains the same.

Realm

Two possible representations of realm were considered. The first being an individual object class realm. For an object class to stand alone, it must be contextual and defined to make sense. That is, an action, is contextual, a cyber, is not. Therefore, the second representation of realm was employed, this being a sub class of an object class. That is, an action within the cyber realm being a cyber action. Sub classes for each object class were developed, such as an agent cyber network, `cicmod:AgentCyberNetwork`, and an initial physical action, `cicmod:InitialPhysicalAction`.

- Analysis An analysis class was originally developed to represent a cognitive filter function as shown in the center box of Figure 1. The cognitive filter is an individual's analytical process. This analytical process is highly complex and beyond the scope of this study. We abstract the process into three yes or no questions: 1) Does the agent **agree** with the message or content of the action? 2) Is the agent **motivated** to take further action? 3) If motivated, in which **realm** will the agent take action? The output of the analysis determines the follow-on action, and influence can be determined by observing the action which is explained further next.

- Influence Quantifying influence is highly complicated (and specific to every use case) for a number of reasons: 1) An agent's disposition with respect to a CIC cannot be assumed, i.e., does the agent already support or oppose the theme of the CIC? 2) How did an action achieve influence? However, the influence can be relatively easily determined by observing the follow-on action of an agent. Our model quantifies influence by observing the state of a follow-on action, i.e., retweets indicate support, while downvotes, dislikes, and thumbs down indicate opposition. No prior information is required for this assessment of influence as positive or negative; rather, this assessment can be based on native metadata.

0.4.1 Model Flow

The cicmod ontology is designed to be cyclical, see Figure 1: Actions are applied to networks of agents that in turn take more actions. These iterate forever until all agents in a network do not take any further action. The following is an example of one cycle of the model:

- *:Initiate* A group decides to begin a Cyber Influence Campaign and engages an agent.
- *:Action* The engaged agent posts 'Elect John for President' to their Twitter account.
- *:Network* The agent's followers on Twitter are delivered the post.
- *:Agent* An individual is part of the engaged agent's network and consumes the post.
- *Analysis* The individual makes a decision whether or not to act.
- *:Action* The individual re-posts the initial post with a 'thumbs up'.
- *:Influence* Positive influence is inferred due to the 'thumbs up' associated with the repost.
- *:Informs* The positive action taken informs the group that the campaign is working as desired.

In this example, we can observe influence flowing from agent to network to agent to action. Therefore, we can observe the model representing the action, network, agent, and influence. The influence is captured by observing the nature of the action taken. That is, if the follow-on action is supportive, then the influence was positive.

0.4.2 Refinement of the Model

We tested our cicmod ontology with small datasets from TrackMyHashtag tmh (2019), which are discussed further in Section 0.4.3. To ensure that our model continued to reflect reality, we made the following refinements:

- 1. The analysis class was not required. The decision process does not change the outcome, nor does it provide any additional insight into the influence assessment of the action. Hence, the cognitive filter (`cicmod:Analysis`) was removed.

- 2. Cross platform indicators were removed. There is no additional value in knowing which platform the action is taken on, as for this study's purposes, all action have the same potential influence.

- 3. Initially, influence was assessed at the agent. As mentioned, only an action has influence, hence, the influence must connect the action to the agent, not the agent to the agent.

- 4. `cicmod:Following` and `cicmod:Followers` were changed to data properties of the network. As this allows for the networks to have scale and for an agent to have multiple networks in different realms.

0.4.3 Data Pipeline

In order to test the ontology using real-world data, we needed a real-world dataset from a campaign. This was a complicated process, as there are a number of steps required to take raw data from a social media platform and turn it into triples (i.e., semantic objects) for a functioning ontology. This was achieved through the following steps:

Campaign selection

We identified that the campaign needed to have two key elements. First, the campaign needed a strong physical timeline of actions and events that were easily distinguishable and consistent in reporting. Second, the cyber activity had to be of significant scale, i.e., above the noise floor. We began this process by considering a number of well-known CICs. We discovered that a suitable campaign should be bipartisan, as this reduced complexity. Also, the involvement of a military resulted in reporting being somewhat consistent and readily available. Thus, we selected two campaigns, namely the euromaidan protests during the Crimea crisis of 2013/2014 and the Balakot Airstrike during the Indian-Pakistan hostilities in 2019.

Data Collection

In order to achieve a complete understanding of social influence propagation through the network we require full-take or "fire hose" twitter datasets. Without loss of generality, we focused on the twitter platform as it is simplistic and consistent, also obtained example data showed that twitter metadata contained network detail, hashtags, and user generated content. Text logs from any of the other SNS platform, e.g., Instagram, WeChat, and Facebook, would be equally suitable for our cicmod ontology model. To test the ontology, small trial datasets were used. These are manageable percentages of the full-take twitter stream and were easily obtainable. Trial twitter data came from two different sources. 1) TrackMyHashtag tmh (2019) which only provided 100 tweet samples, and 2) Spritzer style twitter logs from the Internet Archive ia (????). These were suitable for testing and also helped confirm the nature of each criterion, i.e., hashtag, date range, agents etc. From testing with the ia (????) logs we confirmed the suitability of the two campaigns, then employed a third-party website TweetBinder TB (2019) to access the developer.twitter API API (2019) and provide the key hashtags over the date ranges. The number of tweets were confirmed by cross-referencing the quotes provided by twitter academic support staff and TB (2019). #euromaidan resulted in over a million tweets and more than 3 million tweets with first-tier related hashtags. Similar numbers were achieved for Balakotstrike.

Physical Events

The following timeline details the physical events from the #euromaidan campaign which we have been translated into triples for our ontology. We built the timeline using numerous conventional media sources on the conflict Tracker (2014); Britannica (2014); Agency (2014). However, when building this timeline, a decision to define an event as either an initial physical action or a physical reaction had to be made. Unfortunately, the definition of initial physical action versus physical reaction can be individually interpreted and potentially introduce inconsistencies. Therefore, to ensure consistency, we interpreted only the first physical event as an initial physical action; all subsequent physical events are interpreted as physical reactions. Therefore, the timeline for the #euromaidan campaign is as follows and a similar timeline was built for the Balakotstrike campaign.

- November 21, 2013: Cessation of EU agreement discussions by President Viktor Yanukovych
- November 21–23, 2013: Small demonstrations in Kiev in response to failed EU association agreement.
- November 30, 2013: Ukraine special police, Berkut, beat unarmed peaceful

protesters.

- December 01, 2013: Ukraine anti-government protesters have smashed their way into Kiev's city hall.
- December 13, 2013: Parliament passes restrictive anti-protest laws as clashes turn deadly.
- December 16, 2013: Protesters begin storming regional government offices in western Ukraine.
- December 28, 2013: Prime Minister Mykola Azarov resigns.
- February 14, 2014: 234 protesters arrested since December are released.
- February 18, 2014: Clashes erupt, with reasons unclear: 18 dead.
- February 21, 2014: Crimean parliament members called for an extraordinary meeting.
- February 22, 2014: Vote to remove President Yanukovych and Putin holds meeting to regain the Crimean peninsula.
- February 24, 2014: Parliament votes to ban Russian as the second official language.
- February 26, 2014: Large scale clashes during opposing rallies in Simferopol.
- February 27, 2014: Undeclared Russian Troops enter Crimean Parliament and Russia commences military training exercise in vicinity of Crimea peninsula.
- March 1, 2014: Aksyonov declared head of police, immediately requests support from Russia to maintain order.

- March 16, 2014: Public vote held to align with Russia.
- March 17, 2014: The EU and US impose travel bans and asset freezes on several officials from Russia and Ukraine over the Crimea referendum.
- March 18, 2014: President Putin signs a bill to absorb Crimea into the Russian Federation.

Data Ingest

In order to translate the data from the raw collection into triples, a unique data translation script was designed, written, and tested. The following points detail key design elements.

- 1. Inspecting the raw data. By inspecting the data before progressing, we ensured that the fields and meta data contained enough detail to populate the ontology. Moreover, the date range of actions covered the course of our specific campaign. We decided to extend the date range by 10 percent before and after the expected campaign dates to ensure that we caught the initial and final actions. API (2019) uses Java Script Orientation Notation (JSON) format to output the raw data. This was advantageous as the JSON dictionary format allowed for simple inspection and the JSON tool set within Python allowed for easy manipulation and processing.
- 2. The code was designed to loop through the JSON file building agents, Cyber Agent Networks, actions, and hashtag networks as triples from each tweet which was contained within a JSON dictionary. Moreover, the code used "mention" and "retweet" information to build additional networks and agents as they were referenced.

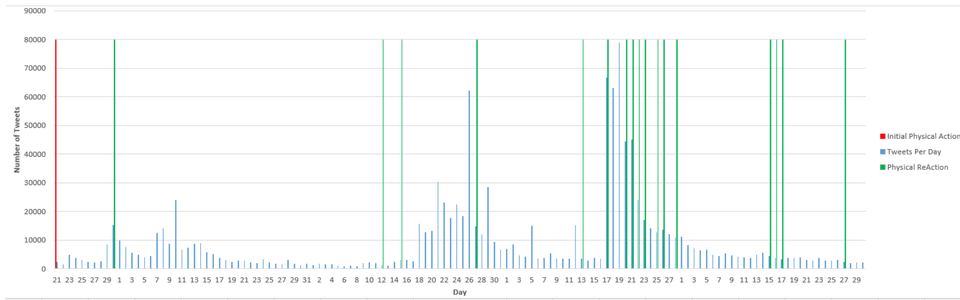

Figure 2: Total Number of Cyber Actions (tweets) Per Day of #euromaidan Campaign.

- 4. Privacy issues. Whilst the publication of tweets is public, the agent's user name and alias are not important to our research. Hence, the script anonymized agent names and aliases.

- 5. Additional information. Our real dataset also included artifacts of the actions and networks, such as favourite, location, and language, which were not present in the ia (????) JSONs. These are all highly valuable search criteria for the ontology and needed to be included. Whilst not in our initial test data, including these elements made our ontology richer and more valuable.

Triple Store

Once the ontology and dataset triples had been built, the file contained in excess of 20 million triples. Therefore, careful consideration of a triple store was required, as many stores cannot handle datasets of this size. We initially had used Protege pro (2019) and WebVOWL web (2015) to build and view the ontology; however, these were not capable of handling the large dataset. We selected the application Stardog which has been stable and user friendly and included a GUI, the Stardog Studio.

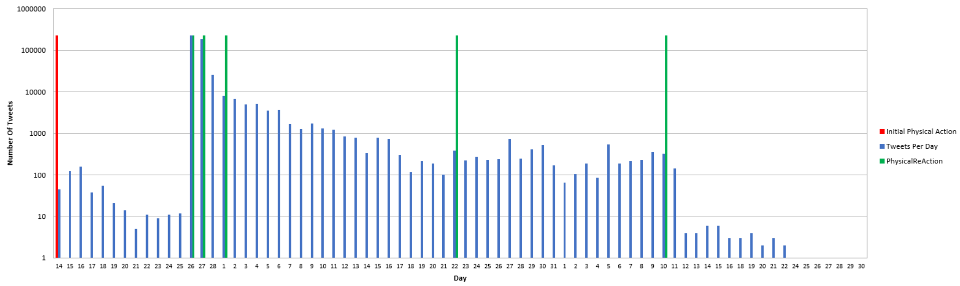

Figure 3: Total Number of Cyber Actions (tweets) Per Day of #Balakotstrike campaign.

0.4.4 Analysis of Case Study CICs

The cicmod ontology is applicable to all conceivable CICs, as the object relationships and causal connections remain the same. The graph-based structure of cicmod allows for unique connections to be made though ontological reasoning or inference rules. This section presents the specific analysis of the two selected case study datasets to showcase the evolution from intuitive results through to a deep analysis of behaviors across both the physical and cyber realms of a CIC. Each query has been specifically designed to demonstrate various possible types of analysis. We have analyzed the two selected CICs, #euromaidan and Balakotstrike, to compare and contrast key metrics, such as the size of the campaign, influence actions achieved, and key artifacts of the engaged networks.

It is also important to reinforce the point that our CIC ontological modeling and feature extraction has been tailored to the CIC use case. Our deep understanding of these specific case study CICs and their corresponding physical events provided insights that enabled us to extract suitable features for our analysis. We then used the extracted features for the database and model. Without this deep understanding, there is a potential for misinterpretation which could result in false positives or a model failure.

0.4.5 SPARQL

The SPARQL Protocol and RDF Query Language is the semantic query language used with data stored in an RDF dataset. Hence, to access the novel information generated as part of the ontology, specifically designed SPARQL queries must be written. Therefore, a unique query is written for each element of our analysis in order to extract the detail from our ontology. Each query is published with the dataset and hosted together for ease of reference and use as the IEEE DataPort Cyber Influence Campaign Ontology dataset (DOI 10.21227/70kc-yx38).

0.4.6 Actions Per Day

The number of cyber actions (tweets) per day is an elementary quantitative metric for the comparison of campaigns and gives an initial appreciation for the overall volume of the campaign. The number of actions per day does not directly indicate influence; however, provides some initial insights into the scale and behaviour of the CIC and potential time periods that require further analysis. The #euromaidan and Balakotstrike campaigns cover a period of 131 days and 46 days, respectively. The differences in timeline do not impact the numbers of actions per day, which still reflect the relative volume of interest in the issue over time. The SPARQL query first searches all actions, physical or cyber, and then sorts the actions by day and counts the number of actions per day. Figures 2 and 3 show the results of these queries with the physical events represented as vertical lines. This is because each day has a maximum of one event per day in both campaigns, except for February 24th 2014, when there were four physical events attributed to the #euromaidan campaign.

From Figures 2 and 3, we can identify clusters of physical events that correspond to increases in actions per day. This is an intuitive result that demonstrates that our

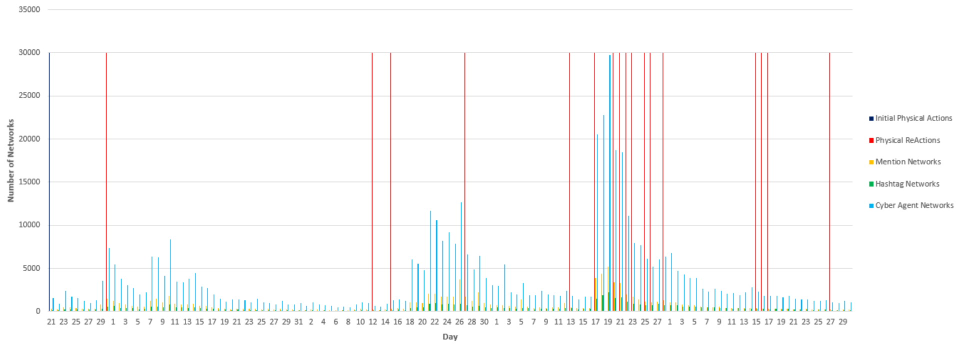

Figure 4: Total Numbers of Cyber Agent Networks, Mentions Networks, and Hashtag Networks by Day of #euromaidan Campaign.

data is accurate and our model reflects reality. Of note, in Figure 2, there are some offsets, as the physical events that were reported on the 13th and 16th of January 2014 did not have an immediate SNS response; the SNS response began to increase on the 19th of January 2014, potentially due to details of the physical events being released. The drop in tweet activity on the Balakotstrike campaign in Figure 3 on the 12th of April 2019 likely indicates the calming influence of the independent inspection of the airstrike location on the 10th of April. In Figure 2, the actions per day for the #euromaidan campaign peaked at around 80,000 tweets per day at the height of the hostilities between protesters and the Ukraine government. Whilst a shorter campaign, the Balakotstrike SNS activity spiked to almost 230,000 tweets per day in Figure 3 on the 26th of February 2019, the day of the retaliatory Indian strike against Pakistan.

0.4.7 Networks Per Day

The number of networks per day is a similar quantitative metric as the number of actions per day, however, the number of networks per day metric considers the volume of networks being engaged. The number of networks per day metric represents the diversity of how these actions are applied to the SNS platform. The Cyber

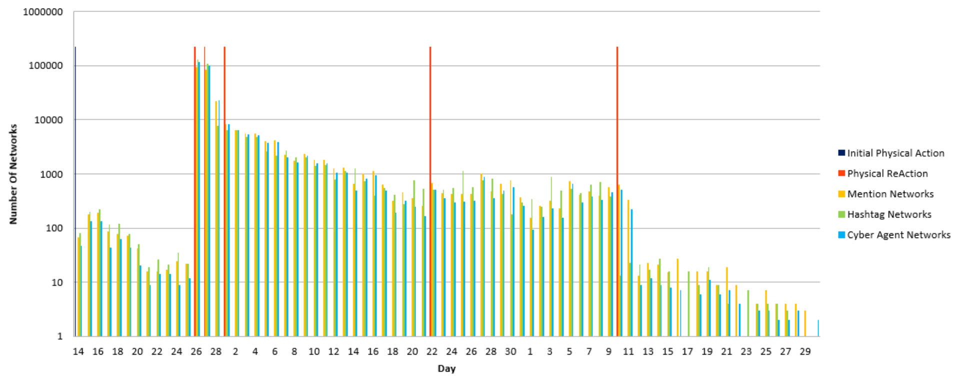

Figure 5: Total Numbers of Agent Networks, Mentions Networks, and Hashtags Networks by Day of Balakotstrike Campaign.

Agent Networks are the only networks that individual agents can post to and control. The predicate used in this situation was `cicmod:ControlledBy`. An agent can also "mention" another agent in an action. By mentioning another agent, a relationship represented by the predicate `cicmod:Mentions`, connects another agent's network to the action. The hashtag networks connected to the action are captured with the `cicmod:appliedTo` predicate, as a hashtag network is not controlled by an agent. This means that a hashtag network can be manipulated by any agent or narrative. Our raw data did not contain the follower or following metric for the hashtag networks; future research may consider hashtag networks with the follower and following metric.

For both campaigns, our first observation is that the numbers of networks per day in Figures 4 and 5 correlate closely with the number of actions per day in Figures 2 and 3. This is logical as the number of actions taken by agents is expected to be similar to the number of unique networks, because most agents will first post to their own network. Figure 4 shows limited hashtag employment compared to Figure 5. Potentially due to the limited public awareness of hashtags, only a small number of hashtags were used throughout the #euromaidan campaign.

Generally, the smaller the number of agents, the more limited the distribution of information, which curtails the dilution and manipulation of the information and

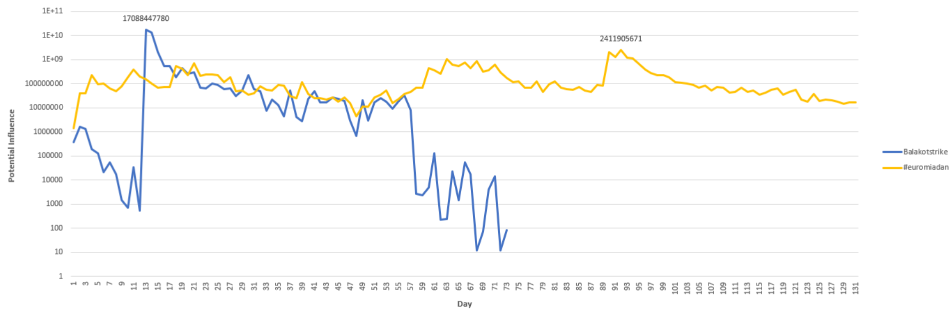

Figure 6: Potential Influence $\Pi(d)$ of the #euromaidan and Balakotstrike Campaigns

details. For organizing events, a single source of truth is preferable for an organizer if the priority is to coordinate demonstrations; however, the limited distribution restricts the exposure of the campaign. We observe relatively low numbers of Cyber Agent Networks in the early phase of the #euromaidan campaign in Figure 4, which helps of information. These low numbers of Cyber Agent Networks may also contribute to the very low numbers of hashtag networks in the left part of Figure 4.

Generally, an action needs to be taken in the cyber realm in order to enable the subsequent "Mention". That is, "Mention" networks are predominately retweets, which we have corroborated in additional data analysis that is not included here to keep the analysis presentation concise. Our additional data analysis has suggested that the CyberReActions are connected to all three types of networks, giving them the most exposure.

0.4.8 Potential Influence

Quantifying potential influence is possible by using the `cicmod:AppliedTo` predicate and accumulating the followers (agents) of the total networks that an action influences (`cicmod:Influences`) on a given day d. We define a quantitative potential influence metric $\Pi(d)$ which represents the number of end nodes of our graph. Formally, we denote $a(n, d)$ for the number of actions on a given network n on a given day d and denote $f(n, d)$ for the number of followers of a given network n on a given day d. Furthermore, we define $\tau \in \mathcal{N} = \{\mathtt{a}, \mathtt{h}, \mathtt{m}\}$ as an indicator variable for the network type, which can take on values from the set $\mathcal{N}$ of network types, specifically agent ($\mathtt{a}$) networks, hashtag ($\mathtt{h}$) networks, and mention ($\mathtt{m}$) networks in the context of twitter. We define $N_\tau(d)$ to denote the number of networks of type τ on a given day d. We then define the potential influence score $\Pi(d)$ on a given day d as:

$$\Pi(d) = \sum_{\tau \in \mathcal{N}} \sum_{n=1}^{N_\tau(d)} a(n,d) f(n,d). \tag{1}$$

The potential influence metric $\Pi(d)$ is akin to assessing the magnitude of a campaign by summing the numbers of network followers (agents) which are influenced by campaign actions. Thus, the $\Pi(d)$ metric allows for a fair comparison of two CICs. The values for Π quickly become massive, reaching orders of 10^9; these numbers reflect the potential end nodes, not the actual agents engaged.

Figure 6 displays the potential influence Π by day for both campaigns on a logarithmic scale. Using this quantitative Π metric we can see that the potential influence of the retaliatory strike from India had a large influence on the twitter population and by extension the world, reaching Π values above 10^{10}, which are higher than for any of the #euromaidan events. However, the #euromaidan campaign had a sustained influence over time, while the Balakotstrike campaign covered a much shorter time period. This comparison based on the potential influence metric Π as defined in Eqn. (1) provides a unique and novel ability to measure and compare one physical or cyber event against another with a consistent metric.

0.4.9 Active Agents over Time

Having established the potential influence of a campaign, the following example has been selected to show the ability to focus on specific details of the dataset and showcase the flexibility of the ontology. The agent behaviour over time showcases the

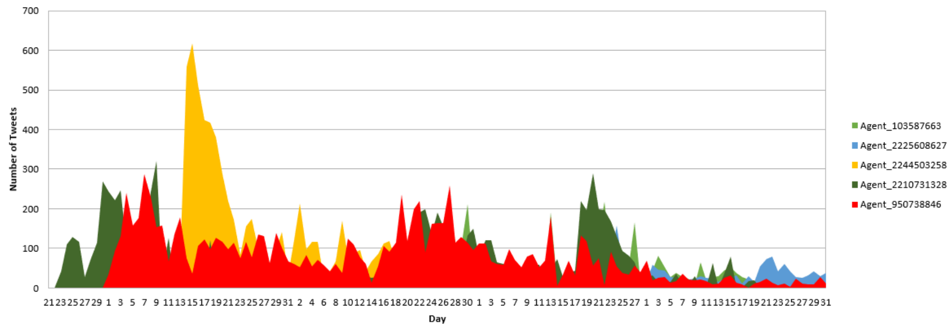

Figure 7: Number of Tweets Per Day of Top 5 Agents Within #euromaidan Campaign

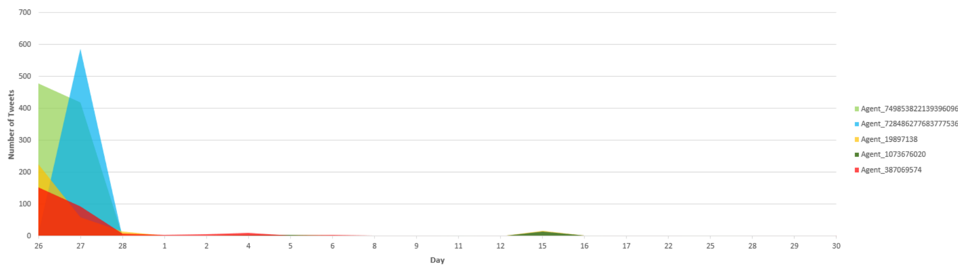

Figure 8: Number of Tweets Per Day of Top 5 Agents Within Balakotstrike Campaign

flexibility of the ontology. This analysis gives a quantitative appreciation of which agents were most active and when. The ability to not only identify key agents within CICs, but to also confirm human or automated behaviours is highly advantageous for the operational analysis of CICs. We conducted this analysis with two sequential queries: the first query discovering the most active agents over time; the second query grouping agent actions over time.

In Figure 7 for the #euromaidan campaign, Agent-950738846 in the data set is observed as part of an initial intense activity along with Agent-2210731328. Their activity peaks at over 300 tweets per day in early January; potentially organizing or reporting on the euromaidan demonstrations. However, their activity is quickly surpassed in mid January by Agent-2244503258, who peaks at over 600 tweets in one day, but then rather suddenly ceases all activity by mid February. This discontinuation of activity warrants further investigation, as it may provide insights into potential

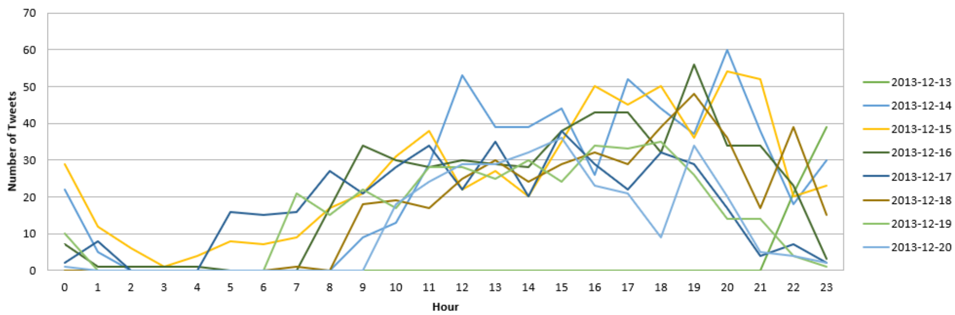

Figure 9: Agent-2244503258 Actions as a Function of Hour of the Day for Eight Day Period.

automated or state sponsored activity.

Similarly, in Figure 8, Agent-728486277683777536 peaks at just under 600 tweets per day early in the Balakotairstrike campaign, but then ceases any action. This dynamic suggests that this individual or account was only interested in the initial physical action and not in the subsequent physical events that happened in response.

The actions of Agent-2244503258, who tweeted over 600 times in a day as shown in Figure 7, are shown in Figure 9 per hour, over a six day period. From Figure 9 we can observe that the activity of Agent-2244503258 maintains periodicity with normal patterns of life for a human agent. This means, sleep patterns are maintained at night as well as peak periods of activity around early evening each day. Other fields within the dataset can also be leveraged to provide evidence of a human or automated agent. The device used for all actions by Agent-2244503258 was a desktop based interface of VK.com and each action came from the same location within the Ukraine. The combination of a single interface device and location supports the theory of a human agent using a desktop interface to publish a high number of tweets over a sustained period.

0.4.10 Outcome

The quantitative analysis using the CIC ontology has validated the mechanics of our semantic model and shown that the linkages and relationships of objects within the dataset reflect real life. The modeling and ontological representation is novel and provides the basis for future work in the field of cyber influence and the investigation of CICs. The application of our ontology to other research will allow for the integration of physical and cyber events in feature extraction and other machine learning (ML) techniques used in social media analysis. Having established an influence flow semantic model as the basis of our ontology, it is now possible to track and identify influence across realms through leveraging established ontologies.

0.4.11 Limitations

The presented case study analyses only represent samples of the types of analysis possible with our cicmod ontological model. With the flexibility of the SPARQL query language and graph based cicmod ontology, key insights can be gained into the behaviours and nature of CICs and cyber influence in general. The use of the location and language fields within the dataset is highly versatile for operational and thematic analysis in conflict. The intent of this research was to provide a novel and flexible ontology to progress the field of cyber influence.

We acknowledge that to the best of our knowledge, a theoretical analysis of failure or error bounds of the introduced CIC ontology model is intractable. From an empirical research perspective, two independent CICs have been assessed with the developed CIC ontology in this article. Future research should explore additional CICs in order to determine if there are any scenarios or types of CICs that do not fit the introduced CIC model or cause it to fail. In order to support future research, the ontology, datasets, code for the data pipeline, and SPARQL queries have been hosted as the IEEE DataPort Cyber Influence Campaign Ontology dataset (DOI 10.21227/70kc-yx38).

ML can be employed to recognise and define indicators of activity that may lead to physical events. For example, determining the preconditions that result in physical demonstrations or potentially a change of leadership within a state. With the ability to compare physical events against each other, we can also use the SNS activity to suggest when activity reaches a threshold to cross domain into the physical realm.

This study has focused on developing and evaluating an ontology model for analyzing cyber influence campaigns in conflicts conducted in social media networks. Social media networks can also give indications of emerging cyber security threats Bose *et al.* (2019); Michel and King (2019); Riesco and Villagrá (2019); Simran *et al.* (2019); Syed (2020). One interesting future work direction is to adapt our ontology model to uncover the sources and agents behind emerging cyber threats. Moreover, social media can be used to spread misinformation to wide audiences. In future research, our model could be adapted to identify the sources of potential misinformation.

0.5 THEMATIC CAMPAIGN CLASSIFICATION FRAMEWORK

The TCC Framework begins with raw data from any SMP and ends with classification of thematic campaigns within a specific data set Johnson *et al.* (2022b). Our TCC framework accepts data sets from a SMP and detects types of campaigns within the data sets. This has significant value to entities tracking specific themes in unfamiliar networks or platforms. The frameworks includes a data ingest and pipeline. The first stage extracts the data from raw JSON files that are then normalised for parsing to the linear regression model. Normalisation is important to remove the bias of individuals in a network. We use a common measurement of influence, namely

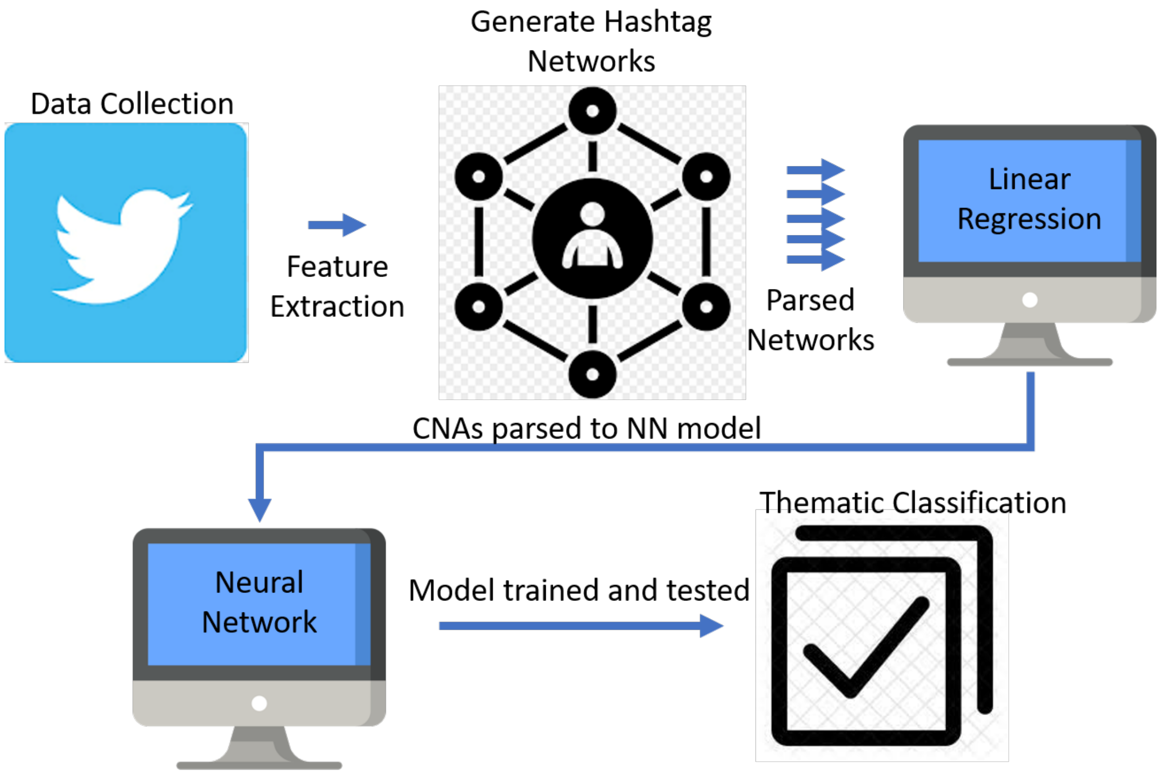

Figure 10: Visual representation of the Thematic Campaign Classifier (TCC) framework

cumulative likes. Generally, the more likes are received for an action, the more influential is the action. However, likes are based on followership, hence, cumulative likes is skewed, and needs to be normalised to remove the individual bias before conducting network analysis. The normalisation process focuses the importance on the nature of the network not the account Jin (2020).

The linear regression model predicts the normalized likes based on the network specific features of the data set. The Coefficient of Importance (CI) then weights each network attribute based on which one effected the likes goal the most. For example, if retweets are most popular in a network, the CI process will rank retweets as the highest in the CNA scale. Alternatively, if photos relate to likes more closely in a network, then the CI will rank photo content higher than retweets. From there, the aggregated CNAs within a parent campaign are parsed to the Neural Network

(NN) classifier. An NN model is built to predict the nature of a CNA based on its supervised learning of other CNAs. Figure 10 is a graphical representation of the process. The remainder of this section describes each step in the framework in detail.

0.5.1 Data Ingest and Pipeline

Data collection

When experimenting with social media analysis, researchers must specifically define the boundaries of their data collection. This is for a number of reasons: 1) The size of using the entire platform becomes unfeasible without the support of high-performance computing Gundecha and Liu (2012). 2) Search parameters build sub spaces within the platform and as such completeness cannot be assumed when using sub spaces for search purposes Shahi (2020).

3) Once the data of a large enough proportion is considered, the behaviors are no longer random, but rather can be modelled by various probability distribution techniques.

4) Data sets are time sensitive. Social media data can be represented as a time-series and multivariate data set, with each piece of data in a chronological order that contains multiple data features. However, each piece of data can be changed by agents at any time, e.g., user comment on actions changes the importance or visibility of that data. This means, that whilst the data is cumulative, the time t of collection reflects a 'snapshot' of the platform at $t = 0$. If another collection of data is taken at $t = 1$, this becomes another 'snapshot' and any action that occurred between the two snapshots will change the nature of the data and potentially the results of an experiment or analysis.

The study Arora *et al.* (2019b) showed how seed data can be used to query a SMP data set so as to obtain a subset of the platform data. This is common practice when extracting content based features, as the time t is not important and homogeneous data is assumed. However, in our research use case, these assumptions cannot be made and data sets have to essentially stop evolving in order to achieve static results.

Other studies accept large data sets without investigating the origins or nature of collection Ma *et al.* (2013). However, this trend has been superseded by targeted collection in recent years. For a campaign, firstly, the campaign itself must be defined and then a data set must be collected to the best of the researcher's ability. The best form of data set for this type of research is a "historic" snapshot of campaigns that are no longer of interest. Such campaigns are almost static in nature. We have used data sets from well-established historic campaigns. These data sets where purchased from the third party provider TweetBinder twe (2021) that interacts directly with the Twitter and Instagram APIs. TweetBinder provides full-take Twitter and Instagram data sets. The authors purchased two holistic SMICs, as well as five 50,000 tweet campaign subsets to use in the experimentation.

Feature Extraction

By default, social media data sets are captured in Java Script Object Notation (JSON) format. A JSON is a dictionary of dictionaries. This dictionary structure is ideal for vast social media data and allows to iterate through each file for simple processing. For the pre-processing stage of our TCC framework, each JSON is read in its entirety (these are files between 50,000 to 2,000,000 tweets (50 Mb – 2 Gb) which is approximately 1000 tweets per Mb.

Pre-Processing Our pre-processing extracts feature fields of each tweet that specifically relate to network interactions. That is, the pre-processing extracts any feature field that indicates how the network reacts to an action, such as, replies, likes, favorites, or mentions. This relates back to the semantic interpretation of social media network behavior being action-network-reaction Johnson *et al.* (2020).

The feature fields that are available for extraction are unique to both the SMP and the method used for scraping the data. Our data sets are historical campaigns provided by 'TweetBinder' twe (2021). A historical campaign gives a full-take retrospective view of the campaign. That is, the time t value is at a maximum and unlikely to change. Moreover, the historical view shows the end result of the interactions with cumulative metadata. In contrast, data that is scraped at $t = 0$, would not reflect any interactions with the network.

Following the extraction, the feature values are normalized. A normalizing technique is used in Arora *et al.* (2019b) to correlate homophily, confounding, and influence in social media as well as finding the importance of influencers in social media, respectively. Our approach differs from Arora *et al.* (2019b) as they scale down the values to a specific range. Our technique removes the scale of the number of followers of an agent. This allows direct comparison of action influence, i.e., if a tweet from an agent with 1000 followers gets 10 retweets (1% influenced), then this tweet is actually less influential than a tweet from an agent with 100 followers that gets 10 retweets (10% influenced).

Notation for Social Media Data We define the set of campaign types $\mathcal{C} = \{\text{Conflict, Entertainment, Environment, Political, Sports}\}$. For each tweet (post) i in a given JSON, we define $\mathcal{H}_i$ as the list of hashtags used in the post. Moreover, we define a set $\mathcal{F}$ of quantitative features (characteristics) that are considered for a given tweet i; specifically, we define for this study $\mathcal{F} = \{\text{Favorites or Likes, Hashtags, Links, Mentions, Replies, Retweets, } \ldots\}$. We note that this set of features $\mathcal{F}$ is an example and could be varied. For a given tweet i and feature $f \in \mathcal{F}$, we define the function $F_f(i)$ to return the total number (raw value) of feature f for tweet i. For instance, suppose a tweet $i = 4$ has 12 retweets, then $F_{\text{Retweets}}(i = 4) = 12$. We define the normalized value $\bar{F}_f(i)$ of a given feature f of a tweet i as the raw value $F_f(i)$ divided by the corresponding number N of followers of the user that posted tweet i, i.e., $\rho_{i,f} = \bar{F}_f(i) = F_f(i)/N$.

Network Generation The definition of sub sets or

sub networks is the next stage of the TCC framework. If the data set is taken in as a whole, only one set of attributes would be extracted. This would be the highest level of abstraction of a campaign and would provide a singular network perspective. As alluded to in Section **??**, forming sub sets will allow for insights into the sub networks associated with a campaign. Three specific ways to define sub networks can be considered.

1. Time: Considering the campaign tweets by time allows for the tracking of network attribute changes over time. Time deltas also allow for periodicity, e.g., breaking down the time epoch into minutes, hours, days, weeks, and months, as required.

2. Hashtags: This method takes each individual tweet within the campaign and allocates it to an individual "hashtag network" Ξ_h. This process has been called binning the data **?**.

   If a tweet has two hashtags, e.g., h_1 and h_2, then the tweet will be allocated to both hashtag networks or bins Ξ_{h_1} and Ξ_{h_2}. Conceptually, this means that two networks have been exposed to the one tweet.

3. Blocks of Tweets: This simplest form of division processes a set block of tweets

in chronological order. Whilst potentially random in nature, this does have utility. This methods provides a way of conducting parity checks to ensure that the overarching nature of the campaign is consistent.

We focus on the hashtag network as an input in this study. The hashtag network method presents a unique and indicative sub network.

Moreover, this can then be cross applied to other data sets, as the hastag sub network is an independent observation of the network behavior.

0.5.2 CNA Design

We define the CNA using linear regression and coefficient importance. At this point in the framework, the process has extracted normalized network features $\rho_{i,f}$ and generated hashtag networks Ξ_h, see Fig. 10. Conceptually, the data is now an ensemble of samples (hashtag networks) of the network behavior. These samples are aggregated to form a unique representation of a campaign's underlying behavior. Each sample (hashtag network) with its normalized feature set is fed into a linear regression model. The target or independent variable is the normalized likes, with all other features being the dependant variables. The more complex the data set, the higher the dimensionality.

Linear Regression Model and Importance Coefficient

We investigated several regression models and selected linear regression because it returned the best results in the classification phase, see Section 0.5.4. Once the regression model has been developed, the importance of each feature $f \in \mathcal{F}$ to the LR model is evaluated using an Importance Coefficient $IC(f)$.

Figure 11 shows an example of a Campaign Network Attribute CNA_h from the #Euromaidan data set. The Importance Coefficient IC as a function of the features

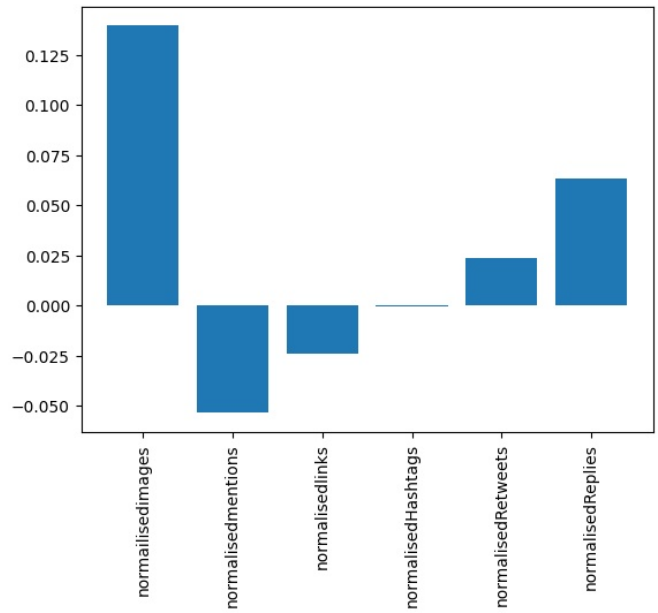

Figure 11: Example of Campaign Network Attribute (CNA) from #Euromaidan campaign: Importance coefficient $IC(f)$ as a function of features $f \in \mathcal{F}$.

$f \in \mathcal{F}$ indicates how important each feature f is to that hashtag h and it's respective hashtag network. In the CNA in Figure 11, the image feature is highly important, while the mentions and replies features are moderately important, and the links and retweets features have low importance, and the number of hashtags has essentially no importance.

CNA pseudocode

Fig. 12 presents the pseudocode for generating the CNAs. For a given JSON of tweets, let i denote an individual given tweet in the JSON. Let $f \in \mathcal{F}$ denote a given network feature, such as likes, retweets, mentions, and hashtags as described in Section 0.5.1.

For example, $F_{\text{likes}}(i)$ is an unnormalized specific feature, namely the total number of likes of tweet i. $\rho_{i,f}$ is set to the corresponding normalized feature f value from tweet i. As summarized in the pseudocode in Figure 12, for each feature $f \in \mathcal{F}$, the code iterates through, placing the normalized feature value $\rho_{i,f}$ for each tweet i. For

Figure 12: Pseudocode for generating Campaign Network Attribute (CNA) characterization and Thematic Campaign Classification (TCC).

for $f \in \mathcal{F}$ **do** ▷ For all features f

  for i in JSON **do** ▷ For all tweets i

    $\rho_{i,f} \leftarrow \bar{F}_f(i)$ ▷ Normalized feat. f value of tweet i

    if $t_m \leq t_i \leq t_n$ **then** ▷ Tweet i in time range?

      for $h \in \mathcal{H}_i$ **do** ▷ Hashtags in tweet i

        $\Xi_h \leftarrow \ \rho_{i,f}$ ▷ Add to hashtag netw.

      end for

    end if

  end for

end for

$\mathcal{S} = \{h : H_{\min} \leq |\Xi_h| \leq H_{\max}\}$ ▷ HT netw. size in range

for $h \in \mathcal{S}$ **do** ▷ Create stack of $|\mathcal{S}|$ CNAs

  $IC_h(f),\ f \in \mathcal{F} \leftarrow$ Lin. Regression (Ξ_h)

  $\text{CNA}_h \leftarrow\ IC_h(f),\ f \in \mathcal{F}$; Plot $IC_h(f)$ as a fct. of f

end for

TCC: NNmodel $\leftarrow$ $\text{CNA}_h,\ h \in \mathcal{S}$

TCC: Hyp. param. tun. $\leftarrow$ NNmodel

each hashtag value h in i, we create a hashtag network that is a list of the associated tweets - this process known as binning. Specifically, for each unique hashtag $h \in \cup_i \mathcal{H}_i$, we create a list Ξ_h that contains the indices i of the tweets that contain the hashtag h. The list Ξ_h represents the hashtag network for hashtag h and we denote $|\Xi_h|$ for the hashtag network size, i.e., the number of tweets that contain hashtag h.

The hashtag networks $h \in \mathcal{S}$ that satisfy the hashtag network size selection criteria, i.e., have hashtag network size $|\Xi_h|$ between a prescribed lower bound $H_{\min}$

and upper bound $H_{\max}$, which are examined in Section 0.5.4, are then parsed to the linear regression. The resulting standardized regression coefficients for the considered features $f,\ f \in \mathcal{F}$, are assigned to the Importance Coefficient vector, i.e., to $IC(f)$. The completed $IC(f)$ vector can be plotted against the features $f \in \mathcal{F}$ and represents the CNA_h for a given hashtag $h \in \mathcal{S}$ as shown in Figure 11.

0.5.3 TCC: Thematic Classification with Neural Network

In preparation for the classification phase of the framework, the network themes for various SMIC categories must first be defined. We took a "newspaper" approach to categories, i.e., each SMIC theme should correlate to a section of a newspaper, such as Conflict, Politics, Finance, Sport, Entertainment, and Environment. As such, when collecting data, a campaign should have a clear affiliation with one of these sections. We took precautions to avoid selecting SMIC themes that could be incorrectly associated, thus negatively impacting the performance of the classification model.

The stack $\text{CNA}_h,\ h \in \mathcal{S}$ of CNAs is a unique encoding of a social media network and can be represented as a vector in high-dimensional space. These CNA vectors can be grouped and sorted mathematically. As summarized in the last two lines of the pseudocode in Figure 12, the TCC classification provides the $\text{CNA}_h,\ h \in \mathcal{S}$, i.e., the Importance Coefficients $IC(f)$ for the features $f \in \mathcal{F}$ for the hashtags $h \in \mathcal{S}$ to the neural network (NN) model. We refer to the number $|\mathcal{S}|$ of CNAs that are considered in the TCC as the support of the classification. Also, the TCC classification conducts the NN model optimization, i.e., tunes the hyperparameters.

We employed a feed forward Neural Network (NN) model to discover the underlying pattern of the CNAs. Specifically, the supervised NN model was selected based on its ability to maintain the dimensionality of the CNA data. In other words, the supervised NN model can find and group our CNA features without losing detail.

Much of the implementation of NN models in Python is provided by Python's 'sklearn' package. To train the NN model, it must be provided true vectorized data to train the model and a sample to test the model. Using Python's 'split' function allows for division of the data into training and testing pools, respectively. In this research, a 70% train and 30% test ratio was maintained for the training and testing phases, respectively. Generally, the 70:30 train to test split ratio has been found to be optimal for training NN models in a wide range of scenarios. This split ratio has demonstrated the high performance of the model without biasing towards major or minor classes. Additionally, the model achieved the stability of results regarding detection accuracy and false alarm rates, neglecting the overfitting and underfitting issues. To avoid these issues, the model was trained and validated using enough data (70%) feeding with suitable numbers of class instances. We verified that the 70:30 train to test split ratio achieved the best performance and therefore adopted the 70:30 train to test split ratio for this study.

0.5.4 CNA-TCC Framework Tuning

Overview

Initial testing indicated that there were input parameters and hyperparameters that effected the classification performance of the NN model. These included:

- Network size – the number $|\Xi_h|$ of tweets per hashtag network used to create the CNAs.
- Linear Regression Model – used to generate the CNAs.
- Hyperparameters of the NN model itself.

Hence, framework tuning was conducted to evaluate and refine each of these elements as described in this section, before the actual testing in Section 0.5.5.

Individual tests were conducted to evaluate and tune each input parameter and hyperparameter. The setup for these tests was kept consistent and only a single variable was changed at a time in order to observe input and response. In order to test an input parameter, it must be within a range that allows the code to execute through to classification. It must also be noted that the model is not deterministic, not always achieving the same level of precision. Hence, each experiment was run 10 times and the average results are reported.

Data Sets

We used the following data sets, which were provided by TweetBinder and contained the full API level networking features, for this study (available from DOI 10.21227/01e7-pj58):

- #FIFAWWC - Fifa Word Cup – Sport
- #MTVEMA - Music TV awards – Entertainment
- #zagrebearthquake - Croatian Earth Quake – Environment
- #ARRESTTRUMPNOW – Political
- #Euromaidan – Conflict
- #BalakotAirStrikes – Conflict
- #IndiaStrikesBack – Conflict

For the framework tuning we used two small data sets of approximately 50,000 tweets each from the #BalakotAirStrike and #IndiaStrikesBack campaigns, referred to as 'Conflict 1' and 'Conflict 2', respectively. We designed the tuning evaluation

Table 13: Results of NN classification of two small similar campaigns.

Campaign Type	Precision	Recall	F1-score	Support $\|\mathcal{S}\|$
Conflict 1	0.40	0.35	0.38	17
Conflict 2	0.67	0.71	0.69	31
Arithm. Mean	0.53	0.53	0.53	48
Weighted Average	0.57	0.58	0.58	48

to be very challenging for the classifier, having to decide between two highly similar campaigns patterns within a given campaign theme. More specifically, we provided the training data set of Conflict 1 (with the #BalakotAirStrike labels) to the CNA-TCC framework for NN learning. Then, the trained NN was presented with the Conflict 1 testing data set (without labels) and the Conflict 2 data set, and the NN was tasked with classifying the data sets as Conflict 1 or Conflict 2, which is a harder classification problem than classifying a data set as a conflict thematic data set. As such, any application to more distinct data sets, e.g., data sets of campaigns with different themes, will tend to yield better results, as examined in Section 0.5.5. Figure 0.5.4 shows the resulting confusion matrix and loss curve, while Table 13 gives the Precision, Recall, and F1-score metrics. The Support metric is the number $|\mathcal{S}|$ of CNAs (hashtag sub networks) considered in the classification. We observe an average precision of approximately 50%. The loss curve exhibits a linear relationship between iterations and costs after approximately 10 interactions. This indicates that the NN model is highly challenged when having to classify two similar campaigns.

Hyperparameter Tuning

This section focuses on the hyperparameters of the NN model itself, as opposed to input parameters to the LR considered in Section 0.5.4. There are many ways to

Table 14: Results of hyperparameter tuning, given an initial model precision of 0.40

Optimizer	LR model	Layers	Iterations	Precision
Grid	DT	120, 80, 40	150	0.46
Random	DT	150, 100, 50	50	0.51

tune the hyperparameters of an NN model for the best performance. Both the grid and random methods are implemented and allow for simple execution. The grid method, combines both the number of hidden layers and the number of iterations together, e.g., if the hidden layers for the model are $\{30, 60, 90\}$ and the iterations are $\{50, 100, 150\}$, then the grid hyper parameter optimisation will consider the settings $X_a = [30(50, 100, 150),\ 60(50, 100, 150),\ 90(50, 100, 150)]$. The random method is similar, however, selects random combinations within the established boundaries. The grid method may miss global optima due to the regularity of sampling. Whilst the random method can potentially discover optima closer to the global optima. The typical start point for the hidden layer sizes are $[(150, 100, 50),\ (120, 80, 40),\ (100, 50, 30)]$ and for maximum iterations of $[50, 100, 150]$. These hyperparameters were kept the same for the tuning phase.

Optimising the hyperparameters leveraged the same test setup as Section 0.5.4. However, this time, the optimal results from the previous experiments where used as inputs to start with the best-case scenario. This results in the best of both situations, where the input parameters gave the best result and the NN model itself was tuned to achieve the best result. Table 14 contains the results of this testing and shows that with an initial classification precision of 0.40, the hyperparameter tuning was able to achieve a 0.06–0.11 increase in precision.

The initial precision before the optimization for this comparison was 40%. As shown in Table 14, there was an increase of 6% for the grid optimizer and 11% for

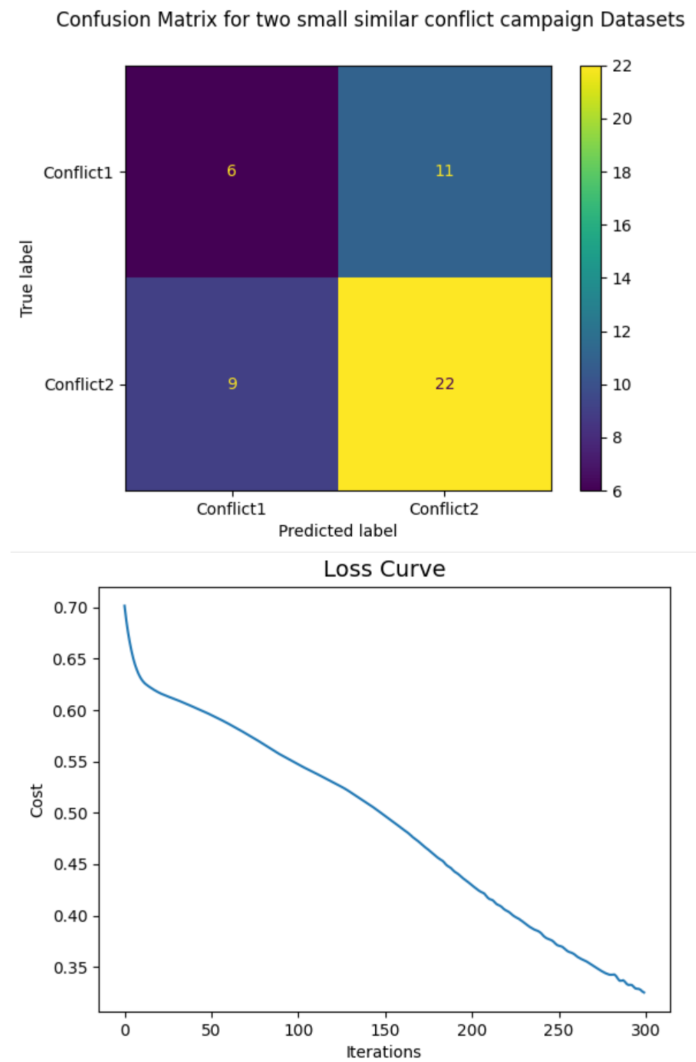

Figure 13: Confusion matrix and Loss Curve for initial classification trial.

the random optimizer. This is a significant improvement that is achieved by tuning the hyperparameters and well worth the computational cost. The random optimizer was used in the final experiments, however, given the stochastic nature of optimizing, both methods will be run in the final experiments regardless to observe which one performs better.

Tweets per Hashtag Network Tuning

We initially hypothesized that the larger the hashtag network becomes, the closer its behavior reflects the overall campaign. This would result in a higher classification precision for a larger hashtag network. However, as the number of tweets in a sub network grows, it introduces behaviours of its own. Therefore, the sub network being observed will begin to have independent underlying behaviors of its own which reduces precision. Based on these considerations, we proceeded to hypothesize that an optimal hashtag network size exists which reflects the parent campaign and is dependant on the size of the overall campaign. Therefore, boundaries for the maximum and minimum number of tweets per hashtag network can be found, e.g., network sizes could be optimal between 0.1% and 0.5% of the overall data set size. Typically, a campaign data set of 50,000 tweets translate to suitable hashtag networks between 50 and 250 tweets as further examined in the following section.

Method

We increased the acceptable network size by a discrete amount (step size) and recorded the classification precision. Specifically, we increased the size of the hashtag networks via a step size of 50, which resulted in maximum network sizes of $|\Xi_h| = \{50, 100, 150, \dots\}$ being parsed to the LR and CI model.

Given the independent nature of campaigns, whilst this would keep the network sizes consistent in relation to the overall campaign, the classifier has to handle the comparison of inconsistent numbers of CNAs for the campaigns.

In our evaluations, we examine the impact of the stack size by reporting the weighted average precision, i.e., the mean of the precision for the different theme classes obtained by weighting by the number of CNAs (support) for a given theme class. We next investigate the impact of the hashtag network size $|\Xi_h|$.

Results

The results of this experiment were as anticipated, the classifier struggled to delineate between the two similar campaigns, indicating a strong similarity in the CNAs extracted from the two campaigns. The Loss Curve as shown in Figure 0.5.4 shows little to no 'elbow' after a few initial iterations, with a linear relationship out to 300 iterations. This suggests an initial gain, however, there is no obvious convergence of the classifier, and incremental benefits over iterations.

The intent of test one was to assess the number of tweets per hashtag network to generate the CNAs, evaluated in terms of precision achieved by the NN model. Table 15 indicates consistently high precision (≥ 0.60) for hashtag network sizes $|\Xi_h|$ between 100 and 250. Hence, these are acceptable tweet numbers per has tag sub network. In additional evaluations that are not included due to space constraints, we observed that outside of this range the precision becomes chaotic and not suitable for classification with the NN model.

Overall, these results indicate that as the size of the network increases, the underlying pattern remains consistent for the campaign and independent behaviours do not become apparent for these hashtag network sizes $|\Xi_h|$ in the 100–250 range.

Discussion

The problem with having a finite data set for tuning is that as the campaign size decreases, the number of available sub networks also decreases. During both the tuning and experiments, to avoid introducing artifacts, a minimum and maximum number of networks for classification had to be determined. As shown in the thematic classification experiment in Section 0.5.4, this was achieved by increasing the minimum

Table 15: NN classification of two small similar campaigns increasing hashtag network size $|\Xi_h|$ in increments of 50 tweets.

$\lvert\Xi_h\rvert$	Campaign	Prec.	Recall	F1-score	Support $\lvert\mathcal{S}\rvert$
50	Conflict1	0.58	0.39	0.46	116
50	Conflict2	0.66	0.81	0.72	170
100	Conflict1	0.71	0.21	0.33	149
100	Conflict2	0.62	0.94	0.75	207
150	Conflict1	0.63	0.19	0.29	155
150	Conflict2	0.63	0.93	0.75	228
200	Conflict1	0.60	0.13	0.21	162
200	Conflict2	0.61	0.94	0.74	232
250	Conflict1	0.66	0.17	0.27	171
250	Conflict2	0.60	0.96	0.74	233
300	Conflict1	0.55	0.21	0.30	171
300	Conflict2	0.61	0.87	0.72	239

number of networks acceptable for classification until artifacts disappeared And then, continuing to increase the number of networks until artifacts were observed again.

It is shown that for these data sets of approximately 50,000 tweets, the relationship between network size and precision is noisy, but linear until it breaks down around $|\mathcal{H}_i| = 300$ for this data set size. This means that a hashtag network should be less than 0.6% of the total campaign data set and greater than 0.02% for the behavior to reflect that of the parent campaign. Input parameters outside of these for the hashtag network size result in the network exhibiting chaotic or non-linear behavior.

Table 16: NN classification of two small similar campaigns for different linear regression tuning models.

LR model	Campaign	Prec.	Recall	F1-score	Support $\|\mathcal{S}\|$
LS	Conflict1	0.36	0.13	0.19	211
LS	Conflict2	0.46	0.76	0.58	206
DT	Conflict1	0.45	0.36	0.40	211
DT	Conflict2	0.46	0.55	0.50	206
Ridge	Conflict1	0.47	0.82	0.60	211
Ridge	Conflict2	0.25	0.06	0.10	206

Linear Regression Tuning

We evaluated different types of linear regression models for the CNA extraction, specifically: 1) Least squares, 2) Decision tree, and 3) Linear ridge **?**. The hypothesis was that the most efficient least squares would be the most suitable. Each algorithm can be called individually, making comparison simple and direct.

The scenario was run and the different models were called to extract the CNA. These were then parsed to the NN model for classification. The NN model classification precision was the metric used to determine the success of each model. The results in Table 16 indicate that overall, the decision tree regression model achieved the highest precision values.

Similar to the results observed in the tweets per hashtag network tuning, regardless of the linear regression model, the NN model has difficulty in discriminating the two similar campaigns. Table 16 shows that the Decision Tree regression preformed the best for our purposes. Therefore, in the hyperparameter tuning and thematic classification experiments, the LR and CI models will use the Decision Tree algorithm.

Thematic Classification of Two Campaigns

The first testing experiment is the culmination of our our framework development and built upon the results achieved from the framework tuning. Having optimized input parameters and examined the effectiveness of the hyperparameter tuning, the experiments aim to answer the research question: can the TCC framework thematically classify SMCIs?

Data sets

The data sets required to answer our research question had to be of unique and distinguishable themes, as noted in Section 0.5. The news paper section approach has broad and individual themes with strong underlying patterns which allow the NN model to delineate the campaigns. For this experiment, we conducted pairwise comparisons of campaigns with the themes Sport, Entertainment, Environment, Political, and Conflict. All data sets were of similar size to avoid any potential bias or weighting issues and all were extracted from the Twitter API using Tweetbinder which scraped the exact same network metadata. These data sets were #FIFAWWC (Sport), #MTVEMA (Entertainment), #zagrebearthquake (Environment), and #ARRESTTRUMPNOW (Political).

Experiment Method

Having determined boundaries for the hashtag network sizes, the TCC framework was applied to the data sets. The data sets are subject to all the pre-processing, CNA extraction, as well as the NN modelling and hyperparameter tuning for each trial. For a given comparison in Table 17, the CNA-TCC framework was provided with the training data set of the first campaign, e.g., Politics (#ARRESTTRUMPNOW) in the first row of Table 17 for NN training. Then, the trained CNA-TCC framework was provided with the test data set of the first campaign and the data set of the second campaign and tasked with classifying the data sets into the first campaign theme or a different campaign theme (e.g., for the first row in Table 17 the classification is into politics or non-politics). Each comparison was conducted with both the grid method and the random method for the hyperparameter tuning.

Results

The results in Table 17 indicate that the NN model classifier can achieve precision levels ranging between 0.68 to 0.90, which is inline with and in excess of the comparable literature Chu *et al.* (2012). The the lowest optimized precision of 0.68 is for the comparison between Sport and Environment, while highest optimized precision of 90% is achieved for the comparison between Entertainment and Environment.

We also observe from Table 17 that the random hyperparameter tuning method tends to generally achieve slightly higher precision than than grid method. However, for the first comparison Politics vs. Sport, the random method performs poorly (0.57 precision) compared to the grid method (0.69 precision). The grid method appears therefore to give generally more consistent campaign classification results.

Thematic Classification of Five Campaigns

In the previous experiment, the NN model classifier was only presented with two campaign types to decide from. Essentially, this is a 50% chance of guessing correctly, regardless of the learning process. Therefore, we conducted two experiments that involved five campaigns to increase the complexity of the problem by presenting multiple campaigns to the model.

Table 17: NN model classification precision for two campaigns.

Campaign themes	Default (non-opt)	Grid	Rdn.	% Gain
Polit. vs. Sport	0.71	0.69	0.57	−0.02
Polit. vs. Entert.	0.82	0.82	0.826	+0.007
Polit. vs. Environ.	0.80	0.80	0.828	+0.03
Sport vs. Entert.	0.82	0.82	0.825	+0.006
Sport vs. Environ.	0.68	0.67	0.669	−0.01
Entert. vs. Environ.	0.90	0.90	0.902	+0.002

Classifying Five Campaigns

Using five campaign data sets creates a classification decision between five different campaigns. Rather than a 50% chance of being correct, the classifier has a one in five (20%) chance, and any value significant higher than 0.20 demonstrates higher than guess-level performance of the NN classifier.

Experiment Method The CNA-TCC framework was trained with the training data sets of the campaigs #FIFAWWC (Sport), #MTVEMA (Entertainment), #zagrebearthquake (Environment), #ARRESTTRUMPNOW (Political), and #BalakotAirStrike (Conflict). The trained CNA-TCC was tasked with classifying the testing data sets of these five campaigns.

Results Table 18 gives the precision, recall, F1-score, and support numbers, while Figure 14 shows the confusion matrix and loss curve, which exhibits a pronounced elbow function. We observe from Table 18 and the confusion matrix that the entertainment, politics, and sport campaigns are classified with a precision of over 40%. In contrast, the classification fails for the environment campaign in this 5-campaign classification, mainly due to the weak support for the environment campaign. The

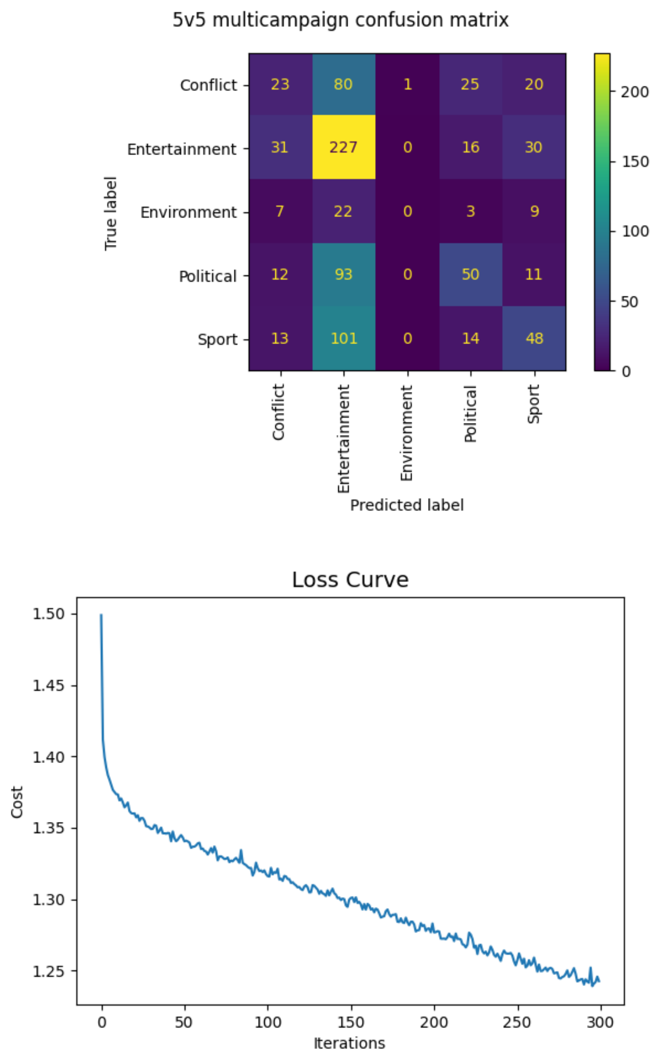

Figure 14: Confusion matrix and loss curve of multi campaign thematic classification

weighted average precision of 0.38 compared to the arithmetic mean precision of 0.31 further underscoring the importance of classification based on a strong support. A key outcome of this 5-campaign classification evaluation is that provided strong support, the CNA-TCC framework can achieve classification precision levels in the 30-40% range and thus substantially reduce the search space compared to a random guess classification with a 1 in 5 (20%) chance of correctness.

Table 18: Optimized NN model classification performance for five campaigns.

Campaign theme	Prec.	Recall	F1-score	Support $\|\mathcal{S}\|$
Conflict	0.27	0.15	0.20	149
Entert.	0.43	0.75	0.55	304
Environ.	0.00	0. 00	0.00	41
Polit.	0.46	0.30	0.36	166
Sport	0.41	0.27	0.33	176
Arithm. Mean	0.31	0.30	0.29	836
Weighted Avg.	0.38	0.42	0.38	836

Search for a Known Campaign Among Five Campaigns

Finally, we conducted a second experiment of thematic classification using multiple campaigns. We retained the campaign identifier for a known campaign and removed the campaign identifiers of the other four campaigns in the data sets. Specifically, for the training, the CNA-TCC framework was provided with the training data set of the one known campaign (with the corresponding campaign label), as well as the training sets of the other four campaigns (with the campaign labels removed). The trained CNA-TCC framework was tasked with classifying the testing data sets of all five campaigns into either the one known campaign theme or an ”other” campaign theme. Thus, we gave the NN model essentially a Boolean choice, simulating the concept of searching for a known campaign in generic social media data. Table 19 shows that the known campaigns were found with a precision ranging from about 25% to over 60% for the known target campaign for the campaigns with strong support of on the order of 150 or more hashtag networks. We also note that the upper range of the precision range of 67% was achieved for the Entertainment campaign with the very strong support of $|\mathcal{S}| = 304$ hashtag networks for the campaign characterization.

Table 19: Optimized NN model classification performance for finding a known campaign among data for five campaigns.

Campaign	Prec.	Recall	F1-score	Support $\lvert\mathcal{S}\rvert$
Conflict	0.25	0.01	0.01	149
Other	0.82	1	0.90	687
Entert.	0.67	0.03	0.05	304
Other	0.64	0.99	0.78	532
Environm.	0	0	0	41
Other	0.95	1	0.97	795
Political	0.59	0.06	0.11	166
Other	0.81	0.99	0.89	670
Sport	0.59	0.17	0.26	176
Other	0.81	0.97	0.88	660

This evaluation Table 19 corresponds to the common scenario of seeking to identify a specific SMIC in generic data, i.e., the CNA-TCC classifier is given multiple (four) campaigns without labels and trained for a specific (fifth) campaign theme. The CNA-TCC classification achieved a precision of 25% to over 60% for the campaigns with strong support, i.e., a reasonably large ensemble of hashtag networks for the campaign characterization. This significantly reduces the search space and would generate a highly comparative list of SMICs that closely reflect the original or trained SMIC.

Ablation Study of CNA-TCC

This section presents an NN ablation study of CNA-TCC, utilizing the components of NN, their hyperparameters, and evaluation criteria. In the ablation study, the overall performance by removing one NN component at a time has been estimated. Here, we consider various runs of the proposed TCC, which correspond to different analytical perspectives:

- **Performance of TCC with learning rates**: To analyze the effect of learning rates on the TCC, we initiated the NN with a 0.5 learning rate. When we increased the learning rate, the model's performance, in terms of precision, recall, f-measure and precision decreased.

- **Precision of TCC with optimizers**: To study the effect of using random and grid optimizers on the TCC, grid optimisation was replaced by the random optimizer, and the precision of the model has slightly enhanced.

- **Performance of TCC with hidden layers**: We examined the effect of the hidden layers on the TCC performance. For this analysis, the TCC was trained and evaluated with a cost function, where the cost was reduced and increased with the number of iterations.

To summarise, the full version of TCC using NN achieved better performance with the choice of the tuned learning, random optimizer, and adaptable hidden layers. This performance gain is because the TCC with the linear regression can effectively define the correlation between data features and accurately adopt the most representative features to be encoded into the unified representation using NN, thereby enhancing the performance. Hence, TCC can effectively model campaigns and elicit their thematic characteristics from chosen features' values in the social media data. Overall, we found that the combination of linear regression and NN asserts the best outcome in classifying the campaigns

Comparisons

Overall our testing evaluation results are very similar to the performance achieved by other studies that employed NN type classifiers with social media data. There are several studies that have used these techniques for similar but distinct purposes. The study Kumari *et al.* (2021) employed a Convolutional Neural Network to classify aggressive content in social media, a complicated task with non-English and non-text-based content. Their framework was able to achieve a precision of 74% which also included the employment of a Binary Particle Swarm Optimization algorithm for feature selection.

The study Al-Garadi *et al.* (2021) proposes a text-based classifier to detect non-medical prescription medication use from social media, i.e., using social media data to detect prescription drug abuse in the community. In the performance comparison of BERT, BERT-like, BiLSTM, and Fusion models, Al-Garadi *et al.* (2021) achieved a precision range of 68%–91% across four different data sets. Finally, a method for hateful meme classification was presented in Aggarwal *et al.* (2021). Text was extracted from the meme, then text captioning was employed to convert the image into text, allowing Aggarwal *et al.* (2021) to use text-based features for classification. Three different implementations of LSTM were compared with precision between 55% and 64%.

Overall, our CNA-TCC framework and results are well aligned with recent research using social media data and NN model classifiers. However, these comparisons imply that the aims of these works are the same or strongly related to this, whereas they are using similar techniques for distinct purposes. Our results are difficult to find comparison for, as our research aims are unique.

0.6 Conclusion and Future Work

We developed the framework CNA-TCC framework that employs a NN model instead of heuristics to thematically classify Social Media Influence Campaigns (SMICs) based on network features. We introduced the novel Campaign Network Attribute (CNA) feature extraction as a basis for the Thematic Campaign Classification (TCC). We conducted extensive NN model training, tuning, and testing. Our evaluations quantitatively demonstrated that SMICs can be thematically classified using their CNA signatures. Thus, we demonstrated that the CNA-TCC framework can encode an SMIC and then classify it based on a learned model. This has important implications for search applications used by commercial, political, and national security organisations where heuristics are unavailable. The next stage for this research is to apply the framework to massive data sets with higher dimensions as well as exponential numbers of campaigns to reflect more realistic search and deployment scenarios. This would also include investigating full optimisation of the hyperparameters to increase the classification performance of the modelling.

URL

"Emarsys", (2001).

URL http://xmlns.com/foaf/spec/ (2011).

URL http://sioc-project.org/ (2011).

URL http://vowl.visualdataweb.org/webvowl/old/ (2015).

URL https://www.w3.org/RDF/ (2017).

URL https://www.w3.org/wiki/Good Ontologies (2017).

"'study on media literacy and online empowerment issues raised by algorithm driven media services"', (2017).

URL https://www.trackmyhashtag.com/ (2019).

URL https://www.tweetbinder.com/ (2019).

URL https://developer.twitter.com/en (2019).

URL https://protege.stanford.edu/ (2019).

"Facebook Reports Third Quarter 2019 Results", (2019).

"IAF pilots in Balakot air strikes say op was over in 90 seconds families didn't even know", URL https://theprint.in/ (2019).

"Social media statistics", (2021).

"www.tweetbinder.com", (2021).

Abu-Salih, B., B. Bremie, P. Wongthongtham, K. Duan, T. Issa, K. Y. Chan, M. Alhabashneh, T. Albtoush, S. Alqahtani, A. Alqahtani *et al.*, "Social credibility incorporating semantic analysis and machine learning: A survey of the state-of-the-art and future research directions", in "Proc. Workshops of the Int. Conf. on Adv. Inform. Netw. and Applications", pp. 887–896 (2019).

Agarwal, A. and D. Toshniwal, "Face off: Travel habits, road conditions and traffic city characteristics bared using twitter", IEEE Access **7**, 66536–66552 (2019).

Agency, M. F. B., "Everything you need to know about the ukraine crisis", Online URL https://www.vox.com/2014/9/3/18088560/ukraine-everything-you-need-to-know (2014).

Aggarwal, A., V. Sharma, A. Trivedi, M. Yadav, C. Agrawal, D. Singh, V. Mishra and H. Gritli, “Two-way feature extraction using sequential and multimodal approach for hateful meme classification”, Complexity **2021**, Art. ID 5510253 (2021).

Aiken, L. R., *Psychological Testing and Assessment* (Pearson Education India, New Delhi, 2009).

Akkari, N. and N. Dimitriou, “Mobility management solutions for 5G networks: Architecture and services”, Computer Networks **169**, 107082 (2020).

Al-Garadi, M. A., Y.-C. Yang, H. Cai, Y. Ruan, K. O’Connor, G.-H. Graciela, J. Perrone and A. Sarker, “Text classification models for the automatic detection of nonmedical prescription medication use from social media”, BMC medical informatics and decision making **21**, 1, 1–13 (2021).

Alalewi, A., I. Dayoub and S. Cherkaoui, “On 5G-V2X use cases and enabling technologies: a comprehensive survey”, IEEE Access (2021).

Alduaiji, N., A. Datta and J. Li, “Influence propagation model for clique-based community detection in social networks”, IEEE Transactions on Computational Social Systems **5**, 2, 563–575 (2018).

Alhabash, S. and M. Ma, “A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students?”, Social Media + Society **3**, 1, 1–13 (2017).

Ali, M. S., E. Hossain and D. I. Kim, “LTE/LTE-A random access for massive machine-type communications in smart cities”, IEEE Communications Magazine **55**, 1, 76–83 (2017).

Alizadeh, M., J. N. Shapiro, C. Buntain and J. A. Tucker, “Content-based features predict social media influence operations”, Science advances **6**, 30, eabb5824 (2020).

Alkaff, S. H. B. O. and R. Mahzam, “Islamic state after the fall of Mosul and Raqqa”, Counter Terrorist Trends and Analyses **10**, 1, 57–61 (2018).

Alturki, A., N. Alshwihi and A. Algarni, “Factors influencing players’ susceptibility to social engineering in social gaming networks”, IEEE Access **8**, 97383–97391 (2020).

Altuwairiqi, M., E. Arden-Close, N. Jiang, G. Powell and R. Ali, “Problematic attachment to social media: the psychological states vs usage styles”, in “Proc. IEEE Int. Conf. on Research Challenges in Information Science (RCIS)”, pp. 1–6 (2019).

American Psychological Association, https://dictionary.apa.org/social-influence (1985).

Anderson, E., *Code of the streets – Decency, Violence, and the Moral Life of the Inner City* (W. W. Norton & Company, New York, 1999).

Anderson, K., "Cubs of the caliphate: The systematic recruitment, training, and use of children in the islamic state", International Institute for Counter-Terrorism. Forrás: www. ict. org. il/UserFiles/ICT-Cubs-of-the-Caliphate-Anderson. pdf (2018. 05. 03.) (2016).

Angele, K., D. Fensel, E. Huaman, E. Kärle, O. Panasiuk, U. Şimşek, I. Toma and A. Wahler, *Semantic Web Empowered E-Tourism*, pp. 1–46 (Springer International Publishing, Cham, 2020), URL `"https://doi.org/10.1007/978-3-030-05324-6`$_2 2-1$".

Arora, A., S. Bansal, C. Kandpal, R. Aswani and Y. Dwivedi, "Measuring social media influencer index–Insights from Facebook, Twitter and Instagram", Journal of Retailing and Consumer Services **49**, 86–101 (2019a).

Arora, A., S. Bansal, C. Kandpal, R. Aswani and Y. Dwivedi, "Measuring social media influencer index-insights from facebook, twitter and instagram", Journal of Retailing and Consumer Services **49**, 86–101 (2019b).

Assenmacher, D., L. Clever, J. S. Pohl, H. Trautmann and C. Grimme, "A two-phase framework for detecting manipulation campaigns in social media", in "International Conference on Human-Computer Interaction", pp. 201–214 (Springer, 2020).

Awan, I., "Cyber-extremism: ISIS and the power of social media", Society **54**, 138–149 (2017).

Balaji, T., C. S. R. Annavarapu and A. Bablani, "Machine learning algorithms for social media analysis: A survey", Computer Science Review **40**, 100395 (2021).

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker and R. Bonneau, "Tweeting from left to right: Is online political communication more than an echo chamber?", Psychological Science **26**, 10, 1531–1542 (2015).

Bayrakdar, S., I. Yucedag, M. Simsek and I. A. Dogru, "Semantic analysis on social networks: A survey", Int. Journal of Communication Systems **33**, 11, e4424.1–e4424.30 (2020).

Belhadi, A., Y. Djenouri, J. C. Lin and A. Cano, "A data-driven approach for twitter hashtag recommendation", IEEE Access **8**, 79182–79191 (2020).

Benzaïd, C. and T. Taleb, "AI for beyond 5G networks: a cyber-security defense or offense enabler?", IEEE Network **34**, 6, 140–147 (2020).

Bindu, V. and C. Thomas, "Knowledge base representation of emails using ontology for spam filtering", in "Advances in Artificial Intelligence and Data Engineering", edited by N. N. Chiplunkar and T. Fukao, pp. 723–735 (Springer Singapore, Singapore, 2021).

Bischof, S., A. Karapantelakis, C. Nechifor, A. P. Sheth, A. Mileo and P. Barnaghi, "Semantic modelling of smart city data", URL `https://corescholar.libraries.wright.edu/knoesis/572` (2014).

Boer, M., G. W. Stevens, C. Finkenauer, M. E. de Looze and R. J. van den Eijnden, "Social media use intensity, social media use problems, and mental health among adolescents: Investigating directionality and mediating processes", Computers in Human Behavior **116**, 106645 (2021).

Bonacchi, C. and M. Krzyzanska, "Digital heritage research re-theorised: ontologies and epistemologies in a world of big data", Int. Journal of Heritage Studies **25**, 12, 1235–1247 (2019).

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization", Nature **489**, 7415, 295–298 (2012).

Borden, J., X. A. Zhang and J. Hwang, "Improving automated crisis detection via an improved understanding of crisis language: Linguistic categories in social media crises", Journal of Contingencies and Crisis Management **28**, 3, 281–290 (2020).

Bose, A., V. Behzadan, C. Aguirre and W. H. Hsu, "A novel approach for detection and ranking of trendy and emerging cyber threat events in twitter streams", in "Proc. IEEE/ACM Int. Conf. on Adv. in Social Netw. Analysis and Mining (ASONAM)", pp. 871–878 (2019).

Boyd, D. M. and N. B. Ellison, "Social network sites: Definition, history, and scholarship", Journal of Computer-Mediated Communication **13**, 1, 210–230 (2007).

Brangetto, P. and M. A. Veenendaal, "Influence cyber operations: The use of cyberattacks in support of influence operations", in "Proc. Int. Conf. on Cyber Conflict (CyCon)", pp. 113–126 (2016).

Breuker, J., A. Valente and R. Winkels, "Legal ontologies in knowledge engineering and information management", Artificial Intelligence and Law **12**, 4, 241–277 (2004).

Britannica, E., "Ukraine crisis", Online URL `https://www.britannica.com` (2014).

Brodaric, B., T. Hahmann and M. Gruninger, "Water features and their parts", Journal of Applied Ontologies **14**, 1, 1–42 (2019).

Bromander, S., A. Jøsang and M. Eian, "Semantic cyberthreat modelling", in "Proc. STIDS", pp. 74–78 (2016).

Brown, J. D., "What is construct validity", Available at jalt. org/test/bro_8. htm (accessed 10 June 2013) (2000).

Cadwalladr, C., "I created Steve Bannon's psychological warfare tool: Meet the data war whistleblower", The Guardian. Retrieved from http://www. theguardian. com/news/2018/mar/17/data-war-whistleblower-christopherwylie-faceook-nix-bannon-trump (2018).

Campbell, J., “Ex Machina: Technological Disruption and the Future of Artificial Intelligence in Legal Writing, University of Denver Sturm College of Law, Legal Research Paper Series, Working Paper No. 20-04”, (2020).

Carlson, M., “Fake news as an informational moral panic: The symbolic deviancy of social media during the 2016 us presidential election”, Information, Communication & Society **23**, 3, 374–388 (2020).

Carr, C. T. and R. A. Hayes, “Social media: Defining, developing, and divining”, Atlantic Journal of Communication **23**, 1, 46–65 (2015).

Castillo, C., M. Mendoza and B. Poblete, “Information credibility on twitter”, in “Proceedings of the 20th international conference on World wide web”, pp. 675–684 (2011).

Chakraborty, A., T. Dutta, S. Mondal and A. Nath, “Application of graph theory in social media”, International Journal of Computer Sciences and Engineering **6**, 10, 722–729 (2018).

Chen, M., S. Mao and Y. Liu, “Big data: A survey”, Mobile Networks and Applications **19**, 171–209 (2014).

Chen, S., X. Yin, Q. Cao, Q. Li and H. Long, “Targeted influence maximization based on cloud computing over big data in social networks”, IEEE Access **8**, 45512–45522 (2020).

Cheng, X., S. Fu and G.-J. de Vreede, “Understanding trust influencing factors in social media communication: A qualitative study”, International Journal of Information Management **37**, 2, 25–35 (2017).

Chu, Z., I. Widjaja and H. Wang, “Detecting social spam campaigns on twitter”, in “International Conference on Applied Cryptography and Network Security”, pp. 455–472 (Springer, 2012).

Chung, W. and D. Zeng, “Dissecting emotion and user influence in social media communities: An interaction modeling approach”, Information & Management **57**, 1, 103108 (2020).

Cialdini, R. B. and N. J. Goldstein, “Social influence: Compliance and conformity”, Annual Review of Psychology **55**, 1, 591–621 (2004).

Cisco Inc., “Cisco Annual Internet Report (2018–2023) White Paper”, URL `https://www.cisco.com/c/en/us/solutions/collateral.html` (2020).

Cohen, S. B., URL `https://www.theguardian.com/technology/2019/nov/22` (2019).

Coleman, J., E. Katz and H. Menzel, “The diffusion of an innovation among physicians”, Sociometry **20**, 4, 253–270 (1957).

Cresci, S., M. Petrocchi, A. Spognardi and S. Tognazzi, "On the capability of evolved spambots to evade detection via genetic engineering", Online Social Networks and Media **9**, 1–16 (2019).

Decker, S. H. and D. C. Pyrooz, "Gangs: Another form of organized crime?", in "Oxford Handbook of Organized Crime", edited by L. Paoli, pp. 1–9 (Oxford University Press, Oxford, UK, 2013).

Deepa, M. D. *et al.*, "Bidirectional encoder representations from transformers (bert) language model for sentiment analysis task", Turkish Journal of Computer and Mathematics Education (TURCOMAT) **12**, 7, 1708–1721 (2021).

Deparis, E., M. Abel and J. Mattioli, "Modeling a social collaborative platform with standard ontologies", in "Proc. IEEE Int. Conf. on Signal Image Techn. Internet-Based Systems", pp. 167–173 (2011).

Depoux, A., S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith and H. Larson, "The pandemic of social media panic travels faster than the COVID-19 outbreak", Journal of Travel Medicine **27**, 3 (2020).

Deutsch, M. and H. B. Gerard, "A study of normative and informational social influences upon individual judgement", Journal of Abnormal Psychology **51**, 3, 629–636 (1955).

Dhiman, A. and D. Toshniwal, "An approximate model for event detection from twitter data", IEEE Access **8**, 122168–122184 (2020).

Dhiraj, B., A, "The 20 top most used social networking sites and apps in the world, 2019", https://ceoworld.biz/2019/03/03/the-20-top-most-used-social-networking-sites-and-apps-inthe-world-2019 (2019).

Dimock, M., "Defining generations: Where Millennials end and Generation Z begins", (????).

Doan, T. V., G. T. Nguyen, M. Reisslein and F. H. Fitzek, "Fast: Flexible and low-latency state transfer in mobile edge computing", IEEE Access **9**, 115315–115334 (2021).

Dollarhide, M., "Social media definition", Investopedia. Available online: http://billscomputerpot. com/menus/windows/SocialMedia. pdf (accessed on 20 July 2020) (2019).

Dou, M., Y. Gu and G. Xu, "Social awareness of crisis events: A new perspective from social-physical network", Cities **99**, 102620 (2020).

Dubois, E., S. Minaeian, A. Paquet-Labelle and S. Beaudry, "Who to trust on social media: How opinion leaders and seekers avoid disinformation and echo chambers", Social Media + Society **6**, 2, 2056305120913993 (2020).

Durr, B., *ISIS: The use of Social Media*, Ph.D. thesis, Utica College, NY (2016).

Eden, C., “Cognitive mapping”, European Journal of Operational Research **36**, 1, 1–13 (1988).

Edosomwan, S., S. K. Prakasan, D. Kouame, J. Watson and T. Seymour, “The history of social media and its impact on business”, Journal of Applied Management and Entrepreneurship **16**, 3, 79–91 (2011).

El Asikri, M., J. Laassiri, S. Krit and H. Chaib, “Contribution to ontologies building using the semantic web and web mining”, in “Proc. Int. Conf. on Engineering MIS (ICEMIS)”, pp. 1–5 (2016).

El Kassiri, A. and F.-Z. Belouadha, “A FOAF ontology extension to meet online social networks presentation and analysis”, in “Proc. IEEE Int. Conf. on Power, Control, Signals and Instrumentation Eng. (ICPCSI)”, pp. 3056–3061 (2017).

Eltantawy, N. and J. B. Wiest, “The Arab spring— social media in the Egyptian revolution: reconsidering resource mobilization theory”, International Journal of Communication **5**, 1–18 (2011).

Fang, M., Y. Li, Y. Hu, S. Mao and P. Shi, “A unified semantic model for cross-media events analysis in online social networks”, IEEE Access **7**, 32166–32182 (2019).

Feldkamp, J., “The rise of TikTok: The evolution of a social media platform during COVID-19”, in “Digital Responses to Covid-19: Digital Innovation, Transformation, and Entrepreneurship During Pandemic Outbreaks”, pp. 73–85 (Springer International Publishing, New York, 2021).

Ferrara, E., O. Varol, C. Davis, F. Menczer and A. Flammini, “The rise of social bots”, Commun. ACM **59**, 7, 96–104 (2016).

Fogg, B., “Persuasive computers: Perspectives and research directions”, in “Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems”, pp. 225–232 (1998).

Fogg, B. J., “Mass interpersonal persuasion: An early view of a new phenomenon”, in “Proc. Int. Conf. on Persuasive Techn.”, pp. 23–34 (2008).

Fogg, B. J. and H. Tseng, “The elements of computer credibility”, in “Proceedings of the SIGCHI conference on Human Factors in Computing Systems”, pp. 80–87 (1999).

Garmazhapova, A., “Where do trolls live. and who feeds them, Novaya Gazeta”, (2013).

Gefen, D., E. Karahanna and D. W. Straub, “Trust and tam in online shopping: An integrated model”, MIS Quarterly **27**, 1, 51–90 (2003).

Gerasimov, V., “The ’Gerasimov Doctrine’ and Russian non-linear war”, (2013).

Glesne, C., *Becoming Qualitative Researchers: An Introduction* (ERIC, 2016).

Grainger, “Broadcasting, co-regulation and the public good 1999 spry memorial lecture”, (1999).

Gray, C. H. and Á. J. Gordo, “Social media in conflict: Comparing military and social-movement technocultures”, Cultural Politics **10**, 3, 251–261 (2014).

Gregory, P. R., “Putin’s New Weapon In The Ukraine Propaganda War: Internet Trolls, Forbes.com”, (2014).

Guadagno, R. E. and R. B. Cialdini, “Preference for consistency and social influence: A review of current research findings”, Social Influence **5**, 3, 152–163 (2010).

Gundecha, P. and H. Liu, “Mining social media: a brief introduction”, New directions in informatics, optimization, logistics, and production pp. 1–17 (2012).

Gunning, D., M. Stefik, J. Choi, T. Miller, S. Stumpf and G.-Z. Yang, “Xai—explainable artificial intelligence”, Science Robotics **4**, 37 (2019).

Hair, J. F., J. J. Risher, M. Sarstedt and C. M. Ringle, “When to use and how to report the results of pls-sem”, European business review (2019).

Halawi, B., A. Mourad, H. Otrok and E. Damiani, “Few are as good as many: An ontology-based tweet spam detection approach”, IEEE Access **6**, 63890–63904 (2018).

Han, S. N., G. M. Lee and N. Crespi, “Towards automated service composition using policy ontology in building automation system”, in “Proc. IEEE Int. Conf. on Services Computing”, pp. 685–686 (2012).

Harakawa, R., S. Takimura, T. Ogawa, M. Haseyama and M. Iwahashi, “Consensus clustering of tweet networks via semantic and sentiment similarity estimation”, IEEE Access **7**, 116207–116217 (2019).

Hart, S. W. and M. C. Klink, “1st Troll Battalion: Influencing military and strategic operations through cyber-personas”, in “Proc. IEEE Int. Conf. on Cyber Conflict (CyCon US)”, pp. 97–104 (2017).

He, W., F.-K. Wang and V. Akula, “Managing extracted knowledge from big social media data for business decision making”, Journal of Knowledge Management **21**, 2 (2017).

He, Y., C. Wang and C. Jiang, “Mining coherent topics with pre-learned interest knowledge in twitter”, IEEE Access **5**, 10515–10525 (2017).

Heidari, M., H. James Jr and O. Uzuner, “An empirical study of machine learning algorithms for social media bot detection”, in “2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)”, pp. 1–5 (IEEE, 2021).

Heidari, M. and J. H. Jones, “Using bert to extract topic-independent sentiment features for social media bot detection”, in “2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)”, pp. 0542–0547 (IEEE, 2020).

Herrick, D., “The social side of ’cyber power’? Social media and cyber operations”, in “Proc. Int. Conf. on Cyber Conflict (CyCon)”, pp. 99–111 (2016).

Hoekstra, R., J. Breuker, M. Di Bello and A. Boer, “The LKIF core ontology of basic legal concepts”, in “Proceedings of LOAIT, II Workshop on Legal Ontologies and Artificial Intelligence Techniques”, pp. 43–63 (2007).

Hoeschele, T., C. Dietzel, D. Kopp, F. H. Fitzek and M. Reisslein, “Importance of internet exchange point (ixp) infrastructure for 5g: Estimating the impact of 5g use cases”, Telecommunications Policy **45**, 3, 102091 (2021).

Horwood, S. and J. Anglim, “Self and other ratings of problematic smartphone use: The role of personality and relationship type”, Computers in Human Behavior **116**, 106634 (2021).

Hui, P.-M., K.-C. Yang, C. Torres-Lugo and F. Menczer, “Botslayer: Diy real-time influence campaign detection”, in “Proceedings of the International AAAI Conference on Web and Social Media”, vol. 14, pp. 980–982 (2020).

Hutchinson, A., “How TikTok’s algorithm works - as explained by TikTok”, https://www.socialmediatoday.com/news/how-tiktoks-algorithm-works-as-explained-by-tiktok/580153/ (June 18, 2020).

Jackson, M. O., *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2010).

Jamali, M., A. Nejat, S. Moradi, S. Ghosh, G. Cao and F. Jin, “Social media data and housing recovery following extreme natural hazards”, International Journal of Disaster Risk Reduction **51**, 101788 (2020).

Jennings, W., G. Stoker, H. Willis, V. Valgardsson, J. Gaskell, D. Devine, L. Mckay and M. C. Mills, “Lack of trust and social media echo chambers predict COVID-19 vaccine hesitancy”, medRxiv (2021).

Jin, X., “Exploring crisis communication and information dissemination on social media: Social network analysis of hurricane irma tweets”, Journal of International Crisis and Risk Communication Research **3**, 2, 179–210 (2020).

Johnson, N., B. Turnbull, T. Maher and M. Reisslein, “Semantically modeling cyber influence campaigns (CICs): Ontology model and case studies”, IEEE Access **9**, 9365–9382 (2020).

Johnson, N., B. Turnbull and M. Reisslein, “Social media influence, trust, and conflict: An interview based study of leadership perceptions”, Technology in Society **68**, 101836 (2022a).

Johnson, N., B. Turnbull, M. Reisslein and M. Nour, “CNA-TCC: Campaign network attribute based thematic campaign classification”, IEEE Transactions on Computational Social Systems, in print (2022b).

Kafeza, E., C. Makris and G. Rompolas, "Exploiting time series analysis in twitter to measure a campaign process performance", in "2017 IEEE international conference on Services Computing (SCC)", pp. 68–75 (IEEE, 2017).

Kalemi, E. and E. Martiri, "FOAF-academic ontology: A vocabulary for the academic community", in "Proc. IEEE Int. Conf. on Intelligent Netw. and Collaborative Systems", pp. 440–445 (2011).

Kane, B., *One Million Followers: How I built a massive social following in 30 days* (BenBella Books, Dallas, TX, 2018).

Karami, A., M. Lundy, F. Webb and Y. K. Dwivedi, "Twitter and research: A systematic literature review through text mining", IEEE Access **8**, 67698–67717 (2020).

Karlsen, R. and B. Enjolras, "Styles of social media campaigning and influence in a hybrid political communication system: Linking candidate survey data with twitter data", The International Journal of Press/Politics **21**, 3, 338–357 (2016).

Kaura, V., "India's Pakistan policy: from 2016 'surgical strike' to 2019 Balakot 'airstrike"', The Round Table **109**, 3, 277–287 (2020).

Kemp, S., "Digital trends 2020: Every single stat you need to know about the internet", URL `https://thenextweb.com/growth-quarters/2020/01/30/` (2019).

Kent, M. L. and C. Li, "Toward a normative social media theory for public relations", Public Relations Review **46**, 1, 101857 (2020).

Ketter, E. and E. Avraham, "The social revolution of place marketing: The growing power of users in social media campaigns", Place Branding and Public Diplomacy **8**, 4, 285–294 (2012).

Khan, S., R. Umer, S. Umer and S. Naqvi, "Antecedents of trust in using social media for E-government services: An empirical study in Pakistan", Technology in Society **64**, 101400 (2021).

Kilcullen, D., "Counter-insurgency redux", Survival **48**, 4, 111–130 (2006).

Kim, D. J., D. L. Ferrin and H. R. Rao, "A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents", Decision Support Systems **44**, 2, 544–564 (2008).

Köppen, M., "The curse of dimensionality", in "5th Online World Conference on Soft Computing in Industrial Applications (WSC5)", vol. 1, pp. 4–8 (2000).

Kuang, G. and Y. Du, "Building semantic business web services based on ontology", in "Proc. Int. Conf. on New Trends in Information Science and Service Science", pp. 634–637 (2010).

Kumar, S., J. Cheng, J. Leskovec and V. Subrahmanian, "An army of me: Sockpuppets in online discussion communities", in "Proc. Int. Conf. on World Wide Web", pp. 857–866 (2017).

Kumari, K., J. P. Singh, Y. K. Dwivedi and N. P. Rana, “Multi-modal aggression identification using convolutional neural network and binary particle swarm optimization”, Future Generation Computer Systems **118**, 187–197 (2021).

Kursuncu, U., M. Gaur, C. Castillo, A. Alambo, K. Thirunarayan, V. Shalin, D. Achilov, I. B. Arpinar and A. Sheth, “Modeling islamist extremist communications on social media using contextual dimensions: Religion, ideology, and hate”, Proc. ACM on Human-Computer Interaction pp. 1–22 (2019).

Lazarsfeld, P. F. and R. K. Merton, “Friendship as a social process: A substantive and methodological analysis”, Freedom and Control in Modern Society **18**, 1, 18–66 (1954).

Lazer, D. M. J., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts and J. L. Zittrain, “The science of fake news”, Science **359**, 6380, 1094–1096 (2018).

Lee, K., J. Caverlee, Z. Cheng and D. Z. Sui, “Content-driven detection of campaigns in social media”, in “Proceedings of the 20th ACM international conference on Information and knowledge management”, pp. 551–556 (2011).

Lee, K., J. Caverlee, Z. Cheng and D. Z. Sui, “Campaign extraction from social media”, ACM Transactions on Intelligent Systems and Technology (TIST) **5**, 1, 1–28 (2014).

Lee, O., H. L. Nguyen, J. E. Jung, T. Um and H. Lee, “Towards ontological approach on trust-aware ambient services”, IEEE Access **5**, 1589–1599 (2017).

Lee, S., S. Lee and H. Baek, “Does the dispersion of online review ratings affect review helpfulness?”, Computers in Human Behavior **117**, 106670 (2021).

Li, D., H. Chaudhary and Z. Zhang, “Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining”, International Journal of Environmental Research and Public Health **17**, 14, 4988 (2020).

Li, J. and C. Cardie, “Early stage influenza detection from twitter”, arXiv preprint arXiv:1309.7340 (2013).

Linguaglossa, L., S. Lange, S. Pontarelli, G. Rétvári, D. Rossi, T. Zinner, R. Bifulco, M. Jarschel and G. Bianchi, “Survey of performance acceleration techniques for network function virtualization”, Proceedings of the IEEE **107**, 4, 746–764 (2019).

Lundberg, S. M. and S.-I. Lee, “A unified approach to interpreting model predictions”, in “Proceedings of the 31st international conference on neural information processing systems”, pp. 4768–4777 (2017).

Ma, Z., A. Sun and G. Cong, “On predicting the popularity of newly emerging hashtags in t witter”, Journal of the American Society for Information Science and Technology **64**, 7, 1399–1410 (2013).

MacFarquhar, N., "Inside the russian troll factory: Zombies and a breakneck pace", The New York Times (2018).

Maheshwari, L., "India's February 2019 strike in Pakistani territory: A Jus ad Bellum analysis", Https://www.lawfareblog.com (2019).

Maistrelli, A., *Palestinian Transnational Advocacy Network: A comparative analysis of the three most powerful network actors and their discursive practices on Twitter*, Ph.D. thesis, Bournemouth University, Poole, England (2019).

Masolo, C., A. B. Benevides and D. Porello, "The interplay between models and observations", Applied Ontology **13**, 41–71 (2018).

McLennan, D. and J. Miles, "A once unimaginable scenario: No more newspapers", The Washington Post **21** (2018).

McQuail, D. and S. Windahl, *Communication models for the study of mass communications* (Routledge, Milton Park, UK, 2015).

Mejias, U. and N. Vokuev, "Disinformation and the media: The case of Russia and Ukraine", Media, Culture and Society **39**, 7, 1027–1042 (2017).

Michel, M. C. K. and M. C. King, "Cyber influence of human behavior: Personal and national security, privacy, and fraud awareness to prevent harm", in "Proc. IEEE Int. Symp. on Techn. and Soc. (ISTAS)", pp. 1–7 (2019).

Miller, A., "To win the hearts and minds: The combined action program during the Vietnam war", Https://ir.vanderbilt.edu/handle/1803/9474, Last accessed July 21, 2020 (2019).

Mullah, N. S. and W. M. N. W. Zainon, "Advances in machine learning algorithms for hate speech detection in social media: A review", IEEE Access (2021).

Nie, X., W. Zhang, Y. Zhang and D. Yu, "Method to predict bursty hot events on twitter based on user relationship network", IEEE Access **8**, 44031–44040 (2020).

Ning, Z., Y. Liu, J. Zhang and X. Wang, "Rising star forecasting based on social network analysis", IEEE Access **5**, 24229–24238 (2017).

Novak, T., "The rise of TikTok and understanding its parent company ByteDance", https://turner.substack.com/p/the-rise-of-tiktok-and-understanding (May 20, 2020).

Nusratullah, K., S. A. Khan, A. Shah and W. H. Butt, "Detecting changes in context using time series analysis of social network", in "2015 SAI Intelligent Systems Conference (IntelliSys)", pp. 996–1001 (IEEE, 2015).

Oinas-Kukkonen, H. and M. Harjumaa, "A systematic framework for designing and evaluating persuasive systems", in "Proc. Int. Conf. on Persuasive Technology", pp. 164–176 (2008).

Olca, E. and O. Can, “Extending FOAF and relationship ontologies with consent ontology”, in “Proc. IEEE Int. Conf. on Computer Science and Eng. (UBMK)”, pp. 542–546 (2018).

Oluchukwu, O. P., “The use of social media in resolving marital conflict”, Social Sciences & Psychology **69**, 1, 1–12 (2021).

Papachristos, A. V., D. M. Hureau and A. A. Braga, “The corner and the crew: The influence of geography and social networks on gang violence”, American Sociological Review **78**, 3, 417–447 (2013).

Patton, D. U., J. Lane, P. Leonard, J. Macbeth and J. R. S. Lee, “Gang violence on the digital street: Case study of a south side chicago gang member’s twitter communication”, New Media & Society **19**, 7, 1000–1018 (2017).

Penas, P., R. del Hoyo, J. Vea-Murguia, C. Gonzalez and S. Mayo, “Collective knowledge ontology user profiling for twitter – automatic user profiling”, in “Proc. IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence (WI) and Intelligent Agent Techn. (IAT)”, pp. 439–444 (2013).

Perešin, A., “Fatal attraction: Western muslimas and ISIS”, Perspectives on Terrorism **9**, 3, 21–38 (2015).

Philip, S. A., “Inside story of attack on balakot – from IAF officer who planned and executed it”, https://theprint.in/ (2020).

Pitas, I., *Graph-based social media analysis*, vol. 39 (CRC Press, 2016).

Qin, L., Q. Sun, Y. Wang, K.-F. Wu, M. Chen, B.-C. Shia and S.-Y. Wu, “Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index”, International Journal of Environmental Research and Public Health **17**, 7, 2365 (2020).

Quan-Haase, A. and A. L. Young, “Uses and gratifications of social media: A comparison of facebook and instant messaging”, Bulletin of Science, Technology & Society **30**, 5, 350–361 (2010).

Rauta, V., “Towards a typology of non-state actors in ’hybrid warfare’: Proxy, auxiliary, surrogate and affiliated forces”, Cambridge Review of International Affairs, in print pp. 1–20 (2020).

Ray, L., “Google’s algorithm undergoes enormous shifts to prioritise content focused on expertise, authoritativeness and trustworthiness”, Journal of Digital & Social Media Marketing **7**, 2, 128–136 (2019).

Razis, G. and I. Anagnostopoulos, “Semantifying twitter: The influence tracker ontology”, in “Proc. IEEE Int. Workshop on Semantic and Social Media Adaptation and Personalization”, pp. 98–103 (2014).

Razis, G., I. Anagnostopoulos and S. Zeadally, "Modeling influence with semantics in social networks: A survey", ACM Computing Surveys (CSUR) **53**, 1, 7.1–7.38 (2020).

Redondo, R. P. D., C. Garcia-Rubio, A. F. Vilas, C. Campo and A. Rodriguez-Carrion, "A hybrid analysis of lbsn data to early detect anomalies in crowd dynamics", Future Generation Computer Systems **109**, 83–94 (2020).

Resende de Mendonça, R. R. d., D. F. d. Felix de Brito, F. d. F. de Franco Rosa, J. C. dos Reis and R. Bonacin, "A framework for detecting intentions of criminal acts in social media: A case study on twitter", Information **11**, 3, 154.1–154.40 (2020).

Ribeiro, M. T., S. Singh and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier", in "Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining", pp. 1135–1144 (2016).

Riesco, R. and V. A. Villagrá, "Leveraging cyber threat intelligence for a dynamic risk framework", Int. J. Information Security **18**, 6, 715–739 (2019).

Romeo, A. V., S. M. Edney, R. C. Plotnikoff, T. Olds, C. Vandelanotte, J. Ryan, R. Curtis and C. A. Maher, "Examining social-cognitive theory constructs as mediators of behaviour change in the active team smartphone physical activity program: a mediation analysis", BMC Public Health **21**, 1, 1–11 (2021).

Rosyiq, A., A. R. Hayah, A. N. Hidayanto, M. Naisuty, A. Suhanto and N. F. A. Budi, "Information extraction from twitter using DBpedia ontology: Indonesia tourism places", in "Proc. IEEE Int. Conf. on Informatics, Multimedia, Cyber and Information System (ICIMCIS)", pp. 91–96 (2019).

Saif, H., T. Dickinson, L. Kastler, M. Fernandez and H. Alani, "A semantic graph-based approach for radicalisation detection on social media", in "European semantic web conference", pp. 571–587 (Springer, 2017).

Sakaki, T., M. Okazaki and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors", in "Proc. of the 19th Int. Conf. on World Wide Web", pp. 851–860 (2010).

Saldana, J., *The Coding Manual for Qualitative Researchers, 2nd edition* (Sage Publications, New York, 2013).

Salloum, S. A., N. M. N. AlAhbabi, M. Habes, A. Aburayya and I. Akour, "Predicting the intention to use social media sites: a hybrid sem-machine learning approach", in "International Conference on Advanced Machine Learning Technologies and Applications", pp. 324–334 (Springer, 2021).

Sara, A. and H. Mostafa, "Exploring persuasive systems using comparative study between actual technologies", in "Proc. Int. Conf. on Big Data and Smart Digital Environment", pp. 369–379 (Springer, 2018).

Sarwar, R., A. Zia, R. Nawaz, A. Fayoumi, N. R. Aljohani and S.-U. Hassan, "Webometrics: Evolution of social media presence of universities", Scientometrics **126**, 2, 951–967 (2021).

Schaner, E. X., "What is military information power?", Marine Corps Gazette **104**, 4, 17–19 (2020).

Schirch, L., *Social Media Impacts on Conflict and Democracy: The Techtonic Shift* (Routledge, Milton Park, UK, 2021).

Schwär, H., "Tiktok breaks all download records - thanks to boredom in isolation", https://www.businessinsider.de/gruenderszene/media/tiktok-download-rekord-corona/?interstitial (20 Apr 2020).

Sediyono, E., Suhartono and C. Nivak, "Measuring the performance of ontological based information retrieval from a social media", in "Proc. IEEE European Modelling Symp.", pp. 354–359 (2014).

Shahi, G. K., "Amused: an annotation framework of multi-modal social media data", arXiv preprint arXiv:2010.00502 (2020).

Shane, S., "The fake americans russia created to influence the election", The New York Times **7**, 09 (2017).

Shantharama, P., A. S. Thyagaturu and M. Reisslein, "Hardware-accelerated platforms and infrastructures for network functions: A survey of enabling technologies and research studies", IEEE Access **8**, 132021–132085 (2020).

Sharma, S. and A. Jain, "Cyber social media analytics and issues: a pragmatic approach for twitter sentiment analysis", in "Advances in Computer Communication and Computational Sciences", pp. 473–484 (Springer, Singapore, 2019).

Shi, K., C. Gong, H. Lu, Y. Zhu and Z. Niu, "Wide-grained capsule network with sentence-level feature to detect meteorological event in social network", Future Generation Computer Systems **102**, 323–332 (2020).

Shibuya, Y. and H. Tanaka, "Using social media to detect socio-economic disaster recovery", IEEE Intelligent Systems **34**, 3, 29–37 (2019).

Shumanov, M. and L. Johnson, "Making conversations with chatbots more personalized", Computers in Human Behavior **117**, 106627 (2021).

Simran, K., P. Balakrishna, R. Vinayakumar and K. Soman, "Deep learning approach for enhanced cyber threat indicators in twitter stream", in "Proc. Int. Symp. on Security in Computing and Commun.", pp. 135–145 (2019).

Singer, P. and E. Brooking, *LikeWar: The Weaponization of Social Media* (Houghton Mifflin Harcourt Books, Rancho Cucamonga, CA, 2018a).

Singer, P. W. and E. T. Brooking, *LikeWar: The Weaponization of Social Media* (Eamon Dolan Books, New York, 2018b).

Speckhard, A., *Bride of ISIS* (Advances Press LLC, 2015).

Stengel, R., *Information Wars: How We Lost the Global Battle Against Disinformation and What We Can Do about It* (Grove/Atlantic, Incorporated, New York, 2019).

Storrod, M. and J. Densley, "'Going Viral' and 'Going Country': The expressive and instrumental activities of street gangs on social media", Journal of Youth Studies **20**, 6, 677–696 (2017).

Strudwicke, I. J. and W. J. Grant, "# junkscience: Investigating pseudoscience disinformation in the russian internet research agency tweets", Public Understanding of Science p. 0963662520935071 (2020).

Syed, R., "Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system", Information & Management **57**, 6, 103334.1–103334.17 (2020).

Taleb, T., K. Samdanis, B. Mada, H. Flinck, S. Dutta and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration", IEEE Communications Surveys & Tutorials **19**, 3, 1657–1681 (2017).

Tench, R. and B. Jones, "Social media: The Wild West of CSR communications", Social Responsibility Journal **11**, 2, 290–305 (2013).

Thomson, K.-L., "Information security obedience: A definition", Computers and Security **24**, 69–75 (2005).

Thyagaturu, A. S., Y. Dashti and M. Reisslein, "SDN-based smart gateways (Sm-GWs) for multi-operator small cell network management", IEEE Transactions on Network and Service Management **13**, 4, 740–753 (2016).

Tracker, G. C., "Ukraine crisis: Timeline", Online URL `https://www.bbc.com/news/world-middle-east-26248275` (2014).

Tyagi, R. R., F. Aurzada, K.-D. Lee, S. G. Kim and M. Reisslein, "Impact of retransmission limit on preamble contention in LTE-advanced network", IEEE Systems Journal **9**, 3, 752–765 (2013).

Tyagi, R. R., F. Aurzada, K.-D. Lee and M. Reisslein, "Connection establishment in LTE-A networks: Justification of Poisson process modeling", IEEE Systems Journal **11**, 4, 2383–2394 (2015).

Van Noort, G., M. L. Antheunis and E. A. Van Reijmersdal, "Social connections and the persuasiveness of viral campaigns in social network sites: Persuasive intent as the underlying mechanism", Journal of Marketing Communications **18**, 1, 39–53 (2012).

Ver Steeg, G. and A. Galstyan, "Information transfer in social media", in "Proceedings of the 21st international conference on World Wide Web", pp. 509–518 (2012).

Vyas, P., M. Reisslein, B. P. Rimal, G. Vyas, G. P. Basyal and P. Muzumdar, "Automated classification of societal sentiments on Twitter with machine learning", IEEE Transactions on Technology and Society, in print (2021).

Wang, M., Q. Xu and N. He, "Perceived interparental conflict and problematic social media use among Chinese adolescents: The mediating roles of self-esteem and maladaptive cognition toward social network sites", Addictive Behaviors **112**, 106601 (2021).

Watts, D. J. and P. S. Dodds., "Influentials, networks, and public opinion formation", Journal of Consumer Research **34**, 441–458 (2007).

Weibel, S. L. and T. Koch, "The Dublin core metadata initiative", D-lib Magazine **6**, 12, 1082–9873 (2000).

Weimann, G., "The Influentals: Back to the concept of opinion leaders?", Public Opinion Quarterly **55**, 2, 267–279 (1991).

Weitzen, J., N. Sutter, R. Wakim, A. Alkhatabih, W. Wang, Y. Tian, C. Ahn, S. Mylonakis and H. Eghbali, "First measurements of cloud-RAN LTE smallcells in an indoor stadium", Journal of Selected Areas in Telecommunications **6** (2016).

Williams, G., *Chaos Theory Tamed* (CRC Press, Boca Raton, FL, 1997).

Wu, L., F. Morstatter, K. M. Carley and H. Liu, "Misinformation in social media: definition, manipulation, and detection", ACM SIGKDD Explorations Newsletter **21**, 2, 80–90 (2019).

Wu, L., F. Morstatter, X. Hu and H. Liu, "Mining misinformation in social media", in "Big data in complex and social networks", pp. 135–162 (Chapman and Hall/CRC, 2016).

Wylie, C., *Mindfuck: le complot Cambridge Analytica pour s' emparer de nos cerveaux* (Grasset, 2020).

Xie, Y., M. Chen, H. Lai, W. Zhang, Z. Zhao and C. Anwar, "Neural basis of two kinds of social influence: Obedience and conformity", Frontiers in Human Neuroscience **10**, 51.1–51.8 (2016).

Xu, X., B. Shen, X. Yin, M. R. Khosravi, H. Wu, L. Qi and S. Wan, "Edge server quantification and placement for offloading social media services in industrial cognitive IoV", IEEE Transactions on Industrial Informatics, in print (2020).

YouTube, "Youtube", (2005).

Yu, D., D. Xu, D. Wang and Z. Ni, "Hierarchical topic modeling of twitter data for online analytical processing", IEEE Access **7**, 12373–12385 (2019).

Zannettou, S., T. Caulfield, W. Setzer, M. Sirivianos, G. Stringhini and J. Blackburn, "Who let the trolls out? Towards understanding state-sponsored trolls", in "Proc. ACM Conf. on Web Science", pp. 353–362 (2019).

Zeitzoff, T., "How social media is changing conflict", Journal of Conflict Resolution **61**, 9, 1970–1991 (2017).

Zou, Y., X. Yao, Z. Chen and M. Zhao, "Verifiable keyword-based semantic similarity search on social data outsourcing", IEEE Access **7**, 5616–5625 (2019).