

ISBN: 978-9934-564-18-5



DIGITAL HYDRA: SECURITY IMPLICATIONS OF FALSE INFORMATION ONLINE

Digital Hydra: Security Implications of False Information Online

Project director: Giorgio Bertolin

Researchers:

Nitin Agarwal, Professor of Information Science, University of Arkansas at Little Rock
Kumar Bandeli, Doctoral Cand., Information Science, University of Arkansas at Little Rock
Giorgio Bertolin, Social Scientist, NATO Strategic Communications Centre of Excellence
Nora Biteniece, Software Engineer, NATO Strategic Communications Centre of Excellence
Katerina Sedova, Project Assistant, NATO Strategic Communications Centre of Excellence

Text Editor: Anna Reynolds

The NATO StratCom Centre of Excellence, based in Latvia, is a multinational, cross-sector organization which provides comprehensive analyses, advice and practical support to the alliance and allied nations. This report is a product of the NATO Strategic Communications Centre of Excellence (NATO StratCom COE). It is produced for NATO, NATO member countries, NATO partners, related private and public institutions and related individuals. It does not represent the opinions or policies of NATO.

© All rights reserved by the NATO StratCom COE. Reports may not be copied, reproduced, distributed or publicly displayed without reference to the NATO StratCom COE.

The views expressed here are solely those of the authors in their private capacity and do not in any way represent the views of NATO.

Riga, November 2017

NATO Strategic Communications Centre of Excellence
Riga, Kalnciema iela 11b, Latvia LV1048
www.stratcomcoe.org
Ph.: 0037167335463
info@stratcomcoe.org

ISBN 978-9934-564-18-5

Digital Hydra: Security Implications of False Information Online

OUR LATEST PUBLICATIONS



NEW TRENDS ON SOCIAL MEDIA



DAESH RECRUITMENT



STRATCOM LAUGHS:
IN SEARCH OF AN ANALYTICAL FRAMEWORK



STRATEGIC COMMUNICATIONS:
INSIGHTS FROM THE COMMERCIAL SECTOR



ROBOTROLLING



ACADEMIC JOURNAL
"DEFENCE STRATEGIC COMMUNICATIONS"

TABLE OF CONTENTS

Introduction.....	5
Platforms: Geography, Culture, Language.....	11
Introduction.....	12
Geography.....	12
Facebook.....	13
Twitter.....	15
Other platforms.....	17
The Arab world and the MENA region.....	19
Social media on RuNet.....	23
Concluding remarks.....	29
Blogs, Fake News, and Information Activities	31
Introduction.....	32
Methodology.....	32
Typical characteristics of disinformation riddled blogs.....	35
Tracking the origins of misleading blog content.....	38
Mixed-media vs. Cross-media approaches.....	40
Tracking how an antagonistic narrative travels.....	42
Conclusions.....	45
Third-party Services and Fake News Media Sites	47
Background.....	48
News sources and third-party services.....	53
Conclusions and implications.....	59
Conclusions and Recommendations.....	61
Glossary.....	66
Endnotes.....	73

INTRODUCTION

The study investigates misinformation and disinformation on social media in the context of the rise of ‘fake news’ and the birth of the ‘post-truth’ era. Are these concerns substantiated by facts? What are the consequences of these phenomena for the information environment? Most importantly, do these phenomena pose a threat for our societal security? This study will provide actionable knowledge by answering to these questions.

This introduction is an attempt to position the emergence of ‘fake news’ in a wider societal context. Particular emphasis is placed on the cognitive biases that enable information manipulation. In turn, this will lead to a discussion about the tactics employed by adversarial actors to carry out information activities.

DEFINITIONS

A glossary with definitions is provided as in the appendix (page 66). However, some terms deserve to be defined from the very beginning. ‘Disinformation’ and ‘misinformation’ are not officially defined in the NATO terminology. For these key terms we adopt the following definitions:¹

Disinformation: *The dissemination of false information with the deliberate intent to deceive or mislead.*

Misinformation: *The dissemination of false information, either knowingly or unknowingly.*

Throughout this study, we will focus on the malicious use of information, when information is used to mislead and deceive. In practice, misinformation is often understood as only the unintended dissemination of false information. In order to avoid confusion, this is how the term is used throughout this study. Disinformation is comprised of two elements: the falsehood of the information, and the clear intention to mislead.² The term is modelled after ‘dezinformatsiya’, a Russian term first coined by the KGB³ to refer to the use of false or otherwise misleading information that is purposely provided to selected audiences to influence their behaviour.⁴ This was part of the broader set of tactics called ‘active measures’, which frequently involved ‘attempts to deceive the target (foreign governmental and non-governmental elites or mass audiences), and to distort the target’s perception of reality’.⁵

However, contemporary attempts at disinformation are not just the revival of an old Soviet strategy. The means offered by contemporary communication practices magnify the effects of disinformation. Old-style disinformation devoted considerable care to the crafting of false stories, while today the focus is on quantity rather than quality. Contemporary disinformation is like a Lernaean Hydra: one story may be discredited, but many more will appear.

CONTEMPORARY CHALLENGES: MISINFORMATION, DISINFORMATION, AND INFORMATION ACTIVITIES

False information on social media⁶ has gained enormous popularity over the last year, but it has rarely been framed in terms of information activities by the mainstream media. This issue affects audiences from all facets of the political spectrum.⁷ The expression most widely used to refer to the phenomenon in this context is 'fake news'. The expression quickly gained popularity, so much so that even those who were accused of spreading fake news in the first place started using the term to describe the accusations themselves.⁸ Media outlets have been accused of spreading disinformation from their inception. It is therefore unsurprising how many observers have resisted contemporary concerns about the emergence of fake news.⁹ However, the threat today is qualitatively different.

This has led to social media platforms, as well as private companies and, occasionally, governments, taking action. Google is particularly involved in this effort. It does so through direct initiatives (e.g. Google NewsLab and the introduction of a fact-checking snippet)¹⁰ and by funding fact-checking networks (e.g. the Poynter International Fact-Checking Network and the First Draft News network).¹¹ However, some studies point out how debunking and fact-checking may be ineffective and sometimes even counterproductive.¹² These initiatives are quick and easy, but they do not get at the root causes of the issue.

It is for this reason that social media companies are exploring other ways to counter misinformation, disinformation, and other information activities. For example, Facebook recognized that 'social media platforms can serve as a new tool of collection and dissemination for [information activities]'. In fact, '[t]hrough the adept use of social media, information operators may attempt to distort public discourse, recruit supporters and financiers, or affect political or military outcomes.¹³ Because of this, Facebook states that countering information activities is a priority for the platform. It claims to be doing that by 'collaborating with others to find industry solutions [...], disrupting economic incentives [...], and] building new products to curb the spread of false news'.¹⁴

THE SOCIAL CONTEXT

In NATO doctrine, the information environment is composed of three domains: the physical, the virtual, and the cognitive/psychological. Our perception of the world is constructed in these domains; as noted by R. Waltzmann, 'the Internet and social media provide new ways of *constructing realities* for actors, audiences, and media'.¹⁵

The social context that underlies the rise of false and misleading information on social media is labelled by some to be a 'post-truth' environment. As outlined by *The Economist*, '[t]here is a strong case that, in America and elsewhere, there is a shift towards a politics in which feelings trump facts more freely and with less resistance than used to be the case'.¹⁶ Italian semiotician Umberto Eco famously claimed that, while social media can support the democratization of authoritarian regimes, they also give voice to 'legions of imbeciles'.¹⁷ The spread of disinformation over social media would not be possible without a suitable habitat. Eco's statement is a provocation that highlights how information on social media is left without qualified gatekeepers, people who can take responsibility for what is published.

The threat posed by misinformation and disinformation may affect social stability. In its 2016 report on global risks, the World Economic Forum described the phenomenon of the '(dis)empowered citizen': 'individuals feel empowered by

changes in technology that make it easier for them to gather information, communicate and organize', while at the same time feeling increasingly 'excluded from meaningful participation in traditional decision-making processes'.¹⁸ Disinformation aims at destabilizing society by exploiting these emerging dynamics, many of which take place on social media.

According to a recent study published by Al Jazeera, 'The explosion of social media can be both a blessing and a curse for journalists. It has made anyone [...] a potential witness or source; it has allowed people to tell stories from places where journalists are not present, or where they cannot easily go. Yet it comes with its own problems, problems that can be boiled down to a single question: How can you trust what you see online?'¹⁹ Western audiences rely heavily on social media to get their news. This is evidenced by a recent survey highlighting how the majority of American adults use social media as their primary source of information.²⁰ While the democratization of the informational space brings indisputable positive effects, it also leaves society more vulnerable to manipulation. Tailored social media content generates 'information bubbles'²¹ where voters see only stories and opinions suiting their preferences and biases—ripe conditions for [...] disinformation campaigns'.²² Manipulation is carried out by influencing the way information is processed by the human brain. In the absence of qualified gatekeepers, these techniques can be exploited to their full extent. The dissemination of false

information is inherently linked to wider dynamics, such as ‘increasing mistrust of people with respect to institutions, to the increasing level of functional illiteracy²³ [...], as well as the combined effect of confirmation bias’.²⁴

Confirmation bias, i.e. the tendency to interpret new information as evidence that confirms one’s existing beliefs, is the underlying mechanism that allows misinformation and disinformation to flourish. Social psychology indicates a number of other cognitive biases that adversarial actors can capitalize on. Among other things, information is perceived to be valid when:

- The subject is **repeatedly** exposed to a specific statement²⁵
- The information has been encountered **previously** by the subject²⁶

Moreover, a number of factors make a subject less likely to analyze a piece of information carefully before making a decision regarding its validity. Among them:

- The subject’s perceived **familiarity** with the subject at hand²⁷
- The level of **interest** in the topic: the less a subject is interested in the topic, the less likely he/she is to accurately analyze information²⁸

These cognitive biases inform the tactics that enable acts of disinformation in cyberspace.

TACTICS OF DISINFORMATION ON SOCIAL MEDIA

Broadly speaking, tactics used to spread disinformation on social media share the same desired outcome, i.e. manipulating public discourse. Facebook lists three major tactics employed by malicious actors to conduct operations on their platform:

1. Targeted data collection
2. Content creation (false or real)
3. False amplification (coordinated activity by inauthentic accounts)²⁹

This study focuses on the first and second point.

For journalists, the distinction between true and false statements is important. However, stories are often valued more for their psychological impact and than for their intrinsic value. When stories are designed to be part of a broader effort, the primary objective can be influencing selected audiences, which is achieved via the following activities:³⁰

- Increasing the target’s suggestibility
- Gaining control over the target’s environment
- Creating doubt
- Creating a sense of powerlessness
- Creating strong emotional responses in the target
- Heavy intimidation

Tactics	Platforms	Desired outcome
Broad data collection	Blogs, Friendship-based networks	Retrieving public information in order to conduct audience analysis and deliver targeted content
Targeted data collection	Friendship-based networks	Retrieving non-public information on selected individuals in order to expose it ³¹
False content creation and spreading	Friendship-based and Follower-based networks	Inject selected narratives in public discourse, confuse, (possibly) reflexive control ³²
Emotional content ³³ creation and spreading	Friendship-based and Follower-based networks	Inject selected narratives in public discourse, ³⁴ confuse, (possibly) reflexive control
Saturating the information environment, informational fog	Mainly Follower-based networks	Silence targeted discussions, confuse, divert attention ³⁵
False amplification	Mainly Follower-based networks	Increase reach and perceived credibility of selected content ³⁶
Impersonation (people)	Mainly friendship-based networks	Psychological manipulation of selected targets into performing actions or divulging confidential information ³⁷
Impersonation (organizations)	Mainly Follower-based networks	Inject selected narratives in public discourse, confuse, (possibly) reflexive control

Manipulation on social media can be channelled into any of these activities. It is important to highlight that these activities are not compartmentalized; on the contrary, several activities can be pursued at the same time and in synergy. This is done through social media-specific tactics, summarized in the matrix above.

The outlined tactics can be countered by NATO and its Allied countries by adapting established procedures to the evolution of disinformation. The manipulation of public discourse through social media stands out among the challenges emerging in the information environment. Countering

disinformation on social media is a subset of the general counter-propaganda effort. This highlights how we should look primarily at the social implications rather than the technical details. A considerable number of counter-strategies are currently focused on the latter, sometimes neglecting social dynamics. For example, targeted counter-efforts³⁸ don't take into consideration the fact that, when compared to the appeal of emotional content, logical argumentation has little power when it comes to countering the spread of disinformation online.³⁹ Analogously, automatic fact-checking applications⁴⁰ invariably stumble on the same obstacle, i.e. the fact that the

people who will download and install these applications are generally not those who are most vulnerable to propaganda in the first place.

STUDY OUTLINE

This study provides a look into what can be done to counter the problem of disinformation on social media by analysing more closely the various facets that compose it. The study is organized as follows. Chapter 1 frames the issue of false information on social media in the context of the existing military doctrine on disinformation. Chapter 2 outlines a conceptual map describing how disinformation differs across various social media platforms. The following chapters take a look at what may be the Achilles' heel of any strategy involving the use of so-called fake news, i.e. the link between social media and external content. Chapter 3 looks at blogs specifically, and how they are used in concert with social media to spread misinformation and disinformation online. Chapter 4 explores the third-parties tracking user behaviour on internet outlets associated with the spread of false and misleading information. The conclusion brings together the findings of the study, highlighting recommendations and delineating possible counter-strategies. The study is complemented by a glossary that incorporates both NATO-approved definitions and, for those terms not currently present in NATO doctrine, definitions developed by subject-matter experts and other sources.

1

PLATFORMS: GEOGRAPHY, CULTURE, LANGUAGE

Giorgio Bertolin, Katerina Sedova

Different platforms dominate different cultural-geographical areas. Different networks lend themselves to different exploitation tactics. Social media companies are aware of the impact that disinformation planted online has on public discourse, and have come up with some countermeasures. However, it remains to be seen the extent to which these countermeasures are effective. Russian-speaking internet users prefer Russian-made social media platforms. These platforms are qualitatively different from their Western counterparts, and can be used more effectively in disinformation campaigns by malicious actors.

INTRODUCTION

In this section we will describe the specificities of misinformation and disinformation across different social media platforms. We will provide an overview of the challenges encountered by major social media platforms, and of what the platforms themselves are currently doing to curb the spread of mis- and disinformation.

A recent study found that '[t]he rapid growth of social networks caused them to become ideal platforms for spreading disinformation campaigns (...) [t]o spread fake news, it is necessary to promote it to social media users'.⁴¹ This chapter will adopt a 'microscopic' perspective, looking at the characteristics of current major social networks.

We must remember that the social media landscape has a transient nature.

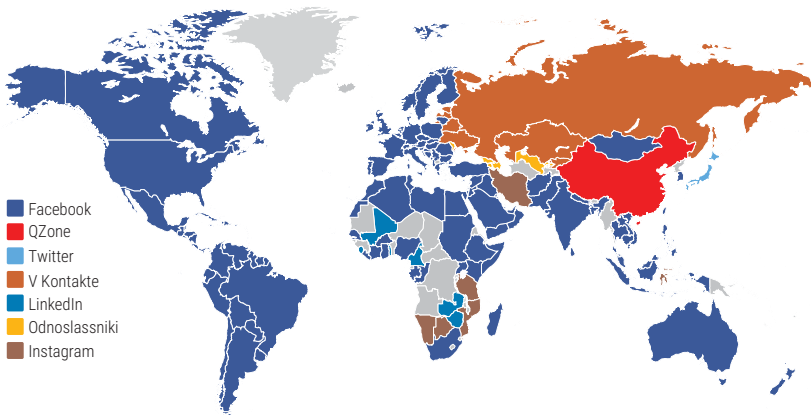
The platforms analysed in this chapter are those that are most relevant today, but they will not necessarily retain their positions in the future. One just needs to consider the fall of social media giants, such as MySpace⁴² to be reminded that the popularity of social media platforms is not set in stone.

GEOGRAPHY

A mere quarter century since the World Wide Web entered the public domain, 3,77 billion—more than half of the world's population—is online. As of 2017, 2,8 billion people are using social media, and the pace of growth continues to accelerate.⁴³ The following maps show the regional nature of the world's leading platforms.⁴⁴ They depict, respectively, the first and second most popular platforms in the countries surveyed. Facebook is dominant in the Western

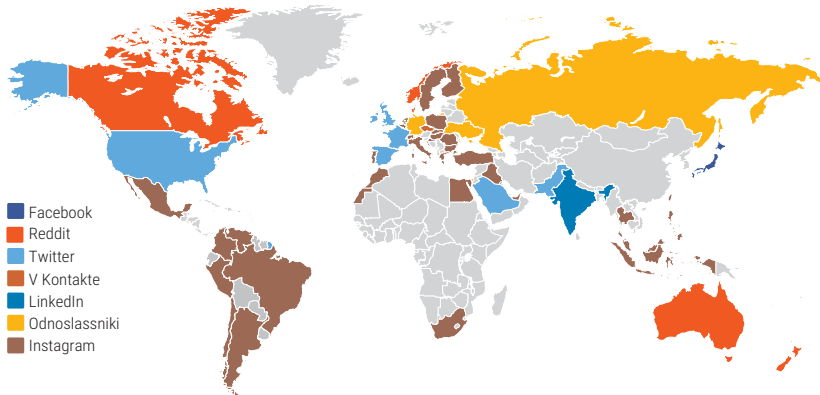
WORLD MAP OF SOCIAL NETWORKS

January 2017



WORLD MAP OF SOCIAL NETWORKS

Ranked 2nd - January 2017



world, South America, the Middle East and North Africa (MENA) region, and all English-speaking countries. Reddit is particularly popular in English-speaking countries such as Canada, Australia, and New Zealand. And while Twitter has considerable traction in the West, its role is less important in other world regions.

FACEBOOK

Facebook is currently regarded as the leading friendship-based network, at least in the Western world. In this capacity, this platform is the most valuable target for disinformation campaigns. Recent concerns about the weaponization of false information were focused on Facebook. Many observers pointed to Facebook's role in exacerbating the negative aspects of the kinds of social dynamics that facilitate the spread of mis-/disinformation. In particular, the algorithms behind the selection of stories on Facebook's homepage were accused of worsening the echo chamber

phenomenon, where the Facebook user is offered information of a nature similar to what he himself produces.⁴⁵ To the use of its platform to spread mis- and disinformation, Facebook responded by developing reporting and flagging procedures.⁴⁶ These efforts have received mixed appraisals. Most of the measures simply don't work, mainly because they do not take into account cognitive biases such as the 'continued influence effect'.⁴⁷ One critic pointed out that exposing false or misleading information in stories and/or accounts is useful only as a whitewashing maneuver: 'It's ultimately a kind of PR move. It's cheap to do. It's easy. It doesn't actually require them to do anything.'⁴⁸ Yet, it is important to note that 'disinformation campaigns happen largely outside of Facebook's control'⁴⁹—what happens on Facebook is a symptom of broader dynamics in society at large. While the company's efforts might curb the spread of disinformation,



Researchers can monitor only the tip of the iceberg, i.e. public pages and user groups.


it cannot fight it directly. Facebook cannot be tasked with countering disinformation: this task does not fall within the responsibility or the competence of a social media company. As Bounegru writes, 'Facebook's architecture poses challenges to the study of circulation of content due to the nature of its access and permissions system'.⁵⁰ Therefore researchers can monitor only the tip of the iceberg, i.e. public pages and user groups. Users who subscribe to groups and public pages receive updates whenever new posts are published. In this context two scenarios are possible:

1. Users subscribe to pages spreading mis-/disinformation because the content resonates with them.
2. Users subscribe to pages that share 'neutral' content that resonates with them. These pages can potentially spread disinformation at a later stage.

However, the biggest impact is that of stories that quickly gain popularity and

are casually encountered by users, either because their Facebook contacts are sharing these stories or because they are being promoted by Facebook's algorithms, or a combination of both. These stories that 'go viral', either genuine or misleading, can have a considerable effect in promoting selected agendas, especially when the content is tailored to specific target audiences. This practice is blossoming, especially in regard to the concerns the commercial sector, although some companies claim to have applied the same methodology to fit the specific needs of political campaigning.⁵¹ Regardless of the veracity of this claim—which is difficult to assess, since said companies refuse to share their methodologies and measures of effect—the possibility of applying commercial principles to political purposes must be considered. Adversarial actors can take advantage of this while hiding behind business-driven and relatively anonymous entities.

As with all friends-based networks, users perceive Facebook as a familiar environment. This presents a clear risk:



On Twitter, manipulation of the information environment has occurred through the trending topics feature and hijacking or clotting hashtags.


if a Facebook user's contact shares a story, the user will likely assume that the contact is vouching for that piece of information, which may not be entirely true. Moreover, some users share stories without double-checking their veracity, particularly if they are not tech- or fake news-savvy.

Grassroots attempts to counter misinformation and disinformation on Facebook have focused on debunking false stories. Websites like Snopes feature efforts to investigate the veracity of posts shared on social media, primarily Facebook. However, research shows that these efforts not only do not reach their stated goals, but might actually make the situation worse. First, very few of the users exposed to unsubstantiated claims 'actively interact with the corrections'.⁵² Second, these users seem to be more active within their own echo chamber after they have come in contact with a correction, suggesting a hardening of their initial beliefs.⁵³

TWITTER

Twitter has become particularly important for political manipulation, since it is the platform of choice for many traditional media outlets, politicians, and other opinion leaders. A considerable number of quantitative studies on social media use Twitter as their testing ground since, compared to other platforms, Twitter presents more publicly available data.

As early as 2010, Chamberlain observed how 'the proliferation of disinformation capabilities represented by Twitter will almost guarantee that users of social networks will be exposed to disinformation'. Users are at risk of being 'manipulated by any organisation that cares to develop an information operations capability'.⁵⁴ The same author attempted an explanation for why Twitter is a favourable environment for disinformation:



The world of social media is in constant flux, and it is therefore necessary to monitor developments in order to stay current with emerging platforms and cross-platform trends.

Twitter messages can seem credible without containing any references to support their claims. The short length of tweets encourages short declarative statements absent of supporting arguments and thus users do not become suspicious of unreferenced assertions. The fact that in some instances Twitter has been the primary source of news about a currently unfolding event also gives it some inherent credibility.⁵⁵

On Twitter, manipulation of the information environment has occurred through promoting a certain idea via the trending topics feature, and through suppressing certain ideas or discussions through hijacking⁵⁶ or clotting hashtags⁵⁷. Both forms of manipulation oftentimes make use of networks of fake Twitter accounts. These fake accounts are often automated or partly automated—robotic activity is what plagues Twitter most. While ‘bot

accounts’ are present on other platforms, it is here that this potentially malicious technique demonstrates its reach.⁵⁸ Up to 15% of all Twitter users are in fact automated scripts that mimic human behaviour with growing sophistication.⁵⁹ This number can grow considerably in certain specific contexts. As recently highlighted by our Centre of Excellence, nearly 70% of all Russian-language Twitter accounts posting about NATO in Poland and the Baltic states are in fact automated scripts.⁶⁰ These networks of automated accounts (or ‘social bot networks’) can considerably boost the reach of disinformation.

Any efforts by Twitter to curb the diffusion of false and misleading stories are likely to be channelled towards reporting abusive content and ‘fake news’, as Facebook already does.⁶¹ However, no such measures are in place as of this writing. The criticisms directed toward Facebook’s strategy can be applied to Twitter as well. Moreover, the main problem affecting this platform is the proliferation of automated activity.



Micro-platforms exacerbate the echo chamber phenomenon.

Any measure aimed at curbing it is likely to have a sanitizing effect on the overall Twitter environment.

Just like Facebook, Twitter is a battleground for companies selling various degrees of targeted messaging fuelled by audience analysis. However, Twitter offers less personal information than Facebook. The main reason is that profile descriptions on Twitter are not as codified as they are on Facebook. A special category of metadata is dedicated only to geo-localization and external URL links. Therefore, collecting information on these profiles is a process that revolves around various types of secondary analysis,⁶² meaning that the final result is less refined what can be obtained from Facebook.

OTHER PLATFORMS

Beyond Facebook and Twitter, the reach of any other platform in the Western information environment is limited. However, the impact of the major platforms is not necessarily directly proportional to their reach. The world

of social media is in constant flux, and it is therefore necessary to monitor developments in order to stay current with emerging platforms and cross-platform trends. Moreover, existing platforms are not separate worlds. They exist in interconnection with each other—a story that originates on YouTube can be shared on Facebook, then picked up by a newspaper website; the article can then be shared on Twitter and might generate a thread on Reddit, and so on. In this context, YouTube occupies a distinctive space.

Youtube it is a fundamental element in the virtual space where social media exists—'Web 2.0'. A considerable number of stories shared on social media originate here, and misleading stories are not an exception. Conspiracy theories have been thriving on YouTube since the early days of the platform.⁶³ This is because YouTube was and is regarded as a medium that allows the distribution of 'alternative' information, while at the same time being a mainstream information source for millions of

follow in the wake of conspiracy theories. Politically slanted channels that distribute false content prosper on YouTube and are shared on social media, where they reach and attract larger audiences. It must be noted that, as is the case for the creation of content on other mediums, most creators are motivated by financial gain. In order to fight the spread of misleading content YouTube fosters media literacy campaigns.⁶⁵ YouTube has also modified its terms of use by implementing a new review process for its Partner Program. Since April 2017, YouTube channels cannot generate revenue until their videos reach 10,000 views.⁶⁶ This higher threshold is supposed to give YouTube 'enough information on the validity of a channel'.⁶⁷ However, this cannot do much to deter state actors, or proxies of state actors who get their funding from other sources or who are not motivated by financial gain.

While Instagram seems to be relatively immune to the worst aspects of disinformation-spreading on social media, it is by no means a safe space. Among the main issues affecting this platform are impersonation and spamming.⁶⁸ Instagram's guidelines prohibit 'artificially collecting likes, followers, or shares, posting repetitive comments or content, or repeatedly contacting people for commercial purposes without their consent'.⁶⁹ The fact that guidelines are tailored for spam marketing suggests that manipulation of the information environment for political purposes is not yet an issue for Instagram.

Recently, self-styled 'alternative' platforms have burgeoned, aiming to circumvent the policies of mainstream platforms, in particular the perceived censorship carried out by Facebook and Twitter towards inflammatory and controversial speech. These micro-platforms⁷⁰

JANUARY 2017

DIGITAL IN THE MIDDLE EAST

KEY STATISTICAL INDICATORS FOR THE REGION'S INTERNET, MOBILE, AND SOCIAL MEDIA USERS



246
MILLION

68%



147
MILLION

60%



93
MILLION

38%



312
MILLION

127%



83
MILLION

34%

Data and design from: *Digital in 2017: Global Overview (We Are Social, 2017)*

exacerbate the echo chamber phenomenon. Their reach is limited, but not so limited as to be negligible: as mentioned above, stories easily jump from one platform to another. In a disinformation campaign, targeting groups that are particularly receptive⁷¹ on an alternative platform can serve as an intermediate stage through which selected narratives can be passed on to mainstream networks. Micro-platforms can act as accumulators, where hostile narratives are free to flourish because of the absence of any significant obstacle.⁷²

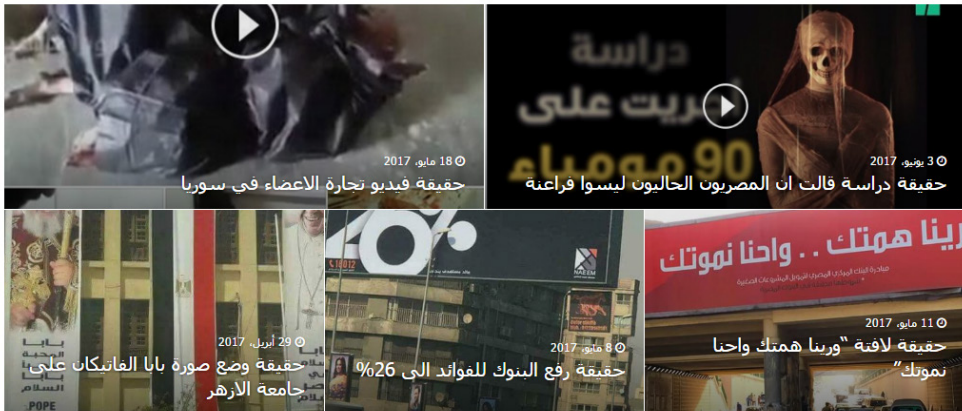
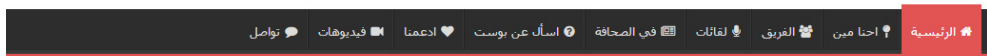
THE ARAB WORLD AND THE MENA REGION

Arabic is the fourth most common language online;⁷³ the MENA region produces a wide range of Arabic-language material, but, as previously noted, this material is not shared on native platforms. Rather, Arabic-

language users are particularly active on Facebook, Twitter, and Instagram.

Although revolutions do not take place on social media, social media played a prominent role in the Arab Spring⁷⁴ in 2010. Tunisian protesters famously used Twitter to make their voices heard in the Arab world and beyond.⁷⁵ As of January 2017, there are 93 million active social media users in the Middle East alone (see Figure on page 18).⁷⁶

Facebook is the most popular social network in the Arab world. 87% of social media users have a Facebook account. Of these, 89% access the platform on a daily basis.⁷⁷ As evidenced in a recent report published by Al-Jazeera, Facebook is ‘the first platform to consider for newsgathering, but also for *storytelling and audience engagement*’.⁷⁸



SCREENSHOT OF DA BEGAD'S HOMEPAGE (14 JUNE 2017)

البوست

قد تكون حقيقة صادمة! .. المصريون الحاليون ليسوا فراعنة!
صدق أو لا تصدق لكن هذا ما تقوله دراسة علمية على 90 مومياء



الحقيقة

الدراسة ماقلتش كدة تماما.

هما توصلوا ان المصريين اللي موجودين حاليا عندهم جينات افريقية اكثر ب 8% من الجينات الأفريقية اللي كانت عند قدماء المصريين ، و ان قدماء المصريين كان فيهم جينات اوروبية و شرق اوسطية اكثر من الافريقية وده عكس ما كانوا متوقعين.



Throughout the MENA region, social media are still perceived as an alternative, relatively independent source of communication.

Content creation and content diffusion are tactics that bring best results when they are used in concert. Analogous to what happens in other parts of the world, in the MENA region there is a 'range of services available to anyone looking to distribute fake news and launch public opinion manipulation campaigns'.⁷⁹ Companies like 'Dr Followers' and 'CoolSouk' offer a wide range of services aiming at boosting the popularity of social media posts.⁸⁰ These services target popular platforms including Twitter, Facebook, Instagram, and YouTube. The activity of these companies can considerably increase the visibility of counterfactual stories, which, as elsewhere, can be spread for political purposes.

Fact-checking efforts are present in the MENA region as well. The Egypt-based website Da Begad (see the screenshot of Da Begad's homepage from 14 June 2017 on page 20) debunks false stories found on social media.⁸¹ Analogous to similar initiatives in other parts of the world, Da Begad is

maintained by a team of volunteers that claim to be independent from any political affiliation; this team relies on crowdsourcing for reporting.⁸²

Da Begad's graphic concept is very basic. First, the disputed story is introduced in a section called 'The Post' (since all of these stories are found in social media). Then a brief analysis of the contested claims is given in a section called 'The Facts' together with the necessary references.

There are features to suggest that the spread of false and misleading information over social media is aided by automated activity. During the 2017 diplomatic crisis involving Qatar, researchers collected evidence pointing to automated activity in support of information attacks against Qatar: the Twitter hashtags that advocated cutting off relations with Qatar 'originated in Kuwait and spread fast, suggesting heavy bot usage', while the response hashtag, in defence of Qatar, 'increased gradually, without the kind of significant peak its

rival hashtag experienced. Anti-Qatar hashtags seem to be more organized and suggest advance preparation'.⁸³

Non-state actors are particularly apt at combining audience characterization with aggressive information activities. Groups like Daesh are 'particularly successful in targeting tech savvy, impatient and respect-seeking millennials (...). They know how and what they think and feel, how they want to be perceived and how they wish to receive information'.⁸⁴ This allows the terrorist organization to spread emotional content and biased stories, mostly focused on praising the Caliphate utopia.⁸⁵ These activities are carried out prominently on social media, both in the MENA region and

among Arabic speakers in the West. It is for this reason that some Arab States decided to strengthen the government's ability to monitor and curb the use of social media by violent groups. Bahrain, Egypt, Lebanon, and Kuwait have enforced legislation to address this issue.⁸⁶ In some cases, this entails directly targeting the most vulnerable demographic group, i.e. young men. The Kuwaiti government collaborates actively with the Kuwait University's Media Department in a research project aimed at detecting early signs of youth radicalization on social media.⁸⁷ These efforts are complemented by those of supra-national entities. The Global Coalition's Information Cell developed an audience characterization system that



exploits interactive videogames to detect potential supporters of Daesh.⁸⁸

The issue is not limited to Arab countries alone. As part of its fight against pro-Palestinian violent political extremism, Israel is working towards compelling social media companies (namely Google and Facebook) to curb support for groups and pages that spread extremist messages. Israel's new counterterrorism law 'established a new criminal offense for demonstrating solidarity with a terrorist organization or with an act of terrorism, and incitement to terrorism, *including via the internet and social media*'.⁸⁹ Since social networks, through extremist propaganda, are widely regarded as catalysers for radicalization, governments throughout the world are pressuring social media companies to take action, but whether this will result in concrete actions is questionable.⁹⁰

Throughout the MENA region, social media are still perceived as an alternative source of communication, relatively independent from the constraints imposed on traditional media by state authorities (on page 22, a cartoon published by Al-Jazeera's website light-heartedly illustrates this point - TV screen is captioned with the label 'Authority's Media').

However, governments in the region are quickly weaponizing new media. For example, Iran is believed to be creating bogus online personas to carry out targeted attacks.⁹¹ As evidenced above, the online environment in the MENA region suffers from the same issues encountered in the West, not least because the most

popular social media platforms are the same. The following sections will analyse a region where the social networks that rank highest in popularity are relatively unknown to the rest of the world.

SOCIAL MEDIA ON RUNET

RuNet (or the 'Russian Internet') continues to be dominated by home-grown social networks, as populations in Russia and many nations of the former Soviet Union are mainly active on VK, Odnoklassniki, and MoiMir. To understand RuNet's social media space, one must understand the domestic origin of Russian disinformation, its weaponization for geopolitical goals, and the consequences for Russian-owned social networks.

[VK.com \(VKontakte.com\)](#)⁹²

As of January 2017, VK reports 90 million monthly active users with almost 70% of them living in Russia.⁹³ Founded in 2006, VKontakte intended to connect university students. The network continues to attract a younger audience in comparison to other social networks, with its largest user demographic group under the age of 35.⁹⁴

VK is known for its uncluttered user interface, focus on communities, and entertainment—the source of its high audience **engagement**. While a typical friends-based network, several features differentiate it from Facebook. The rich built-in image modification features allow VK users to easily overlay images with text for meme creation. Music and

video sharing—sometimes in violation of existing regulations—are central to VK’s success and continued active audience engagement.

The detailed information codified in its profile questionnaire enables a powerful search function, which makes it easy to find specific people, and locale- and interest-based groups. A phenomenon specific to VK, that does not exist in most Western social media, is relying on local groups to spread information. Towns and other geographical areas have local group pages with thousands of participants, who share pictures, videos, and eyewitness accounts.⁹⁵ Such fine-grained social connectivity presents fertile ground for disinformation.⁹⁶

VK’s weak **privacy and security** settings make its users vulnerable to disinformation. Its API and user data protection allow micro-targeting. In 2017, VK added several features that enhance its advertising platform and make it more vulnerable to misinformation and disinformation. One of these features allows advertisers to display shortened web addresses to streamline the appearance of their ads by obfuscating the destination address. This hinders a user’s ability to identify the source of the posting and to critically evaluate the link before clicking it, thus increasing users’ vulnerability by encouraging them to unknowingly navigate to sites that may be malicious. Lax **security** measures further enable registration of mass accounts, making VK the cheapest platform for the creation of bots, which are used to amplify disinformation.⁹⁷

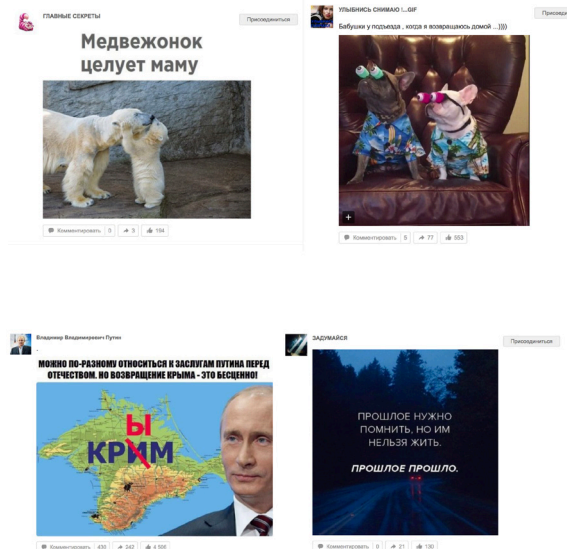
OK.ru (Odnoklassniki.ru)⁹⁸

Odnoklassniki (OK) is the second most popular social media network in Russia and the nations of the former Soviet Union, with 40 million registered users in Russia and 65 million overall.⁹⁹ OK.ru users are more likely to be women and over 30 than users of any other Russian social network.¹⁰⁰ OK has the typical features of other friends-based networks—personal profiles, chats, discussion boards, and the functions that make status updates, sharing pictures, and searching for friends possible.

OK’s focus is on **engagement** through entertainment. Personal feeds are flanked by social games, ads, and banners showing the most popular songs and videos trending on OK.¹⁰¹ To compete with VK, whose foray into multimedia content contributed to its early lead in popularity, Odnoklassniki launched its own video, TV, and cinema service. Unlike VK, the service allows users to watch TV shows online from STS and TNT—two popular Russian entertainment TV broadcasters. This is a key method driving continued user engagement within the network.¹⁰² The mastery of state-controlled Russian TV in blending the Kremlin’s narratives with entertainment is well documented.¹⁰³

Image consumption and manipulation are central to the user experience. Numerous options facilitate custom framing ‘postcards’ and rating pictures. Before users log in, the default feed displaying trending content showcases the essence of OK—moralistic memes,

TYPES OF MEDIA TRENDING ON OK.RU




jokes, cute pictures, and the ubiquitous cat videos, which are occasionally interlaced with propagandistic imagery of the Russian leader.¹⁰⁴ This feature relies on a vulnerability that can be exploited: if the feed can be manipulated to push content, public opinion may be influenced through the juxtaposition of propagandistic imagery with emotional content, just as purchasing habits may be influenced through advertisement.

Odnoklassniki supports an extensive set of search parameters, which enhance **discoverability**. The network is home to many large user-generated communities, counting millions of followers from the RuNet area.

Moi Mir (my.mail.ru)

The third most popular Russian social network is Moi Mir,¹⁰⁵ a property of the Mail.Ru web service, akin to the Google+ and Gmail ecosystem. As of 2016, Moi Mir claimed 25 million users.¹⁰⁶ Moi Mir users set up their accounts through a mail.ru e-mail address, which serves as a single sign-on into the mail.ru universe and Moi Mir. The lack of an SMS verification requirement to complete registration on Moi Mir exposes the site to anyone who wants to easily create non-genuine accounts.¹⁰⁷ The key to Moi Mir's popularity is the rich social gaming and video-sharing experience it provides. The platform was recently augmented with digital TV content from STS and TNT. This move is consistent



If the feed can be manipulated to push content, public opinion may be influenced through the juxtaposition of propagandistic imagery with emotional content.


with the goals of Mail.Ru Group, which owns both Odnoklassniki and Moi Mir¹⁰⁸ Once the integration has been put in place, Moi Mir will be vulnerable to the tactic of blending entertainment with disinformation, masterfully honed by Russian State TV.¹⁰⁹

The search functionality of Moi Mir has many features in common with VK and Odnoklassniki. However, unlike the other two platforms, many of Moi Mir's search parameters are based on physical appearance, or the user's 'chronotype', signifying waking and sleeping patterns. Like most dating sites (and VK), Moi Mir offers a 'last active/last visited' indicator to reveal the levels of engagement.¹¹⁰ While this facilitates making connections, such information exposes personal details that a malicious actor might use. The 'dating-like' atmosphere can create an illusion of intimacy which may make users vulnerable to a specific type of hybrid troll, those featuring attractive individuals.¹¹¹

The RuNet evolution

Nationwide, Russian Internet penetration grew from 5% in 2002 to 73% in 2017.¹¹² Why are Western social media platforms lagging so far behind in Russian-speaking communities? In a comparative study of Facebook and VK, researchers explained this phenomenon through the concepts of platform and culture. The users interviewed expressed their appreciation for the user-friendly minimalistic interface unencumbered by advertisements, access to engaging content through audio and video sharing, trustworthiness, and fun. Although VK exists in 70 languages, it dominates among Russian speakers, who can connect with each other in Russian, on the basis of entertainment, cultural humour, and pride for being on the 'made in Russia' platform.¹¹³

Since 2011, Facebook has slowly gained share in certain demographic segments. First, Facebook users in Russia are on average older, more educated, and earn a higher income than VK users.¹¹⁴ Second, Facebook's foothold in Russia



The 'dating-like' atmosphere can create an illusion of intimacy which may make users vulnerable to a specific type of hybrid trolls, those featuring attractive individuals.

is increasing as it is used for business communication and maintaining professional contacts outside of the RuNet ecosystem.¹¹⁵ However, the most significant factors contributing to the increase of genuine Facebook users in the RuNet ecosystem are the domestic political climate in Russia and the geopolitical adventurism of the Kremlin, as dissidents and the politically engaged are fleeing increased government control.

RuNet, and VK in particular, was once a pivotal medium for political engagement and a dynamic platform for political discourse, organizing protests, and other activism. The 'colonization of RuNet' by the state political technologists since the 2000s occurred gradually in several stages.¹¹⁶ As traditional media became the target of government control,¹¹⁷ serious political reporting moved to social media that were free from interference, particularly blogs,¹¹⁸ where the culture of skepticism required elaborate proof of one's assertions.¹¹⁹ Coupled with the boom in the technology

sector encouraged by the Kremlin,¹²⁰ RuNet was allowed to flourish, empowering the development of local social networks. VK, Odnoklassniki, and other popular online media were founded by mid-2000s. President Putin's early disinterest in RuNet ensured its freedom through 2010,¹²¹ when the first large scale disinformation campaign in Russia was launched in support of Medvedev's re-election campaign. It deployed pro-government botmasters and trolls, recruited from pro-Kremlin youth groups such as 'Nashi', to comment on opposition blogs and to repost pro-government messages.¹²² The trolls responded to the highly skeptical blogging culture by faking detailed, believable proofs in support of their false narratives.¹²³

These early attempts at disinformation, augmented by a rapidly maturing spam industry and search optimization, emphasized tactics that went beyond reposting and retweeting to manipulate popular rankings with engaging, viral content.¹²⁴ In their disinformation efforts, the leaders of 'Nashi' became



RuNet, and VK in particular, was once a pivotal medium for political engagement and a dynamic platform for political discourse, organizing protests, and other activism.

fixated on creating professionally produced, visually engaging content that could go viral and beat the trending topic algorithms.¹²⁵ In 2011, the political environment changed; in reaction to Putin's announcement that he intended to run for president, mass street protests occurred, and continued for several years, but by that time the infrastructure for a disinformation machine was in place.

RuNet entered a pivotal stage in 2011, one of increasing government censorship fueled by the Kremlin's push to silence domestic opposition. During this stage, the local bloggers bore the brunt of the state's displeasure as it expanded the legal definitions of 'extremist' content. New censorship empowered arbitrary banning of local and foreign websites without explanation, and required bloggers with audiences of over 3,000 readers to register with the government as mass-media outlets. As Odnoklassniki and Moi Mir were already properties of Mail.ru and owned by Alisher Usmanov, a close Kremlin ally, compliance to government

pressure likely occurred quietly. The storm over Pavel Durov's VK illustrates the Kremlin's interest in social networks. In a move reportedly orchestrated by Putin's close ally Igor Sechin, the founder and president of VKontakte sold his shares and the platform was taken over by allies of the Kremlin.¹²⁶

The case of Ukraine

Since the Euromaidan protests of 2014, Ukraine has been on the front line of multiple disinformation campaigns.¹²⁷ The most successful planted stories used emotionally compelling images and video content of unrelated events and geographic locations as 'evidence' of Ukrainian misdeeds.¹²⁸

In 2014, a quarter of the Ukrainian public had an account on VK, with a large percentage using social media as its main news source.¹²⁹ The Ukrainian government prohibited these networks,¹³⁰ declaring them tools of warfare and banning access to them in Ukraine.¹³¹ Weeks after the announcement of the ban and before the block was fully implemented, 2.2 million

Ukrainians moved from VK and OK to Facebook.¹³² In Russia, VK's change of ownership and the Kremlin's control over the social media through its close allies sent the educated, politically engaged, and surveillance-weary social media users to Facebook, Twitter, and encrypted messaging platforms such as Telegram.¹³³ The Kremlin continues to threaten shutdowns¹³⁴ and bans on VPNs (Virtual Private Networks).¹³⁵ As the Russian state exerts control over social media companies and coopts them as tools of statecraft, these trends in increased censorship are likely to accelerate.

CONCLUDING REMARKS

All social media platforms share the same vulnerability: their users trust the online environment. They are surrounded by 'friends', and they feel as though they have control over the information they are given access to. However, the threshold for critical evaluation of the information received is considerably lower than for traditional media.¹³⁶ This means that adversarial actors encounter fewer obstacles in the execution of disinformation campaigns than was the case in the past. Many of the tactics that can be applied to disinformation campaigns are based on standard business models. If social media are used in hybrid warfare, escalation from guerrilla marketing to guerrilla warfare becomes a tangible possibility.

A large part of the efforts to counter misinformation and disinformation online consists of debunking initiatives supported by the platforms themselves

and by private organizations. Research shows that these approaches are, at best, well intentioned but ineffective. This is a dead end also for countering structured disinformation campaigns. It is, therefore, vital that countermeasures be grounded in radically different procedures.

Facebook is currently the undisputed leading social media platform in the Western world, Latin America, the Middle East and North Africa, India, as well as a number of other regions. Facebook was recently accused of unintentionally facilitating the spread of disinformation. The company has shown interest in tackling the issue, but it remains to be seen how fruitful these efforts will be.

The most popular followers-based network, Twitter, is just as likely to be a vehicle for disinformation. On Twitter, the use of automated bot accounts demonstrates the full potential of the medium for spreading false information. Some of the structural characteristics of Twitter—concise messages and metadata tags—can be exploited by malicious actors to magnify the reach of selected narratives.

The world of social media is constantly changing and evolving. New platforms emerge, and the old ones keep re-inventing themselves, adding new interface features. Younger audiences are active on platforms that are virtually unknown to their parents. Moreover, different social networks are popular in different regions of the world. Chinese and Russian audiences are active on 'home-grown' social media, over which

their respective governments have a considerable degree of control. In the Middle East and North Africa, the most popular social networks are those used in the West. While the political dynamics are profoundly different, we see evidence of the same vulnerabilities.

Compared to Facebook, RuNet's social media networks have fewer security and privacy protections, and they offer more robust capabilities for discovering people and groups. The features on RuNet drive engagement through entertainment with rich visual content, and video and audio sharing. The integration of TV programming enables passive consumption, bringing platforms closer to the well-honed disinformation techniques of Russia's state-controlled media. Their advertising frameworks, which shape consumer choices, can also be used to influence public opinion with nuanced audience targeting. When weaponized, these legitimate platform features are powerful vehicles for disinformation.

Yet in today's Russia, the ownership of social media platforms and the ability of their policy teams to withstand pressure from the state, are the decisive factors determining the platforms' vulnerability to exploitation. With Russian social media consolidated in the hands of the Kremlin's allies, the lines between the state and social media technology companies have become blurred. The ownership of the networks, their susceptibility to state pressure, enhanced by 'engagement through entertainment' platform model, leave RuNet social media networks—and their users—uniquely vulnerable to mis-/disinformation.

2

BLOGS, FAKE NEWS, AND INFORMATION ACTIVITIES

Nitin Agarwal, Kiran Kumar Bandeli

Blogs provide fertile ground for framing narratives. This chapter demonstrates that aside from the blog post itself, reader comments can make the narrative more persuasive. However, the absence of a social network structure for blogs inhibits the dissemination of these narratives. Social media platforms such as Twitter and Facebook are used as vehicles to disseminate the content using cross-media and mixed-media tactics. The link between blogs and social media platforms is vital for understanding contemporary disinformation campaigns.

INTRODUCTION

Blogs have ushered in an era of citizen journalism that has irreversibly changed the way we consume information, partly supplanting traditional journalism. Blogs have endowed citizens with the power and freedom to express their opinions or frame narratives for a greater audience; readers' comments on blogs afford greater inclusiveness and dialog. Blogs cater to the needs of the public to receive information in manageable chunks, tailored to their individual preferences. They can provide intimate details and live accounts featuring compelling, on-the-ground-style coverage of an event. Together, these two capabilities—news chunking and first-person reporting—can create the capability to orchestrate highly biased, partial, and distorted information, i.e. an information campaign.

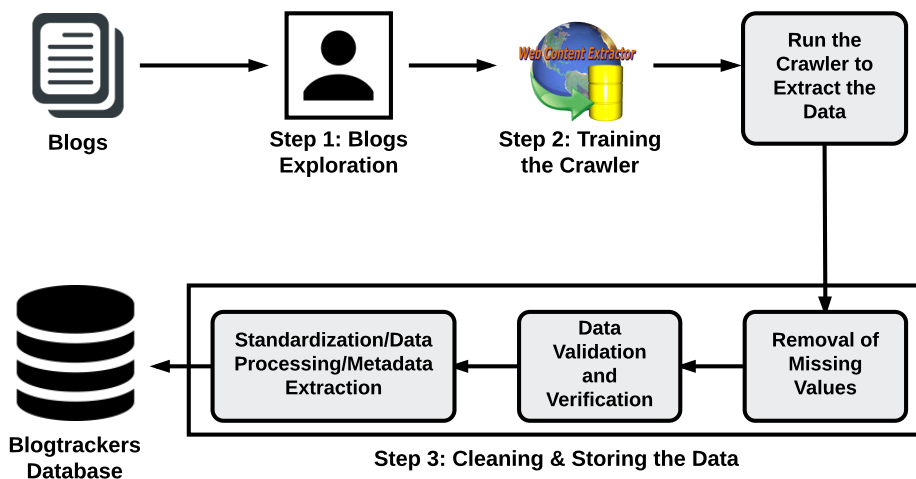
Blogs alone are not effective in conducting information campaigns. Blogs provide fertile ground for framing narratives, but the absence of a social network structure inhibits dissemination. Various social media platforms, such as Twitter, Facebook, and VK, are then used as vehicles to disseminate the content. Nine out of ten bloggers have Facebook accounts. 78% of bloggers use Twitter to promote their content. This percentage is higher, i.e. almost 90% for professional and full-time bloggers.¹³⁷ In addition to bloggers promoting their content, studies have widely reported the exploitation of computer programs,¹³⁸ also known as social bots, to massively amplify content dissemination via Twitter. The ability to

embed YouTube videos, SoundCloud files, and Internet-based memes in blogs has led to unprecedented convenience in framing narratives, disseminating them widely, and driving online traffic to generate a rich conversation around a chosen topic. In addition to content promotion, prolific media integration helps boost search rankings artificially—a technique known as *link farming*, which is a well-known strategy for search engine optimization. Gaming search engines by using prolific linking to blogs has become part of modern information activity. By further examining the information flows within the media networks, we attempt to understand the sources of mis-/disinformation and their reach; if we can detect how far and how quickly the mis-/disinformation can travel, we can also understand the extent to which information is being manipulated. This chapter will present an in-depth examination of the social media networks using a social-network-analysis-based methodology to identify the prominent information brokers and leading coordinators of disinformation campaigns. A methodological section will describe how the data is fetched from different sources, and the approach we propose for studying information flows. The analysis and findings below provide a deep dive into the research questions we set out to answer in this study.

METHODOLOGY

For the purposes of our analysis, we examined several blogs and identified common attributes among them, such as *title, date and time of posting, author/*

THE DATA COLLECTION PROCESS FOR BLOGS



blogger, blog post content, comments, and permalinks. We collected and indexed all blog content from four different blog datasets into our Blogtrackers database. The database can be accessed at <http://blogtrackers.host.ualr.edu/>. The dataset consists of 372 blog sites, 7576 bloggers, and 196940 blog posts riddled with false and misleading information. To crawl these blogs from different sources, we setup crawler(s) for each blog to extract all the required attributes. There are three main steps in crawling data from a blog: (1) exploring the blog site, (2) crawling the blog site, and (3) cleaning and storing the data in a database for analysis and retrieval. Figure above represents the data crawling process for the blogs.¹³⁹

For this study, data was collected from four diverse sources. The descriptions associated with the attributes in these four types of datasets are as follows:

- *Fake news dataset from kaggle.com.* This dataset has 244 blogs, 2236 bloggers, 12,999 posts, and 20 attributes. Some of the key attributes in this dataset are: domain name, site_url, author, post title, text, published date, language, comments, replies_count, shares, and likes. The dataset is available at <https://www.kaggle.com/mrisdal/fake-news>.
- *Dr. Melissa Zimdars' compiled list of fake news blogs.* Dr. Melissa Zimdars, a professor from Merrimack College (<http://bit.ly/2wTMIUb>), compiled blogs featuring fake news. These blog sites are available at <http://bit.ly/2ezvFbV>. This dataset has 37 blogs, 971 bloggers, 96,056 posts, and 79 attributes. The key attributes are: blog name,

blogger, blog post title, blog post, posting date, location, and language.

- *Blogs containing disinformation regarding the Baltic States.* This dataset has 21 blogs, 728 bloggers, 16,667 posts, and 79 attributes. The key attributes are: blog name, blogger, blog post title, blog post, posting date, location, and language.
- *Blogs containing disinformation regarding NATO exercises/activities.* This dataset contains blogs collected by the Blogtrackers tool that posted mis-/disinformation during various exercises conducted by NATO, such as the Trident Juncture 2015, Brilliant Jump 2016, and Anakonda 2016. This dataset has 70 blogs, 3641 bloggers, 71,218 posts, and 79 attributes. The key attributes are: blog name, blogger, blog post title, blog post, posting date, location, and language.

The characteristics of these four datasets are presented in Table below. Next we present the research methodology used to analyse these blogs in order to examine the spread of disinformation.

In this study, we plan to answer the following research questions:

- What are the typical characteristics of mis-/disinformation-riddled blogs?
- Can we track the origins of the content, such as memes, images, etc., appearing in these blogs?
- What strategies are common in disseminating the content (e.g. mixed-media and cross-media)? And, can we identify the other media sites that are predominantly used to disseminate the original blog posts?
- How do antagonistic narratives travel?

Dataset	Number of Blogs	Bloggers	Number of Posts	Attributes
<i>Fake news from Kaggle.com</i>	244	2236	12,999	20
<i>Prof. Melissa Zimdars' compiled fake news blogs</i>	37	971	96,056	79
<i>Blogs containing disinformation regarding the Baltic States</i>	21	728	16,667	79
<i>Blogs containing disinformation regarding NATO exercises/activities</i>	70	3641	71218	79

TYPICAL CHARACTERISTICS OF DISINFORMATION-RIDDLED BLOGS

What are the typical characteristics of mis-/disinformation-riddled blogs? Based on our observations and the work of other experts, we provide a set of heuristics to identify blogs that are potentially riddled with mis-/disinformation.¹⁴⁰ These heuristics are:

- 1. Pay attention to the 'contact us' section of the page to validate and verify site authors.** The contact information sections of these blogs do not provide real contact information for the author. For instance, one such real-looking contact URL is <http://abcnews.com.co/>.
- 2. Do not read just the headline; instead, skim the body content** to familiarize yourself with the details of the story. For example, the headline *'Obama Signs Executive Order Declaring Investigation into Election Results; Revote Planned for Dec. 19th – ABC News'* is a false story with a catchy headline. But, reading through the content will enable the reader better to evaluate the story.
- 3. Pay close attention to the URLs, sources, images, and editorial standards of the writing.** For instance, the URL bloomberg.ma is used to imitate the well-known site bloomberg.com.
- 4. Always crosscheck the story with fact-checking websites,** such as snopes.com, factcheck.org, mediabiasfactcheck.com, or politifact.com for the credibility of the story. For example, a blog post titled *'The Amish In America Commit Their Vote to Donald Trump; Mathematically Guaranteeing Him a Presidential Victory – ABC News'* is a fake story reported by the well-known fact checking website snopes.com.
- 5. Search for the post in well-known search engines,** such as Google, Bing, Yahoo, etc., to see if the same post or content is repeated on other sites using mix/cross media approaches to disseminate the narrative. For instance, the blog post *'Obama Signs Executive Order Declaring Investigation into Election Results; Revote Planned for Dec. 19th – ABC News'* has been shared on many websites, indicating the use of a mixed-media approach.
- 6. Check if the article has been previously published and if it is being reused to affect perceptions about an event.** For example, a blog post title *'Muslims BUSTED: They Stole Millions in Govt Benefits'* published in 2016, contained an image that was reused from the year 2013.
- 7. Check if the post is disturbing or controversial.**

Fake stories usually appear under sensational headlines. For instance, the blog post titled *'EU NATO Commit Adultery, Prince Charles Saudi Trade & More'* presents disturbing information. Disinformation narratives are often embedded in such stories.

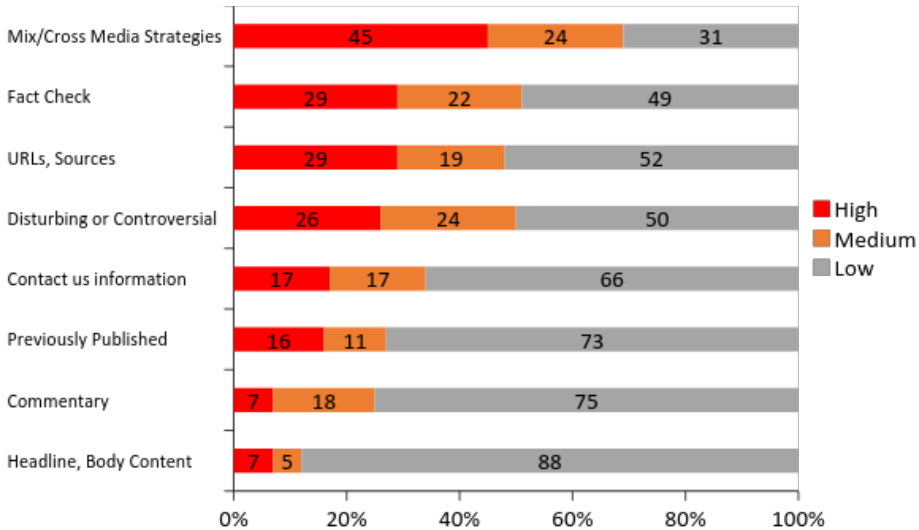
- 8. Check if the post has any 'likes', 'replies', or 'comments'.** This will indicate how interested readers are in a given story, and whether they agree or disagree. The *sentiment* can be used to infer this. For example, a blog post titled *'NASA Confirms—*

Super Human Abilities Gained' has a lot of comments, many of which were debunking the story.

To evaluate the efficacy of these eight criteria, we conducted a survey. We randomly selected 96 blogs featuring mis-/disinformation and asked survey participants to rate (low, medium, high) how effective each of the eight criteria was in assessing whether the blog site contained misleading or false information.

After collecting the survey data, we constructed a stacked bar for each of the criteria where the X-axis represents values (0%–100%) indicating participant

EFFECTIVENESS OF EACH OF THE 8 CRITERIA IN IDENTIFYING BLOGS CONTAINING MISINFORMATION OR DISINFORMATION.



* The criteria are sorted in decreasing order of effectiveness. The smaller the gray bar the more effective the criterion is. Numbers on the colored bars indicate the number of blog sites identified as containing misleading or false information with a confidence of High, Medium, and Low.

confidence in the 96 blogs rated low, medium, or high, and the Y-axis denotes the eight criteria. Looking at Figure on page 36, it is clear that the best criterion is the use of mix/cross media strategies by the blog site in disseminating the content. This can be used as the superlative feature for assessing the mis-/disinformation contained in any blog post. The next best feature is fact-checking websites.

Next, we present some empirical observations vis-à-vis mis-/disinformation heuristics on the fake news dataset collected from kaggle.com. Incidentally, most of the posts had very few comments or none, which might imply that the stories were mainly disseminated but not discussed much on these sites. We also found that during the US elections many posts were primarily intended to reach large numbers,

draw their attention, and direct them to non-factual stories with the intention of influencing readers. For example, 96% (12,468 of 12,999) of the posts had zero 'likes' and 94% (12,304 of 12,999) of the posts had zero 'replies'. These posts were primarily intended to be disseminated to reach more people and mislead. We also observed that the majority of the stories originated from a set of domains that are usually reported as containing false information by snopes.com.

We further examined the website structure disseminating these false stories. We found, in many cases, that the 'contact us' page does not provide any real contact information or redirects readers to another website, usually a social media site, e.g. Facebook or Twitter. The example below illustrates how a

ILLUSTRATION: THE 'CONTACT US' PAGE REDIRECTS TO ANOTHER WEBSITE

The illustration shows a sequence of three screenshots connected by orange arrows. The first screenshot is an ABC News article page with a 'CONTACT US' button. The second screenshot shows the result of clicking that button: a page with social media icons for Facebook, Twitter, YouTube, Instagram, and LinkedIn, along with an email icon. The third screenshot shows the CNN 'Contact' page, which is reached after clicking one of the social media links. This demonstrates how a 'contact us' page can be used to redirect users to another website.

site redirects to another website when readers look for contact information for the author. The example provided here refers to the site name – ABC NEWS with the URL <http://abcnews.com.co/>. The contact information link is present at the bottom of the page for this site. If the reader clicks on ‘contact us’, he is redirected to another site named CNN with the URL <http://cnn.com.de/contact/>. The <http://cnn.com.de/> website closely mimics the CNN News website (<http://www.cnn.com/>), even using the CNN logo, website structure, etc. However, cnn.com.de is riddled with false stories and conspiracy theories. When posted on Facebook, an article from cnn.com.de would bear the CNN logo and appear as if the article were actually published by the genuine CNN.com. This deception

tactic is highly effective in disseminating disinformation originating on blogs via other social media channels.

TRACKING THE ORIGINS OF MISLEADING BLOG CONTENT

Can the origins of misleading content, such as memes, images, etc., which appear on these blogs, be tracked? We began our analysis with a ‘reverse image search’ (i.e. searching for the URL of a given image on Google Images to identify other sources that have used the image) and found that the images were not unique for each article and not relevant to the context they are used for. The same image was reused with different narratives, as shown below. Images lend credibility to a narrative

REVERSE IMAGE SEARCH SHOWS THE USE OF ONE IMAGE WITH DIFFERENT NARRATIVES

Google Song

All **Images** Maps Shopping More Settings Tools

About 25,270,000,000 results (1.19 seconds)

Image size: 896 × 478
No other sizes of this image found.

Pages that include matching images

Fadexadii Kabayare oo la dagaashay daawadayaashii ... - Opify.net
https://opify.net/.../Fadexadii_Kabayare_oo_la_dagaashay_daawadayaash...
 320 × 180 - Thumbnail: HANUUNIYE HEESTA GOBANIMO WADANI SONG Official video HD - Thumbnail: GABAR YAR ... Thumbnail: Best Somali song - Thumbnail: Farxiya ...

Fadexadii Kabayare oo la dagaashay daawadayaashii Showgeeda ...
<https://www.youtube.com/watch?v=6qXeEz0U49Y>
 320 × 94 - Jan 1, 2015 - HANUUNIYE HEESTA GOBANIMO WADANI SONG Official video HD - Duration: 5:49 - xusuusonline 168,094 views - 5:49 - Fadeexo: Khadra ...

Lacago childcare lagu khiyaano - YouTube
<https://www.youtube.com/watch?v=63MzN8Kq0o>
 168 × 94 - Mar 4, 2013 - Lacagta sida khiyaanada ah xanaanada caruurta lagu qaato iyo

and are more effective than text alone for fabricating perceptions. The use of images and videos in framing narratives is effective because multiple modalities are exploited to influence thinking.¹⁴¹ We also observed a pattern in which

a post shared on Twitter was actually linked to a blog post using hashtags and links. This pattern is common across various social media channels, i.e. the origin of the content is generated on a blog and later disseminated

BLOG POST USES HASHTAGS AND LINKS TO REFER TO TWITTER

The screenshot shows the Global Research website interface. At the top, there is a navigation bar with links for About, Contact, Membership, Store, and Donate, along with regional options like USA, Canada, Latin America, Africa, Middle East, Europe, Russia, Asia, and Oceania. The main header features the Global Research logo and search bars. A sidebar on the left lists 'Latest News & Top Stories' with various headlines. The main content area displays a blog post titled 'Syria. Analyzing Madaya's Starvation Falsification. Western Media Propaganda in Support of US-NATO War Crimes' by Paul Antonopoulos, dated January 11, 2016. The post includes a map of Syria and a quote: 'Public outcry and condemnation against the Syrian government spread like wildfire across mainstream news and social media when the horrific photos of starved children and civilians from the besieged town of Madaya emerged. No one could understand why Assad would allow this to happen to his own people, especially since videos emerged (that can be seen on my last Madaya article) under a month ago that displayed rallies against the occupying terrorist forces and in support of the Syrian government.' Below the article, there is a tweet from @SMO_SYRIA with the text: 'Official statement by the #IRCC on #Madaya #Zabadani #Fouaa #Kafaraya #Nabul #Zahraa #Break_Hunger_Siege'. The tweet includes a 'Follow' button and a timestamp of 17:03 - 9 Jan 2016.

TWEET USES HASHTAGS/LINKS

The screenshot shows a Twitter interface. A tweet from a user with a blacked-out profile picture is highlighted. The tweet text reads: '#break_hunger_siege #FreeSyria #SaveSyriasChildren and save humanity on #InternationalChildrensDay'. Below the main text, there is a quote: '#children under #siege eating grass an attempt to survive a war...'. The tweet is dated '2:54 PM - 1 Jun 2016' and has '2 Retweets'. The background shows the Twitter search filters and 'New to Twitter' section.

through social media channels. Figures below depict this pattern. Initially the content is generated on blog posts where the use of hashtags and links serve as the vehicles connecting to other social media channels, in this case to Twitter.

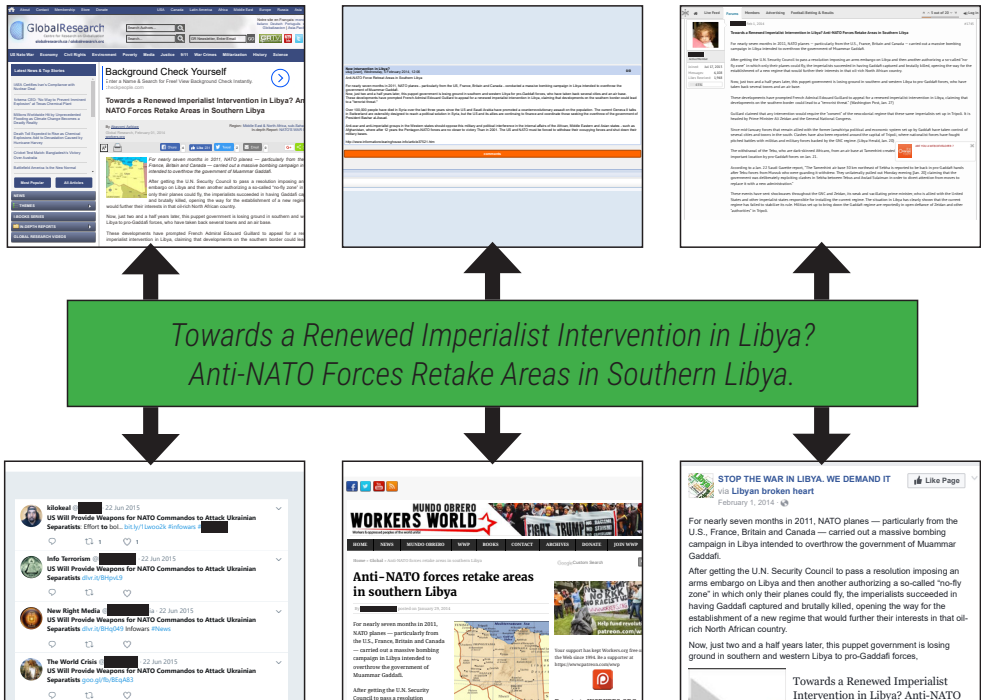
MIXED-MEDIA VS. CROSS-MEDIA APPROACHES

A **mixed-media information dissemination campaign** uses multiple social media channels to perpetuate

a narrative. More precisely, the information campaign can be observed on multiple social media sites through the use of text, images, and audio and video content. Although the content may not be strictly identical on the various social media channels where it appears, it clearly pertains to a single information campaign.

A **cross-media information dissemination campaign** is characterized by a central channel around which the campaign is built. More precisely, the information is

MIXED-MEDIA STRATEGY FOR DISSEMINATING MISINFORMATION OR DISINFORMATION ON DIFFERENT WEBSITES.



The image above shows the mixed-media dissemination campaign for 'Towards a Renewed Imperialist Intervention in Libya? Anti-NATO Forces Retake Areas in Southern Libya'

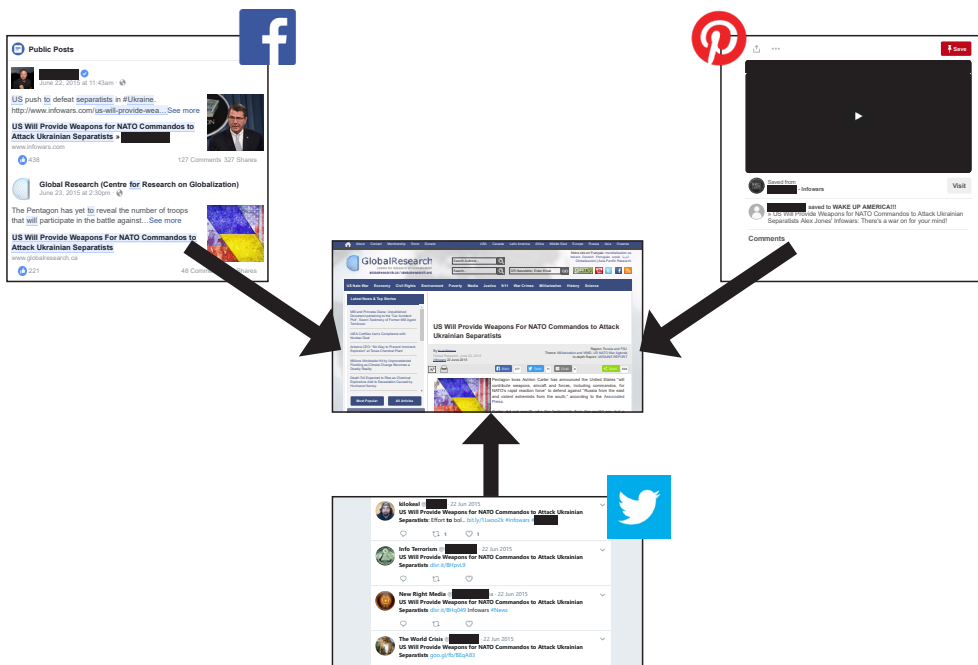
hosted on a website (e.g. text on a blog or video on a YouTube channel) and is widely distributed through other social media channels that provide established social network structures, such as Twitter, Facebook, etc.

First we investigated the use of the mixed-media approach in disseminating stories. In this study, we encountered cases where an article was shared on different sites as shown on page 40. For instance, a story titled *'Towards a Renewed Imperialist Intervention in Libya? Anti-NATO Forces Retake Areas*

in Southern Libya' was disseminated on multiple sites, i.e. [facebook.com](https://www.facebook.com), [ooroom.org](https://www.ooroom.org), twitter.com, [globalresearch.ca](https://www.globalresearch.ca), [hotnews.ro](https://www.hotnews.ro), and [workers.org](https://www.workers.org).¹⁴²

Next, we examined the cross-media information dissemination approach. This tactic was observed to a good effect in our dataset. There were many sites that shared links to specific social media channels such as Twitter, Facebook, and Reddit sites. For instance, a blog site named 'globalresearch.ca' had a post entitled *'US Will Provide Weapons For NATO Commandos to Attack Ukrainian*

CROSS MEDIA INFORMATION DISSEMINATION STRATEGY FOR DISSEMINATING MISINFORMATION OR DISINFORMATION ON SOCIAL MEDIA



Separatists' with the link – <http://bit.ly/2ewVTg7>. This post was shared on Twitter (<http://bit.ly/2xEQxnU>), Pinterest (<http://bit.ly/2x02sQ0>), and Facebook (<http://bit.ly/2wrlhZD>) as depicted below. This clearly indicates a cross-media pattern.

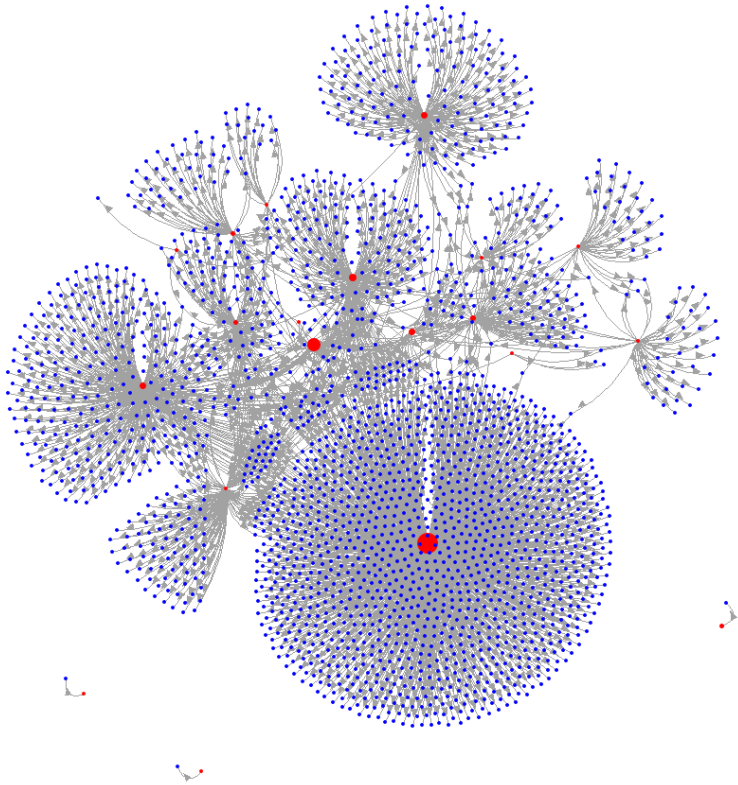
TRACKING HOW AN ANTAGONISTIC NARRATIVE TRAVELS

To analyze how a narrative travels, we examined the 'likes' and 'comments' features available on blogs. A higher number of retweets, shares and comments at blog level show that posts have been circulated widely, demonstrating that media integration strategies do help in disseminating the narratives. Readers can like the content and comment on the post. Note that the 'like' feature on the blogs embeds various social plugins from Twitter, Facebook, Reddit, etc. These social plugins allow readers to like the page simultaneously on the different social media platforms, thereby disseminating the content on a variety of platforms simultaneously. For instance, a blog site, 21stcenturywire.com, published a blog post on September 18, 2016 entitled '*Syria: No "Dusty Boy" Outrage for 7 yr old Haider, Sniped by NATO Terrorists in Idlib Village of Foua*'. This blog post received 65 comments in which the audience presented their views. Moreover, the article was shared on other social media channels such as Twitter, where it got 19 retweets, 5 likes, and 2 replies. The same post on Facebook got 6 reactions, 3 comments,

and 2 shares. Also, many groups posted this article to disseminate to an intended audience. The same blog, i.e. 21stcenturywire.com, published another blog post on September 27, 2016 entitled '*EU NATO Commit Adultery, Prince Charles Saudi Trade & More*' that again presented factually incorrect information. As we did with the previous example, we tracked how this post was disseminated through different social media channels. This blog post, however, received no comments. The article was shared on Twitter, but it got only 1 retweet, 1 like, and no replies. The same post was also shared on Facebook, where it received 27 reactions, 1 comment, and 11 shares. But all the shares were coming from the same group, 21stcenturywire.com. No other Facebook group posted this article. Since not many individuals or groups showed interest in spreading this information, it is clear that this article did not get any traction on blogs and not much on other social media platforms.

Next, we analyzed the effects network of blogs have on content dissemination. Unlike social media platforms, blogs do not have a social network structure, i.e. there is no follow-follower relation among blogs. However, it is still possible to observe the information flow network in blogs based on who links to whom. More specifically, we examined the hyperlinks in the blogs to extract the blog network. We used this approach to extract the network of the blogs containing disinformation regarding Baltic States. We used specific software to visualize the network¹⁴³, as depicted in on page 43. The network

NETWORK* OF BLOGS AND SHARED HYPERLINKS



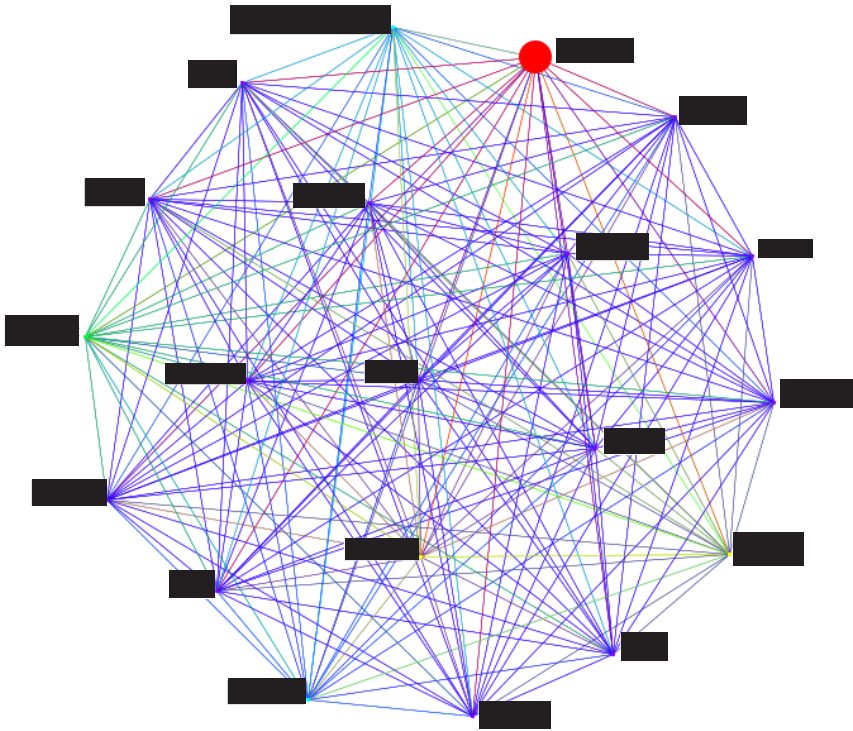
** The network contains 21 blogs (red nodes) and 2321 hyperlinks (blue nodes). Size of a node is proportional to the number of shared hyperlinks (i.e. out-degree centrality). Edge thickness is proportional to the number of times a blogs shared a hyperlink.*

contains 21 blogs (red nodes) and 2321 hyperlinks (blue nodes). Further analysis of the blog network helps in identifying 5 blogs out of 21 that were the most resourceful (having the most hyperlinks), as well as the most exclusive in resources (i.e. they shared hyperlinks that no other blogs shared). 10 out of 2321 hyperlinks were the most shared and most exclusively shared, i.e., these hyperlinks were shared by only a few blogs. Most of these top ten shared hyperlinks have a domain suffix from the Baltic nations, i.e. 'ee' for Estonia, 'lv' for Latvia, and 'lt' for Lithuania.

The exclusivity of resource sharing by a few blogs hints at information campaign coordination. To dig deeper, we construct a blog network based on the commonly shared hyperlinks. The blog network thus identified is depicted on the next page. The network is fully connected, i.e. every blog connects to every other blog. This suggests that every blog in this set shared the same hyperlinks. This confirms our conjecture that there is intensive campaign coordination among these blogs. Further investigation is required to know if these blogs belong to

A NETWORK* OF BLOGS BASED ON COMMONLY SHARED HYPERLINKS

Meta Network



powered by ORA-NetScenes

** The network is fully connected, i.e. a clique, where every blog is connected with every other blog. This depicts massively coordinated information campaign.¹⁴⁴*

or are controlled by the same individual or a group.¹⁴⁴

Next, we analyzed the role of blogs in providing a persuasive dimension to the narrative. We examined how 'exemplified accounts'¹⁴⁵ in the user comments to a story may influence audience perceptions.¹⁴⁶ We provide an example where

commentary lends a persuasive dimension to the blog post.

On page 45 it is possible to see how exemplified accounts in users' comments for a post may influence the audience perceptions.¹⁴⁷ After reading through the comments, we can actually observe that some of the commenters' accounts help in developing a persuasive discourse.

Global Research (Centre for Research on Globalization) July 2, 2014 · Like Page

"I have met allies who can report that Russia, as part of their sophisticated information and disinformation operations, engaged actively with so-called non-governmental organisations — environmental organizations working against shale gas — to maintain European dependence on imported Russian gas," said Rasmussen, the former Prime Minister of Denmark.



NATO Accuses Moscow of Covertly Funding Western Anti-Fracking Activists
At a June 19 speaking event at London's Chatham House, North Atlantic Treaty Organization (NATO)...
GLOBALRESEARCH.CA

89 Likes 26 Comments 50 Shares

Like Comment Share

89 Top Comments ▾

50 shares 26 Comments

I wish it was true, 😊, because Fracking is CRIMINAL and must be stopped!
There are enough GREEN Energy options left and wide open for further exploration.
It is also "TYPICAL" NATO, to accuse ANYONE not agreeing with their policies for "Wild West" (read East) expansions.

It is the NATO that have caused a lot of tension in and around Eastern Countries.
One should never drive a Car in a Corner, as reaction is UNPREDICTABLE.
So, this gambling game as the NATO plays it, goes over the head and at Cost of European and non European citizens.

<https://www.facebook.com/SolarEnergyglobalInfo?ref=hl>
Like · Reply · 21 · July 2, 2014 at 2:29pm · Edited

In addition to being an imbecile, Rasmussen is also a liar.
Besides, European environmental organizations don't need Russia's alleged "sophisticated information & disinformation operations" to know that shale fracking is a huge environmental hazard.
Like · Reply · 20 · July 2, 2014 at 3:01pm · Edited

Rasmussen, Obama's pet parrot...
Like · Reply · 6 · July 2, 2014 at 2:12pm

wonder what they could think of next to blame on Russia lol... gee they try everything to try and make something stick lol
Like · Reply · 5 · July 2, 2014 at 2:37pm

If only it was true. Is this because NATO has a stake in Slavyansk? After all Shell Oil signed a \$10 BILLION dollar drilling contract for the region that is now the most bombed in Ukraine since the Nazi's in WW2.....
Like · Reply · 2 · July 2, 2014 at 2:45pm

Why refer to NATO as if it were anything more than a puppet and mouthpiece of the United States?
Like · Reply · 2 · July 2, 2014 at 2:31pm

WOW. they have gone THAT low, for oil money, now they are accusing people who had to drink oil in their drinking water because of fracking of being traitors.
despicable filth.
Like · Reply · 4 · July 2, 2014 at 2:50pm

An example illustrating exemplified accounts in comments may shape the discourse to form a persuasive dimension

We observe that user comments actually augment the narrative presented in the blog post. We can see a lot of users commenting about the post to further strengthen the narrative. At the same time, we can see patterns such as linking this content to other websites or pages (such as Facebook fan pages), sharing to other channels (50 shares) to further raise discussions.

CONCLUSIONS

Blogs are becoming virtual town halls that are shaping the public perceptions and narratives of regional events. Narratives are first framed on the blogs, then they are disseminated through other social media channels. The key findings include the identification of massively coordinated information campaigns among blogs by applying social network analysis concepts—and demonstrating that commentary on blogs lends a persuasive dimension to the discourse.

In our research, we highlighted the role that blogs can have in weaponizing narratives and conducting disinformation campaigns, suggesting that action be taken towards developing countermeasures. The major contributions of this chapter include: assessment of guidelines for detecting blogs containing misinformation or disinformation; tracking the origins of the content on blogs such as memes, images, videos, etc.; evaluating mixed-media and cross-media narrative dissemination strategies; tracking how the narratives originating in blogs travel

in the social media ecosystem; and analyzing campaign coordination from blog networks. We studied four different blog datasets consisting of 372 blog sites, 7576 bloggers, and 196,940 blog posts riddled with misleading or false information. Social network analysis of the blog network revealed most resourceful blogs and blogs that were most exclusive in sharing resources. Furthermore, a massive misinformation coordination campaign was discovered.

Acknowledgements

This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Fund at the University of Arkansas at Little Rock. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

3

THIRD-PARTY SERVICES AND FAKE NEWS MEDIA SITES

Nora Biteniece

This chapter discusses how user data collected by third-party services coupled with online advertising technologies can be exploited for targeted information activities. This chapter also presents findings from our study of online news sources mentioned in a discussion regarding the NATO presence in the Baltic States and Poland on Twitter. The discussion is referred to as the Enhanced Forward Presence (eFP) discussion, using NATO's name for its defence and deterrence posture in Eastern Europe.

We identified 933 unique news sources linked from Tweets mentioning eFP; 43% of these sources can be classed as fake news media sites. We also observed the systematic use of particular third-party services across the fake news sites in our data set. Two of these services raised concerns as they loaded a variety of opaque third-party services, set unreasonable cookie expiration dates and had been associated with malicious behaviour in the past such as creating spam links. We also identified three social media third-party services present on several fake news media sites that load additional advertising and site analytics services potentially allowing these parties to tie visitors to specific online personas.

User behaviour online such as visiting websites, reading articles, watching videos, searching for keywords, and 'sharing' and 'liking' content on social media can reveal a lot about them. This insight has been effectively used by online advertisers to target specific consumer groups with relevant advertisements. The ability to target specific groups is the main goal of online advertising systems, and ad-providers are willing to pay for these services. Hence, the importance of online advertising services continues to grow, as do their revenues. In 2013, for instance, companies paid \$42.8 billion to US online advertising services.¹⁴⁸ As online advertising has grown, there has been a corresponding rise in exploitation of the advertising ecosystem by cybercriminals seeking to locate victims. According to the online security firm Symantec, more than a half of website

publishers have suffered a malware attack through advertisements.¹⁴⁹ This is just one of the ways in which online advertising technologies are exploited. In September 2017 an article was published on the Facebook Newsroom website reporting on geographically targeted advertisements purchased by inauthentic accounts and pages that originated in Russia.¹⁵⁰ These 'ads and accounts appeared to focus on amplifying divisive social and political messages' including LGBT matters, race issues, immigration, and gun rights.¹⁵¹ This suggests that user data coupled with online advertising technologies can be used for targeted information activities. To gain insight into how online advertising enables actors to target individuals or groups, it is necessary to understand two processes—how user data is collected, and how online advertisements are delivered using these data. We will begin by describing the mechanisms used to collect user data online followed by a brief overview of behavioural advertising technologies and the vulnerabilities of the entire ecosystem. The second section will present findings from our own study of online news sources mentioned in the eFP discussion on Twitter. In conclusion we will discuss the implications of our findings.

BACKGROUND

Behavioural Tracking and Profiling

The purpose of collecting online behavioural data is to track users over time and build profiles containing




User data coupled with online advertising technologies can be used for targeted information activities.

information about their characteristics (such as gender, age, and ethnicity), interests, and shopping activities.¹⁵² This is known as behavioural tracking and profiling, and it has been effectively used in online advertising. Companies use behavioural data to display advertisements that closely reflect users' interests.¹⁵³ User behavioural tracking and profiling occur across three of the most popular Internet services, i.e. websites, location-based services, and social media sites.¹⁵⁴ Each service has different tracking mechanisms. For example, social media platforms are designed to track content accessed by users, what they 'like' and 'share', and what they engage with. This is achieved through requiring all users to create a personal profile, providing platform features such as creating a post, sharing an article, liking content, etc.¹⁵⁵

Web tracking,¹⁵⁶ however, is mainly performed by monitoring IP addresses, and by using cookies, Javascripts,¹⁵⁷ and supercookies¹⁵⁸.

Cookies are small text files that web servers can set and read from a user's browser. When a user navigates to a particular website for the first time, the website may call a script to set a cookie, containing a unique ID, on the user's machine. The browser will attach the cookie to all subsequent communication between the client and the web server until the cookie expires, is reset by the server, or deleted by the user. The most basic function of a cookie is to identify a device, and by extension unique visitors to a website. Cookies help websites to provide services such as visitor counters for website owners, customized web pages, and anti-fraud provisions. Note that cookies are sent only to the websites that set them or to servers in the same domain. However, a website might host content, e.g. images, links, or IFrames¹⁵⁹, stored on servers in other domains.¹⁶⁰ Cookies that are set during the retrieval of this content are third-party cookies, whereas first-party cookies are set by the website that the user is actually accessing. To illustrate this, let us say that an internet user navigates to a



The collected data are stored with records of all the websites the user has visited in the previous minutes, months, and even years.

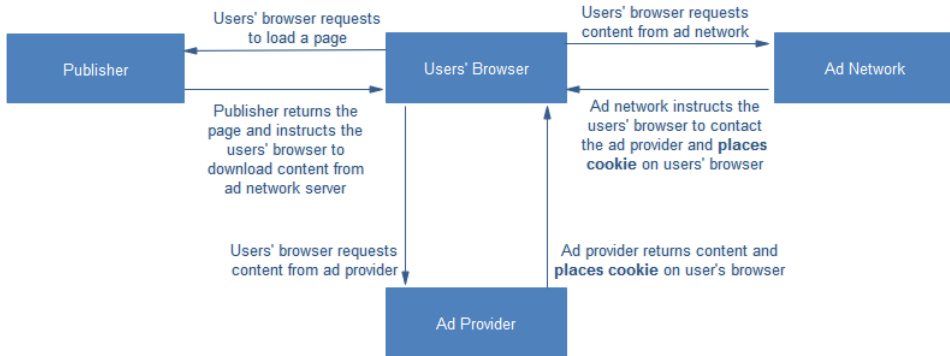
website that also loads advertisements from a third-party server. Because the third-party server has established a connection with the user's computer, it is also able to set a cookie containing an ID unique to that user's machine.

Companies use cookie technology to track user activities online. The collected data are stored with records of all the websites the user has visited in the previous minutes, months, and even years. This information can be augmented with contextual data provided by websites (the content of the website) and/or by data from large data brokers.¹⁶¹ Often, companies deepen their connection with users by planting cookies on several sites to gather additional information regarding online behaviour. The more data they collect from different websites about a particular user, the better the inferences they can draw. Let us say that a user visits a cooking website; the company can read the content of the website and infer that the user is interested in cooking. It can cross-reference this information with any other website visited by that user

detected through its cookie network, for instance a visit to an e-commerce website for gluten-free products. Knowing this, the company can target that user for gluten-free product advertisements. Adjusting advertisements for each user based on previous online activity is known as behavioural targeting, and is enabled by the current online advertising technologies.

Behavioural Targeting

Online advertising systems are typically composed of three main entities: the ad-provider, the ad-publisher, and the ad-network.¹⁶² The ad-provider is the entity wishing to advertise its product or service; the ad-publisher is the website that hosts/displays advertisements; and ad-networks are companies that aggregate available ad space across a large collection of publishers, code their inventory, and sell it on to ad-providers. In the process of coding the available ad spaces, ad-networks segment their audience based on the online behavioural and contextual data they



have collected, and any inferences that can be drawn, thus allowing ad-providers to carry out both contextual advertising and behavioural targeting for various audience segments.¹⁶³


To illustrate how an online advertising system might work, let us say that a user visits a website/ad-publisher that uses ad networks, the ad-publisher instructs the user's browser to contact the network. The ad-network, in turn, retrieves whatever user-cookie-identifiers it can. Using those identifiers, the ad-network can access its own database to see what other information about the user's history it has in order to identify that user's interests and demographic information. The ad-network can then decide which advertisements to display for that particular user.¹⁶⁴ Although the ad-network decides which advertisements should be displayed, it often does not deliver the actual advertisements. Instead, the ad-network instructs the user's browser to contact the actual ad-providers' server (See Figure above).

Note that ad-networks have a built-in opportunity to plant cookies every time they deliver an ad, thus, their cookie network is as large as the pool of sites for which they service ads. This is due to the fact that the host website's server must contact the ad-network every time it needs an ad. In the next section, we will describe the vulnerabilities of the online advertising ecosystem and the problems with behavioural targeting.

Vulnerabilities

The online advertising ecosystem assumes that each entity (ad provider, publisher, ad network) when given the connection to a user's machine, will not compromise that machine and, when gathering data, will gather only what it needs, store it safely, and use it to enhance the user's web experience. In reality, the ecosystem is very complex and each layer is vulnerable to malicious exploitation.

First, a publisher itself may be a phishing site—a website that looks similar to genuine companies or financial services,



The data collection that makes online advertising possible allows advertisers and other entities to target and possibly influence specific user or audience segments.

but is set up to mislead users into entering important details such as their usernames and passwords. This may be done with the purpose of stealing user data and/or fraud. Second, the advertisements delivered through ad networks are not under the control of the publisher; this means that it is not the users who decide which entity is allowed to connect with their machine and which is not. Third, online advertisements can deliver files and entire programs to a user even if the advertisement itself appears to be just an image. This means that ad providers are able to transmit advertisements with embedded executable scripts—a key vulnerability,¹⁶⁵ such scripts would be able to download malware on the user's computer without any clicks or other actions being taken by the user. Ad networks usually perform some kind of quality control on the advertisements they service; however, the actual file at a given URL can be changed after the initial quality control check has taken place.¹⁶⁶ In addition, an advertisement passes through several networks before it actually reaches the user's browser. Each time it passes through another network, there

is an opportunity for the introduction of malware.

Moreover, behavioural targeting enabled by user data collection and the ad delivery systems can be used to take advantage of vulnerable users. For example, information about a user's health, financial condition, or age can be inferred from online tracking and used to target that person for payday loans, sub-prime mortgages, or bogus health cures.¹⁶⁷ Users' behavioural profiles can be used to offer certain customers products at a higher cost or deny them access to goods altogether ('online redlining').¹⁶⁸ In the absence of clear privacy laws and security standards, these behavioural profiles leave users vulnerable to identity theft and information activities. Our study did not focus on detecting malicious ads or identifying phishing sites. Instead we looked at the risks of online advertising technology being used for targeted information activities. The data collection that makes online advertising possible allows advertisers and other entities to target and possibly influence specific

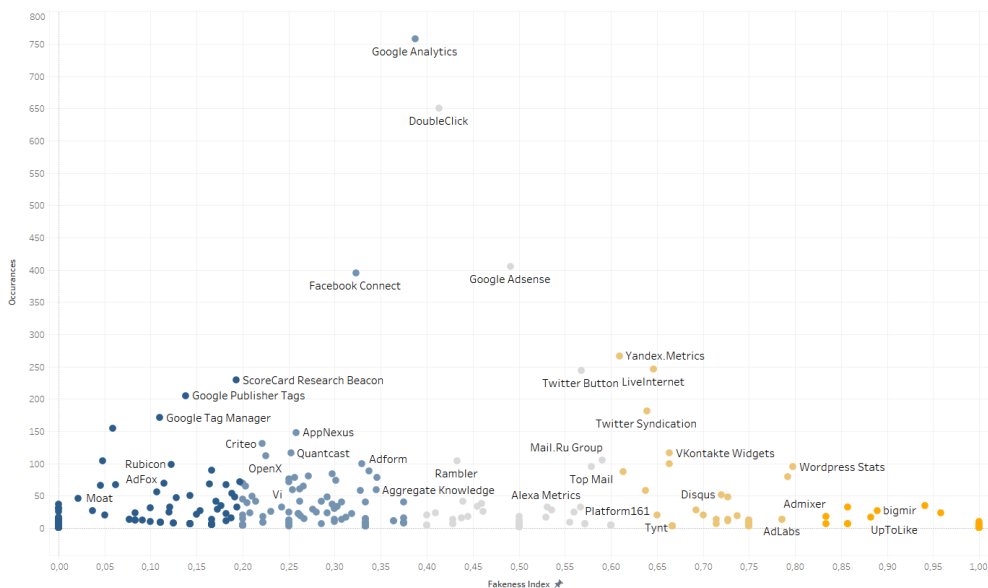
user or audience segments. A striking example is the case of Facebook carrying ads from linked and inauthentic accounts and pages originating in Russia and essentially targeting socially divisive messages at the US public.¹⁶⁹

We set out to understand whether fake news media sites mentioned in the eFP discussion on Twitter use the described technology to collect data that would enable them to target and possibly influence individual users and user groups. Although we do not know if the collected data is used for targeted information activities, we demonstrate that the information available could be used for that purpose. It is not within the scope of this chapter, or of our study, to detect targeted information activity or attribute such activity to anyone. The next section will present findings from our study.

NEWS SOURCES AND THIRD-PARTY SERVICES

We examined 933 online news sites mentioned in the eFP discussion on Twitter. We found 588 unique third parties that receive data about visitors to these sites. 43.1% of the websites in our dataset were classified as fake news media sites, and 71 of the identified third parties were found mostly on these sites.¹⁷⁰ The observed third-party services included ad networks, web analytics, and social media services. When examining the use of third-party services by legitimate and fake media sites, we observed:

- Both classes of news media sites share popular third-party services such as Google Analytics, Double Click, Google AdSense, and Facebook Connect;



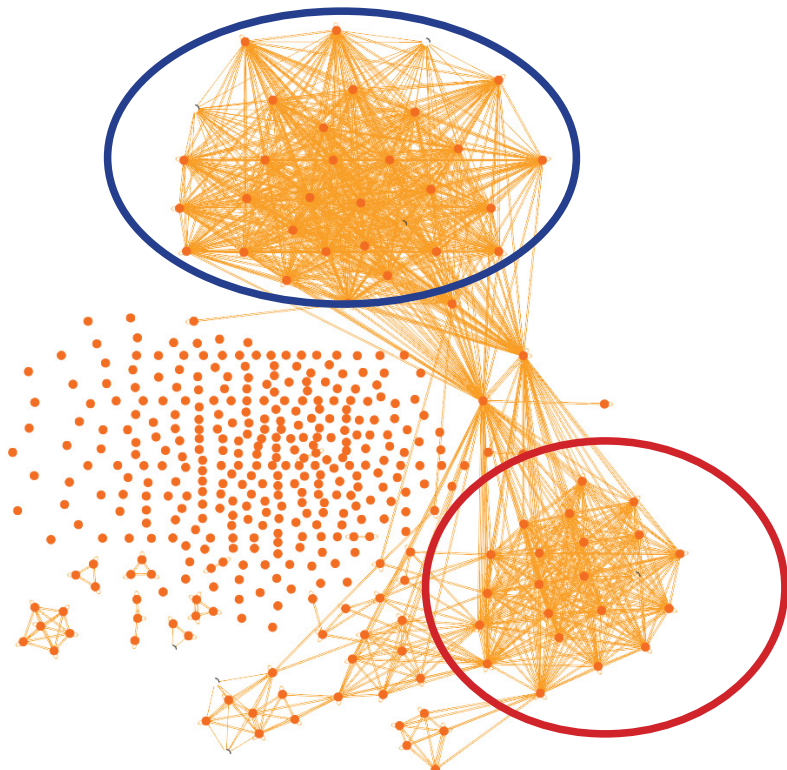
- The legitimate news media, however, seems to use a greater variety and number of third-party services (See Figure on page 53).

The Y axis in Figure on page 53 shows the total number of times a third-party service appeared on any website in our dataset; the X axis shows the calculated **Fakeness Index**¹⁷¹ for each third-party service. We then developed a network graph (see Figure below) to examine which services with a fakeness index 0.9 or above were shared across fake news media sites. The individual nodes are websites classed as fake; the edges between them are third-party services. If there's an edge

connecting two nodes, a particular third-party service is present on both of these websites.

In Figure below you can see two distinct clusters of sites; the larger one consists of sites that all use the Google AdSense Asynchronous service, the smaller one consists of sites that use the Artificial Computation Intelligence service. The 4 much smaller clusters are sites that use MarketGid, i.ua, Whos.amung.us, and Clicky services respectively. We discuss some of these services in more detail in the next section.

When examining the patterns in the data collected by both fake and legitimate news media outlets, we observed three distinct categories of data being collected:



- Anonymous (analytics, user agent details, cookie data, date/time, ad views, etc.)
- Pseudonymous (device ID, search history, IP address, and other location based data)
- Person-identifiable information or PII (address, email address, name, login, phone number)

We found that there is no significant difference in the pattern of data collected by services present on mostly legitimate and mostly fake news media sites. However, a number of services across the fake and legitimate news media sites collect IP addresses (Pseudonymous), addresses, names, and email addresses. Collecting this information allows third parties to later target specific users via means that go beyond online advertising, such as e-mails and IP addresses.¹⁷² Thus, both legitimate and fake news media sites in our dataset collect data that enable them to target specific users and/or audience segments.

Investigating the services shared across mostly fake news media sites further, we found that the services were legitimate: Google Adsense Asynchronous is a well-known ad network, while Clicky, i.ua, and whos.amung.us are legitimate web analytics services. The criteria we used included transparency of ownership, privacy policy, cookie expiration date, Web Of Trust (WOT)¹⁷³ rating, and the additional services required to load a website. To test the latter, we bought a domain name and

hosting, embedded service scripts, accessed the page, and examined the network traffic. Two more suspicious third-party services are discussed below.

Artificial Computation Intelligence Service

As already mentioned, we found that the third-party service Artificial Computation Intelligence (called through loading a JavaScript file—acint.js) was used across several fake news media sites. When investigating the company behind the service, we found that it claims to be a web analytics service for the largest RuNet websites, and is supposedly collecting user IP addresses, operating system information, browser details, and the number of visits. However, according to WOT it produces spam and malware links to .ru domains, and when placing acint.js on our website, we observed that it loaded 13 other third-party services to our site.¹⁷⁴ In addition, it attempted to load a resource from stat.sputnik.ru (See next page).

Another third-party service Artificial Computation Intelligence loads comes from the sape.ru domain. Sape.ru itself is a web analytics and backlink service,¹⁷⁵ however, in the past it has produced unwanted links and injected acint.js script¹⁷⁶ on websites its service was installed. This suggests two things: first, Artificial Computation Intelligence and Sape services are related. Second, Artificial Computation Intelligence uses sape.ru as a proxy to enroll websites into their ad framework without web developer/admin consent.

IN MAY:

This screenshot shows the network headers for a request in May. The 'Connection: keep-alive' header is highlighted with a red box. Other visible headers include 'Content-Type: image/gif', 'Date: Sun, 30 Apr 2017 09:56:56 GMT', 'Expires: Mon, 01 May 2017 09:56:56 GMT', 'Last-Modified: Mon, 28 Sep 1970 06:00:00 GMT', 'P3P: policyref="/w3c/p3p.xml"', 'Server: nginx', and 'Set-Cookie: amber_id=Zr.GAPrtF50555a7YY; domain=stat.sputnik.ru; exp=2038 00:00:01 GMT' and 'Set-Cookie: visid=CvLurVkFt0iDqm6jNMLRAg==; expires=Tue, 30-Apr-19 09: in=sputnik.ru; path=/'. The 'Request Headers' section also shows 'Accept: image/webp,image/*,*/*;q=0.8', 'Accept-Encoding: gzip, deflate, sdch, br', 'Accept-Language: en-GB,en-US;q=0.8,en;q=0.6', 'Cache-Control: no-cache', 'Pragma: no-cache', and 'Referer: http://www.acint.net/mc/?dp=10'. The 'Query String Parameters' section is also visible.

IN AUGUST:

This screenshot shows the network headers for a request in August. The 'Status Code: 404 Not Found' is highlighted with a red box. Other visible headers include 'Request URL: https://stat.sputnik.ru/amber.gif?id=PhuxeVrhA5M5555T5AA', 'Remote Address: 5.143.224.43:443', 'Reterrer Policy: no-referrer-when-downgrade', 'Connection: keep-alive', 'Content-Length: 13', 'Content-Type: text/plain', 'Date: Tue, 29 Aug 2017 11:29:12 GMT', and 'Server: nginx'. The 'Request Headers' section shows 'Accept: image/webp,image/png,image/*,*/*;q=0.8', 'Accept-Encoding: gzip, deflate, br', 'Accept-Language: en-US,en;q=0.8', 'Connection: keep-alive', and 'DNT: 1'. The 'Host: stat.sputnik.ru' and 'Referer: http://www.acint.net/mc/?dp=10' are also highlighted with red boxes. The 'Query String Parameters' section shows 'Id: PhuxeVrhA5M5555T5AA'.



Collecting this information allows third parties to later target specific users via means that go beyond online advertising, such as e-mails and IP addresses.

MarketGid

MarketGid was also used across several fake news media sites in the form of i.js, a Javascript file. Investigating MarketGid further we found that it is an ad network; however, according to WOT it produces spam and malware links. When placing MarketGid on our website, we observed that it loaded two other cookies,¹⁷⁷ from a targeted advertising company and a news agency respectively. All three service cookies are set to expire in 2038, which is, according to EU privacy policy, an unreasonable cookie expiration date. This also means they could collect data about the user for that entire period of time unless the individual cookies are deleted.

Social Media Third-Party Services

As mentioned before, we found several social media third-party services in our dataset. These services in most cases facilitated interactions **from** external sources **to** a social media platform, e.g. liking, sharing, commenting, etc., and

were used by both legitimate and fake news media sites. By enabling a website visitor to like or interact with the content on social media, websites and services are able to tie the visitor to a specific online persona. This online persona will have a lot more information associated with it than just browser details, IP addresses, or referral data.

Most of the social media third-party services identified in our data set were provided by the social media companies themselves; however, several integrated widgets¹⁷⁸ from different platforms supported interactions across a variety of social media services. Companies that provide such widgets free of charge most likely monetize their use by collecting user data. AddThis, for example, has profiles for 1.9 billion people. This suggests that the core business for AddThis does not lie in providing free social media widgets, but rather in selling user profile data to third parties.

To examine whether the third-party services identified in our data set expose

any social-media-related user data, we bought a domain name and hosting, and embedded the third-party services used on fake news media sites only. The user data would have to be in the HTTP header when interacting with the third-party service for website owners like us to access it. See table below for our observations

In short, when calling a social media service from a website, no user data is passed or exposed during this communication. Instead, the social media service handles the aftersteps

of this request, i.e. retrieving the user ID from cookies present in the browser or redirecting to a pop up window with a login screen. Although some of the information is visible from the developer tools in Google Chrome (Facebook user ID in a cookie), it is not accessible to the website owner. As explained in the previous sections, cookies can be read by the domain from which they originate. Thus, for a website to read the Facebook cookie containing a user ID, it would have to be from the same domain as Facebook. In addition, we observed that four of the identified third-party services loaded

Widget	Supported Functionality	Comments
Facebook Connect	Authorization through Facebook	Retrieves a browser cookie with users' Facebook IDs
Facebook Social Graph	Querying Facebooks' Social Graph ¹⁷⁹	
Facebook Social Plugins	Like/Share/Comment on Facebook	Loads Facebook Connect and Facebook Impressions
LinkedIn Widgets	Post on LinkedIn	
Lockerz Share	Share content across social media platforms of choice	
Pinterest		
Pluso	Share content across social media platforms of choice	Loads several advertising services (adapt.tv, advertising.com, DoubleClick, Eyeota, FACETz, rutarget, Vi)
Reddit	Post on Reddit	
Share42	Share across social media platform of choice	
Tumblr Buttons	Post on Tumblr	Loads Cedexis radar, Google Analytics, and ScoreCard Research Beacon
Twitter Button	Tweet on Twitter	Loads Twitter Syndication
UpToLike	Like on social media platform of choice	Loads Yandex.Metrics and Mail.ru Group
VKontakte Widgets	Like/Share/Comment on VK	When logged in on VK, it also loads Mail.ru group and Top Mail

other third-party services (advertising and website analytics services). This again increases the number of third parties that record user browsing habits, allowing them to cross-tabulate this information with their other records and infer more about the user. However, as already mentioned, by allowing users to share content from a website on their social media profiles, it enables website owners to then 'backtrack' the users who shared their content as well as the platform they shared it on. For example, by navigating to facebook.com/search.php, pasting a link, and then clicking on the option 'Posts by everyone' a list of Facebook users who have shared that particular link will be displayed. There are even services that aggregate this information across the different social media platforms.¹⁸⁰ One can also backtrack people who liked or commented on a social media page through those social media APIs that allow page owners to query the list of users who liked or commented on their page.

In 2009 Krishnamurthy and Wills identified several ways in which social media sites leak person-identifiable information to third-party services.¹⁸¹ They observed that information that could lead to a user profile (user name, user ID, or email address) was leaked through the 'referrer' and 'request' URL fields in the HTTP header when accessing external content from various social media sites (MySpace, Facebook, Twitter, LiveJournal, LinkedIn, Hi5, Imeem, Orkut, and Xanga). Using

Krishnamurthys' and Wills' methodology, we also examined whether navigating to an article linked from Facebook, VK, Twitter, LinkedIn, Tumblr, and Reddit leaks any social media user information. We observed no user information in the requested URL for any of the social media sites we looked at. The referrer URL for Facebook and Tumblr was the respective social network domain (facebook.com and tumblr.com); for Twitter it is a shortened URL to the article; for Reddit it is the full URL to the article, for LinkedIn there is no such field;¹⁸² and for VK it is a URL that links to the reader's profile (or login page if they do not have a profile). Since 2009 there has been a tremendous improvement in user data privacy when interacting with external content on social media sites. However, in some cases (Facebook, Tumblr, VK, and Reddit) it is still possible for websites to track which social networks a particular visitor uses.

CONCLUSIONS AND IMPLICATIONS

On social media, with everything packaged as URLs linking to external sites, new and unpoliced parts of the internet are visited. Consequently, the way people get their news has also changed. A recent study has shown that 62% of US citizens get their news through social media sites.¹⁸³ This, however, has lowered the barrier of access to non-traditional, possibly untrustworthy, news media. We also saw this in our study on the eFP discussion on Twitter where 43.1% of the linked news sites were fake news media sites. When

examining the third-party services on news sites in our dataset, we observed that both legitimate and fake news media sites use social media services to provide additional functionality such as 'liking' or 'sharing' on a platform. This has several implications:

- The external sources can track which social media platforms their visitors use through the referer field in the HTTP header.
- The external sources can backtrack their own content shared on social media platforms, together with information about any user who shares it. This allows third-parties, or anyone who utilizes this data from third-parties, to target specific individuals on social media sites.
- Companies that provide social media widgets free of charge most likely monetize their use by collecting user data and selling it to third-parties.

In addition, several of the identified social media third-party services loaded other web analytics or advertising services. This raises some concerns, since it allows additional third-parties to collect information on visitors solely on the grounds that they shared an article on their Facebook profile, for example.

When examining other third-party services present mostly on fake news media sites, we observed the systematic

use of Artificial Computation Intelligence and MarketGid. These services load content from several other opaque third-party services, enabling them to place cookies on users' machines and obtain data such as IP addresses, user agents, and the sites they visited. As explained in the first part of this chapter, these data and any information inferred from them can be employed to target user groups based on interests, demographics, and/or geographical location. Moreover, because of the widespread cross-interaction between websites and social media sites, third parties present on these fake news media sites (or their web admins for that matter) can tie a visitor to a specific online persona, and thus target them individually and with a lot more insight. In addition, in the past both of these services have been associated with malicious behaviour such as creating spam links and injecting Javascripts. This suggests that Artificial Computation Intelligence and MarketGid act as proxies to spread spam and malware and to plant cookies from other third parties enabling these parties to collect user data without consent.

CONCLUSIONS AND RECOMMENDATIONS

This publication highlights how false information online brings about a number of security implications. We likened false information to the Lernaean Hydra, the mythical creature that could generate two new heads for each head it lost to the axe. According to the myth, Heracles slayed it by thinking outside the box, burning the stumps of the severed heads, and smashing the only true mortal head the monster had with a rock. Analogously, anyone who is battling disinformation online must think beyond simply debunking single stories.

Social media platforms are popular because they cater to the basic human need for building and maintaining social interactions. It is for this reason that new media are extremely valuable for Strategic Communications, and can be dangerous vehicles for disinformation. Today's disinformation shows continuity with the past at the strategic level, and discontinuity at the tactical level. The contemporary media landscape is characterized by informality and reciprocity. As the relevance of the traditional gatekeepers of information is fading, print media for example, the online environment is becoming less regulated than its offline counterpart.

Contemporary disinformation is more quantitative than qualitative. The majority of false stories shared on social media are rudimentary, and in some cases so improbable that authorities are reluctant to even address them. Yet, these stories can have strategic-level effects on public discourse.

Social media platforms offer unprecedented levels of sophistication to malicious actors who aim at influencing a political conversation through the use of false or misleading information. Social media users often trust the online information environment more than traditional media. This is due to the structure of the platforms: information comes from friends, acquaintances, and sources that resonate with the user's beliefs and values. Given these circumstances, information is rarely evaluated critically. The cognitive biases we all fall into from time to time are what enables malicious actors to manipulate online audiences, but technological innovations make it easier for them to exploit these mechanisms.

There is wide scope for capitalizing on the social media environment to fight disinformation. Social media can generate informational bubbles, but can also pierce them.

Chapter 2 highlights how different social media providers cater to different world regions. The Russian-language internet is, in many respects, a galaxy of its own. Russian-made social media platforms are qualitatively different from their Western counterparts, and can be used more effectively in disinformation campaigns. Western analysts should familiarize themselves with these platforms. This will enable them to understand the narratives that are being pushed through these channels and, potentially, interact with them. The platforms that are popular in Arabic-speaking regions

are mostly those that are common in the West, but the stories being shared reflect the different social and political issues affecting the region.

The discussion of blogs in Chapter 3 proves that social media is a channel for dissemination of narratives, rather than the place where they originate. False stories often originate in blogs and are shared on social media only at a later stage. Disinformation campaigns coordinate the activity of several channels, in order to reach the largest audience possible. Blogs are among the most important environments where narratives are crafted and propagated. Aside from the blog post itself, the comments below the post reinforce the persuasiveness of the narrative.

Buying and selling user profile data has become big business. The discussion of user data collection in Chapter 4 demonstrates that this new reality brings about significant security implications. External actors can monitor content they post to social media platforms together with information about the users who share it, paving the way for tailored messaging—specific groups, even specific individuals, can be targeted on social media with political content designed specifically for them. Several firms are engaged in the analysis of social media audiences. These services are used by for-profit companies, political adversaries, and, potentially, malicious actors aiming at influencing selected audiences.

RECOMMENDATIONS

The chapters highlighted a number of common themes that cut across the topics. These common themes are: data awareness, channel identification, dialogue with the social media industry, and regulation. The recommendations below address these themes.

Data awareness

As the means to collect user data grow in sophistication, users are more and more vulnerable to this kind of activity. Users should be aware of these risks. This is particularly true for those social media users whose work is of a delicate nature, i.e. military/security personnel and civil servants.

Moreover, we must keep in mind that algorithms can discover attributes not explicitly expressed by the user.¹⁸⁴ Despite our efforts, malicious actors can still find ways to use the data we leave behind to target us with tailored messaging that is more likely to influence our behaviour. Understanding this is an important part of data awareness.

The general public needs to be educated on how their online behaviour is being tracked and how this information can be used. There have been a number of efforts in this direction, mostly by citizen-journalists and browser-extension developers.

Channel identification

As the Lernaean Hydra had a single mortal head, so contemporary disinformation campaigns waged on multiple channels have a single 'backbone'. Detecting this backbone helps us understand the context in which a specific group of false stories has originated, and is, therefore, a fundamental step towards assessing whether a specific case should be considered misinformation or disinformation. Western analysts must leave their comfort zones and explore channels they are unfamiliar with. This means those platforms that are distant from their socio-cultural context, be it because of geography or language (as is the case with Russian- or Arabic-language social media) or because they cater to different demographics (as is the case with emerging platforms targeting younger audiences).

False information does not exist in a vacuum, it needs a context and a medium. Different audiences have different interests and are active in different virtual spaces. Malicious actors know this, and adapt their messaging campaigns to the audiences they want to target.

Dialogue with the industry

The use of false information for malicious purposes can be likened to traffic violations. While responsibility for misbehaviour rests solely on the drivers, highway authorities can help the police in making roads safer. The same is true for social media: those

who are most knowledgeable about the platforms' vulnerabilities are the social media companies themselves. For this reason governments (and, in particular, the security sector) should dialogue with social media companies.

Social media companies need this dialogue as much as governments do. They have been facing considerable criticism over the use of their products in spreading misinformation, and have responded with in-house solutions, as outlined in Chapter 2. However, in order for countermeasures to be effective relevant authorities should be involved, so actions can be based on exchange of relevant information. Some steps in this direction are already coming from the industry, as demonstrated by Facebook's self-accusation regarding Russian interference in the 2016 US elections.¹⁸⁵ It is in the companies' self-interest to collaborate with authorities on these matters, as users are likely to respond positively to actions aimed at sanitizing the social media environment whilst protecting their privacy.

Browser providers should assume a more active role in educating their users about behavioural tracking online. Information about what kinds of user data is being collected and by whom should be a standard part of browser functionality. Apple Inc., for example, has restricted several tracking mechanisms in their newest Safari browser.¹⁸⁶ However, this does not solve the problem of users being unaware that their data is being collected and what it will likely be used

for. Social media companies should be encouraged to tighten their data sharing policies. Targeting people on social media is so easy and effective because social media companies have gathered a considerable amount of insight on social media users, their interests, and their attitudes. They provide the mechanisms for targeting. After significant ad sales to a network of inauthentic accounts and pages that disseminated socially divisive messages, Facebook has made their ad review process more rigorous. However, much more can be done to curb access to technologies that enables third-parties to tap into the information Facebook has on users.

Regulation

Regulation is intended to *prevent* the suppression of uncomfortable voices by authoritarian regimes. It is in the users' interest that the virtual spaces where they voice their opinions are kept safe so that they can be truly free. This entails deterring abusive behaviour online, protecting users' privacy, and limiting the intentional spread of false information. Individually, false or misleading stories are easy to falsify, and even easier to create. More work and creative solutions are needed in order to tackle the root causes that make it so cheap to spread misinformation and disinformation.

An area that deserves particular attention is the protection of personal data. Some companies are already self-regulating to support user privacy. However, self-regulation can achieve only limited results, systemic regulation

must come from the institutions. In May 2018, the EU will enforce a new regulation regarding user data protection,¹⁸⁷ which aims to give ownership of personal data back to the users through several key requirements. The companies collecting data on EU citizens, regardless of where the company is registered or where it stores its data, will have to abide by the new regulations. Every user will have the right to be forgotten or for their data to be moved to another data controller. It should be clear to the user who is collecting their data and for what reason, as well as how to opt-out of the data collection process.

These new regulations will be a significant improvement in the protection of user data and user privacy. However, enforcing the regulations must be combined with efforts to educate the general public on user data collection and their rights to own their own data. Moreover, because the entire online tracking process is opaque, the new regulations will still only affect the companies that interact directly with the user. The largest data brokers still collect user data in the background and, in most cases, without the knowledge of the user.

GLOSSARY

The entries presented here are intended to help the reader understand the key terms that are discussed throughout the research product. This unofficial terminology, updated as of 1st November 2017, is aimed at serving further research. The list is inclusive, i.e. it includes terms that are not used in the study, but are central to the discussion. Moreover, some of the entries were not intended to be descriptions in the original context: when this is the case, the “comments” section points it out. While the list is inclusive, only one definition is given for each term, in order to keep this glossary simple and easy to use.

TERM	DEFINITION	SOURCE	COMMENTS
Audience	An individual or group that witnesses an event or information conveyed through social audiovisual or printed media.	AJP 3.10 Allied Joint Doctrine for Information Operations	
Blog	Websites where information is posted on a regular basis. Content varies widely, from personal diary-type minutiae to sustained discussion of politics, hobbies or other interests. Some blogs are a “grab bag” of topics, while others focus on a particular subject.	PAO Handbook 2014	
Blog client	Software to manage (post, edit) blogs from operating system with no need to launch a web browser. A typical blog client has an editor, a spell-checker and a few more options that simplify content creation and editing.	PAO Handbook 2014	
Blogger	Person who runs a blog. Also blogger.com, a popular free website for blog hosting.	PAO Handbook 2014	
Counter-propaganda	A multidiscipline effort led and coordinated by Info ops function to analyse an adversary's information activities, its source content, intended audience, media selection, and effectiveness.	MC 402/2 NATO Military Policy on Psychological Operations	
Disinformation	Dissemination of false information with the deliberate intent to deceive or mislead.	Oxford Dictionary of Media and Communication	
Echo Chamber	An ideological environment in which ideas and opinions are amplified and reinforced by their repetition, creating a mainstreaming effect of like-mindedness.	Oxford Dictionary of Media and Communication	Akin to the concept of filter bubble (see definition).

Fake News	News articles that are intentionally and verifiably false, and could mislead readers.	H. Allcott & M. Gentzkow (2017) "Social Media and Fake News in the 2016 Election", Journal of Economic Perspectives 31 (2)	The source does not aim at giving a definition. This is a working definition in the context of a journal article.
False Amplifiers	Coordinated activity by inauthentic accounts with the intent of manipulating political discussion (e.g., by discouraging specific parties from participating in discussion, or amplifying sensationalistic voices over others).	J. Weedon, W. Nuland, A. Stamos (2017) "Information Operations and Facebook", Facebook	The source does not aim at giving a definition. This is a working definition in the context of a report.
Filter Bubble	A phenomenon whereby the ideological perspectives of internet users are reinforced as a result of the selective algorithmic tailoring of search engine results to individual users (as reflected in recorded data such as search history, click data, and location).	Oxford Dictionary of Social Media	Akin to the concept of echo chamber (see definition).
Homophily	A widespread tendency of human beings to be drawn to others with whom they see themselves as having much in common. This is reflected in the folk wisdom that 'birds of a feather flock together' or 'like attracts like' (in contrast to heterophily). We seek out that which supports our social identity in terms of major social characteristics, such as age, sex, socioeconomic status, and ethnicity. This even applies to parasocial relations with characters represented in texts (in any medium).	Oxford Dictionary of Media and Communication	

Influence	The capacity to have an effect on the character, development, or behaviour of someone or something, or the effect itself.	Oxford Online Dictionary	
Information	Unprocessed data of every description which may be used in the production of intelligence.	AAP-06 NATO Glossary of Terms and Definitions	
Information Activities	Actions designed to affect information and/or information systems. They can be performed by any actor and include protection measures.	MC 0422/5 NATO Military Policy on Information Operations	
Information Effects	A desired condition created in the information environment as a result of information activities. Information effects should be measurable to enable analysis, planning, execution and assessment of related activities and the effects them self.	MC 0422/5 NATO Military Policy on Information Operations	
Information Environment	the information itself, the individuals, organizations and systems that receive process and convey the information, and the cognitive processes that people employ, including the virtual and physical space in which this occurs.	MC 0422/5 NATO Military Policy on Information Operations	
Information Objective	A desired condition to be created in the information environment. It should be measurable to enable analysis, planning, execution/management and assessment/evaluation of related actions and effects.	MilStratCom Practitioners Handbook 2016-08-22	

Information Operations	A staff function to analyse, plan, assess, and integrate information activities to create desired effects on the will, understanding, and capability of adversaries, potential adversaries, and NAC-approved audiences, in support of Alliance mission objectives.	AJP-3.10 Allied Joint Doctrine for Information Operations	
Information Systems	Information systems are socio-technical systems for the collection, processing and dissemination of information. They comprise personnel, technical components, organisational structures and processes that create, collect, perceive, analyse, assess, structure, manipulate, store, retrieve, display, share, transmit and disseminate information.	AJP 3.10 Allied Joint Doctrine for Information Operations	
Information Warfare	Warfare that integrates electronic warfare, cyberwarfare, and psychological operations (PSYOPS) into a single fighting organisation.	D. Stupples (2015) "The next war will be an information war, and we're not ready for it", The Conversation	The source does not aim at giving a definition. This is a working definition in the context of a magazine article.
Media Operations	All activities pertaining to managing the interaction with the news media; can refer to the function responsible for such activities, such as the 'media operations section'. For use in this handbook, the terms media operations is synonymous with media relations.	PAO Handbook 2014	
Misinformation	The dissemination of false information, either knowing it to be false (see disinformation), or unknowingly.	Oxford Dictionary of Media and Communication	

Propaganda	Information, especially of a biased or misleading nature, used to promote a political cause or point of view.	AJP-3.10.1 Allied Joint Doctrine for Psychological Operations	In common speech, the term refers exclusively to false information.
Psychological Effect (in PSYOPS)	A statement of a measurable response that reflects the desired attitude or behaviour change of a selected target audience as a result of psychological operations.	AJP-3.10.1 Allied Joint Doctrine for Psychological Operations	
Receptivity (in PSYOPS)	The vulnerability of a target audience to particular psychological operations media.	AAP-06 NATO Glossary of Terms and Definitions	
RuNet	Russian-speaking Internet	http://dic.academic.ru/	
Social Media	Web-based technologies used for social interaction and to transform and broadcast media monologues into interactives, social dialogues'	NATO ACO Directive on Social Media, 16 September 2014	
Spamming	Sending unsolicited and unwanted e-mails in bulk for advertising purposes. The proliferation of such material, which now accounts for some 85% of all e-mails sent, has become a serious nuisance to business users.	Oxford Dictionary of Business and Management	
Susceptibility	The anticipated acceptance or rejection of a target audience to a particular psychological operations approach.	AAP-06 NATO Glossary of Terms and Definitions	
Target Audience Analysis (TAA)	Examining selected groups of people across a host of psycho-social research parameters, to determine how best to change those groups' behaviour	From S. Tatham, Target Audience Analysis, The Three Swords Magazine 28 (2015)	Additional information can be found on the NATO StratCom COE's Target Audience Analysis course.

Troll	Somebody who disrupts an on-line or social media community by posting abusive or irrelevant material, normally while hiding their identity behind one or more user-names.	Oxford Dictionary of Journalism	
Trolling	Posting of incendiary comments with the intent of provoking others into conflict.	M. Brandel (2007) "Blog trolls and cyberstalkers: How to beat them", Computerworld	The source does not aim at giving a definition. This is a working definition in the context of a magazine article.
Web 2.0	The web seen as a platform for participation in which the consumer is also a producer. This was enabled by multiple software applications that supported user-generated content.	Oxford Dictionary of Media and Communication	The term 'Web 2.0' was coined in 2003 by Tim O'Reilly and Dale Dougherty of O'Reilly Media as a marketing response to the 'dot-com' crash of 2000-02. It is intended to be seen in contrast to a selective framing of 'Web 1.0', which characterized the web of the 1990s.

ENDNOTES

1. Both entries are taken from the Oxford Dictionary of Media and Communication.
2. B. Nimmo, *Identifying Disinformation: an ABC*, Institute for European Studies (2016).
3. Perhaps ironically, the term itself is deceptive, as it was made to be vaguely French-sounding, and was even given a false French etymology in the Soviet Encyclopedia.
4. R. Godson, *Written Testimony to the Senate Select Committee on Intelligence, Open Hearing, March 30, 2017: Disinformation: A Primer in Russian Active Measures and Influence Campaigns* (2017), p. 1.
5. *Ibid.*, p. 11.
6. We define social media as “websites and applications that enable users to create and share content or to participate in social networking” (Oxford Online Dictionary).
7. K. Starbird, interviewed by L. Garcia Navarro, ‘How Misinformation Spreads On The Internet’, NPR (2017), available at <http://www.npr.org/2017/04/09/523170115/how-misinformation-spreads-on-the-internet-and-how-to-stop-it>, accessed on 04/07/2017.
8. See for example A. Jamieson, ‘You are fake news: Trump attacks CNN and BuzzFeed at press conference’, *The Guardian*, 11 January 2017 <https://www.theguardian.com/us-news/2017/jan/11/trump-attacks-cnn-buzzfeed-at-press-conference>
9. Neil Durkin, ‘Don’t Believe The Hype Around Fake News’, *Huffington Post UK*, 17 March 2017 http://www.huffingtonpost.co.uk/neil-durkin/fake-news_b_15387590.html; see also C. Archetti, ‘The Future of Social Media: Strategic Communication, Politics & Context’, unpublished seminar paper (Trends in Social Media and their Further Development seminar, Riga, 20 March 2017)
10. Google NewsLab <https://newslab.withgoogle.com>; ‘Fact Check now available in Google Search and News around the world’, *Google Blog* <https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world>
11. <https://www.poynter.org/tag/international-fact-checking-network/>; <https://firstdraftnews.com>
12. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, et al. (2017) Debunking in a world of tribes. *PLOS ONE* 12(7); Peter, C., & Koch, T. (2016) When Debunking Scientific Myths Fails (and When It Does Not) The Backfire Effect in the Context of Journalistic Coverage and Immediate Judgments as Prevention Strategy. *Science Communication*, 38(1); Nyhan, B., & Reifler, J. (2010) When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2).
13. J. Weedon, W. Nuland, A. Stamos, *Information Operations and Facebook*, Facebook (2017), p. 4 <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and->

- information-operations-v1.pdf
14. Ibid., p. 5
 15. R. Waltzman, *The Weaponization of Information: The Need for Cognitive Security*, p. 2. Emphasis added
 16. 'Yes, I'd lie to you', *The Economist*, 10 September 2016 <http://www.economist.com/news/briefing/21706498-dishonesty-politics-nothing-new-manner-which-some-politicians-now-lie-and>
 17. 'Umberto Eco e i social: Danno diritto di parola a legioni di imbecilli', *La Repubblica*, 11 June 2015 <http://video.repubblica.it/tecnologia/scienze/umberto-eco-e-i-social--danno-diritto-di-parola-a-legioni-di-imbecilli/203952/203032>
 18. 'The Global Risks Report 2016', *World Economic Forum* (2016), p.40.
 19. R. Younes and E. Mackintosh, 'Trusting What You See Online - It's Not Just About the Tools', in *Finding the Truth amongst the Fakes*, Al Jazeera Media Institute (2017), p. 44
 20. J. Gottfried and E. Shearer, 'News Use Across Social Media Platforms 2016', *Pew Research Center* <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/#fn-55250-1>
 21. In this context, the term is synonymous with the more common "filter bubble".
 22. C. Watts, *Statement Prepared for the U.S. Senate Select Committee on Intelligence hearing: Disinformation: A Primer In Russian Active Measures And Influence Campaigns* (US Senate, 30 March 2017), available at <https://www.intelligence.senate.gov/sites/default/files/documents/os-cwatts-033017.pdf>, accessed on 28/06/2017. P. 7.
 23. Functional illiteracy is the inability to understand information correctly.
 24. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, et al. (2017) *Debunking in a world of tribes*. *PLOS ONE* 12(7), p. 9.
 25. H. M. Claypool et al. 'The effects of personal relevance and repetition on persuasive processing', *Social Cognition* 22/3 (2004), pp. 310-335; see also J. W. Alba and H. Marmorstein 'The effects of frequency knowledge on consumer decision making', *Journal of Consumer Research* 14/1 (1987), pp. 14-25
 26. L. A. Henkel and M. E. Mattson 'Reading is believing: The truth effect and source credibility' *Consciousness and cognition* 20/4 (2011), pp. 1705-1721
 27. T. Garcia-Marques and D. M. Mackie 'The feeling of familiarity as a regulator of persuasive processing' *Social Cognition* 19/1 (2001), pp. 9-34
 28. H. M. Claypool et al. 'The effects of personal relevance and repetition on persuasive processing', pp. 310-335
 29. J. Weedon, W. Nuland, A. Stamos, *Information Operations and Facebook*, p. 6
 30. *Bullet points taken from C. Hadnagy, Social Engineering*, Wiley (2011), pp. 233-234
 31. This in turn is done to defame, ridicule, and threaten the targets. See S. Svetoka, 'Social Media as a Tool of Hybrid Warfare', *NATO StratCom COE* (2016), p. 20
 32. Reflexive control is defined as "a means of conveying to a partner or an opponent specially prepared information to incline him to voluntarily make the predetermined decision desired by the initiator of the action". T. Thomas, 'Russia's Reflexive Control Theory and the Military',

- Journal of Slavic Military Studies 17/2 (2004), p. 237
33. Emotional content is not necessarily false, although it can be used in association with false content.
 34. These narratives are likely to feature predominantly negative sentiment: “[p]ropaganda seeks out and exploits the most powerful emotions (...), it is primarily in the negative emotions that propaganda activities reside. In psychological terms, we understand what we hate better than what we like”. N. O’Shaughnessy, ‘Putin, Xi, and Hitler – Propaganda and the paternity of pseudo democracy’, *Defence Strategic Communications 2* (2017), p. 123
 35. S. Svetoka, ‘Social Media as a Tool of Hybrid Warfare’, p. 20
 36. This is done through the use of automated and semi-automated accounts.
 37. S. Svetoka, ‘Social Media as a Tool of Hybrid Warfare’, p. 20
 38. Such as Google’s targeted counter-radicalization initiatives. See B. Quinn, ‘Google to point extremist searches towards anti-radicalisation websites’, *The Guardian*, <https://www.theguardian.com/uk-news/2016/feb/02/google-pilot-extremist-anti-radicalisation-information>. Other initiatives, like Quilliam’s #NotAnotherBrother campaign, are different, because they make use of emotional content. See <https://www.youtube.com/watch?v=ljIQ0ctzyZE>
 39. B. Heap, ‘Strategic Communications: Insights from the Commercial Sector’, *NATO StratCom COE* (2017), p. 17.
 40. See, for example, the charity Full Fact: <https://fullfact.org/automated>
 41. L. Gu, V. Kropotov, and F. Yarochkin, *The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public* (TrendMicro, 2017), available at: <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/fake-news-cyber-propaganda-the-abuse-of-social-media>, accessed 14 June 2017, p. 6.
 42. F. Gillette, *The Rise and Inglorious Fall of Myspace* (Bloomberg, 2011), available at: <https://www.bloomberg.com/news/articles/2011-06-22/the-rise-and-inglorious-fall-of-myspace>, accessed 13 June 17.
 43. S. Kemp, ‘Digital in 2017: Global Overview’, available at: <https://wearesocial.com/special-reports/digital-in-2017-global-overview>, accessed 12 July 2017
 44. R. Hutt, *The World’s Most Popular Social Networks Mapped*, (World Economic Forum, 2017), available at: <https://www.weforum.org/agenda/2017/03/most-popular-social-networks-mapped/>, accessed on 08 June 2017.
 45. The debate was popular on mainstream media. See, for example, K. Hosanagar, *Blame the Echo Chamber on Facebook. But Blame Yourself, Too* (2016), available at: <https://www.wired.com/2016/11/facebook-echo-chamber/>, accessed on 25 July 2017.
 46. A. Mosseri, *News Feed FYI: Addressing Hoaxes and Fake News* (Facebook, 2017), available at: <https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/>, accessed on 13 June 2017.
 47. ‘Continued influence’ refers to the widespread tendency among the general populace to believe in misinformation after corrections of

- false statements have been issued: in other words, misinformation is proved to be resistant to correction. See S. Lewandowsky et al., *Misinformation and Its Correction Continued Influence and Successful Debiasing* (2012), *Psychological Science in the Public Interest* 13 (3). See also F. Zollo, A. Bessi, M. Del Vicario, A. Scala, G. Caldarelli et al. (2017) *Debunking in a world of tribes*. *PLOS ONE* 12(7), p. 9.
48. Melissa Zimdars, assistant professor of communication and media at Merrimack College, in S. Levin, *Facebook Promised To Tackle Fake News But The Evidence Shows It's Not Working* (The Guardian, 16 May 2017), available online at: <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>, accessed on 16 June 2017.
 49. *Ibid.*
 50. L. Bounegru et al., *A Field Guide to Fake News* (Public Data Lab, 2017), p. 16.
 51. This topic became particularly popular over the last year, when a number of companies claimed to have applied target audience analysis (variously paraphrased) to steer the results of major political events worldwide. This study does not name these commercial entities.
 52. F. Zollo et al. (2017) *Debunking in a world of tribes*. *PLOS ONE* 12(7), p. 8.
 53. *Ibid.*
 54. P. Chamberlain, 'Twitter as a Vector for Disinformation', *Journal of Information Warfare* 9/1 (2010), 6.
 55. *Ibid.*, 4.
 56. See glossary.
 57. G. Moraetes, 'Information Security in the Age of Disinformation', *IBM Security Intelligence* (2017), available at: <https://securityintelligence.com/information-security-in-the-age-of-disinformation/>, accessed on 30 June 2017.
 58. Regarding automated activity, the StratCom COE is in the process of launching a regular product focused on robotic trolling.
 59. O. Varol, E. Ferrara, C.A. Davis, F. Menczer, & A. Flammini, (2017). *Online human-bot interactions: Detection, estimation, and characterization*. arXiv preprint arXiv:1703.03107.
 60. R. Fredheim (2017) *Robotrolling 1*. Available at <http://stratcomcoe.org/robotrolling-20171>, accessed on 03/10/2017.
 61. E. Dwoskin, 'Twitter is looking for ways to let users flag fake news, offensive content' (The Washington Post, 29 June 2017), available at: https://www.washingtonpost.com/news/the-switch/wp/2017/06/29/twitter-is-looking-for-ways-to-let-users-flag-fake-news/?utm_term=.79db7791edca, accessed on 30 June 2017.
 62. Such as network analysis, temporal analysis, sentiment analysis, etc.
 63. See for example Jennifer Keelan et al. 'YouTube as a source of information on immunization: a content analysis' *Jama* 298.21 (2007): 2482–2484. Despite the fact that YouTube had been launched merely two years before, 'anti-vaxxers' had already discovered its potential for sharing their content in a more permissive environment than that of traditional media. Nowadays, virtually all conspiracy theories are represented in the YouTube galaxy, but it is beyond the scope of this study to list them.
 64. Just in the US, 10% of adults

- reported in 2016 that they got their news from YouTube: J. Gottfried and E. Shearer, 'News Use Across Social Media Platforms 2016', Pew Research Center <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/#fn-55250-1>
65. 'YouTube to offer fake news workshops to teenagers', BBC Newsbeat (21 April 2017), available at: <http://www.bbc.co.uk/newsbeat/article/39653429/youtube-to-offer-fake-news-workshops-to-teenagers>, accessed on 27 June 2017.
 66. Introducing Expanded YouTube Partner Program Safeguards to Protect Creators, YouTube Creator Blog (06 April 2017), available at: <https://youtube-creators.googleblog.com/2017/04/introducing-expanded-youtube-partner.html>, accessed on 28 April 2017.
 67. Ibid.
 68. Instagram Help Centre, available at: <https://help.instagram.com/370054663112398>, accessed on 28 June 2017.
 69. Instagram Community Guidelines, available at: <https://help.instagram.com/477434105621119>, accessed on 28 June 2017.
 70. One of the most successful of said platforms, Gab, counts little more than 180,000 users in mid-2017: <https://gab.ai/a/posts/8106308>, accessed on 30 June 2017.
 71. See glossary.
 72. It must, however, be noted that obstacles such as debunking of false stories et similia might not be obstacle at all, as previously noted.
 73. A. Elsheikh, 'Finding Your Story: Which Platform and Where?', in Finding the Truth amongst the Fakes, Al Jazeera Media Institute (2017), p. 138.
 74. See 'Social media and its influence on the Arab Spring' (Al Jazeera America, 2015), available at: <http://america.aljazeera.com/watch/shows/live-news/2015/12/social-media-and-its-influence-on-the-arab-spring-movement.html>, accessed on 15 June 2017.
 75. M. Esseghaier, 'Tweeting Out a Tyrant: Social Media and the Tunisian Revolution', Wi Journal of Mobile Media 11 (1), available at: <http://wi.mobilities.ca/tweeting-out-a-tyrant-social-media-and-the-tunisian-revolution/>, accessed on 15 June 2017.
 76. Data from: Digital in 2017: Global Overview (We Are Social, 2017), available at: <https://wearesocial.com/special-reports/digital-in-2017-global-overview>, accessed on 14 June 2017.
 77. A. Elsheikh, 'Finding Your Story', p. 139.
 78. Ibid., emphasis added.
 79. L. Gu et al, The Fake News Machine, p. 9.
 80. Ibid., p. 34.
 81. The name can be translated as 'Is this serious?/ 'Is this real?'. Website available at: <https://dabegad.com>, accessed on 24 August 2017.
 82. Da Begad's Website, 'Who are we?' [in Arabic]: <https://dabegad.com/%D8%B9%D9%86-%D8%AF%D9%87-%D8%A8%D8%AC%D8%AF>, accessed on 14 June 2017.
 83. H. A. Unver, 'Can Fake News Lead To War? What The Gulf Crisis Tells Us', War on the Rocks, available at: <https://warontherocks.com/2017/06/can-fake-news-lead-to-war-what-the-gulf-crisis-tells-us/>, accessed on 16 June 2017.

84. B. Heap, 'Strategic Communications: Insights from the Commercial Sector', NATO StratCom COE (2017), p. 17.
85. Ibid.
86. US Department of State, Country Reports on Terrorism 2016 (Country Reports: Middle East and North Africa) (2017), available at: <https://www.state.gov/j/ct/rls/crt/2016/272232.htm>, accessed on 21 July 2017.
87. Ibid.
88. As explained by a representative of the Global Coalition during the Foreign Terrorist Fighters Working Group Meeting (15 March 2017).
89. US Department of State, Country Reports on Terrorism 2016 (Country Reports: Middle East and North Africa) (2017), available at: <https://www.state.gov/j/ct/rls/crt/2016/272232.htm>, accessed on 21 July 2017, emphasis added.
90. See 'Fighting the Cyber-Jihadists', The Economist (10 June 2017).
91. T. Fox-Brewster, 'With Fake News And Femmes Fatales, Iran's Spies Learn To Love Facebook', Forbes (2017), available at: <https://www.forbes.com/sites/thomasbrewster/2017/07/27/iran-hackers-oilrig-use-fake-personas-on-facebook-linkedin-for-cyberespionage/#56002c2e49af>, accessed on 28 July 2017.
92. Note: VK was formerly Vkontakte, ВКонтакте, 'In Contact' in Russian; both names are still encountered.
93. Note: The platform is also popular in Belarus, Kazakhstan, and Ukraine, with 11.9 million users in the last-named, although this number is expected to diminish following recent action by the government of Ukraine to ban the Russian-owned social networks. Kemp, Simon, 'Digital in 2017: Global Overview' Report, We Are Social, January 2017. Accessed 9 August 2017: <https://wearesocial.com/special-reports/digital-in-2017-global-overview>
94. Fanteev, Frank, '300+ Million Users: Understanding Russia's VK Social Network,' Digital Marketing Magazine, 2015. Accessed 9 August 2017: <http://digitalmarketingmagazine.co.uk/social-media-marketing/300-million-users-understanding-russia-s-vk-social-network/2564>; Pavelek, Ondrej, 'Vkontakte Demographics', Havas Worldwide, 2013, Accessed: <https://www.slideshare.net/andrewik1/vkontakte-demographics>
95. Note: Aric Toler of Bellingcat points to granular search, detailed information about military service and communities for military units as critical in Bellingcat's open source research. Toler, Aric, 'The Open Source Guidebook to RuNet'. Accessed 9 August 2017: <https://medium.com/1st-draft/how-to-get-started-investigating-the-russian-language-internet-3a934b9d55e2>
96. Note: Local and military service groups figured prominently in the spread of disinformation about MH17 downing and, ironically, in the digital forensic analysis tracking down its perpetrators. Bellingcat, 'MH17: The Open Source Investigation Three Years Later'. Accessed 9 August 2017: <https://www.bellingcat.com/wp-content/uploads/2017/07/mh17-3rd-anniversary-report.pdf>; Bellingcat Investigation Team, 'Pre-MH17 Photograph of Buk 332 Discovered', June 5, 2017. Accessed 9 August 2017: <https://www.bellingcat.com/news/uk-and-europe/2017/06/05/pre-mh17-photograph-buk-332>

- discovered/
97. Zhdanova, Mariia & Orlova, Dariya, 'Computational Propaganda in Ukraine: Caught between external threats and internal challenges,' in Samuel Woolley and Philip N Howard, eds, Working Paper 2017.9, Oxford, UK: Project on Computational Propaganda. Accessed 9 August 2017: <http://comprop.oii.ox.ac.uk/2017/06/19/computational-propaganda-in-ukraine-caught-between-external-threats-and-internal-challenges/>
 98. Odnoklassniki (Одноклассники) or 'Classmates' in Russian.
 99. Ghedin, Guido, 'Odnoklassniki: Users, Features and the Power of Communities', Digital in the Round, 13 December 2013. Accessed 9 August 2017: <http://www.digitalintheround.com/odnoklassniki-users-features-communities/>
 100. Russian Search Tips, 'Top social networks in Russia: latest numbers and trends', 20 January 2015. Accessed 9 August 2017: <http://www.russiansearchtips.com/2015/01/top-social-networks-russia-latest-numbers-trends/>
 101. Ghedin, 'Odnoklassniki'.
 102. Sivertseva, Ekaterina, 'Odnoklassniki and MoiMir Bring TV Shows to Russian Internet Users', Digital in the Round, 22 May 2014. Accessed 9 August 2017: <http://www.digitalintheround.com/odnoklassniki-moimir-tv-russia/>
 103. Pomerantsev, P and Weiss, M, 'The Menace of Unreality: How the Kremlin Weaponizes Information, Culture, and Money', Institute of Modern Russia. Accessed: https://imrussia.org/media/pdf/Research/Michael_Weiss_and_Peter_Pomerantsev_The_Menace_of_Unreality.pdf
 104. Note: Translated from Russian. Clockwise from top left: "Little bear kisses mom", "Grandmothers at the door when I get home", "It is necessary to remember the past, but not to live it. The past is past", "It is possible to question Putin's merits towards the Fatherland. But the return of the Crimea is priceless"
 105. Moi Mir (Мой Мир) or 'My World' in Russian.
 106. Note: There are also around 25 million Russian-speaking Facebook users.
 107. Toler, Aric, 'What you Need to Know About Russian Social Networks to Conduct Open-Source Research,' GlobalVoices, 21 October 2015. Accessed 9 August 2017: <https://globalvoices.org/2015/10/21/what-you-need-to-know-about-russian-social-networks-to-conduct-open-source-research/>
 108. Sivertseva, 'Odnoklassniki and MoiMir'.
 109. Pomerantsev & Weiss, 'The Menace of Unreality'.
 110. Toler, 'What you Need to Know About Russian Social Networks',
 111. Fedor, Julia., & Fredheim, Rolf. (2017). 'We need more clips about Putin, and lots of them:' Russia's state-commissioned online visual culture. Nationalities Papers 45(2), 161–181
 112. Figure 8. 2017 Digital Yearbook by We Are Social Singapore, available at: <https://www.slideshare.net/wearesocialsg/2017-digital-yearbook?ref=http://www.digitalstrategyconsulting.com/intelligence/russia-digital-marketing/>
 113. Baran, Katsiaryna, & Stock,

- Wolfgang, 'Facebook has Been Smacked Down. The Russian Special way of SNSs: Vkontakte as a Case Study' in ECSM 2015 - The Proceedings of the 2nd European Conference on Social Media. Accessed 9 August 2017: http://www.isi.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Informationswissenschaft/Dateien/Wolfgang_G._Stock/Baran_2015_ECSM_2015_Proceedings-276.pdf
114. Note: Facebook users in Russia tend to be university graduates and young professionals fall within the age group 24–40. Sikorska, Olena, 'VKontakte vs. Facebook: How Russians consume social networks? (Infographic)', Digital EastFactor, 12 May 2014. Accessed 9 August 2017. <http://www.digitaleastfactor.com/vkontakte-vs-facebook-russians-consume-social-networks-infographic/>
115. Anna Lubov, 'Top social networks in Russia: latest trends, winter 2015–2016', Russian Search Tips. Accessed 9 August 2017: <http://www.russiansearchtips.com/2016/03/top-social-networks-in-russia-latest-trends-in-winter-2015-2016/>
116. Fedor & Fredheim, 'We need more clips about Putin', 161–181.
117. Lipman, M. 'Media manipulation and political control in Russia.' Chatham House, 2009. Accessed 9 August 2017: <https://www.chathamhouse.org/sites/files/chathamhouse/public/Research/Russia%20and%20Eurasia/300109lipman.pdf>
118. Sanovich, Sergey. 'Computational Propaganda in Russia: The Origins of Digital Misinformation' in Samuel Woolley and Philip N Howard, Eds Working Paper 2017.9, Oxford, UK: Project on Computational Propaganda Accessed 9 August 2017: <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Russia.pdf>
119. Ibid.
120. Note: It was in the interest of the Kremlin and encouraged by the Kremlin that the viable competitive technology sector should take advantage of the engineering talent pool, as is evidenced by Putin's early promise, which surprised the technology executives: 'Whenever we'll have to choose between excessive regulation and protection of online freedom, we'll definitely opt for freedom'; this is how as the late Anton Nossik, CEO of the prominent online news outlets, recalled the meeting. In Nossik. Anton, 'I helped build Russia's Internet. Now Putin wants to destroy it', The New Republic, 15 May 2014. Accessed 9 August 2017: <http://www.newrepublic.com/article/117771/putinsinternet-crackdown-russias-first-blogger-react>
121. Etling, B., Alexanyan, K., Kelly, J., Faris, R., Palfrey, J. G., & Gasser, U. (2010). 'Public discourse in the Russian blogosphere: Mapping RuNet politics and mobilization', Berkman Center, Research Publication No. 2010–11. Accessed 9 August 2017: <http://papers.ssrn.com/abstract=1698344>
122. Note: In search of his own power base leading up to the presidential re-election campaign, Dmitry Medvedev, in contrast to Putin, turned to RuNet and to the educated middle-class professionals who consumed and engaged in it. Shortly after assuming his post as President, he established

- a social media presence, famously earning himself a reputation as 'Blogger-in-Chief.' See Sanovich, 'Computational Propaganda in Russia'.
123. Ibid.
 124. Ibid.
 125. Fedor & Fredheim, 'We need more clips about Putin', 161–181.
 126. Toor, A. (2014) 'How Putin's cronies seized control of Russia's Facebook': <http://www.the-village.ru/village/business/story/150063-kak-otbirali-vkontakte>, accessed on 14/09/2017.
 127. The full spectrum of disinformation in Ukraine is documented by numerous StratCom Center of Excellence studies. See: <http://www.stratcomcoe.org/analysis-russias-information-campaign-against-ukraine-1>; <http://www.stratcomcoe.org/framing-ukraine-russia-conflict-online-and-social-media>
 128. The following are examples of fake news unmasked by the Stop Fake organisation. 'Video with Russian "GRAD" volleys aimed at South Ossetia was presented as events in Sloviansk': <http://www.stopfake.org/en/video-with-russian-grad-volleys-aimed-at-south-ossetia-was-presented-as-events-in-sloviansk/>; 'Photo from China Dated 1989 Presented as the Actual Events in Donbass': <http://www.stopfake.org/en/photo-from-china-dated-1989-presented-as-the-actual-events-in-donbass/>
 129. Note: The current population of Ukraine is 45 million. 63% of the adult population are active internet users, and 21% of these use social media as the main source of news.
- The most popular social networks among Ukrainians were VK (11.9 million users), Facebook (over 8 million), Odnoklassniki (5.7 million) and Twitter (2.5 million). Detector Media. (2017). Як російська пропаганда впливає на суспільну думку в Україні (дослідження) (How Russian propaganda influences public opinion in Ukraine [research findings]). Retrieved from: http://osvita.mediasapiens.ua/mediaprovsvita/research/yak_rosiyska_propaganda_vplivae_na_suspilnu_dumku_v_ukraini_doslidzhennya/
130. Note: Moi Mir is the property of the government-approved Mail.Ru conglomerate.
 131. Указ Президента України №133/2017Ж Про рішення Ради національної безпеки і оборони України від 28 квітня 2017 року 'Про застосування персональних спеціальних економічних та інших обмежувальних заходів (санкцій)' (Decree of the President of Ukraine № 133/201ж 'On the decision of the Council of National Security and Defense of Ukraine dated April 28, 2017 "On the Application of Personal Special Economic and Other Restrictive Measures [Sanctions]"'). <http://www.president.gov.ua/documents/1332017-21850>
 132. 'In Ukraine Facebook Grows at the Expense of Russian Competitors', Gemius Global, 14 July 2017, <https://www.gemius.com/all-reader-news/in-ukraine-facebook-grows-at-the-expense-of-russian-competitors.html>; Oleg Dytrenko, 'Facebook обійшов ВКонтакте вже в перший тиждень після введення санкцій проти російських соцмереж' ('Facebook has surpassed VK on the first week

- after the introduction of sanctions against Russian social networks'), Watcher, <http://watcher.com.ua/2017/06/07/facebook-obiyshov-vkontakte-vzhe-v-pershyy-tyzhden-pislyya-vvedennya-sanktsiy-proty-rosiyskyh-sotsmerezh/>
133. Note: Telegram is an encrypted messaging platform, launched in 2013 by Pavel Durov. It has been under pressure to cooperate with the Russian government or face shutdown.
 134. <https://www.cnet.com/news/telegram-registers-in-russia-wont-share-user-data/>
 135. Note: VPNs (Virtual Private Networks) enable users to bypass blocks and navigate to censored sites. <http://www.reuters.com/article/us-russia-internet-idUSKBN1AF0QI>
 136. This is exemplified by the case of a Norwegian Facebook group opposed to immigration that mistook a picture of empty bus seats for burqa-clad women. The picture was commented on, liked, and shared by a considerable number of angered users before the mistake was detected: <https://www.thelocal.no/20170731/norwegian-anti-immigrant-facebook-groups-confuses-empty-bus-seats-with-terrorists>, accessed on 02 August 2017.
 137. Technorati, "State of the Blogosphere 2011," 2011, <http://technorati.com/state-of-the-blogosphere-2011/>.
 138. Nitin Agarwal et al., "Examining The Use Of Botnets And Their Evolution In Propaganda Dissemination," *Defence Strategic Communications 2* (2017): 87–112.
 139. Muhammad Nihal Hussain, Saaduddin Ghouri Mohammad, and Nitin Agarwal, "Blog Data Analytics Using Blogtrackers" (International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation, July 2017).
 140. Tim O'Reilly, "How I Detect Fake News," December 2016, <http://www.kdnuggets.com/2016/12/oreilly-detect-fake-news.html>; Melissa Zimdars, "False, Misleading, Clickbait-y, and Satirical 'News' Sources," Google Docs, 2016, https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QG5ewC3VAL6pLkT53V_81ZyitM/preview?usp=embed_facebook; Krishna Bharat, "How to Detect Fake News in Real-Time," NewCo Shift, April 27, 2017, <https://shift.newco.co/how-to-detect-fake-news-in-real-time-9fdae0197bfd>.
 141. Scot Macdonald, *Propaganda and Information Warfare in the Twenty-First Century: Altered Images and Deception Operations* (Routledge, 2006).
 142. This story was reported as conspiracy theory by bsdetector.tech
 143. We used ORA-Lite (Organization Risk Analyzer), available at <http://www.casos.cs.cmu.edu/projects/ora/software.php>
 144. The names of the blog sites have been smudged to keep the identity of the bloggers anonymous.
 145. This concept is widely studied in communication literature under the heading Exemplification Theory.
 146. Patric R. Spence et al., "That Is So Gross and I Have to Post About It: Exemplification Effects and User Comments on a News Story," *Southern Communication Journal* 82, no. 1 (2017): 27–37.
 147. Zillmann, D. (2002).

- Exemplification theory of media influence. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 19–41). Mahwah, NJ: Lawrence Erlbaum Associates
148. <https://www.iab.com/2013-internet-ad-revenues-soar-to-42-8-billion-hitting-landmark-high-surpassing-broadcast-television-for-the-first-timemarks-a-17-rise-over-record-setting-revenues-in-2012/>
 149. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v18_2012_21291018.en-us.pdf
 150. <https://newsroom.fb.com/news/2017/09/information-operations-update/>
 151. *Ibid.*
 152. http://www.springer.com/cda/content/document/cda_ *Ibid.*
 153. *Ibid.*
 154. *Ibid.*
 155. http://www.springer.com/cda/content/document/cda_downloaddocument/9789400729025-c1.pdf?S-GWID=0-0-45-1302338-p174266596
 156. Tracking users across different visits and/or across different sites.
 157. Snippet of JavaScript code or executable file. In the context of websites, used to implement behaviour, change page content etc.
 158. https://www.nccgroup.trust/globalassets/our-research/us/whitepapers/isec_cleaning_up_after_cookies.pdf
 159. An HTML document embedded inside another HTML document on a website. IFrames are often used to insert content from another source, such as an advertisement.
 160. http://www.springer.com/cda/content/document/cda_downloaddocument/9789400729025-c1.pdf?S-GWID=0-0-45-1302338-p174266596
 161. Company that collects personal information about consumers from public and non-public sources and sells that information to other organizations. Data brokers create profiles on users for marketing purposes and sell them to businesses who want to target their advertisements.
 162. http://www.springer.com/cda/content/document/cda_downloaddocument/9789400729025-c1.pdf?S-GWID=0-0-45-1302338-p174266596
 163. Advertising on a website that is targeted to be relevant to the page's content.
 164. https://otalliance.org/system/files/files/resource/documents/report_-_online_advertising_hidden_hazards_to_consumer_security_date_privacy_may_15_20141.pdf
 165. https://otalliance.org/system/files/files/resource/documents/report_-_online_advertising_hidden_hazards_to_consumer_security_date_privacy_may_15_20141.pdf
 166. *Ibid.*
 167. <https://www.eff.org/files/onlineprivacylegprimersept09.pdf>
 168. *Ibid.*
 169. <https://newsroom.fb.com/news/2017/09/information-operations-update/>
 170. Sites were the Fakeness Index was 0.9 or higher.
 171. Fakeness Index – a value between 0 and 1 expressing the ratio of a cookie occurrences on fake news media sites against its occurrences on legitimate news media sites. The higher the index, the “faker” the cookie. This index was developed to narrow down the cookies of interest.
 172. See Turlas' watering hole cam-

- campaign that delivered fingerprinting scripts based on the IP address range requests were coming from: <https://www.welivesecurity.com/2017/06/06/turlas-watering-hole-campaign-updated-firefox-extension-abusing-instagram/>
173. WOT is a website reputation and review service that helps people make informed decisions about whether to trust a website or not (see <https://www.mywot.com/>).
 174. Admerge, x01.aidata.io, ads.betweendigital.com, digitaladsystems.com, dmg.digitaltarget.ru, doubleclick.net, mail.ru, marketgid.com, otm-r.com, sync.dmp.otm-r.com, rambler.ru, republer.com, rtb.com.ru, rutarget.ru, ssp-rtb.sape.ru, targeterra.info, targetix.net, upravel.com
 175. Service for creating backlinks to sites to increase web traffic; the company is owned by Butko
 176. <https://stackoverflow.com/questions/23411188/hidden-malicious-script-insertinga-code-into-html-web-page-how-to-remove-clean>
 177. tovarro.com and lentainform.com
 178. An interface component that enables a user to perform a function or access a service.
 179. Facebook data structure. Explained here: <http://www.businessinsider.com/explainer-what-exactly-is-the-social-graph-2012-3>
 180. <https://muckrack.com/whoshared/>
 181. <http://conferences.sigcomm.org/sigcomm/2009/workshops/wosn/papers/p7.pdf>
 182. Some web servers have security software installed which strips the referrer from all requests.
 183. <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>
 184. This example, about Jewish users, is particularly significant: S. A. O'Brien and D. O'Sullivan, How Facebook knows you're Jewish, CNN (2017), available at <http://money.cnn.com/2017/09/21/technology/business/facebook-rosh-hashanah-ad-targeting/index.html>, accessed on 22/09/2017.
 185. C. Leonnig, T. Hamburger and R. Helderman (2017) 'Russian firm tied to pro-Kremlin propaganda advertised on Facebook during election', available at https://www.washingtonpost.com/politics/facebook-says-it-sold-political-ads-to-russian-company-during-2016-election/2017/09/06/32f01fd2-931e-11e7-89fa-bb822a46da5b_story.html?utm_term=.dd26689bcd74, accessed on 25/09/2017.
 186. A. Hern (2017), 'Apple blocking ads that follow users around web is 'sabotage', says industry', available at <https://www.theguardian.com/technology/2017/sep/18/apple-stopping-ads-follow-you-around-internet-sabotage-advertising-industry-ios-11-and-macos-high-sierra-safari-internet>, accessed on 21/09/2017.
 187. See the dedicated website: <http://www.eugdpr.org/>.