Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-020-00659-9 OPEN

Measuring the scope of pro-Kremlin disinformation on Twitter

Yevgeniy Golovchenko [1]✉

This article examines the scope of pro-Kremlin disinformation about Crimea. I deploy content analysis and a social network approach to analyze tweets related to the region. I find that pro-Kremlin disinformation partially penetrated the Twitter debates about Crimea. However, these disinformation narratives are accompanied by a much larger wave of information that disagrees with the disinformation and are less prevalent in relative terms. The impact of Russian state-controlled news outlets—which are frequent sources of pro-Kremlin disinformation—is concentrated in one, highly popular news outlet, RT. The few, popular Russian news media have to compete with many popular Western media outlets. As a result, the combined impact of Russian state-controlled outlets is relatively low when comparing to its Western alternatives.

[1] Department of Political Science, University of Copenhagen, Copenhagen, Denmark. ✉email: yg@ifs.ku.dk

Introduction

Following Russia's annexation of Crimea in 2014, both scholars and authorities from a wide range of countries have raised concern with the Kremlin's strategic use of (dis) information (Abrams, 2016; Department of National Intelligence, 2017; European Parliament News, 2016; Renz, 2016). Literature from various disciplines, ranging from security studies to computational science, offers many examples of how Russian state-affiliated institutions use social media to target societies with disinformation—both domestically and abroad (Badawy et al., 2018; Bjola and Pamment, 2016; DiResta et al., 2018; Howard et al., 2018; Linvill et al., 2019; Zannettou et al., 2019b). 'Disinformation' refers to content that is *intentionally* misleading, whereas 'misinformation' is not necessarily intended to mislead (Fallis, 2015; Søe, 2016).

Scholars within security studies, political science, and similar fields often describe these campaigns as information warfare (Darczewska, 2014; Thornton, 2015). The term refers to strategic and manipulative use of information for the purpose of achieving a political or military goal (Myriam Dunn Cavelty, 2008; Taylor, 2003; Thornton, 2015) and it is often referred to as "hybrid warfare" when combined with military operations (Lanoszka, 2016; Reisinger and Gol'c, 2014; Thiele, 2015; Woo, 2015).

Russia's use of information warfare in Crimea—to mobilize support from the local population and to sow confusion in the international scene—is often regarded as a turning point and a foreshadowing of Russian interference in Western societies through social media. Authorities in both US and European countries have responded to Russian "information warfare" through a wide series legislatures and initiatives. For example, the EU has established the EU StratCom Task Force in 2015 to monitor and address disinformation from the east (EU, 2018; European Parliament, 2016), while the US authorities have indicted at least 13 Russian individuals in 2018 for interfering in the 2016 presidential elections through an online disinformation campaign (Department of Justice, 2018).

These policies and debates are in part driven by the notion that Russia is highly capable of influencing societies abroad through information warfare (Unver, 2019) and cyber warfare (Jamieson, 2018), while some commentators even argue that the West is losing the informational struggle to Russia (Lockie, 2017; Lockwood, 2018; Torossian, 2016; Wallance, 2018). At the same time, there is no consensus among scholars on the potential effects of Russian interference in the West through hacking and fake online accounts (Lawrence, 2019; Nyhan, 2018). For example, Jamieson (2018) draws on existing communication literature to present an account of how and when such campaigns are likely to affect US elections. In a more recent empirical study based on a longitudinal survey, Bail et al. (2020) find no evidence that interaction with fake Russian profiles from the Internet Research Agency (IRA) on Twitter had an effect on political attitudes or behavior among American users. Although these profiles use fake online personas—often presenting themselves as American or local news outlets—their actual tweets often draw on news from credible news media and are far from always misleading (Yin et al., 2018).

While (Vosoughi et al., 2018) show empirically that the false content about different topics (not limited to Russia) spreads "... significantly farther, faster, deeper, and more broadly than the truth" on Twitter (Vosoughi et al., 2018, p. 1146), an increasing number of studies across multiple disciplines show that misinformation and disinformation—although highly problematic—comprises only a small proportion of online content consumed by regular users (Allcott and Gentzkow, 2017; Guess et al., 2018). For example, Grinberg et al. (2019) find that only 5% of all exposures to political URL's on Twitter during the 2016 US elections campaign were tied to "fake news sources". They also find that only "...1% of individuals accounted for 80% of fake news source exposures" (Grinberg et al., 2019, pp. 1–2)—findings that are very similar to Guess et al. (2019) study of exposure to misinformation on Facebook. Despite of the vast focus on Russian disinformation campaigns and information warfare, little is known about the breadth of Russian disinformation on social media.

This article seeks to nuance this notion empirically, by using social network—and content analysis to the examine the debate on Twitter about the Crimean crisis and Crimea more generally. The analysis is centered on the following research question: *How prevalent is Russian disinformation about Crimea on Twitter?*

This research does not rely on a hypothesis-driven approach or test whether Russia achieved a specific strategic goal or whether the Kremlin is "winning" the information war on Twitter. Many researchers have argued that the overall strategy behind Kremlin-driven disinformation towards foreign audiences in similar cases is often to sow doubt and confusion about real events (Lucas and Nimmo, 2015; Thornton, 2015). However, the exact objectives used to achieve this end-goal in Twitter conversations about Crimea are not known. For example, The purpose behind the disinformation campaign could be to sow doubt about events in Crimea by (1) targeting a specific demographic group, (2) dominating the overall stream of information among a broader audience, or (3) merely establishing a limited foothold in a Western-dominated media environment.[1] As a result, this article does not offer objective criteria for evaluating whether the disinformation campaign has failed or whether the disinformation had an effect on political attitudes and behaviors. Instead, I use a relatively explorative and descriptive approach to mapping the scope of disinformation, while using alternative narratives and the impact of competing Western media on Twitter as a point of comparison. The findings may serve as a stepping stone for future research on both the strategies and the scope of pro-Kremlin disinformation in other information domains.

This paper focuses on the supply of disinformation on Twitter by examining what proportion of tweets contain disinformation content related to Crimea. In addition to this, the study measures the impact of disinformation sources—in terms of visibility—in the broader network of retweets. Understanding the supply of disinformation is highly important, because misleading content needs to be prevalent and visible before it can influence politics and world views among broader audiences. Existing research suggests that misleading content is more likely to spread, if it is repeated (Pennycook et al., 2018) or comes from multiple sources (Lewandowsky et al., 2012), which is why larger volumes of disinformation as well as its visibility pose an even greater challenge to the online ecology.

I find that pro-Kremlin disinformation did penetrate the Twitter debate about Crimea, however it was accompanied by a much larger wave of tweets that disagree with the disinformation by intentionally or unintentionally contradicting and undermining the misleading narratives. Similarly, I show that the Kremlin-controlled news sources gained much less visibility than their Western counterparts. The Russian government did gain relative impact on Twitter through RT (formerly known as Russia Today). However, their ability to generate visibility in the retweet network is concentrated in just one popular Russian outlet, which has to compete with many popular Western media in the fight for the "truth" about Crimea.

I focus on Crimea for two reasons. Firstly, the Russian government has a strong, strategic interest in shaping the global opinion about the situation in Crimea. This is the case since the Crimean crisis serves as a catalyst for Western sanctions against Russia. Secondly, The Crimean annexation offers a scenario where pro-Kremlin disinformation is likely to thrive. Scholars and

authorities often use the Crimean crisis as an example of Russia's successful use of disinformation and hybrid warfare *in Ukraine* (Cimbala, 2014; Lanoszka, 2016; Snegovaya, 2015; Thornton, 2015). For example, General Breedlove, NATO commander in Europe at the time, described the Russian operation as "...the most amazing information warfare blitzkrieg we have ever seen in the history of information warfare" (Vandiver, 2014). Little is known about whether the Kremlin had similar success in spreading disinformation about Ukraine outside of the country's borders, for instance, among Western audiences. Russia may have improved its information capacities since 2014. Nevertheless, one can expect that Western audiences were relatively vulnerable to pro-Kremlin disinformation during the Crimean annexation due to a lack of awareness. Most of the Western government, private sector—as well as the civil society initiatives against Russian disinformation were created after the annexation (AFP, 2019; BBC, 2017; Department of Justice, 2018; EU, 2018; European Parliament News, 2016), suggesting that the general public was not yet prepared. If Russia were to successfully deploy a disinformation campaign online, one would expect to observe this in the online debate about Crimea, where Russia has high stakes in influencing the information sphere abroad and is most likely to succeed.

To be clear, the study does not reveal the prevalence of pro-Kremlin disinformation in Crimea, Russia, or Ukraine, but focuses instead on debates that take place on Twitter. I delimit the analysis to Twitter, due to the platform's focus on news sharing and its ability to facilitate information networks that span beyond national borders. Although Twitter is not equally as popular in all countries and it does not represent a "global" population or a global public opinion, it does facilitate a platform where users from both Russia, Ukraine, and the West can engage with each other across national borders to a greater extent than on VKontakte or Facebook.

The distinction between "truth" and "falsehood" and non-disinformation can be difficult to draw empirically and conceptually in many cases. Instead of exploring the challenges of distinguishing disinformation from non-disinformation, as has been done elsewhere (Fallis, 2015; Søe, 2018), this article focuses on events in Crimea where pro-Kremlin disinformation has later been retracted by the Kremlin itself.

Background: the Crimean crisis. On February 2014, a group of pro-Russian protesters went onto the streets of Crimea to demonstrate against the new, pro-Western government in Kyiv. The protests evolved into something more. Armed soldiers without insignia, popularly known as "little green men", appeared in Crimea.

Often accompanied by pro-Russian protesters as bystanders, the soldiers seized the Crimean airport, municipality buildings, a television transmission station, and other important infrastructure, while surrounding Ukrainian military bases. This raised the tensions and fear of a massacre to a new level.

Both Ukraine and the international scene was in a state of confusion. They key questions at the time were: Should Ukrainian soldiers open fire? Should NATO and the rest of the world put pressure on Russia?

Throughout the military operation, the Russian authorities (including president Putin) deflected any blame and responsibility, claiming that the invading forces were not affiliated with the Russian Federation. Russian state-controlled media described the soldiers as local rebels or a Crimean 'self-defense force', protecting the locals from a fascist, pro-Western 'Junta' in Kyiv.

After being seized by soldiers, the Crimean parliament proclaimed a referendum, where the locals were encouraged to vote on whether their regions should join the Russian Federation.[2] In March, Russia successfully annexed Crimea—less than 4 weeks after the appearance of the 'unidentified' soldiers.

Not long after the annexation, President Putin admitted that the soldiers were Russian (RT, 2014), de facto retracting his own misleading statements. Kremlin's strategic denial of its military involvement in Crimea stands as a clear example of disinformation, since the statements were both misleading and intentional. However, at the time of the operation, the "truth" about the events was far from clear to everyone. The Kremlin's use of disinformation covered the events with a veil of confusion during a critical military stage of the annexation. Citizens around the world were bombarded with competing 'truths' about the turn of the actual events. Russia's strategic use of disinformation to sow confusion may have helped mobilize pro-Russian Crimeans to the streets in order to help Russian troops achieve their military goals (K.N.C., 2019; Snegovaya, 2015).

What we know about Russian strategic use of (dis)information. Following the Crimean annexation, scholars from across different fields have examined both *how* and *why* the Russian government uses strategic communication in an attempt to influence citizens both domestically and abroad. The literature suggests that the Russian government pursues its political goals both through state-controlled Russian news outlets and by exploiting social media (Bastos and Farkas, 2019; Fredheim, 2015; Slutsky and Gavra, 2017; Xia et al., 2019).[3] However, few studies empirically examine to what extent pro-Russian disinformation dominates the cross-national flow of online content.

Numerous scholars within both media—and Russia studies have emphasized how the Russian government, under Putin's leadership, have co-opted major domestic TV channels and news outlets either through the loyalty of shareholders—who are often entangled with the political elite—or through the loyalty of editors to the pro-Kremlin shareholders (Fredheim, 2017; Mejias and Vokuev, 2017). Even though TV remains the main source of news in Russia, Oates (2016) argues that the proliferation of the internet has been accompanied by a new mode of propaganda in Russia. Total control of the information sphere is becoming increasingly difficult in the context of the new media ecology, where online sources can challenge the hegemony of pro-government sources. Even if the pro-government sources can ignore the domestic opposition, they cannot fully ignore international statements by the UN or other events that are covered by the international media (Oates, 2016, pp. 412–413). This has pushed the government to "rewire" its propaganda efforts from a direct control of information to a supplementary use of disinformation and manipulation in an attempt to refute international criticism directed towards the Kremlin (Oates, 2016).

Although these outlets provide a wide range of information, they serve as a frequent source of pro-Kremlin disinformation, according to both commentators, scholars, and Western authorities (BBC, 2019; Bjola and Pamment, 2016; Elliot, 2019; Pomerantsev, 2015; Thornton, 2015). However, the research does not reveal to what extent the Kremlin succeeds in dominating the online flow of news compared to sources that challenge the Russian government.

In line with Oate's (2016) argument, Olimpieva et al. (2015) stress that one cannot see the state-controlled TV and the Russian internet as separate spheres in opposition to each other. On the contrary, Olimpieva et al. (2015) empirically argue that Russian state-controlled television strategically shaped the agenda in the Russian-speaking online sphere through pro-Kremlin framing of the conflict in Ukraine. Both Olimpieva et al. (2015) and

Gaufmann (2015) argue convincingly that the Russian government relied heavily on collective memories of World War 2 to frame either the Maidan movement or the post-Maidan Ukraine as fascists. Despite these insights, the research falls short of analyzing to what extent the Kremlin succeeds in strategically shaping the online agenda or framing of the war in Ukraine outside of the Russian speaking internet sphere.

In line with this, Thornton (2015) argues that the purpose behind Russian information warfare is to sow confusion, doubt, and to blur the boundaries between enemy and non-enemy, war and peace, in order to make the population question who is the enemy and whether they are at war. This viewpoint is particularly relevant in the context of Russian disinformation about the annexation of Crimea. However, existing research suggests that the Russian government also uses (dis)information to sow social discord or to simultaneously support one political party over the other in the context of elections (Howard et al., 2018).

The Russian government carries out its information campaigns both through overt channels, where the source of the information is known, and covert channels, where the government source is concealed. When it comes to overt reach, the Russian government openly funds English-speaking outlets, such as Sputnik News and RT. These outlets serve as a frequent source of pro-Kremlin disinformation both according to scholars, fact-checkers and Western authorities (BBC, 2019; Elliot, 2019; Thornton, 2015). Although the English-speaking channels use both cable and satellite broadcasting, they rely largely on social media to reach their audience abroad. This has led to a series of public debates on the responsibility and role of Western social media platforms in the spread of pro-Kremlin content. Tech firms have addressed these debates during the last few years through a series of initiatives that seek to curb the influence of foreign state-controlled actors. In the most extreme example of this, Twitter has banned advertisement from RT and Sputnik due to their alleged interference in the 2016 US presidential election (BBC, 2017). In 2019, Facebook has temporarily blocked the page of the RT-affiliated, "In the Now", because the page failed to explicitly disclose its affiliation with the Russian state (AFP, 2019).

Scholarly literature on covert disinformation campaigns from Russia, are largely centered on Russia's use of fake accounts on Western social media. Shortly after the 2016 presidential election, tech firms such as Twitter, Google, and Facebook revealed in a US congress hearing that the Russian IRA—having close ties to Russian authorities—used their platforms to reach the US audience. The agency deployed fake accounts, often posing as concerned American citizens or local outlets. The covert activity has been analyzed in increasing number of quantitative studies across multiple disciplines, including political communication and computational social science (Bastos and Farkas, 2019; Slutsky and Gavra, 2017; Xia et al., 2019; Zannettou et al., 2019b).

This literature suggests that the agency used the accounts to engage in a broad spectrum of divisive topics in US politics, ranging from gun control to LGBT rights to conspiracy theories related to vaccines (Broniatowski et al., 2018; Howard et al., 2018).

Although the studies offer invaluable insight on Russian disinformation in a global context, they fall short of explicitly examining the scope of pro-Kremlin disinformation in comparison to competing narratives and news sources. This study therefore seeks to add to the existing body of knowledge by examining the breadth of pro-Kremlin disinformation in debates on Twitter.

Approach. The burgeoning literature on online disinformation offers a wide variety of methods to operationalize and measure the prevalence of misleading content. Scholars predominantly approach this task through (1) a content-centered approach by evaluating the content's message—often with the help of fact checkers and automated tools or (2) a source-centered approach by evaluation the credibility of the sources.

The source-centered approach treats any content from sources known to spread misleading narratives either as a proxy for disinformation, misinformation, fake news, or junk news, without evaluating the individual content of each post itself (Bovet and Makse, 2019; Grinberg et al., 2019; Guess et al., 2018; Howard et al., 2017). The method enables scholars to measure the flows of content on a large scale without the time-costly analysis of each piece of information from the sources. However, the method fails to address the fact that (1) established media may also disseminate misleading content and that (2) far from all information from non-credible sources is misleading.

The content-centered approach (Bode and Vraga, 2017; Margolin et al., 2018; Vosoughi et al., 2018) is narrow in its scope, considering that scholars and fact-checkers can only evaluate the fact-value of limited number of stories. However, it enables scholars to take into account the "credibility" of the content and to capture disinformation both from non-credible sources and established media while taking into account the context of the individual message.

In order to maximize the validity and the robustness of the study, I use both the content-based and the source-centered approaches to examine disinformation. First, I use content analysis of tweets to measure the relative breadth of the most central disinformation narratives. I then broaden the research scope beyond selected narratives, by using social network analysis to measure the general impact of Russian state media in the multilingual Twitter debate on Crimea.

Data and methods

Data. The data used in this study consists of tweets from the 1st of January 2014 to the 9th of December 2016. The tweets have been collected using "arden Hose Streaming API" and contains a random sample of 10% of all the tweets in the period related to Crimea.[4] I identify relevant content by keeping only tweets that contain at least one of the relatively broad hashtags or keywords related to Crimea in latin letters: crimea, crimean, sevastopol, simferopol, crimea, simferopol. Furthermore, I add the following terms in Cyrillic letters: krim, krimsk, sevastopol, sevastoplsk, simferopolsk, krim, krimnash, sevastopol, and simferopol. This yields a multilingual data set of 773.177 tweets and retweets.

The analysis is based on two subsamples. In the first subsample, I randomly sample 14,529 tweets and retweets in English from the aforementioned data set for the purpose of manual content analysis.[5] I delimit the subsample to the period surrounding the Crimea crisis: from 18th of February 2014, 5 days prior to the first pro-Russian protest, to 18th of June 2014, just 3 months after the Russian annexation.

In the second subsample, I use all of the 266.710 retweets for social network analysis to map the flow of content about Crimea. This does not include replies, original tweets and other posts that are not retweets. While the network analysis is not limited to a specific language, I delimit the content analysis to English tweets due to pragmatic reasons, because English is the most popular language in the entire data set. As I will show in the validation section, however, the results remain robust also when analyzing tweets in Russian.

Measuring disinformation content. In order to measure the scope of pro-Kremlin content, two student assistants have manually evaluated 14,529 randomly sampled tweets in English

related to Crimea (the codebook is available in the Appendix). The two students have annotated approximately half of the data set each. Building on a technique common in stance detection literature, the coders were asked to evaluate whether the tweets are topically unrelated, disagree, agree, or have a neutral stance towards the following statement: "*The Russian Federation is not carrying out a military operation in Crimea*".

The inter-coder agreement is substantial, with a Cohen's kappa (Cohen, 1960) of 0.85 for a set of 97 tweets coded by *both* annotators. I focus on this disinformation statement because of its central role in Russia's attempt to deflect international pressure and because "plausible deniability" is identified by other authors as a key aspect of Russian information warfare (Thornton, 2015).

The "disagree" category is used if the tweet *explicitly* or *implicitly* disagrees with the (disinformation) statement by describing the troops in Crimea as Russian. These tweets do not necessarily confront or respond to the disinformation narrative directly. Nevertheless, they (intentionally or unintentionally) contradict Russia's denial of its military presence. This includes, for example: "*Russian troops thwarted in attempt to storm missile base in Sevastopol*".

Although the tweet above does not explicitly confront the credibility pro-Kremlin disinformation narrative, it contradicts the narrative *implicitly* by attributing the armed men without insignia to Russia despite of the Kremlin's denial.

The "neutral" category is used in instances where the tweet is related to the military in Crimea but does not mention the troops' national affiliation or origin, for example: "*Crimea prepares for referendum under heavy military presence*". I operationalize disinformation content as tweets that agree with the statement. The "agree" category is used if the Russian troops are framed in a way that supports the disinformation narrative by describing the soldiers as "rebels", "self-defence" forces, "policemen" or portray the soldiers in other ways that mask their affiliation to the Russian Federation. This category includes, for example:

"*There are no Russian ground troops in Crimea. The folks you see on TV are Crimean Nationalist who have been seeking independence from Ukraine*".

I therefore measure disinformation by using tweets that support an intentionally *misleading* and state-driven narrative, even in cases where the statement in the individual tweets themselves are not factually wrong. In this sense, the data captures not only explicit and direct disinformation, but also *implicit* disinformation (Søe, 2016) that indirectly supports the disinformation narrative or implies the misleading message in the given context of the tweet. For example, local "self-defence" groups did assist Russian forces in capturing Crimea to some extent. However, the tweets can still be seen as a part of a broader disinformation campaign, because they support the Kremlin's misleading narrative that local Crimean groups—and not the Russian soldiers—were a driving force behind the military take-over. This broad approach is more likely to overestimate—rather than underestimate—the scope of misleading content, putting the notion of the prevalence of Russian disinformation to a critical test.

Disinformation sources in the core/periphery. I use standardized in-degree (Freeman, 1978), to measure the impact of news sources in a multilingual retweet network related to Crimea. The network consists of 167,997 nodes (users) and 222,065 non-weighted edges. The standardized in-degree reflects the proportion of users in the network—other than the node itself—that have retweeted the respective source. This includes general tweets about Crimea and not only those that contain disinformation content. *Impact*, in this case, refers to the ability of the news source to generate content that is widely shared by many users. "Impact" therefore reflects the ability to increase the potential *visibility* of one's content in the online network of shares. For pragmatic reasons, I delimit the analysis to top 10 Twitter accounts from Russian state-controlled media with the highest in-degree in the entire network (e.g. RT, Sputnik, RIA). Furthermore, I use top 10 Western news with the highest in-degree as a baseline for comparison. In addition to this, I select a different range of top outlets to test for robustness, as described in the "Methods" section and Supplementary Information.

In the next step, I measure the impact of Russian state-controlled and Western media by computing in-degree at the core and the periphery of the retweet network, respectively. The distinction between core/periphery is theoretically important, because pro-Kremlin sources that are less popular in the network as a whole, may have high impact among the more engaged users at the network core. For example, Pei et al. (2014) show that social media users at the core of a network are better at spreading content than those at the periphery. Following this logic, the popularity at the network core may be more important for the spread of (dis)information than popularity among the more isolated users at the periphery. Using Seidman's k-core method (Seidman, 1983), Golovchenko et al. (2018) show that established news media, such as BBC and CNN, may have high impact at the periphery of a retweet network, but are simultaneously outperformed by RT (formerly Russia Today) or civil society groups among the highly engaged users at the network core.

I use Seidman's k-core approach (Seidman, 1983) to identify the more cohesive core, where a user is considered to be member of specific k-core, if she is connected to at least k number of users in the same sub-graph. Figure 1a illustrates this, where all gray nodes are part of the $k = 2$ core, and all black nodes are part of a $k = 3$ core. The social network analysis literature offers various competing conceptualizations of both core/periphery and influence (Borgatti and Everett, 2000; Forslid and Ottaviano, 2003; Gallagher et al., 2020; Pei et al., 2014). I identify the core/periphery in the analysis using Seidman's k-core approach (Seidman, 1983) due to the method's relative simplicity and transparency.

It is important to note that k-cores are embedded in each other. For example, all members in $k > 3$ cores are also present in the larger, less cohesive $k = 2$, while the latter is also embedded in $k = 1$. Users at a more cohesive core with a higher k either retweet or have been retweeted by many other (highly connected) users, are therefore strongly engaged in the online debate about Crimea. The users outside of the k-core are considered to be in the periphery. Instead of relying on a arbitrary threshold between the core and periphery, I measure the impact of news sources in varying k-cores and their respective peripheries.

Results

The Kremlin's strategic denial. The annotators have identified 2354 tweets that mention the presence of armed troops on the ground. This is equivalent to 16.2% of all 14,529 annotated tweets in the English-speaking sample. Here, only 263 tweets support the Russian disinformation narrative, while 420 are neutral, and 1671 contradict the pro-Kremlin framing by describing the armed troops as Russian. In other words, nearly 1 out of 10 of all the tweets related to the sub-topic support the Russian disinformation narrative.

When seen in isolation, these numbers appear large. Indeed, they indicate that the Russian government has likely succeeded in making one of its most important narratives visible among an English-speaking audience on Twitter. As I will argue in the discussion, the levels of disinformation may be highly

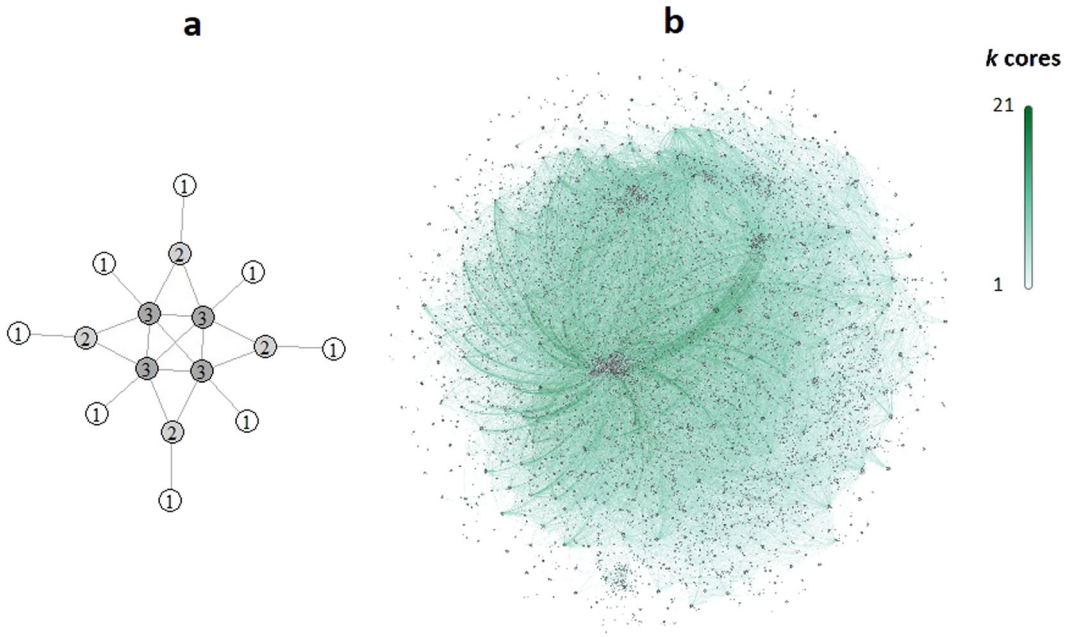

Fig. 1 Network core/periphery. a *k* core example. The numbers and colors reflect the maximum *k* value for each node. **b** The largest component in the Crimean retweet network. Nodes reflect profiles and edges reflect connections through retweets. Node size reflects indegree. Node color reflects maximum *k*-core. *Nnodes* = 134,170, *Nedges* = 199,581.

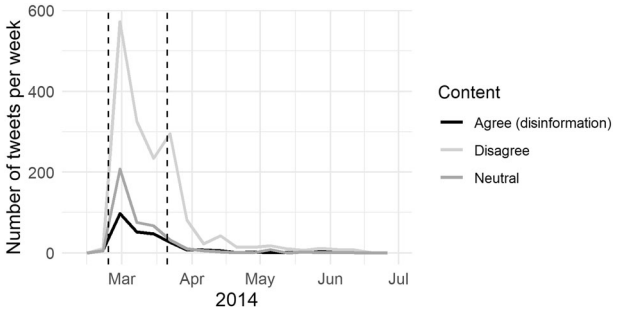

Fig. 2 The framing of Russian military presence in Crimea (in English). The annotators are asked to evaluate whether a tweet is neutral, agrees, disagrees with the following disinformation statement: "The Russian Federation is not carrying out a military operation in Crimea". The dashed lines indicate the beginning and the end of the Russian annexation of Crimea.

problematic, depending on the strategy behind the campaign as well as its effects on attitudes and world views.

However, the findings also show that for every disinformation tweet, there are ~6.4 tweets that disagree with these narratives by implicitly or explicitly contradicting Kremlin's denial of its military presence in Crimea. The pro-Kremlin disinformation narrative is challenged by a wide series of actors, ranging from Western outlets to activists and civil society movements in Ukraine.

Tweets that disagree with the disinformation remain dominant throughout the entire military operation. Figure 2 illustrates the distribution of the relevant tweets on a weekly basis, where the dashed lines indicate the first appearance of Russian troops (27th of February 2014) and the final step of the annexation, the Kremlin's formal admittance of Crimea as a part of the Russian Federation on the 21st of March the same year. As shown in the figure, tweets that contradict the disinformation narrative dominate the online debate from the very first week of the military operation. In other words, tweets that describe the military events in ways that contradict the disinformation narrative (i.e. by presenting the masked troops as Russian) were in the lead even when confirmed information about the event was arguably sparse. This pattern remains throughout the entire military operation.

One must note, that the 2354 tweets about the troops on the ground comprise only a fraction of all the tweets about Crimea: From geopolitics to diplomacy and local protests. In the next section, I will therefore broaden the perspective by analyzing the impact of Russian state-controlled media, a common source of pro-Kremlin disinformation. The analysis below therefore includes all of the retweets related to Crimea in the multilingual data set, including tweets that are not topically related to Russian soldiers.

The impact of disinformation sources. Figure 3a shows the aggregated, standardized in-degree of the top 10 Russian state controlled outlets compared to top 10 Western news outlets. The entire network is denoted as $k = 0$ in the figure, whereas $k = 1$, for example, indicates only users in the $k = 1$ core. In the entire retweet network ($k = 0$), the top Russian outlets have been retweeted by at least 2.2% of all the 167,997 nodes, excluding themselves. In comparison, top Western outlets have been retweeted by at least 6.4% of all the nodes (standardized in-degree of 0.064). In other words, for each user retweeting top Russian outlets, there are nearly three users that retweet top Western news outlets.

Similar to the findings in the previous section, these results indicate that the Russian state-controlled news sources have a limited impact, when comparing to their Western counterparts. As indicated in Fig. 3a, this pattern is consistent when examining the different layers of the network, ranging from the large $k = 1$ core (*Nusers* = 167,945) to the inner $k = 15$ core (*Nusers* = 468) with the most engaged users.

As shown in Fig. 3b the Western outlets are retweeted by at least two users for every user retweeting Russian sources, i.e. a

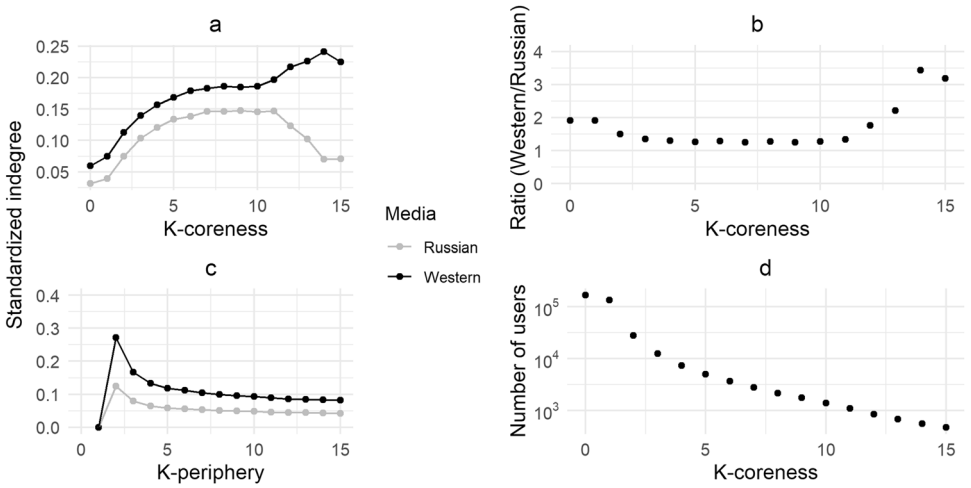

Fig. 3 The impact of Russian state-controlled news media in the core/periphery of the retweet netwrok. a Standardized in-degree for the outlets combined in the core/periphery. **b** Ratio between standardized in-degree for Western and Russian outlets. **c** Standardized in-degree in the periphery (outside of the respective core). For example, $k = 1$ periphery consists of nodes that are excluded from the $k = 1$ core. **d** Number of users in the respective K-core, where $k = 0$ is the entire network.

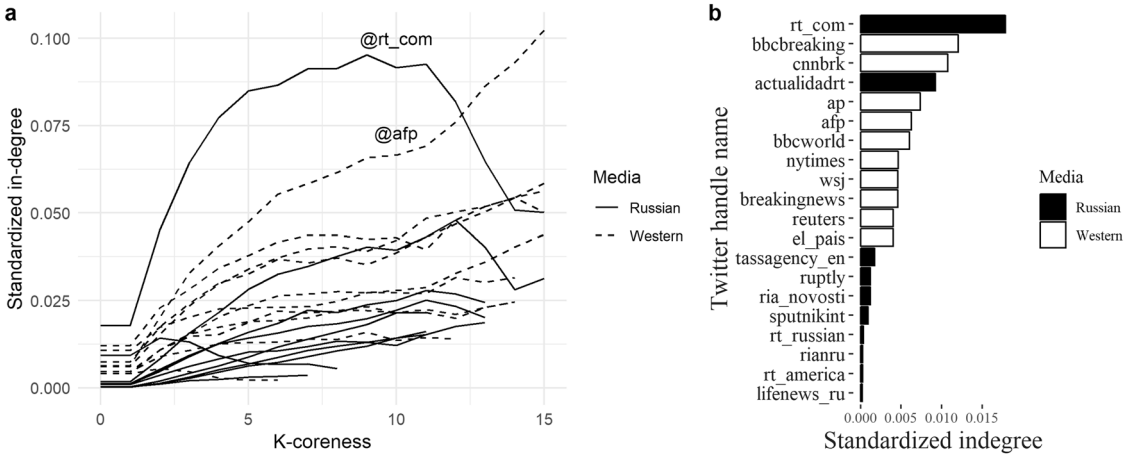

Fig. 4 The impact of top 10 Russian and Western news outlets. a Standardized ind-degree for the individual outlets in the core/periphery and **b** in the entire retweet network ($k = 0$).

ratio of 2:1, throughout the $k = 1$ core. Russian state-controlled news outlets remain outperformed by Western counterparts throughout the different cores/peripheries of the retweet network. However, the relative impact of Russian state-controlled outlets increases among the more engaged users in the inner cores of the retweet network. Beginning from the $k = 2$ core (with 134,143 users), there are up to 1.5 users who retweet Western news outlets for every profile retweeting Russian media. The ratio drops to 1.35 in $k = 3$ among 12,342 highly engaged users, while reaching minimum of 1.24 in the $k = 7$ core of 2770 users, who comprise 1.6% of the entire retweet network. One must note that the number of users continues decreasing when one delimits the analysis to a more cohesive core by increasing k, as shown in Fig. 3d. Although the western outlets gradually regain their relative impact, beginning from $k = 11$ and onward, the total impact (in absolute numbers) becomes small, considering that the core $k = 11$ core consists of only 1091 profiles.

Conversely, the impact of Russian news outlets is weakest at the periphery of the network, where Western sources gain an even greater advantage. This is further illustrated by Fig. 3c. Top Russian outlets have been retweeted by 12.4% in the k-periphery = 2, i.e. among the 25,182 users in the large component who are *outside* of the $k = 2$ core. The proportion is relatively low, when considering that the top Western outlets have been retweeted by 27.2% of the users in the same periphery.

A disaggregated analysis of the top news outlets suggests that the impact of Russian state-controlled outlets is largely concentrated in one highly popular source, that is RT. Figure 4a illustrates the standardized in-degree for top 10 Russian and top 10 Western news outlets in the respective layers of the retweet network. RT's main Twitter account, @RT_com, systematically outperforms any other news outlets through the many layers of the network: From the entire network, $k = 0$, throughout $k = 12$. In the entire retweet network, RT has been retweeted by at least 1.8% of all the other users ($N = 167,997$) or approximately 3000 profiles. @Bbcbreaking ranks second, with a standardized in-degree of 1.2%. BBC holds its second place in $k = 1$ and $k = 2$ cores, respectively. Agence France-Presse's English speaking account, @afp, is RT's main competitor among the more engaged users, beginning from $k = 3$ and onward.

The popularity gap between RT and other Russian outlets further highlights the important role of RT in the Russian government's global reach. Figure 4b shows the impact of the respective outlets in the entire retweet network. The second most

popular Russian news profile after @rt_com is yet another RT profile, its Spanish speaking account, @actualidadrt. Here, the Spanish RT ranks 4th among the 20 news outlets, being retweeted by 0.92% of the users in the full network. In comparison, Russia's third most popular state-controlled source, the English speaking account for TASS (@tassagency_en), ranks 7th and has been retweeted by only 0.17% of all the users in the network. For every user retweeting TASS's English account, there are 10 users who retweet RT's main account alone. Whereas Russian state-controlled alternatives are mainly at the bottom of the 20 news outlets in the sample, the highly popular RT is faced by many popular Western outlets.

Validity and robustness. I have tested the validity of the results presented in the analysis through three steps.

In the first step, I examine the scope of pro-Kremlin disinformation in tweets written in Russian. The same two annotators were asked to manually code 3000 Russian tweets, out of which 170 were related to the subtopic of the (non)presence of Russian troops in Crimea. The proportion of tweets that contain pro-Kremlin disinformation outnumber the tweets that explicitly or implicitly counter the Kremlin's strategic denial of its military involvement in Crimea (see Appendix A for details). The pattern is consistent with the findings presented in the analysis. A similar pattern has been observed by Golovchenko et al. (2018) in their study of disinformation about MH17 on Twitter. While the study does not explicitly examine on the scope of disinformation, the data suggests that counter-disinformation in English outperformed pro-Kremlin disinformation.

In the second step, I compare the impact of top 5 Russian and Western news outlets respectively instead of top 10. The robustness check overemphasizes the impact of Russian outlets, because their reach is concentrated among a few highly popular sources, while the impact of Western outlets is more evenly distributed among the top 10 sources. Even in this case, Russian outlets have a 34.4% lower impact that Western outlets in the entire retweet network, with a standardized in-degree of 0.037 and 0.050, respectively. It is important to note, that the top 5 Russian controlled outlets reach nearly the same impact as the Western ones at the network core, beginning form $k = 2$ (see Appendix B).

In the third and final step, I reiterate the network analysis by narrowing down the time period to the Crimean Crisis: From the 27th of February to 21st of March 2014. Here, I select the top 10 Western and top 10 Russian state-controlled news outlets from this period. The results presented in the analysis remain robust. The impact of Russian news outlets is even smaller during the Crimean crisis, when comparing to their Western competitors (see Appendix C for details). This suggests that the pro-Russian sources did not dominate the Twitter debate about Crimea even during the most critical phase of Russia's military operation in the region.

Limitations. This study is limited to Twitter debates related to Crimea. A large proportion of the users are from the US, while the platform is less popular in Russia and Ukraine (Clement, 2020). The users are not representative of the general population (Barbera and Rivero, 2015; Mellon and Prosser, 2017). For this reason, the study does not reveal the scope of pro-Kremlin disinformation offline or in regions where Twitter is not common. While pro-Kremlin disinformation narratives did not dominate Twitter debates about Crimea, Russia may be more successful when dealing with topics that do not receive as much coverage from competing Western media outlets. It is possible that pro-Kremlin sources have more impact on other online platforms, such as VKontakte or YouTube. More research is needed to map the reach of pro-Kremlin disinformation from a cross-platform perspective as well as offline.

Furthermore, the study does not reveal to what extent pro-Kremlin content shapes the attitudes or behaviors of online audiences or how the individual tweets are received by the "silent" users, who do not respond to the disinformation with tweets. Instead, the paper shows that the pro-Kremlin content is not as visible on Twitter as it is often implied in the popular coverage of Russian information warfare against the West. This is important, because visibility is the first pre-condition for the pro-Kremlin news coverage to have an effect on the attitudes on behaviors of the audience. While exposure to information does not automatically change how the audiences view or interact with politics (Kalla and Broockman, 2018), doing so is even more difficult without significant volume or exposure (Allcott and Gentzkow, 2017). Furthermore, this paper does not examine the speed of the disinformation diffusion, nor does it measure the extent to which the misleading content is more permanent than other information in the data sample. It is possible, for instance, that the audience on Twitter is more likely to remember pro-Kremlin disinformation than content that undermines the misleading narratives. More research is needed to examine the actual exposure as well as consumption of pro-Kremlin disinformation.

The network analysis is static, which is why it does not reveal the change in the impact of Russian state-controlled outlets throughout the different stages of the deteriorating relations between Russia and the West. The comparison of the retweet network from the Crimean Crisis in 2014 and the entire retweet network from 2014 to 2016 (mentioned in the previous section) suggests that Russia's disinformation campaign may have potentially improved over time. More research is needed on the dynamic aspects of pro-Kremlin disinformation to further validate this interpretation. It is important to note, however, that the tactical value of the disinformation about Crimea may change over time. It may be of great importance during the first few weeks of the information operation, as was the case with the Crimean crisis, but have less value the next year. The results in this study show that the Kremlin did not dominate the information space on Twitter during the most critical phase of its military campaign in Ukraine—even before Western authorities and tech companies raised their level of awareness and launched a long series of anti-disinformation initiatives.

Because this research is limited to the online debates on Crimea, it does not reveal whether pro-Kremlin sources are more successful in using disinformation to shape the online debates about other topics. However, the article does show the limitations of pro-Kremlin disinformation in a context where Russia's national interest and international prestige may be at stake.

This paper does not show to what extent the pro-Kremlin sources are promoted by fake Twitter profiles. Existing research emphasizes that the online "popularity" of online sources can be easily faked through shares and retweets from inauthentic accounts, such as automated "bots", manually operated "sock-puppet" accounts or semi-automated profiles (Keller et al., 2017; Monsted et al., 2017; Shao et al., 2018; Varol et al., 2017). Many commentators and researchers have argued the Russian government is either actively deploying or benefiting from at least one of the two strategies (Sanovich, 2017; Zannettou et al., 2019a). However, if a large proportion of the pro-Kremlin reach in the data is strategically driven by fake accounts, this would mean that the impact of pro-Kremlin news outlets among human audiences is even less than what is shown in the analysis.

Discussion and conclusion

The impact of disinformation sources or the scope of disinformation itself should not be seen in isolation alone, but also in relation to

competing information. Pro-Kremlin disinformation is accompanied by a much greater stream of information that offers the Twitter audience an alternative view. Approximately 88.8% of the tweets that are related to the presence of troops in Crimea do not contain pro-Kremlin disinformation—even when using a very broad definition of the term. Each tweet with pro-Kremlin disinformation about Russia's military involvement in the Crimean Crisis is accompanied, on average, by 6.4 tweets that disagree with the disinformation, i.e. posts that contradict the misleading narrative.

The data points toward a similar conclusion when viewing state-controlled disinformation sources in the multi-lingual retweet network. The pro-Kremlin impact is concentrated in RT, while the remaining Russian state-controlled outlets (e.g. Sputnik, TASS, RIA) are retweeted by relatively few users. As a result, the few popular pro-Kremlin news sources have to compete with a wide range of popular Western outlets that offer an alternative view on Crimea. This may offer one potential explanation for the fact that the Kremlin's disinformation about the military operation in Crimea has been greatly challenged by the Twitter-sphere.

When combined, the top 10 Twitter accounts for Russian state-controlled news outlets are retweeted by at least two times fewer users than their top 10 Western competitors in the entire retweet network. The impact of pro-Kremlin outlets is greater at the network core among the more engaged individuals than among the more isolated users at the periphery of the Crimean retweet network. Even though the Russian outlets are capable of competing with Western sources among the relatively few, highly engaged users at the network core, their impact is relatively limited among the many users at the periphery or in the network as a whole. These findings suggest that initiatives against pro-Kremlin disinformation should be focused on the highly engaged users at the network core.

As mentioned in the "Introduction" section, this research does not provide objective criteria for evaluating whether the disinformation campaign on Twitter was successful from Kremlin's point of view. The answer to the question depend on Kremlin's strategic objectives, which cannot be inferred within the scope of this article. As a result, the scope of disinformation can both be interpreted as high or low, depending the theoretical perspective. Below, I will present different theoretical interpretations and their limitations.

While the results show that pro-Kremlin disinformation does not dominate the information flow on Twitter about Crimea, one can critically question whether the Kremlin ever intended to dominate the Twitter-sphere as whole. Such an ambition may be unrealistic, given the solid presence of Western news outlets on the platform. Indeed, the literature on disinformation and propaganda suggests that strategy behind information influence activities is far from always to reach the general audience. In some cases, such campaigns may follow a logic of "sociodemographic targeting", directed towards specific societal groups (Pamment et al., 2018, p. 25). Studies of Russia's use of fake accounts during the 2016 US election indicate that fake Kremlin-controlled profiles on Twitter engaged with both left-winged and right-winged content, while predominantly supporting pro-Republican narratives (Golovchenko et al., 2016). In line with this, Hjorth and Adler-Nissen (2019) show that American users in the 40+ age group and conservatives are more exposed to pro-Kremlin disinformation about Malaysia Airlines Flight 17. While this article does not infer the demographic characteristics of Twitter users, it is possible that the disinformation about Crimea may have been targeted towards a specific demographic group. The disinformation may therefore still be highly visible in some parts of the information network, if it is concentrated among a smaller group of homogeneous users, who are insulated from competing information that challenges the disinformation.

Although the scope of pro-Kremlin disinformation is small compared to tweets that offer a more factual view on events in Crimea, the numbers could still be viewed as alarmingly high even if they are not concentrated in demographic groups. Indeed, the results suggest that the Russian government penetrated—to some extent—the English-speaking Twitter-sphere by deploying a strategic denial of its military presence during the Crimean Crisis. The very fact that ~1 out of 10 English tweets about the subtopic are supporting the pro-Kremlin disinformation narrative can be seen as great challenge. As mentioned earlier, disinformation content has been operationalized relatively broadly both as narratives that directly and indirectly support the Russian government's strategic denial of its military operation in Crimea. A more narrow definition would lead to lower levels of observed disinformation. However, even if the proportion of disinformation content were to be twice as low (i.e. 5%), the volume of the misleading content remains a great concern. This may be particularly the case for users around the world who try to navigate in the streams of competing narratives about real-life events. This is problematic, because disinformation may increase the time and effort needed to evaluate and confirm the contradicting claims. Perhaps even more importantly, the disinformation campaign—even in relatively low volume—could potentially sow distrust in the news outlets and other sources of information. The Russian government may have achieved the overall goal of creating a fog of confusion and doubt among Western audiences (Lucas and Nimmo, 2015; Ramsay and Robertshaw, 2018) even without outnumbering the more factual coverage of events in Crimea. This interpretation holds if the disinformation tweets are capable of changing the attitudes and worldviews among Twitter users. Whether this is the case, remains an open question that should be addressed by future research.

The results based on network analysis can be interpreted in a similar manner. RT has the highest impact in the Crimean retweet network of all the news outlets. Many of the stories posted by RT are not false or even misleading. As a frequent source of pro-Kremlin disinformation, however, these tweets can be used to gain credibility among audiences interested in the topic, which in turn can be used to strategically disseminate disinformation. While Russian state-controlled media do not dominate the multilingual Twitter debate about Crimea, their impact is impressive, when considering that Western audiences, particularly from the US, are relatively dominant on Twitter (Clement, 2020) and that they are competing with news outlets from most parts of the world. Top Russian media outlets, when combined, do not have as much impact as top Western outlets. Yet, their performance is relatively strong, when considering that their budget (Times, 2015) is only a fraction of the resources available to top news outlets from all of the Western countries combined. Similarly, the scope of pro-Kremlin disinformation in the English-speaking part of Twitter, while being greatly outnumbered by posts that disagree with the disinformation, is relatively high when considering the previously mentioned context.

These theoretical interpretations suggest that Russia may be gaining a foothold in the Twitter debate about Crimea by challenging Western news sources with the help of disinformation narratives as well as the highly popular RT channel. Nevertheless, the findings highlight the limits of pro-Kremlin disinformation. It is met by a much greater wave of tweets that strongly conflict and disagree with the disinformation as well as competing Western news sources. Pro-Kremlin disinformation is far from dominant among broader Twitter audiences who are engaged in one of the most strategically important topics for Russia: Crimea. While the general public, tech-firms, and Western authorities are likely to be more aware and prepared to tackle pro-Kremlin disinformation today, it is reasonable to expect that the Russian government also has improved its capabilities following the events in Crimea.

These results call for more research on the breadth of pro-Kremlin disinformation about other topics and on other social media platforms, as well as the mechanisms that may either enable or limit its scope.

Data availability

The IDs and manual labels used by the annotators are available per request.

Received: 26 October 2019; Accepted: 30 October 2020;
Published online: 11 December 2020

Notes

1 See Pamment et al. (2018) for a discussion of the strategies behind foreign influence activities more broadly.
2 The UN has proclaimed in a resolution that the referendum has "no validity, (and) cannot form the basis for any alteration of the status of the Autonomous Republic of Crimea or of the City of Sevastopol" (Charbonneau and Donath, 2014).
3 Please see la Cour (2020) for a more general discussion of online disinformation in international relations.
4 I am very grateful to Professor Alan Mislove, Northeastern University, for access to Crimea tweets based on the Twitter Gardenhose feed.
5 The Codebook has been co-developed with Isabelle Augenstein. It is available online: http://golovchenko.github.io/crimea/crimea_codebook.pdf

References

Abrams S (2016) Beyond propaganda: Soviet active measures in Putinas Russia. Connections 15:5
AFP (2019) Russia's RT fumes after Facebook blocks 'wildly popular' page. AFP
Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research
Badawy A, Ferrara E, Lerman K (2018) Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 258–265). IEEE
Bail CA, Guay B, Maloney E, Combs A, Hillygus DS, Merhout F, Freelon D, Volfovsky A (2020) Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. Proc Natl Acad Sci USA 117:243–250
Barberá P, Rivero G (2015) Understanding the political representativeness of Twitter users. Soc Sci Comput Rev 33:712–729
Bastos M, Farkas J (2019) Donald Trump is my President!: The Internet Research Agency Propaganda Machine. Soc Media+ Soc 5(3):1–13
BBC (2017) Twitter bans RT and Sputnik ads amid election interference fears. BBC
BBC (2019) Russia's RT banned from UK media freedom conference. BBC
Bjola C, Pamment J (2016) Digital containment: Revisiting containment strategy in the digital age. Glob Aff 2(2):1–12
Bode L, Vraga EK (2017) See something, say something: correction of global health misinformation on social media. Health Commun 33(9):1131–1140
Borgatti SP, Everett MG (2000) Models of core/periphery structures. Soc Netw 21:375–395
Bovet A, Makse HA (2019) Influence of fake news in Twitter during the 2016 US Presidential election. Nat Commun 10:7
Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M (2018) Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 108:1378–1384
Charbonneau L, Donath M (2014) UN General Assembly declares Crimea secession vote invalid. Reuters. https://www.reuters.com/article/us-ukraine-crisis-un-idUSBREA2Q1GA20140327. Accessed 1 Dec 2020
Cimbala SJ (2014) Sun Tzu and Salami Tactics? Vladimir Putin and Military Persuasion in Ukraine, 21 February–18 March 2014. J Slav Mil Stud 27:359–379
Clement J (2020) Countries with the most Twitter users 2020. Statista. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/. Accessed 1 Dec 2020
Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46
Darczewska J (2014). The anatomy of Russian information warfare. The Crimean operation, a case study. Ośrodek Studiów Wschodnich im. Marka Karpia
Department of Justice (2018) Grand Jury Indicts Thirteen Russian Individuals and Three Russian Companies for Scheme to Interfere in the United States Political System. Department of Justice
Department of National Intelligence (2017) Assessing Russian activities and intentions in recent US Eelections. Department of National Intelligence
DiResta R, Shaffer K, Ruppel B, Sullivan D, Matney R, Fox R, Albright J, Johnson B (2018) The tactics and tropes of the Internet Research Agency. New Knowledge Report prepared for the United States Senate Select Committee on Russian Interference in the 2016 Election
Elliot R (2019) How Russia spreads disinformation via RT is more nuanced than we realise. The Guardian. https://www.theguardian.com/commentisfree/2019/jul/26/russia-disinformation-rt-nuanced-online-ofcom-fine. Accessed 1 Dec 2020
EU (2018) Questions and answers about the East StratCom Task Force. European Union External Action
European Parliament (2016) EU strategic communications with a view to counteracting propaganda. Technical report, Directorate-General for External Policies
European Parliament News (2016) MEPs sound alarm on anti-EU propaganda from Russia and Islamist terrorist groups News European Parliament. European Parliament
Fallis D (2015) What is disinformation? Libr Trends 63:401–426
Forslid R, Ottaviano GI (2003) An analytically solvable core-periphery model. J Econ Geogr 3:229–240
Fredheim R (2015) Filtering foreign media content: How Russian news agencies repurpose Western news reporting J. Sov. Post. Sov. Polit. Soc. 1(1):37–82
Fredheim R (2017) The loyal editor effect: Russian online journalism after independence. Post-Soviet Aff. 33(1):34–48
Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1:215–239
Gallagher RJ, Young J-G, Welles BF (2020) A clarified typology of core-periphery structure in networks. Preprint at https://arxiv.org/abs/2005.10191
Gaufmann E (2015) World War II 2.0: digital memory of Fascism in Russia in the aftermath of Euromaidan in Ukraine. J Reg Secur 10:17–36
Golovchenko Y, Buntain C, Eady G, Brown M, Tucker JA (2020) Cross-platform state propaganda: Russian trolls on twitter and youtube during the 2016 us presidential election. Int J Press/Politics 25:975–994
Golovchenko Y, Hartmann M, Adler-Nissen R (2018) State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation. Int Aff 94:975–994
Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on twitter during the 2016 US Presidential election. Science 363:374–378
Guess A, Nagler J, Tucker J (2019) Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv 5:eaau4586
Guess A, Nyhan B, Reifler J (2018) Selective exposure to misinformation: evidence from the consumption of fake news during the 2016 US Presidential campaign Eur Res Council 9(3):1–14
Hjorth F, Adler-Nissen R (2019) Ideological asymmetry in the reach of pro-Russian digital disinformation to United States audiences. J Commun 69:168–192
Howard PN, Bolsover G, Kollanyi B, Bradshaw S, Neudert L-M (2017) Junk news and bots during the US election: what were Michigan voters sharing over Twitter. CompProp, OII, Data Memo
Howard PN, Ganesh B, Liotsiou D, Kelly J, François C (2018) The IRA, social media and political polarization in the United States, 2012–2018
Jamieson KH (2018) Cyberwar: how Russian hackers and trolls helped elect a president: what we don't, can't, and do know. Oxford University Press
Kalla JL, Broockman DE (2018) The minimal persuasive effects of campaign contact in general elections: evidence from 49 field experiments. Am Political Sci Rev 112:148–166
Keller FB, Schoch D, Stier S, Yang J (2017) How to manipulate social media: analyzing political astroturfing using ground truth data from South Korea. In: Eleventh international AAAI conference on Web and Social Media
KNC (2019) The west is losing today's infowars and must hit back hard. Economist
la Cour C (2020) Theorising digital disinformation in international relations. Int Politics 57:1–20
Lanoszka A (2016) Russian hybrid warfare and extended deterrence in eastern Europe. Int Aff 92:175–195
Lawrence RG (2019) Cyberwar: how Russian hackers and trolls helped elect a President: what we don't, can't, and do know. Public Opin Q 83:163–166
Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 13:106–131
Linvill D, Boatwright B, Grant W, Warren P (2019) "THE RUSSIANS ARE HACKING MY BRAIN"! investigating Russia's Internet Research Agency twitter tactics during the 2016 United States presidential campaign. Comput Hum Behav 99:292–300
Lockie A (2017) How the US and Russia are fighting an information war—and why the US is losing
Lockwood Alisa (2018) Russia has won the information war in Turkey. Global Risk Insights

Lucas E, Nimmo B (2015) Information warfare: what is it and how to win it. CEPA Infowar Paper (1)
Margolin DB, Hannak A, Weber I (2018) Political fact-checking on Twitter: when do corrections have an effect? Political Commun 35:196–219
Mejias UA, Vokuev NE (2017) Disinformation and the media: the case of Russia and Ukraine. Media Cult Soc 39(7):1027–1042
Mellon J, Prosser C (2017) Twitter and facebook are not representative of the general population: political attitudes and demographics of British Social Media users. Res Politics 4:3
Mønsted B, Sapieżyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using twitter bots. PLoS ONE 12:e0184148
Myriam DC, Mauer V (2008) The role of the state in securing the information age—challenges and prospects. In Myriam Dunn Cavelty, Mauer V, Krishna-Hensel S F (ed), Power and security in the information age, Ashgate Publishing 2008
Nyhan B (2018) Fake News and bots may be worrisome, but their political power is overblown. N Y Times. https://www.nytimes.com/2018/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html. Accessed 1 Dec 2020
Oates S (2016) Russian Media in the digital age: propaganda rewired. Russ Politics 1:398–417
Olimpieva E, Cottiero C, Kucharski K, Orttung RW (2015) War of words: the impact of Russian state television on the Russian Internet. Natl Pap 43:533–555
Pamment J, Nothhaft H, Fjällhed A (2018) Countering information influence activities: the state of the art. Department of Strategic Communication, Lund University
Pei S, Muchnik L, Andrade Jr JS, Zheng Z, Makse HA (2014) Searching for superspreaders of information in real-world social media. Sci Rep 4:5547
Pennycook G, Cannon TD, Rand DG (2018) Prior exposure increases perceived accuracy of fake news. J Exp Psychol 147(12):1865–1880
Pomerantsev P (2015) The Kremlin's information war. J Democr 26:40–50
Ramsay G, Robertshaw S (2018) Weaponising news: RT, sputnik and targeted disinformation. King's College London: The Policy Institute, Center for the Study of Media, Communication, and Power
Reisinger H, Gol'c A (2014) Hybrid war in Ukraine Russiaas intervention and the lessons for the NATO. Osteuropa 64:9–10
Renz B (2016) Russia and 'hybrid warfare. Contemp Politics 22:283–300
RT (2014). Putin acknowledges Russian military serviceman were in Crimea. RT.
Sanovich S (2017) Computational propaganda in Russia: the origins of digital misinformation. Working Paper
Seidman SB (1983) Network structure and minimum degree. Soc Networks 5:269–287
Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9:4787
Slutsky P, Gavra D (2017) The phenomenon of Trump's popularity in Russia: media analysis perspective. Am Behav Sci 61:334–344
Snegovaya M (2015) Putinas information warfare in Ukraine. Soviet origins of Russiaas hybrid warfare. Russia Rep 1:133–135
Søe SO (2016) The urge to detect, the need to clarify: Gricean perspectives on information, misinformation and disinformation. PhD diss., University of Copenhagen
Søe SO (2018) Algorithmic detection of misinformation and disinformation: Gricean perspectives. J Doc 74:309–332
Taylor PM (2003) Munitions of the mind: a history of propaganda from the ancient world. Manchester University Press, Manchester
Thiele RD (2015) Crisis in Ukraine—the emergence of hybrid warfare. ISPSW Strateg Ser 347:1–13
Thornton R (2015) The changing nature of modern warfare: responding to Russian information warfare. RUSI J 160:40–48
Times M (2015) Russia cuts state spending on RT News network. Moscow Times
Torossian R (2016) Russia is winning the information war. Observer
Unver H (2019) Russia has won the information war in Turkey. Foreign Policy
Vandiver J (2014) SACEUR: Allies must prepare for Russia 'hybrid war'. Stars and Stripes. Available via https://www.stripes.com/news/saceur-allies-must-prepare-for-russia-hybrid-war-1.301464. Accessed 1 Dec 2020
Varol O, Ferrara E, Davis C B, Menczer F, Flammini A (2017) Online Human-Bot Interactions: Detection, Estimation, and Characterization. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), pp. 280–289
Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359:1146–1151
Wallance GJ (2018) Why the US is losing the information war to Russia. The Hill
Woo P-K (2015) The Russian hybrid war in the Ukraine crisis: some characteristics and implications. Korean J Def Anal 27:383–400
Xia Y, Lukito J, Zhang Y, Wells C, Kim SJ, Tong C (2019) Disinformation, performed: self-presentation of a Russian IRA account on twitter. Information, Communication & Society, pp. 1–19
Yin L, Roscher F, Bonneau R, Nagler J, Tucker JA (2018) Your friendly neighborhood troll: the Internet Research Agency's use of local and fake news in the 2016 US Presidential election. SMaPP Data Report, Social Media and Political Participation Lab, New York University
Zannettou S, Caulfield T, Setzer W, Sirivianos M, Stringhini G, Blackburn J (2019a) Who let the trolls out?: towards understanding state-sponsored trolls. In: Proceedings of the 10th ACM Conference on Web Science. ACM, pp. 353–362
Zannettou S, Sirivianos M, Caulfield T, Stringhini G, De Cristofaro E, Blackburn J (2019b) Disinformation warfare: understanding state-sponsored trolls on twitter and their influence on the web. In: WWW'19 Companion Proceedings of the 2019 World Wide Web Conference. ACM, New York, pp. 218–226

Acknowledgements

I would like to thank Rebecca Adler-Nissen, Sune Lehmann, Lene Hansen, James Pamment and Alicia Fjällhed for the invaluable feedback. I am very grateful to Isabelle Augenstein for co-developing the codebook, as well as Alona Shevchuk and Valentina Shapovalova for assisting with the annotation process. The research was conducted as part of the 'Digital Disinformation' project (no. CF16-0012), funded by the Carlsberg Foundation and DIPLOFACE, funded by the ERC (project no. 680102). Both projects are directed by Rebecca Adler-Nissen.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1057/s41599-020-00659-9.

Correspondence and requests for materials should be addressed to Y.G.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.