# AN INTELLIGENCE-BASED APPROACH TO COUNTERING SOCIAL MEDIA INFLUENCE OPERATIONS

**Matteo E. BONFANTI***

**Abstract**

*The paper describes an intelligence-based approach to identify, understand and counter influence operations and campaigns that are sponsored, directed, or conducted by State or non-state actors on/through social media platforms. It examines the extent to which the collection and analysis of social media data and further contextual information can provide actionable insight into influence-aimed activities perpetrated by both State and non-state actors. It also discusses how to employ such insight for tailoring effective counter-measures at the operational, tactical and strategic levels. The paper starts by conceptually framing social media platforms and services as tools for exercising influence – or cyber-influence/persuasion – over a target. For this purpose, it examines the notions of "influence operations" and/or "campaigns", presents their featuring elements, and shortly discusses if, how and to what extent they have changed or are changing due to the emergence of the cyber-space and new technological platforms. The study benefits from the review of the recently available literature dedicated to the employment of social media for influence purposes as well as the examination of few selected and documented case studies. Then the paper examines the crafting of intelligence on influence operations form social media data and metadata, in particular on actors engaging in these operations, their employed capabilities, modus operandi, intended goals as well as the operational constraints they face. It outlines the collection and analytical techniques serving the purpose of producing social media intelligence, presents their general advantages and limits, and highlights the challenge of enriching the knowledge on influence operations with further contextual information and intelligence. The paper concludes by proposing recommendations on how to implement the intelligence-based approach to countering influence operations or campaigns over social media in an effective manner. These include the acquisition of the necessary capabilities (human, technological, and organisational) to monitor and analyse social media information. They also include the definition of a framework to better integrate intelligence from social media in to a consistent all-source intelligence system*.

* Senior Researcher dr., ETH Centre for Security Studies, Zurich, Switzerland, matteo.bonfanti@sipo.gess.ethz.ch

**Keywords:** *Influence, Social media, Cyber-influence, Intelligence, Counter-Intelligence*.

## Introduction

On May 27-28th 2018 and in the following days, the Italian President of the Republic was the object of an intensive Twitter messaging campaign asking for his resignation. (Sarzanini, August 5, 2018; Giuffrida, Kirchgaessner and Henley, May 27, 2018)[1] The campaign was triggered by the President's refusal to appoint the recently-formed parliamentary coalition' proposed minister of finances. It endorsed different statements aimed at discrediting the President, accusing him of "high treason", and sustaining his impeachment. Echoed by other new and traditional media, the campaign put a certain political pressure to the holder of Italy's highest institutional post.[2] In the short-term, it polarised part of the Italian public opinion around supporters and opponents of the President, exacerbated some of the prevailing divides between political coalitions, and reinforced existing frictions among social groups and individuals (Riformato, May 27, 2018).[3]

The content, reach and timing of the campaign, its observable dynamics and likely intents, as well as the fact it addressed the President Office prompted an investigation by the Italian authorities. The investigation – which is presently on-going – aims at ascertaining whether the *twits* and further on-line messaging "storm" pursued subversive or other criminal goals.[4] At the same time, it points at identifying the entities who initiated, fuelled, promoted, probably coordinated the spreading of messages, and are

---

[1] #mattarelladimettiti ('#mattarellaresign'), #mattarellavergognati ('#shameonyoumattarella'), #mattarellanonèilmiopresidente ('#mattarellaisnotmypresident'), #mattarellaimpeachement, were the most popular hashtags used in the Twitter messaging campaign.

[2] Thematic pages, posts and videos against the President of the Italian Republic were created and disseminated also through Facebook and Youtube. See for example, https://it-it.facebook.com/pages/category/Community/Sergio-Mattarella-Non-%C3%A8-il-Mio-Presidente-585198994957256/.

[3] #iostoconmattarella is one of the most popular hashtags adopted by Twitter users to express their support to the President.

[4] The investigation is directed by the Anti-Terrorism Prosecutor Office in Rome who can rely on the support of the Italian Postal Police. It is based on art. 277 "Attack on the freedom of the President of the Republic" and art. 278 "Offense to the honor and prestige of the Head of State" of the Italian Criminal Code. Given the national security related implications of the case, the Director of the Italian Department of Information Security was asked to report on it to the Parliamentary Committee for the Intelligence and Security Services.

responsible for that.[5] In this regard, preliminary indicators and limited evidence drew non-domestic entities in to play. They suggested the possible involvement of foreign States or state-sponsored actors who have the capabilities to employ social media platforms and services systematically for achieving pre-defined goals.[6] Actors have already engaged in similar initiatives but in other geographical, political and social contexts, and who might have interests in interfering with Italy's domestic institutional processes. In particular, actors who had resorted to social media content to trigger public pressure on policy-makers abroad and affect their decisions-making processes, undermine their leadership, obtain political paralysis, erode civil society's trust in local institutions, or amplify social divisions. Speculations about the involvement of the above foreign entities were triggered also by some observable patterns of the Twitter campaign against the Italian President, *i.e.* the employed tactical scheme and modus operandi (obviously not the content!) as well as its apparent short-terms results (polarisation and erosion of public trust).

Initially, suspects addressed a Russia-related agency, which had already been reported for using Twitter, Facebook, VKontakte or other social and on-line media in the above-described fashion (Fubini, August 3, 2018; Roeder, 2018).[7] However, this agency is not the only player in the arena. Other Countries run or sponsor similar activities (FireEye, August 2018); and even private organisations could engage in to them, both on the smaller and larger scale.[8] Indeed, results from further technical analysis depict a more articulated

---

[5] See art. 494 "Impersonation" of the Italian Criminal Code. *Cf*. also "Ipotesi illecita: sostituzione di persona (art.494 c.p.)" at https://www.commissariatodips.it/approfondimenti/social-network/approfondimenti-normativi.html.

[6] There is no agreed definition of social media in the literature and among practitioners. In plain language, the expression refers to the Internet-based technologies and practices that allow users to generate and share information in different formats, and to establish relations among them.

[7] As reported by the media, at least twenty Twitter profiles proactively involved in the campaign against Italian President were employed by the Internet Research Agency (IRA) of Saint Petersburg in other propaganda campaigns in favour of populist parties, sovereigntists, and anti-Europeans. These findings emerged from the parallel analysis of a vast repository of IRA-attributed tweets collected by researchers from the Clemson University and provided to the US prosecutor Robert Mueller as part of the investigation of the Russian influence on the 2016 US Presidential elections. The repository includes tweets in Italian, which originated from accounts that were fuelling discussions against government representatives.

[8] See the strategic communication operations run by the private company Cambridge Analytica during the electoral processes in targeted countries. See BBC News, "Cambridge Analytica: The data firm's global influence", accessed 24 September 2018 at https://www.bbc.com/news/world-43476762.

scenario. As far as Twitter is concerned, the user's profile who started the messaging campaign against the President was probably registered in Italy — through the Milan's data hub — but shielded in such a way as to make it seem to originate abroad. Other accounts (at least 150) were suddenly created through servers located in Estonia or Israel. Then, in a very short-time period, about 400 new Twitter profiles generated an anomalous spike in the number of messages against the President. As reported by the media, the campaign originated from a single source, most likely an "Italy-based organisation specialized in this type of activities" (Sarzanini, August 5, 2018).[9] It has still to be clarified whether the organisation acted for its own purposes or as a proxy for other domestic or foreign entities. It goes without saying this is a crucial issue given the implications it may generate from legal and political point of views as well as in terms of the responses that may follow. If ascertained, the involvement of foreign entities will make national security considerations prevailing over those related to the maintenance of the public order and rule of law. In principle, it may prompt diplomatic, economic, military and informational counter actions further to those enacted at the law enforcement and judiciary level.

Pending the above salient issue as well as the results from additional forensic investigation, it is somewhat evident the Twitter campaign promoters' intent to mobilise a segment of the Italian social media users' community and, more broadly, the wider national public opinion. By adopting a certain degree of deception and by exploiting the vulnerabilities of the system of opinion formation within social media, they tried to orient and condition the targeted users' decision-making processes, modify or reinforce their attitudes, alter their perceptions, shape their judgments through persuasion, and induce a specific behaviour (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, pp. 20-21). They acted within the informational environment, especially in its cognitive or semantic dimension, by employing information as a tool for persuasion and influence (Floridi, 2010, p. 3; Durante, 2017, pp 21-37).[10] They orchestrated the dissemination of messages fostering antagonism and confrontation on a specific issue; supported and amplified a non-constructive narrative which can be considered disproportionally

[9] The most probable hypothesis is that the organisation used Tor, a software permitting anonymous online communication by means of encryption, to prevent its identification.

[10] The informational environment is a complex domain in which agents operate through information and interact on the informative level. Also called "infosphere", it is "constituted by all informational entities, their properties, interactions, processes and mutual relations. It is an environment comparable to, but different from cyberspace (which is only one of its sub-regions), since it also includes off-line and analogue spaces of information".

disruptive of social cohesion. At the very first sight, the above suggests they engaged in activities/operations that could be technically considered "inform and influence"; or – given both the domain where these activities/operations occurred and the tools employed to carry them out, i.e. the Internet and social media –"cyber-/social media influence" (Brangetto and Veenendaal, 2016; Szafranski, 1997, pp 395-416; Arquilla and Ronfeldt, 1997, pp. 1-20).[11] These types of activities are different from genuine journalistic, public affairs, civil-society, marketing, and public relations initiatives or lobbying. They are also different from open (on-line or off-line) political campaign and debate, or public diplomacy. Their general lack of transparency coupled with their potential or actual harmful impact on individuals and society put their legitimacy in to question.

How can governments safeguard their interests and tangible or intangible assets from illegitimate cyber- and social media influence operations? What kind of actions should they take to prepare, prevent, respond and recover from these activities? What capabilities do they need? The existing literature on information influence activities provides useful answers to these framework and other more specific questions. The same does the literature covering specifically influence operations through the Internet and social media. In general, they suggest the adoption of multi-discipline (normative, educational, technological, law enforcement), multi-layer (local, national and international), multi-sector (public and private) and multi-stakeholders (governmental institutions and agencies, social media platforms and services providers, users and civil society) coordinated counter-measures. To be effective, these measures should be premised upon a genuine understanding of cyber-social media influence operations, in particular of the actors engaging in the operations (domestic or foreign), the activities they carry out, the strategies, doctrines and principles to which they are informed, the organisational aspects involved and resources needed, the tools, tactics and techniques they employ, their intent and observable results. From a

[11] There are several definitions and connotations of "influence operations". In general, "influence" refers to a broad range of activities including the use of corruption, coercive economic means, public pressure exercised by political parties, think tanks, academic institutions, and the exploitation of ethnic, linguistic, regional, religious and social tensions. Expressions alike "Information Warfare" (IW), "Psychological Operations" (PsyOps), "Strategic Communications" (STRATCOM), "Computer Network Operations" (CNO), and "Military Deception" (MILDEC) refer to operation in the informational domain that are more or less covered by the umbrella term "Information Operation". In addition to the cited labels, terms as diverse as "neocortical warfare" and "net war" have all been used to describe attempts by one or all sides in a conflict or dispute to either change their opponent's position, weaken their resolve or undermine a prevailing narrative.

prevention point of view, such an understanding would support the detection of imminent and on-going cyber- and social media influence operations.

At the tactical and operational level, detection can specifically profit from the collection and analysis of social media data and further contextual information. This intelligence can provide actionable insight into influence-aimed activities perpetrated by both State and non-state actors. The present paper discusses the crafting of such intelligence. It starts by conceptually framing social media platforms and services as tools for exercising cyber-influence over a target. For this purpose, it examines the notions of "influence operations" and/or "campaigns", presents their featuring elements, and shortly discusses if, how and to what extent they have changed or are changing due to the emergence of the cyber-space and new technological platforms. The study benefits from the review of the recently available literature dedicated to the employment of social media for influence purposes as well as the examination of few selected and documented case studies. Then, the paper examines the crafting of intelligence on influence operations form social media data and metadata, in particular on actors engaging in these operations, their employed capabilities, modus operandi, intended goals as well as the operational constraints they face. It outlines the collection and analytical techniques serving the purpose of producing social media intelligence, sketch their general advantages and limits, and highlights the challenge of enriching the knowledge on influence operations with further contextual information and intelligence. The paper concludes by proposing broad recommendations on how to implement the intelligence-based approach to countering influence operations or campaigns over social media in an effective manner. These include the acquisition of the necessary capabilities (human, technological, and organisational) to monitor and analyse social media information. They also include the definition of a framework to better integrate intelligence from social media in to a consistent all-source intelligence system.

## Cyber-influence through social media: is there anything New under the sun?

National governments or sub-national entities have resorted to inform and influence operations or campaigns several times in history – in peacetime, within a rivalry or situation of tension, and during open conflict/warfare

(Costello, 2018).[12] The objectives, targets and techniques involved in these operations are not new.[13] Nor are the strategies and stratagems typically employed to run them (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018). Especially when carried out by or on behalf of foreign entities, influence operations become relevant from a national security perspective because they interfere with the established institutional processes and functions of a State, and invade its sovereign space. Targets of influence may vary from a region's civilian population, selected decision-makers, military personnel, to governments or State institutions. In other words, influence operations can feature in "social" information activities as they target the society as a whole or social groups, in particular those groups' ideas, opinions, motivations and beliefs. They can also entail "individual-oriented" activities as they address selected individuals with a specific psychographic profile. In both cases, they tactically aim at affecting psychological processes, shaping motivations or ideas, conditioning behaviour and choices (Palmertz, 2017).

Whereas the fundamental goals and objectives of inform and influence operations or campaigns have not significantly changed over time, the tools and techniques to achieve them have. They have evolved alongside socio-economic-cultural changes, also because of the development and use of new information and communications technologies, the Internet included. These technologies have progressively transformed the information environment in its constituent elements and inherent dynamics (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, p. 20).[14] They have contributed to generate an additional dimension or space, the so-called cyberspace, within which a wide range of agents (not only States) can use information for several purposes,

[12] For example, during the First World War the Allies engaged in propaganda activities by dropping leaflets over German soldiers, calling on them to surrender. Conflicting parties ran similar actions during the Second World War, the Cold War, the two wars in Iraq, and, more recently, in conflict areas like Libya and Syria. Resort to information operations in peacetime or within hybrid-conflicts is also very common as, for example, the alleged use of media and information in Turkey shows.

[13] The difference between "operation" and "campaign" concerns the scope, reach, objects and time-frame of the influence activities and actions they consist of as well as the level of coordination among them. Generally, while an operation encompasses a smaller and more specific objective to be achieved with tactical actions run within a mid/short time-frame, a campaign entails a coordinated set of operations aimed at achieving strategic objectives on the longer run.

[14] As explained by James Pamment the impact of technological innovations on the information environment is coupled by the commercial reconfiguration of large parts of the media system (especially in Western countries). Such reconfiguration has prioritised commercial imperatives over the reliabilities of the information sources and integrity of the information.

including influence-related ones.[15] Cyber-influence (or persuasion) is the recently coined terminology to refer to inform and influence operations that are run in the cyberspace, leverage this space's distributed vulnerabilities, and rely on cyber-related tools and techniques to affect an audience's choices, ideas, opinions, emotions or motivations.[16]

Compared to "traditional" inform and influence activities, cyber-influence employs context-specific tools and techniques to achieve its goals. It takes advantages of how social media platforms and services are designed and work. It profits from how information is generated, distributed and consumed by social media users, as well as from the way they interact and establish relationships among themselves. To a certain extent, social media platforms make cyber-influence easier, cheaper and faster to carry out.[17]

They present some intrinsic features that make them both the preferred terrain/environment and tool/capability for engaging in cyber-influence. "They enable precision-targeted messaging and advertisement based on psychographic targeting at an unprecedented dimension of contemporary information influence activities" (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, pp. 25, 26, and 28). At the same time, they allow to reach out a wide and geographically distributed audience. They combine text, photos, videos and audios clips (*e.g.* memes) that are amenable to manipulation and misappropriation through simple techniques.[18] They make mechanisms of information scrutiny and source review more complex. By allowing deception and supporting automation, these platforms hinder accountability processes and makes attribution difficult (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, pp. 30-31).[19] Finally, actors proactively

[15] With the advent of the cyberspace, the control and release of information is no longer the purview of few established actors. Every organisations and even ordinary individuals can reach mass audiences through online platforms.

[16] Cyber-influence or persuasion makes use of information and communications technology and networked systems. Unlike "technical" information operations such as computer network operations (CNO) – in their variant of Computer Network Defence (CND), Computer Network Exploitation (CNE), and Computer Network Attacks (CNA) – Cyber-influence does not target an enemy's or opponent's technical assets such as data servers or network nodes. It targets individuals or groups' minds. It goes without saying that cyber influence can be used in combination with technical information operations.

[17] At a very basic level, an actor only needs an internet-enabled device, freely available account-based applications, an elementary knowledge of how to use social media, and an internet connection.

[18] Manipulation can also be quite sophisticated as "deep fake" videos shows.

[19] "Today networked individuals are exposed to fragmented messages, many, like memes, without ascertainable factual content, many without an identifiable source, many only in the flow because they are shared by friends or promoted by algorithms".

engaging in to social media influence can count on the – more or less unconscious – support from users behaving as "useful idiots". These are users who act and process information uncritically, and that amplify the magnitude of the operation. Their active presence and participation in disseminating information makes the detection of the operations even more difficult.

However, the potential of cyber-influence, in general, and social media influence, in particular, should not be overestimated. First, social media penetration is variable across geographical regions and segments of the population. Second, research shows that people rarely rely on the Internet and social media as their unique or primary source of information. Television and other offline media outlets are still the dominant platforms for news. Furthermore, information and news are often sourced from family members, friends, and community leaders. The above also explains why influence operations or campaigns generally entail the synchronised use of different sources of information and involve actors like NGOs, think tanks, individual agents (*e.g*. celebrities, religious leaders) that operate "off-line" too.

### A closer look in to social media influence operations

Social media influence operations and campaigns can be quite complex, rely on extensive planning and preparatory activities, and draw upon different techniques and stratagems. Each operation or campaign has its own specific features. Nevertheless, most of them adopt similar tactical schemes and modus operandi.

Among other things, planning for social media influence operations involves intelligence-gathering activities.[20] It requires identifying exploitable vulnerabilities and targets. Identification may rely on different methodological concepts and approaches as well as on automated or semi-automated techniques and technologies like "crawlers", "spiders" or "monitoring bots".[21] With regard to vulnerabilities, they may concern the platforms' design and functions, their communities of users, and the social-cultural and political contexts to which users belong. Examples of these "technical-technological", "human-related" and "societal" vulnerabilities are: exploits in the platform' provided privacy and security setting or other technical features (API

[20] Intelligence gathering is often conducted prior or in concert with influence operations because it helps in identifying the best courses of action.

[21] "Bots" is a shorter term for robot. It is an automated program which performs repetitive actions along a set of algorithms. It can collect data on the platforms and their users and make these data available for analysis. Simon Hegelich, "Invasion of the social bots" (Berlin: Konrad Adenauer Stiftung, 2016).

applications or widgets); selected users' personal information to be leveraged for malicious purposes (*e.g.* blackmailing or for compromising accounts); and existing socio-economic frictions and political divides.[22] Scouting vulnerabilities is coupled with targets reconnaissance activities. Targets can range from wider audiences, specific groups, to individuals (*e.g.* influencers). They can be identified via psychographic or socio-demographic profiling carried out through more or less advanced technological applications and techniques.[23] The case of Cambridge Analytica is an example of using psychographic targeting to understand users' preferences (Gibney, 2018). Similar techniques were apparently used also with regard to the US presidential election in 2016. According to the FBI, the Russia-related Internet Research Agency (IRA) initiated its influence operation in 2014 by conducting a comprehensive target audience reconnaissance and psychographic mapping of US social media sites dedicated to politics and other social issues.[24] IRA's goal was to determine metrics such as reach, audience engagement, frequency of posts, and nature of content. This baseline knowledge provided a point of departure for the operational design of social media influence activities (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, pp. 79-81).

Coming to the techniques that are suitable for engaging in to social media influence, they are different. They include socio-cognitive and psychographic hacking via dark advertisement, social hacking, band wagering, digital disinformation and fakes, exploitation via bots, botnets and sock-puppets, trolling and flaming (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, pp. 31-79). Most of these techniques are used in combination; they result in stratagems like polarisation, enraging, flooding and laundering

---

[22] In some cases, scanning for human-related vulnerabilities is based on ad hoc activities like the creation of honeypots to solicit information via malicious links. Trend Micro "Hackers Exploit Instagram API Flaw to Steal Information from Verified Users", accessed 24 September 2018 at https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/hackers-exploit-instagram-api-flaw-to-steal-information-from-verified-users; Rob Hornbuckle, "Social Media Vulnerabilities and Considerations for the Corporate Environment", 2016, accessed 24 September 2018 at http://www.infosecwriters.com/ Papers/ RHornbuckle_Social_Media.pdf; Candid Wüest, "The Risks of Social Networking", 2010, accessed 24 September 2018 at https://www.symantec.com/content/en/us/enterprise/media/ security_response/whitepapers/the_risks_of_social_networking.pdf.

[23] While the former concerns the segmentation of users depending on their personality traits, values, preferences and attitudes, or other features, sociodemographic profiling discriminate users according to their age, gender, education, ethnicity. See https://www.cbinsights.com/ research/what-is-psychographics/.

[24] United States of America V. Internet Research Agency LLC, Case 1:18-cr-00032-DLF, accessed 24 September 2018 at https://www.justice.gov/file/1035477/download. See "Intelligence-Gathering to Inform U.S. Operations", par. 29; see also par 37.

(Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, p. 70).[25] For example, polarisation entails supporting opposing views on one or more issues. It aims at exacerbating pre-existing divides and frictions in a targeted community by combining – creatively and opportunistically – social and cognitive hacking, digital disinformation and fakes, sock-puppetry, bots, trolling and further techniques. The use of the cited techniques was especially noticeable during the US presidential election, but also during elections in Europe (including the referendum on Brexit) (Baezner and Robin, 2017). It is expected for other forthcoming electoral terms, when not only foreign States or their proxies but also domestic entities (radical groups or other organisations) are likely to engage in to deceptive influence.[26] Resort to these techniques is also traceable to different non-election-related operations run in Western Countries, the Nordic-Baltic region and Eastern Europe (Helmus et al., 2018; "Lisa Case" in NATO Review Magazine, 2016). Pending the outcomes of the on-going investigation, the Twitter campaign against the Italian President could be seen as a social media influence activity both employing and pursuing polarisation.

Again, some of the cited techniques integrate a certain degree of automation. Bots (especially spammer and impersonator bots) mimic organic behaviour to mislead, confuse and influence the users. Also known as social bots, they are commonly used to engage with political content on social media platforms. They are "highly efficient for amassing virtual social capital online to exploit social pressures and cognitive biases by acting as force multipliers, or false amplifiers in online discussions. Depending on the level of sophistication of a bot, this can be done by, for example, automatically following, re-tweeting, or liking posts from real social media accounts to boost their legitimacy, using spammer bots to reinforce impersonator bots, or using bots to crowd out dissenting opinions to create a false sense of consensus. Bots can repeatedly post and reinforce specific messages via multiple accounts and exploit features such as tags and hashtags to effectively direct content on

[25]In contrast to strategy as a more neutral term, stratagem invokes the meaning of trickery, of outwitting an adversary.

[26] Recently, Denmark has adopted an Action Plan to counter influence campaigns especially in view of the upcoming 2019 parliamentary elections. See Ministry of Foreign Affairs of Denmark, "Strengthened safeguards against foreign influence on Danish elections and democracy", 07.09.2018, accessed 24 September 2018 at http://um.dk/en/news/NewsDisplayPage/?newsID=1DF5ADBB-D1DF-402B-B9AC-57FD4485FFA4. Also, Canada next electoral term is at risk of influence: Communication Security Establishment, "Cyber Threats to Canada Democratic Process", 2018, accessed 24 September 2018 at https://www.cse-cst.gc.ca/en/democratic-process-processus-democratique/table.

social media platforms" (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, p. 57). Similar to bots are sock-puppets, which are not entirely automated, but partially controlled by a human. Semi-automated sock-puppets grant control over multiple false accounts to coordinate content across different community groups and platforms. (Pamment, Nothhaft, Agardh-Twetman, and Fjällhed, 2018, p. 58)

While social bots prove to be efficient tools for engaging in social media influence, they are also vulnerable to exposure. The same goes for other techniques listed above, like sock-puppeting, trolling, or hacking via dark advertisement. In principle, it is possible to detect when they are deployed and check if/how they are used within the scope of an influence operation. There are some methods proving to be useful in that regard.[27] The presence of some (often ambiguous!) indicators can raise a flag; further investigation and analysis would provide the necessary evidence. In general, if properly monitored, social media data offer a certain insight in to influence activities. They deliver information that, if analysed and further combined with other intelligence, can support detection and allow for counteraction. This should not bring to the conclusion that influence operations are easy to spot, especially at their early stage of preparation or enactment. They could be more easily recognised *ex-post*, *i.e.* after their execution. However, *ex-ante* or *interim* detection is in principle possible through the adoption of a social-media and multi-source intelligence-based approach.

## Crafting Intelligence on social media influence operations

### *Ideal standards vs real constraints*

Necessary but not sufficient conditions for countering social media influence are to understand these types of operations and being able to recognise them. Intelligence capabilities can play a significant role in this regard (Treverton, 2017, p. 21). Broadly speaking, they can be deployed by an organisation to collect and analyse social media data as well as other – both openly and covertly sourced – information in order to generate actionable knowledge of potential and actual illegitimate influence actions. To a certain

[27] Social bot detection is premised upon various approaches, which alternate or combine human engagement or algorithmic analysis. Crowdsourcing, social graph analysis, feature analysis are approaches which can be used alone or combined to detect a social bot. *Cf.* Marc-André Kaufhold and Christian Reuter, "Cultural Violence and Peace in Social Media: Interventions by Human and Social Bots", Conference Paper, Cyber-security Conference, Zurich 27-28 September 2018, to be published in proceedings.

extent, the (counter-) intelligence gathering activities aimed at understanding/recognising influence operations mirror those promoted by malevolent actors while preparing and planning for them. They share similar goals (*e.g.* mapping vulnerabilities and potentially targeted groups), sources (*i.e.* social and other media), collection tools (*e.g.* collection of data via bots), and analytical techniques (*e.g.* social network analysis) as well as they face analogous technical constraints. The most evident difference between the two "types" of activities lies in the decision-making-related purpose they aim to serve: influence and counter-influence. Other differences might concern both the technical and non-technical aspects of the intelligence gathering activities.[28]

In principle, intelligence on social media influence should cover the strategic, tactical and operational dimensions of these types of operations.[29] It should be crafted by processing data from social media and other sources of information with appropriate methodologies and tools. It should cover "introspectively" the relevant exploitable vulnerabilities and targets, but also address the "external" threat environment, like actors (both foreign and domestic), their capabilities and modus operandi.[30] Most importantly, it should be actionable *i.e.* afford ground for counter action. The normative nature of the above statements is evident. They suggest what intelligence on social media influence operations should look like. They portray an ideal of intelligence that clashes with what it is often possible to achieve in reality. Indeed, crafting high valuable intelligence on influence operations is a complicated task which requires the allocation of specific resources (human, technological, organisational). Furthermore, the intelligence production process itself is somewhat flawed. It is affected by the general and endemic limitations concerning the crafting of social media and cyber-intelligence (Bonfanti, 2015, pp. 231-262; Omand, Bartlet and Miller, 2012; Bonfanti, , 2018, pp. 105-121).[31] Among the main flaws lie the uncertainty, volatility, scale and anonymity of the data and the sources to be processed (information anarchy). These flaws give to the produced intelligence a high probabilistic connotation, which will in turn affect the scope, nature and effectiveness of possible counterinfluence actions.

---

[28] For example, the actor who engages in counter-intelligence gathering has to circumscribe the scope and reach of the collection according to what is prescribed by law, including privacy and data protection regulations.

[29] There is no clear demarcation from one level to another; they frequently overlap or are combined.

[30] It should adhere to the adage "know other and yourself".

[31] The space limitation of the present contribution does not allow a detailed discussion of the social media intelligence crafting process and it main constrains.

***A (rough) cuboid representation of Intelligence on social media influence***

In light of what has been discussed so far, intelligence on social media influence can be represented through the following cuboid model (Figure 1). Far for being a comprehensive representation of this type of knowledge, the model tries to capture its main ideal features.

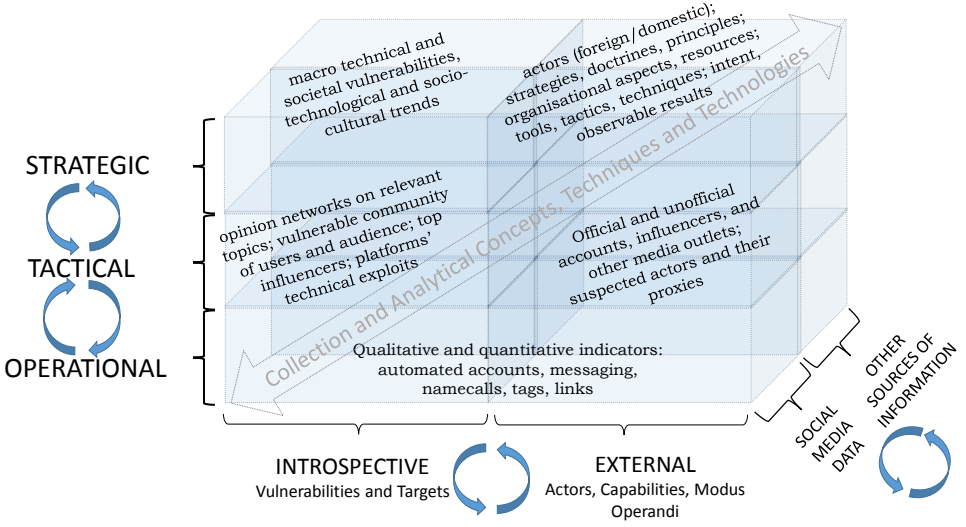


Figure 1: (Counter) Intelligence on Social Media Influence Activities

At the strategic level, intelligence on social media influence should provide insight into present and future threats as well as the risks associated with those threats. It should cover and review the threat landscape for macro trends and identify key threat actors, their goals, capabilities and how these may evolve in the mid-long term (Lin, 2018).[32] Strategic intelligence is generally rich in contextual information. Although focused on social media related influence, it may also frame these operations within broader cyber-persuasion or inform and influence operations and campaigns. With regard to social media, it can examine the actors (foreign or domestic) capable of

[32]As suggested by the author, one element of detecting influence operation is "recognizing parties that might have something to gain from conducting such campaigns. Mere recognition of who gains is not evidence that a party is undertaking an Information warfare and Influence campaign, but a party that does not stand to gain from such a campaign is unlikely to be involved in one".

engaging in to this type of influence, the strategies, doctrines and principles to which current operations are informed, the organisational aspects involved, and resources needed for running these operations, the tools, tactics and techniques they employ, their intent and observable results. Such strategic intelligence can derive from multiple sources. It can also derive from the collection, documentation and analysis of past cases and events.[33] Cases-analysis supports patterns recognition and assist in speculating on future developments or scenarios. Quite important, strategic intelligence should also address context-specific exploitable vulnerabilities and potential targets of social media influence. As discussed above, vulnerabilities lie in the platforms' design and functions, in the community of users, but also in the society as a whole. In a strategic perspective, these latter and the wider-technical ones are the most relevant. Societal vulnerabilities represent those which become also evident in domestic open debates. Most likely, they will be the ones at risk of exploitation from malevolent actors. In monitoring vulnerabilities and potential targets, technological, cultural and socio-economic trends should possibly be considered in order to foresee new potential opportunities for exploiters. Strategic intelligence generally serves apical decision-making processes aimed at achieving an organization's mission and determining its direction and objectives. It can for example support the definition of counter programmes adopted at the policy and legislative level.

At the tactical level, intelligence should concern what happens on selected social media platforms and, to a certain, extent, beyond them. It should be based on the monitoring of selected social media content (data and metadata), their communities of users as well as other online sources like blogs and webpages. With regard to social media, monitoring implies the processing of information which can be further analysed to: mapping opinion networks based on relevant discussion topics; circumscribing community of users and audience which are potentially exposed to influence attempts; identifying top influencers for each community and topics; charting platforms technical exploits. It also entails screening targeted on-line official or unofficial accounts or pages which can fuel influence operations. Information collection and analysis should occur according to pre-defined methodologies; they could benefit from automated or semiautomatic technologies and techniques.[34] With regard to automatic or semiautomatic collection, this task can be carried out

[33] The list of social media influence tools, techniques and stratagem examined above drew on analysis of both the literature and past events.

[34] Both the degree of automation and effectivity of technical collection and analysis could probably improve by the adoption of artificial intelligence-based solutions.

through: software applications designed for retrieving structured set of information from social media "Application Programming Interface (API)" (for example, Facebook Graph API, Twitter Search API);[35] other "scrapers", "crawlers", "spiders" as well as additional tools and algorithms that gather data and metadata to map the information environment of social media platforms; solicited/crowd source information to sense users sentiments. As per the analysis, concepts and approaches may vary. They may include: social network analysis, which involves identifying and visualising social structures and detect communities in large social media data; lexical analysis, to detect structures and patterns in textual data; sentiment analysis, providing insight in to community of users' attitudes, values, feelings; geolocation and geoinferencing, that are methods for determining the geographic origin of a social media messages and making inferences about the geographic location of posts; deep neural networks, to categorise images (Marcellino, Smith, Paul, and Skrabala, 2017). These analytical concepts are then translated into practical approaches and applications like, publics analysis, stance analysis, network description (Marcellino, Smith, Paul, and Skrabala, 2017, pp. 30-35).[36] Overall, they are part of the toolboxes used by different type of actors to generate the so-called social media intelligence (Bonfanti, 2015, p. 241; Omand *et alii*, 2012, p. 15). As mentioned above, the cited approaches, techniques and their application face specific limitations (technical, technological, legal) whose reach should not be underestimated both by those who adopt them to craft intelligence and by the consumer of such product (Bonfanti, 2015, p. 239). Tactically, intelligence on social media operations should also cover blogs, webpages or other on-line media outlets which are potential vehicles of influence. These should be mapped and monitored. The same goes for "off-line" actors like suspected domestic proxies of foreign entities which might fund and sponsor influence operations, or offer other kind of support to domestic actors. Generally more technical in nature, tactical intelligence on social media influence informs the steps and actions an organization can take to design and run counter measures.

Finally, operational intelligence should address imminent or initiated influence operations that pose an actual threat. Drawing upon the knowledge generated at the strategic and tactical level, this intelligence should aim at enabling short-term response, especially on the informative level (shut down

[35] Among this information, the geo-temporal coordinates of the users, the number of their information they generate and share, determine the number of their connections.

[36] Another approach is resonance analysis, which can be applied to measure the spread of specific information and narratives across a social media and identify communities of users. *Cf*. p. 47 ff.

malicious accounts, disseminate counter narratives). It relies on quantitative and qualitative indicators which can raise a flag on an occurring or started-up influence operation. From this point of view, social media enable to listen for malevolent activities by monitoring: accounts, messaging, tags, name calls, links, and attachments. At the very practical level, recognising entails detecting and identifying the tell-tale signs of a planned, ongoing, or concluded influence operations through social media. "The rapid emergence of large numbers of automated social chat bots promulgating similar messages could signal the start of a concerted operation, as automated chat bots do nothing but amplify discourse rather than contributing new content. More generally, one might imagine that some combination of volume (messages per day), content, and platform and so on could identify with probability automated social bots carrying divisive or inflammatory messaging" (Lin, 2018). Event detection or situational awareness techniques can prove to be useful in this regard too). (Abdelhaq, Sengstock, and Gertz, 2013; Atefeh and Khreich, 2015, pp. 132-164) The above described operational intelligence can trigger the deployment of already designed counter measures by response agencies.

Regardless of any clear-cut demarcation between strategic/tactical/ operational intelligence on social media influence operations, the ability to gain knowledge that covers all these levels is paramount. Each level is dependent on and integrated to the others. They overall provide the necessary perspective to look at social media influence operations, and better support recognition and counteraction.

## Conclusion

Social media are relatively new means of information and engagement. They are used by an increasing number of individuals and organisations for creating, sharing, and consuming information, establishing or developing social or other relationships among them, and participating in the on-line (and off-line) community life. Broadly speaking, they are defining a novel space of action whose boundaries are yet to be firmly established.

Among other things, social media have also become both the theatre of and the tool for running illegitimate (cyber-) influence operations. Detecting this type of operations and clearly distinguishing them from genuine political or social engagement and campaigns is tricky. Detection can be pursued through the adoption of an intelligence-based approach. An approach that promises benefits but has its own limitations, and whose adoption faces specific challenges – limitations also include legal constrains. With regard to the limitations and challenges, they generally concern: the collection and

analytical concepts and techniques to be employed for generating tactical/operational intelligence from social media data and further information; the definition of tailored frameworks and procedures to integrate the obtained knowledge with further information and intelligence; the acquisition and deployment of relevant intelligence capabilities (Bonfanti, 2018, pp. 116-117; Bonfanti, 2015).

As per the collection and analytical concepts and techniques, their reliability and effectiveness – in providing valid results – is flawed by the nature of the data they are applied to, *i.e.* social media and, in general, Internet data. The uncertainty, volatility, scales and anonymity of these data undermines the validity/representativeness of obtainable results. This gives to the produced intelligence a high probabilistic connotation, which will in turn affect the scope, nature and effectiveness of possible counterinfluence actions. Another issue concerns the integration of multi-sourced information and intelligence. Technically, consistent integration is difficult to achieve because of the format, nature, and grade of uncertainty of social media and other on-line information. On an organisational level, integration should be achieved within structured collaborative frameworks in which collectors and analysts work together on regular basis. Further research on the topic and more practice/experience can help in improving the approach. With regard to capabilities, the gathering of the above described multi-layers intelligence on social media influence operations requires an organisation to acquire and allocate specific human, technical, organisational, and financial resources. The acquisition and allocation of these resources should prove to be sustainable and assessed for the return they generate.

A final note: it seems that further cases of domestic non-state actors' engagement in social and other online media influence operations (*e.g.* the Twitter campaign against the Italian President) are to be expected in the future – this seems at least the perception of some academics and practitioners. The phenomenon should be monitored and investigated in order to allow more effective law enforcement and security response.

**References:**

1. Abdelhaq, Hamed, Sengstock, Christian, and Gertz, Michael, (2013), "EvenTweet: Online Localized Event Detection from Twitter", *Proceedings of the VLDB Endowment*, Vol. 6. No. 12, accessed 24 September 2018 at http://db.disi.unitn.eu/pages/VLDBProgram/pdf/demo/p809-abdelhaq.pdf.

2. Arquilla, John and Ronfeldt, David, (1997), "A New Epoch — and Spectrum — of Conflict", in John Arquilla, David Ronfeldt, *In Athena's Camp: Preparing for Conflict in the Information Age*, Santa Monica, CA: RAND Corporation, pp. 1-20.

3. Atefeh, Farzindar, and Khreich, Weal, (2015), "A Survey of Techniques for Event Detection in Twitter", *Computational Intelligence*, Vol. 31, No. 1, pp. 132-164.

4. Baezner, Marie, and Robin, Patrice, (2017), "Cyber and Information Warfare in elections in Europe", Hotspot Analysis 2017, accessed 24 September 2018 at http://www.css.ethz.ch/publikationen/risk-and-resilience-reports/details.html?id=/c/y/b/e/cyber_and_information_warfare_in_electio.

5. BBC News, "Cambridge Analytica: The data firm's global influence", accessed 24 September 2018 at https://www.bbc.com/news/world-43476762.

6. Bonfanti, Matteo E., (2015), "Social media intelligence a salvaguardia dell'interesse nazionale. Limiti e opportunità di una pratica da sviluppare", in U. Gori, L. Martino (Ed), "Intelligence e Interesse Nazionale", Roma: Aracne Editrice, pp. 231-262.

7. Bonfanti, Matteo E., (2018), "Cyber Intelligence: In Pursuit of a Better Understanding for an Emerging Practice", *INSS Cyber, Intelligence, and Security*, Vol. 2, No. 1, pp. 105-121.

8. Brangetto, Pascal, Veenendaal, Matthijs A., (2016), "Influence Cyber Operations: The Use of Cyberattacks in support of Influence operations", in N. Pissanidis, H. Rõigas, M. Veenendaal (Eds.) 8th *International Conference on Cyber Conflict Cyber Power*, Tallinn: NATO CCD COE Publications.

9. Communication Security Establishment, "Cyber Threats to Canada Democratic Process", 2018, accessed 24 September 2018 at https://www.cse-cst.gc.ca/en/democratic-process-processus-democratique/table.

10. Costello, Katherine, (2018), "Russia's Use of media and Information Operations in Turkey. Implications for the United States", Perspective, Santa Monica: RAND Corporation, 2018, accessed 15 September 2018 at https://www.rand.org/pubs/perspectives/PE278.html.

11. Durante, Massimo, (2017), *Ethics, Law and the Politics of Information,* Dordrecht: Springer, Ch. 2, pp. 21-37.

12. FireEye, (August, 2018), "Suspected Iranian Influence Operation Leverages Network of Inauthentic News Sites & Social Media Targeting Audiences in U.S., UK, Latin America, Middle East", Report accessed 24 September 2018 at https://www.fireeye.com/blog/threat-research/2018/08/suspected-iranian-influence-operation.html

13. Floridi, Luciano, (2010), *Information: A Very Short Introduction*, Oxford: Oxford University Press 2010, p. 3.

14. Fubini, Federico, (August 3, 2018), "Le manovre dei russi sul web e l'attacco coordinato a Mattarella. In 400 per l'attacco sull'impeachment", *Corriere della Sera*, accessed 24 September 2018 at https://www.corriere.it/politica/18_agosto_03/manovre-russi-web-c6e7d71e-9699-11e8-8193-b4632fd4d653.shtml.

15. Gibney, Elizabeth, (2018), "The scant science behind Cambridge Analytica's controversial marketing techniques. Nature peers into the evidence for 'psychographic targeting'", *Nature*, accessed 24 September 2018 at https://www.nature.com/articles/d41586-018-03880-4.

16. Giuffrida, Angela, Kirchgaessner, Stephanie and Henley, Jon, (May 27, 2018), "New elections loom in Italy amid calls for Mattarella to be impeached", *The Guardian*, accessed 24 September 2018 at https://www.theguardian.com/world/2018/may/27/italys-pm-designate-giuseppe-conte-fails-to-form-populist-government.

17. Helmus, Todd C., Bodine-Baron, Elizabeth, Radin, Andrew, Magnuson, Madeline, Mendelsohn, Joshua, Marcellino, William, Bega, Andriy, Winkelman, Zev, (2018), "Russian Social Media Influence. Understanding Russian Propaganda in Eastern Europe", Santa Monica: RAND Corporation, 2018), accessed 24 September 2018 at https://www.rand.org/pubs/research_reports/RR2237.html.

18. Hornbuckle, Rob, (2016), "Social Media Vulnerabilities and Considerations for the Corporate Environment", accessed 24 September 2018 at http://www.infosecwriters.com/Papers/RHornbuckle_Social_Media.pdf.

19. Kaufhold, Marc-André and Reuter, Christian, (2018), "Cultural Violence and Peace in Social Media: Interventions by Human and Social Bots", Conference Paper, Cyber-security Conference, Zurich 27-28 September 2018, to be published in proceedings.

20. Lin, Herb, (2018), "Developing Responses to Cyber-Enabled Information Warfare and Influence Operations", *Lawfare*, accessed 24 September 2018 at https://www.lawfareblog.com/developing-responses-cyber-enabled-information-warfare-and-influence-operations.

21. "Lisa case" NATO Review Magazine, "The "Lisa case": Germany as a target of Russian disinformation", accessed 24 September 2018 at https://www.nato.int/docu/review/2016/Also-in-2016/lisa-case-germany-target-russian-disinformation/EN/index.htm.

22. Marcellino, William, Smith, Meagan L., Paul, Christopher, Skrabala, Lauren, (2017), "Monitoring Social Media: Lessons for Future Department of Defense Social Media Analysis in Support of Information Operations", Santa Monica: RAND Corporation, accessed 24 September 2018 at https://www.rand.org/pubs/research_reports/RR1742.html.

23. Ministry of Foreign Affairs of Denmark, "Strengthened safeguards against foreign influence on Danish elections and democracy", 07.09.2018, accessed 24 September 2018 at http://um.dk/en/news/NewsDisplayPage/?newsID=1DF5ADBB-D1DF-402B-B9AC-57FD4485FFA4.

24. Omand, David, Bartlet, Jamie and Miller, Carl, (2012), "#Intelligence", London: Demos Publishing.

25. Palmertz, Bjorn, (2017), "Theoretical foundations of influence operations: a review of relevant psychological research", accessed 24 September 2018 at https://www.msb.se/Upload/Om%20MSB/Forskning/Kunskapsoversikt/Theoretical%20foundations%20of%20influence%20operations.pdf.

26. Pamment, James, Nothhaft, Howard, Agardh-Twetman, Henrik, Fjällhed, Alicia, (2018), "Countering Information Influence Activities. The State of the Art", pp. 20-21, accessed 24 September 2018 at https://www.msb.se/RibData/Filer/pdf/28697.pdf.

27. Riformato, Serena, (May 27, 2018), "Governo, firme e tweet di solidarietà a Mattarella. Ma spuntano anche minacce di morte", *Repubblica*, accessed 24 September 2018 at https://www.repubblica.it/politica/2018/05/27/news/governo_dopo_le_minacce_di_impeachment_sui_social_parte_la_solidarieta_a_mattarella-197516727/.

28. Roeder, Oliver, (2018), "Why We're Sharing 3 Million Russian Troll Tweets", *FiveThirtyEight*, accessed 24 September 2018 at https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/.

29. Sarzanini, Fiorenza, (August 5, 2018), "Quirinale, dallo «snodo» di Milano l'account falso per l'attacco web a Mattarella", *Corriere della Sera*, accessed 24 September 2018 at https://roma.corriere.it/notizie/cronaca/18_agosto_05/dallo-snodo-milano-primo-account-falsoper-l-attacco-web-colle-5af8332e-98eb-11e8-9116-c731a1e8fd65.shtml.

30. Szafranski, Richard, (1997), "Neocortocal Warfare? The Acme of Skill", in John Arquilla, David Ronfeldt, *In Athena's Camp: Preparing for Conflict in the Information Age*, Santa Monica, CA: RAND Corporation, pp 395-416.

31. Treverton, Greg, (2017), "Influence Operations and the Intelligence/Policy Challenges", (Stockholm: Swedish Defence University – Centre for Asymmetric Threat Studies, 2017). p. 21, accessed 24 September 2018 at https://www.fhs.se/download/18.1ee9003b162cad2caa5351cf/1524483543405/Influence%20Operations%20and%20the%20Intelligence%20Policy%20Challenges.pdf.

32. Wüest, Candid, (2010) "The Risks of Social Networking", accessed 24 September 2018 at https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the_risks_of_social_networking.pdf.