Turning the Virtual Tables:

Government Strategies for Addressing Online Opposition with an Application to Russia

Sergey Sanovich, Denis Stukal, and Joshua A. Tucker

On December 3, 2014, the Russian news website *Meduza.io* reported that the 100th mirror of another Russian news website, *Grani.ru*, had been banned by the Russian Federal Service for Supervision of Communications, Information Technology, and Mass Media (*Roskomnadzor*). *Grani.ru* was a popular news website with extensive coverage of opposition activity and alternative opinions. It was blocked in the spring of 2014, at the height of Russian-Ukrainian conflict, using a technical system developed by *Roskomnadzor* to block content deemed as extremist, as permitted under a Russian law that was adopted just two months earlier. *Meduza.io* itself was a new Russian media outlet, established in the neighboring Baltic state of Latvia by Galina Timchenko, who had been dismissed as the editor-in-chief of the most popular Russian news website *Lenta.ru* over coverage of the Russian-Ukrainian conflict and moved to Latvia along with most of *Lenta.ru*'s editorial staff. Around the same time, one of the most popular political blogs in Russia, belonging to the Russian opposition leader Alexey Navalny, was also permanently banned on the LiveJournal platform and, in early 2015, authorities began to crack down on its mirrors too.

While one might expect this sort of response in Russia today, it has not always been this way. As late as 2010, a report of the Internet in Russian Society program at the Berkman Klein Center for Internet and Society at Harvard University noted that "the political blogosphere appears to remain a free and open space for Russians of all political stripes to discuss politics, criticize or support government, fight corrupt practices and officials, and to mobilize others around political and social causes."[1] Moreover, as recently as 2009, then President Dmitry Medvedev opened his own blog on LiveJournal and subsequently established a presence on Twitter and Facebook, as did many of his aides. Accordingly, the pro-government youth movements, which were created to confront possible "colored revolutions" on the streets of Moscow, were charged with the duty of competing with oppositional voices in the cyberspace and promoting government-friendly content.[2] In some cases, they had even engaged directly with leading oppositional bloggers on the pressing issues of the day. In more recent years, we have also witnessed a widespread proliferation of pro-government bots in the Russian Twittersphere as well as the notorious "troll factory" in St. Petersburg documented in the pages of the *New York Times*.[3]

Why were the changes in policy so quick and dramatic? What are the options for governments seeking to respond to emerging online challenges? Might a different country (or leader) have responded differently? Inspired by these and similar questions, we seek to (1) argue that these are indeed important questions for political science research to address; (2) introduce an organizational framework for doing so; and (3) provide a Russian case study to illustrate the utility of the framework.

Accordingly, we begin by presenting a classification system for different forms of government response to online opposition in authoritarian or competitive authoritarian regimes such as Russia. We suggest there are essentially three types of responses: *offline responses*, which include legal action and market regulation in addition to more traditional forms of (physical) intimidation; *online restriction* of access to information, which relies on digital tools to filter the information available to end users; and *online engagement* with users that aims to shape online conversations, usually through content generation. In all cases we provide empirical examples of how governments have utilized these strategies.

As an illustration of the utility of this framework, we provide a detailed case study in the appendix[4]—summarized in the text—of the evolution of Internet policies in Russia during Putin's first two terms in office (2000–2008), the Medvedev interregnum (2008–2012), and the period of time since Putin's return to the Kremlin in 2012. In particular, we investigate why the government almost completely ignored the Internet when it was actively imposing its will on traditional media, why this policy of ignoring the Internet changed after Putin left the Kremlin in 2008, and why policy changed yet again when Putin and Medvedev switched offices in 2012.

Finally, we demonstrate the feasibility of trying to identify particular attempts by governments to utilize this third strategy of online engagement. More specifically, we apply digital forensic techniques to identify "bots," or automated (algorithmically controlled) accounts, in the Russian political Twittersphere.[5] We introduce a new, exhaustive framework for classifying Twitter accounts as official accounts, bots, cyborgs, human accounts, or spam, plus sub-categories for bots and humans. We demonstrate five different techniques for identifying bots, all of which prove remarkably adept at finding bots in a collection of politically-relevant Russian Twitter data. These techniques also helped us to locate bots with highly ideologically charged content.

Although this empirical work is largely intended to function as a proof of concept analysis, our initial hand-coding of a small number of accounts identified by our bot-detection methods suggests one interesting finding that warrants further investigation.[6] For the accounts that we can identify as politically oriented bots, only slightly more than half of them appear to have a pro-government orientation; we also find evidence of both pro-opposition and pro-Kiev bot activity. This suggests that simply assuming all political bots are pro-government—and therefore part of a government strategy for online engagement—in an analysis of this type would be a mistake.[7]

Literature

Various forms of government reaction to online activities have been the subject of intensive research in recent years. This literature covers everything from the legal regulation of the Internet in the world's most advanced democracies[8] to online censorship tools used by different types of autocratic governments around the world[9] to establishing an active presence of governments on social media platforms.[10] Freedom House even produces an annual international index of Internet freedom.[11] However, few studies attempt to provide a framework for the systematic analysis of government behavior on the web that would allow us to analyze why particular tools are chosen under any given circumstances, or even what makes up the choice set in the first place.

Notable exceptions are Ronald Deibert et al., Evgeny Morozov, and Margaret E. Roberts.[12] Deibert et al. provide a historical framework that traces the evolution of state-web relationships from initial laissez-faire non-involvement (until 2000) to attempts to either deny altogether (2000–2005) or carefully control access to cyberspace (2005–2010) all the way to the current (after 2010) stage of active contestation between state and corporate censors on the one hand, and cyber-activists on the other. This framework is meaningful in describing global trends, but does not necessarily explain the exact trajectory of each particular country. As we shall see, the Russian government turned to access denial only after it tried and failed both (in Deibert's terms) control and contestation. In addition, while Deibert et al. provide many relevant examples of how governments tried to deny or control access online, they do not provide any systematic classification of the types of action a state may take. Roberts does provide such a classification, but distinguishes between the effects (fear to speak or to listen and either friction or flooding as impediments for access) rather than the tools employed. Indeed, the same tools could lead to both flooding and fear, and different kinds of tools, both online and offline, could be used to increase friction.[13]

Morozov does distinguish between technological and what he calls "sociopolitical" means of controlling online activity, the latter combining technology with online and offline actions by humans.[14] He hypothesizes that if "liberation technologies," such as those promoted by Larry Diamond,[15] were to succeed, embattled governments could turn to potentially more violent methods such as smearing campaigns, criminal prosecutions, or even physical attacks on dissenting bloggers.[16]

Morozov's classification is useful for studying the dangers and promises of "liberation technologies" (what makes a sociopolitical response different is exactly being beyond the reach of these technologies), but it does not fully distinguish between government actions that restrict, or otherwise structure, online media environments and those in which the government actively engages in attempts at shaping the formation of opinions online. This distinction is important for at least three reasons. First, in censoring online media, governments can build on strategies from long before the Internet was created, whereas online propaganda in distributed networks is fundamentally different from a top-down broadcasting of the party line through hierarchical monopolies of traditional media. Second, this part of the government response is experienced differently by users: not as an outcome (e.g., an inaccessible web page), but as a point of interaction with the state (e.g., a paid pro-government troll replying to your tweet). Last but not least, the study of government online activities will increasingly focus on social media, which is simultaneously the most abundant and versatile data source and a key point of contestation between the government and civil society.[17] Since social media data capture exactly the moment of interaction between the user and the state, it is important to understand the place of government action leading to this interaction in the wider menu of options available to regimes.

Therefore, we propose a new classification system for government options to respond to independent online activity in closed societies. In addition to differentiating between "offline" and "online" tools, our classification also distinguishes between online tools aimed at restricting the flow of information and those entailing active engagement with users on behalf of the government. While the former operates largely—although not exclusively—by exerting control over Internet infrastructure, the latter typically involves some content generation. Since users experience each of these possible government actions differently, our classification, effectively, dissects government options from the user's point of view. In the next section, we discuss each option in detail, providing examples and identifying the key resources needed to employ each of these options.

A Classification System for Government Responses to Online Opposition

In this section, we introduce our tripartite system for classifying government responses to online opposition. We begin with offline responses, which primarily refer to changing a country's legal Internet regulations, but also include attempts to change the ownership structure of online media and intimidate particular users. The second category encompasses various ways of technically restricting access to online content, from firewalls to Distributed Denial of Service (DDoS) attacks to sophisticated online censorship systems. The final category also involves online activity, but instead of focusing on restricting access to content, this tactic involves creating content to influence online communications.

Offline Response The first set of options at any government's disposal is based on digital-age implications of traditional governing advantages: nodality ("network centrality"), organizational capacity, legal authority to enforce the law, a monopoly on the legitimate use of violence, the right to regulate human activity, and the ability to expend large financial resources through taxation.[18] The actions facilitated by these advantages could have a huge impact online, but take place offline; thus end-users either observe the consequences online after-the-fact or encounter these actions in person, but offline.

The latter case includes legal prosecution and violence, but can also include actions such as having commenting functionality turned off by their favorite news websites after readers' comments become legally designated as media content.

Another option is to require popular bloggers to register with the government, making each individual blogger responsible for her own content, on par with actual commercial media outlets, as it has recently been done in Russia.[19] As Ackland notes, such a legal designation—as well as other forms of regulating user-generated content—can be attributed by the government to either genuine or fabricated popular demand stemming from concerns over public safety or morality.[20]

Finally, governments can attempt to change the landscape of digital media markets and alter the choice of online platforms available to users. Relying on their authority to regulate commerce, autocrats around the world designate certain companies and industries, including telecommunications, as "strategic," leading them to begin enforcing various restrictions, such as banning foreign ownership and/or investments, appointing state representatives to the board, etc. For example, in late 2013, the publicly-owned, but heretofore relatively editorially independent, major Russian news agency *RIA Novosti* was stripped of its leadership, restructured, renamed, and put under the leadership of a fervent regime supporter.[21] Then, in order to ensure complete control, it was included in the list of "strategic enterprises" in early 2014, along with the second largest Russian news agency, *ITAR-TASS*.[22]

If control over digital media is challenging or costly to legislate or order, especially in the case of private companies, then governments can use other means, in particular their purchasing power and extra-legal pressure they can exert, to assume control over important Internet platforms. The so-called "Russian Google," Yandex, sold a "golden share" to state-owned *Sberbank* in 2009, allegedly after negotiations with Dmitry Medvedev and multiple proposals to designate companies such as Yandex as "strategic," which would have forced them to re-register in Russia[23] and severely diminished their appeal to international capital markets.[24] A similar attempt was made in the case of VKontakte, known as the "Russian Facebook," which resulted in a hostile takeover of the company by business groups loyal to the Russian government,[25] and founder and former owner and CEO Pavel Durov fleeing the country with many members of his team.[26]

Of course, instead of attempting to take control of existing public and private media and communications platforms, governments can try to increase their influence through artificially generated competition. In several countries, including Russia and Turkey, governments reportedly allocated generous funds to the creation of "national" search engines, social networks, and email services.[27] A Russian national search engine has been discussed since at least 2008, and Turkey began a similar project in 2009.[28]

Online Responses: Restrictions The rapid growth in Internet penetration rates and the emergence of the Internet as a principal source of information for increasing numbers of people creates challenges even for autocrats who are able to successfully employ offline tools of control. To begin with, information can be produced and distributed by foreign citizens and entities that are out of reach of the autocrat's security apparatus. Second, some local activists and/or journalists can use their digital proficiency to distribute information anonymously and therefore avoid offline persecution. Moreover, autocrats may simply prefer putting flows of information under their control rather than going after its producers. If, for example, an autocrat wants to avoid taking responsibility for the government's actions, a DDoS attack on a popular oppositional blog can be blamed on "unidentified" hackers, while most types of offline response require at least some involvement of the state apparatus.

Of course, the option always exists to completely monopolize the telecommunication infrastructure inside the country and cut any connections with international networks. North Korea did just that: it maintains *Kwangmyong*, a national intranet, and a national mobile phone service, *Koryolink*, both of which can be controlled and monitored. Communications with the outside world through both channels are prohibited (except for the ruling elite and foreign tourists). However, such a system imposes a heavy toll on the national economy.

A step removed from this extreme approach—albeit still with non-trivial costs—is the highly sophisticated Chinese "Great Firewall," probably the best example of blocking sensitive information without fatally hurting either government communications or commercial activity.[29] This form of targeted Internet-censorship is well documented by King, Pan, and Roberts, who describe the immense Chinese system of monitoring and censoring of user-generated content across the country's dispersed social media platforms and estimate that around 13 percent of all social media posts are censored.[30]

The ability of autocratic governments to filter Internet communications is primarily a function of three factors: control over critical infrastructure; planning ahead and implementing a long-term, comprehensive, but not overly costly solution; and financial and human resources as well as the technical expertise necessary to build filtering tools. The first component is rarely an advantage of autocratic regimes: most of the world's Internet infrastructure (IP address allocation and DNS management, search and social media platforms, transaction services, etc.) is located in advanced democracies and therefore out of reach for autocrats seeking to control them. Consequently, the latter two factors (i.e., preparedness and money) are particularly important.[31]

Two primary technological options for regimes are filtering/blocking particular websites or segments of the web and DDoS attacks. The former has the advantage of being permanent and customizable. China, for example, blocks only certain platforms and content (by keywords), while North Korea famously maintains its local web segment in complete isolation from the outside web. Both policies, though, share a common disadvantage of this approach: high transparency for local users and susceptibility to documentation by outsiders (including other governments, human rights organizations, etc.).[32]

DDoS attacks, on the other hand, are usually hardly traceable, relatively cheap, can be deployed during particularly sensitive political events such as elections or protests, and can be more easily outsourced to loyal but independent groups, such as the Syrian Electronic Army.[33] On the other hand, their ability to break up online communications is limited in time and web space (i.e., a small set of websites at best). Moreover, the most popular platforms, such as Google and Twitter, are highly protected from DDoS attacks.

The most distinctive (and important for our discussion) feature of Internet filtering is that it is observed by users as an end result and does not create any kind of interaction with the state in the course of a user's activity online. If Twitter is blocked in your country (permanently, as in China, North Korea, Iran, and several other countries, or temporarily, as in Venezuela and Turkey in 2014), you either cannot get there, or you can use one of the available tools (anonymizers, proxy-servers, etc.) to restore your access. In either case, users observe government actions as end results. The impact of such actions is either in successful breaking of inter-personal communication or access to websites, or the lack thereof. In other words, the government cannot possibly shape the conversation through these means. To achieve this latter goal, governments have to directly engage with users online.

Online Responses: Engagement Establishing a government presence on the web and using it to promote its agenda constitutes the third option at a government's disposal. This type of government response actually takes place online, and users encounter it in the course of their online activity. Mainly, it includes the government creating content, either through automated generation or real human activity.[34] The most obvious and increasingly popular tool employed to alter political conversations on social media is using either "bots" (i.e., algorithmically controlled accounts) or "trolls" (i.e., real people) to advocate pro-government positions, turn conversation meaningless or prohibitively divisive, or distract users from sensitive political issues altogether.

Bots can perform two key functions: cluttering conversations with "digital dust," which could be pro-government, anti-opposition, or simply aimed at "flooding the zone" with distracting information in order to detract attention from opposition voices;[35] or altering search results, Internet rankings, top lists, and other automated tools for sorting, sharing, discovering, and consuming online content. As such, bots could be used to support real people. For instance, a ranking of the most popular Russian blog posts maintained by Yandex was closed in 2009, after being inundated by bots promoting mostly pro-government posts.[36]

The possible functions of humans acting on the government side are much more diverse. It is useful, therefore, to provide a basic classification of pro-government content producers. This classification is not based on users' honesty, consciousness, or beliefs. Instead, it is primarily focused on formal or informal ties with the government (or the lack of thereof).

To begin with, the government could hire students or other low-paid workers to submit rather simple messages, which would nevertheless pass the human intelligence tests integrated in many modern social media platforms. One particular example of this type of bloggers are the so-called Chinese 50-centers.[37] Russian pro-government youth movements, such as *Nashi* and *Young Guard of United Russia* were sometimes accused of running a similar network of 11-rublers.[38] Leaks released by the Russian arm of Anonymous in 2012 indicated that *Nashi* paid hundreds of thousands of dollars in fees for comments, statuses, Facebook likes, YouTube dislikes, etc.[39]

Cheap bloggers paid per comment are not the only group of friendly users that could be put on the government payroll. Bribing prominent and trusted bloggers, celebrities, or journalists—although potentially much more expensive—could turn out to be a better investment in terms of persuading the public. The same leaks noted previously revealed that along with paying small fees to thousands of low-skilled bloggers, *Nashi* also put aside tens of thousands of dollars to be paid to a small group of popular and, heretofore, supposedly independent bloggers for highly sophisticated positive publicity for the Russian leadership.[40]

The next group consists of government supporters whose social media activity is not paid *per se*, but is facilitated through participation in various political projects or actual employment by the government. These sets of bloggers range from members of various youth political movements to the MPs from the ruling (or affiliated) parties to relatively prominent politicians (ministers, party leaders) who are encouraged to take on the challenge of representing the government's point of view in an often-hostile social media environment.

Finally, the government could also try to mobilize genuine supporters with no formal or informal ties to the government or ruling party. If famous people volunteer to support the government agenda, it could help the autocrat both directly and indirectly through endowing ideas already promoted by the armies of bots and paid bloggers with the weight of fame, reputation, and personal independence.

In the next section, we illustrate the usefulness of this taxonomy by briefly tracing the evolution of Russian government policy regarding the Internet since 2000 (see Appendix B for a substantially more detailed account). This evolution proves to be not a linear increase in censorship efforts, but a complicated process of choosing the optimal strategy, directly reflecting both the political struggles inside the regime and the distinctive challenges associated with the implementation of each of the three options we identify.

The Russian Government Online: A Constantly Evolving Strategy

Russian government activities online first gained serious international attention when they were redirected towards aiding the Russian offensive in Ukraine in the wake of the 2014 Euromaidan Revolution. The resourcefulness and inventiveness of these actions as well as their reach were all the more surprising for Western observers and policy makers since until then Russian authorities were not considered to be particularly artful in their digital operations, even for domestic purposes.

However, a closer examination of the evolution of Russia's Internet policy reveals that perhaps such surprise was unwarranted. Vladimir Putin is famously old-fashioned when it comes to digital tools: he rarely uses a computer and has never had any personal online presence. However, in terms of policy, he has long showed interest both in new technologies and an awareness of potential government strategies in response to these technologies. Even before he became acting President, in late 1999, he convened the leaders of the nascent Russian IT industry and online media and made a clear commitment to protect their freedom and avoiding Chinese-style filtering. While it is unclear whether he was concerned with Russia's image abroad or had other intentions, his choice was politically expedient.

At only 2 percent Internet penetration in 2002 (and 16 percent at the end of Putin's second term in 2008), online media were more a mode of personal communication than of mass persuasion and, as such, were hardly an asset of any political significance. Following an old Soviet tradition, Putin avoided direct interference with personal communication channels.[41] This resulted in the emergence of a thriving and competitive Internet industry, whose leading companies, Yandex and VKontakte, won the competition for local users over Google and Facebook, respectively, and did it without the aid of any protectionist measures, a rare achievement for any country. Years ahead of most Western countries, Russian online news media that had been created from scratch overtook the websites of traditional media in popularity and began doing their own original reporting (instead of relying on existing offline news agencies and outlets). Meanwhile, the Russian public created a vibrant blogosphere that was large enough to completely overtake the major blog platform of the time, LiveJournal, which was eventually purchased by a Russian company.

As Internet penetration continued to rise steadily in the late 2000s and most traditional media became completely sanitized of any alternative opinion, online news media, most of which were rather critical of the regime, became increasingly influential. However, the government first saw this as an opportunity rather than a threat.

To a large extent, this attitude was the result of a change in the government.[42] In 2008, freshly installed into the Kremlin, Dmitry Medvedev and his team were looking for ways to build their own support base. Medvedev published his modernization manifesto "Go Russia" in the online-only liberal newspaper *Gazeta.ru*. Medvedev and his team made a serious attempt to engage the Russian online public in a genuine discussion of the country's future. Both he and his aides created digital presences on multiple platforms, which earned Medvedev the nickname "Blogger-in-Chief."

Most crucially, they sought, received, and responded to critical feedback from the audience, a practice unheard of for years in the traditional media, but necessary to get any attention in the vibrant Russian blogosphere at the time. Pro-government youth movements were mobilized to spread Medvedev's message to every corner of the Russian segment of the Internet. While their activities were not without controversy (more due to corruption and incompetence than ideological zeal),[43] even they had to engage in genuine discussion with bloggers critical of the government, thus facilitating public debate on important issues. While the government did occasionally use DDoS attacks, particularly in relation to the 2008 Russo-Georgian war (see Appendix B.2), the Russian Internet remained remarkably free (in a growing contrast with traditional media) and the government activities there were primarily targeted to mobilize genuine support based on the compelling message and (limited) interaction with the public.

This engagement came to an abrupt end in the wake of the 2011–2012 Russian popular protests, which coincided with Putin's return to the Kremlin.[44] Given the role of media, and social media in particular, in coordinating and sustaining the largest and the longest wave of protests in Russia in two decades, Putin was unlikely to go back to the "disengagement strategy" he used during his first two terms in office.

Instead, Putin began to actively employ both offline means of controlling media production and online means of controlling access to it. The first included pressuring media moguls into either replacing the editorial staff of online media they owned (*Lenta.ru*, *Gazeta.ru,* and *RBK* are the most prominent among dozens of examples) or into selling them to more loyal owners (e.g., Russian Forbes and *VKontakte*).[45] The government also adopted laws making online media liable for the content of comments posted by their readers, thus requiring these websites to either actively police user-generated content or shut commenting tools down altogether. In addition, various laws were adopted to prosecute individual bloggers for alleged extremism and other content deemed inappropriate.[46] Since 2012, these laws have been applied in an increasingly wide-ranging manner, punishing with large fines and real prison terms not only the original authors of messages, but also those who reposted them.[47] Finally, many prominent bloggers and online media journalists have faced threats and (at times life-threatening) assaults, which are rarely, if ever, investigated.

Online tools of controlling access to content included the creation of the Russian Internet Blacklist, maintained by the dedicated government agency, *Roskomnadzor*. Blacklisting initially required a court order, but later was also allowed on a simple request from the Office of the Prosecutor General. While theoretically it is supposed to be easy to exit the blacklist (after removing the content deemed unlawful), after several prominent opposition news websites and opposition leader's blogs were blocked in March 2014, in the midst of the Russian-Ukrainian conflict, they were not informed what content they would have to remove to exit the Blacklist.[48] Instead, the government refused to respond to their requests even after they sued for an answer, and they remain blacklisted to this day, thus illustrating the ways in which formal rules and informal power relations are applied in tandem.[49]

Still, neither offline nor online tools allowed the government to shut down hostile activity online completely. While a long period of unrestricted development of domestic alternatives diminished the market share of Facebook and Google in Russia (which makes the government's job easier, as local platforms are easier to coerce into compliance), Facebook and Google are still used by millions of Russians on a daily basis. And when VKontakte, immediately after getting a request, removed the event page of a pro-opposition rally, Facebook (after some hesitation) refused to comply.[50] Journalists fired by the pressured owners could move abroad and set up a news media there (as *Meduza.io* did).

Therefore, the space for the engagement strategy remains, but instead of playing the leading role, the government is using it to support offline and online restrictions. Rather than trying to engage in a dialog or persuade, the government simply attempts to hammer down the official message, artificially increase the indicators of its take-up (propelling politicians into lists of top bloggers and their messages into lists of top posts), while simultaneously cluttering communication channels used by the opposition. This created a huge demand for various troll and bot factories, which produce pro-government content in volumes, caring about the quantity much more than quality and persuasion capacity. This content requires a new set of tools to study it properly.

Online Engagement: Preliminary Analysis

Online engagement is a complex phenomenon, ranging from completely automated bots producing large volumes of gibberish in order to flood popular communication platforms to high-profile paid bloggers with independent reputations, who send nuanced, targeted messages to different groups of the public. While of course it would be useful to study all forms of this activity, for the sake of space constraints in this manuscript we limit ourselves to bots.[51]

There are three main reasons for this choice. First, bots produce by far the largest volume of content, and without tools to identify (and remove) content produced by bots, studying only the human-generated content would be almost impossible. Second, the only practical way to identify bots is by using automated algorithms; starting the empirical part of our research in this manner has the advantage of therefore creating an objective and replicable approach that can be employed in future analyses. There is, however, another, less methodologically inspired, reason to start with bots, which is that they are both important and interesting objects to study. While social media provide citizens and politicians with new and powerful tools for expressing their political beliefs and preferences, affecting the political agenda, mobilizing supporters, and organizing political actions, they also bring the challenge of differentiating between real political communication on the one hand, and interaction with computer programs that imitate human activity on the other.

Here we seek to identify bots in the strictest sense of the word. *Bots* are accounts that are operated by a computer program that generates account content in an automatic and predefined way.[52] If, on the contrary, an account is operated by a human who also uses some scripts to help herself produce account content, we label this account a *cyborg*, which is an intermediate category between bots and humans. Finally, a troll who is hired to post her own tweets about a certain topic would be a *human* rather than a bot in our classification. Similarly, a real human being who does not post her own tweets, but only retweets accounts she follows, would also be classified as human rather than a bot. We also separate out *official* accounts (such as the account of the president, of a ministry, or a newspaper) and *spam* accounts (those attempting to sell services or items unrelated to politics). For more details on this classification and the coding scheme, see Appendix D.

Data We used the Twitter's Streaming API to collect a large dataset of tweets that contain specific keywords related to Russian politics. We began with a list of politically relevant keywords and hashtags (including major politicians' names, events, and slogans)[53] that spanned the Russian political spectrum, including Putin and United Russia, loyal and radical opposition, Russian nationalists, and others. This allowed us to collect a dataset with more than 14 million tweets posted by approximately 1.3 million Twitter users who listed Russian as their account language, between November 25, 2013 and July 13, 2015.[54]

The Twitter API returns both tweet-level information (i.e., text, date, and time of the tweet, its language, etc.) and metadata (various characteristics of the account sending the tweet including the author's ID and screen name, the number of followers and friends, and official account language). There is a large variation in the number of tweets from different users in our collection, ranging from 1 to almost 97,000.[55]

Analysis Our approach to detecting bots and cyborgs focuses on three account characteristics: the *entropy* of inter-tweeting time intervals, the *followers/friends ratio* of accounts (which produced two different methods of finding bots), and the presence of *identical tweets* in our collection (which we also employed in two different ways).[56]

In the online Appendix C, we justify and describe each method in detail; here, we proceed to assessing the viability of these detection methods. In order to insure the reliability of coding, we had five Russian university students code 512 accounts identified as possible bots by our detection methods. We then examined the inter-coder agreement in each group for each account and imposed a strong requirement for an account to be considered reliably classified: it had to be put into a particular category (bot/cyborg/human/official account/spam) by at least four out of five coders. This is a rather stringent approach that guarantees the inter-coder reliability is at least as high as 80 percent. Stringency of this requirement notwithstanding, more than 80 percent of accounts pass this requirement.

Table 1 shows the results of the verification. Of the accounts classified reliably, 77 percent belong to bots, and a further 1 percent are cyborgs who share more characteristics

Table 1 Results of Suspicious Accounts Verification

	No friends	Low ratio	Entropy	Repeat themselves	Repeat others	Totals
Bots	82	77	83	72	93	394
Cyborgs	1	0	3	1	0	5
Humans	1	0	2	4	1	7
Official accounts	2	0	0	0	0	2
Spam	0	7	1	1	0	9
Unclear	13	51	11	12	8	95
Totals	99	135	100	90	102	512

Note: Entries are frequencies. Row sums do not equal row totals because some methods produced overlapping sets of accounts.

with bots than with humans. Reliably identified humans constitute only 1.4 percent of our dataset. These results reveal an outstanding precision of the five bot-detection methods we have used.

The share of bots is probably even higher than the above calculations indicate. Only twenty-five of the ninety-five unclear accounts were identified as being human accounts by even a single coder. Typically, the discrepancies in the coding that generate unclear accounts are due to attributing accounts to spam or cyborgs by some coders and to bots by others. Hence, even those accounts that belong to the category of unclear are much more likely to be bots than humans. Still, from here forward, we restrict our analysis to the 394 accounts that were reliably classified as bots.

We further disaggregate our bots into seven subtypes. Examining the distribution of bots across subtypes (last column of Table 2) reveals that more than 90 percent belong to just three subtypes: news headlines with and without links and (in a very distant third place) accounts that consist entirely of retweets from other accounts.[57]

Political Orientations in a Subsample of Verified Political Bots In addition to coding each account as bot, human, cyborg, spam, or official account, we also coded their political orientation. Given the rather small sample size, this analysis remains preliminary, but we uncover interesting patterns, suggesting that our bot-detection methods are a promising tool for quantitative studies of the types of online government engagement strategies described in the previous sections of this article.

We distinguish between three different political orientations: *pro-Kremlin*, *pro-opposition*, and *pro-Kiev*. This choice was dictated both by our research framework and by the activity patterns in the Russian segment of the Internet during the time our collection was running. As anecdotal evidence of the Russian government's social media activity continued to mount, our principle interest was in studying accounts that

Table 2 Verified Bots by Type and Method of Identification

	No friends	**Low ratio**	**Entropy**	**Repeat themselves**	**Repeat others**	**Total**
Retweets only	3	8	26	3	3	**9**
Videos only	2	0	0	0	0	**< 1**
Pictures only	2	0	0	0	1	**1**
Text only:						
– News headlines only:						
News headlines with links	38	74	40	41	15	**40**
News headlines without links	48	15	26	36	73	**42**
– Other text	3	2	2	5	1	**3**
Diverse content	3	2	7	15	7	**6**

Note: Entries are column percentages (may not sum up to 100 due to rounding).

post content friendly to the Russian government. However, the Russian government (as well as governments around the world) often claims that it is the victim, nor perpetrator, of bot attacks. Explicitly looking for both pro- and anti-government content in the same collection of tweets provides a valuable opportunity to verify both claims on a level playing field. However, since the timing of our data collection coincided with the political crisis in Ukraine and further Russian involvement there, we decided to distinguish between anti-government content that mostly was focused on Russian domestic politics, as opposed to content focused on the Ukrainian conflict. Thus, the accounts that our coders find to spread content unfriendly to the Russian government are split into two categories: pro-opposition and pro-Kiev.

In this exercise, our focus was on accounts whose political identification is strikingly clear and unambiguously apparent in most of their tweets to even a casual reader.[58] To this end, we deliberately defined the *neutral* category as broadly as possible, so that those accounts that end up classified as non-neutral indeed have a very strong political bias. Furthermore, in the schema we provided to coders, the accounts that do not neatly fit into either of our "camps" (for example, praise Kremlin for its foreign ventures, but loathe it for economic policy) were also left in the neutral category.[59] Table 3 shows the resulting classification, broken down by the method used to identify the bots.

Table 3 highlights several interesting findings. First, more than a half of the bots are neutral, meaning that they do not carry any explicit political message. This does not mean they are not setup with political purposes. As mentioned above, many bots (particularly among those who have no friends and post the same text as many other bots) feature primarily news headlines to promote them in search rankings. These headlines usually come from large media agencies that produce enough routine factual news (who said what, went where, and signed which memorandum) to appear neutral according to our deliberately broad notion of neutrality. This applies even to the state-owned media. It does not mean, however, that if readers go to their websites, they will find politically neutral and objective media.

Second, taken together, pro-opposition and pro-Kiev bots (9 percent) are almost as common as pro-Kremlin ones (11 percent). This result may seem unexpected given the

Table 3 Ideological Distribution of Verified Bots

	No friends	**Low ratio**	**Entropy**	**Repeat themselves**	**Repeat others**	**Totals**
Pro-Kremlin	13	16	8	8	8	**11**
Pro-opposition	1	1	11	1	2	**4**
Pro-Kiev	4	3	12	3	3	**5**
Neutral	65	55	41	51	76	**58**
Unclear	17	26	28	36	11	**23**

Note: Entries are column percentages (may not sum up to 100 due to rounding).

mass media's clamor about Kremlin's social media propaganda campaigns. At the same time, this result might also imply that the Kremlin prefers more sophisticated and expensive online propaganda techniques like paid trolls, whereas the Russian opposition and pro-Kiev users may so far lack resources to employ those techniques at a large scale.[60]

Does the behavior of bots with different ideological orientation vary? In Appendix C, we present a series of statistical analyses that explore this issue in two ways by focusing both on content and the dynamics of tweeting activity. In short, we find that neutral bots seem quite different from both sets of politically oriented bots, but, perhaps not surprisingly, the pro-Kiev and pro-opposition bots exhibit more similarity with one another than with the pro-Kremlin bots we identified.

Conclusion

When social media first burst onto the political scene—and into the writing of political scientists—the overwhelming emphasis was on its potential for allowing citizens to become organized outside of traditional hierarchical arrangements. Nowhere was this more important than in authoritarian and competitive authoritarian regimes, where social media was posited as a way to level the playing field when traditional institutions were largely under the control of the state.[61]

If we want to label this original interpretation of the intersection of Social Media and Politics as "Social Media 1.0," then "Social Media 2.0" could be the story of how these same authoritarian and competitive authoritarian regimes woke up to the possibilities of social media as a threat to their regimes and how they responded.[62] With that in mind, the goal of this article has been three-fold. First, we have provided a classification system for the menu of options at the disposal of regimes seeking to address online opposition including offline responses, online restriction of access to content, and attempts to engage with a user's online experience. Second, we have demonstrated that this classification system allows us to provide a rich qualitative description of the evolution of Russia's Internet policy since the first election of Putin as Russian President in 2000, which hopefully will be its own important contribution to the literature above and beyond demonstrating the value of the theoretical framework. Finally, we have sought to demonstrate that identifying one particular form of online engagement, the use of political bots, is a fruitful area for future academic research by proving that it is indeed possible to use digital forensic techniques to observe and analyze various forms of online engagement.

Our preliminary empirical analysis shows the high precision of our digital forensic techniques that allow us to identify bots with a high level of confidence. However, the fact that a relatively small proportion of the bots we detect are pro-Kremlin, whereas the majority are politically neutral accounts that post news headlines, might imply that government online engagement is multifaceted and includes both direct engagement through government-controlled bots and indirect engagement through the support of pro-regime mass media that promote themselves online by using bots. Still, even this preliminary first "proof of concept" step suggests an important warning for scholars seeking to study the behavior bots: simply applying bot-detection techniques to collections of political tweets and assuming that whatever is identified by the method is government activity has the potential of significantly overestimating the prevalence of such behavior.

Our empirical analysis also demonstrated that different types of bots perform different types of functions and, intriguingly, that different techniques can be used to locate these different types of bots. Thus, although bots are but one tool at the disposal of political agents to try to impact the online experience of citizens, it appears that even this specific category—algorithmically controlled social media accounts—is but an overarching category for a diverse set of tools.

The potential future research that our preliminary work in this article has identified as both possible and important strikes us as vast indeed. At the most focused level, the next steps within Russia should involve harnessing tools of machine learning to be able to code entire collections of political tweets in terms of the likelihood that content was produced by bots and the political orientation of those bots.[63] Once this task is complete, it will then be possible to test specific hypotheses regarding the way in which the Russian state and its political opponents have utilized bots in an effort to shape the online experience of Russians.

More generally, though, both the classification systems and digital forensic tools developed in this article could be applied far beyond Russia. Next steps could include theory development for cross-national predictions about the relative prevalence of offline activity, online suppression of contention and online engagement with users based on political and economic goals and resources, cross-national case studies to compare with the Russian case in this article, and eventually cross-sectional time-series statistical analysis. In addition, the types of bot analysis demonstrated in the previous section and discussed in this one could of course also be applied around the world anywhere researchers have access to social media data. Truly, we are in the very early days of figuring out what we can and should be studying in terms of how regimes combat online opposition, but, then again, so too are the regimes themselves.

NOTES

All three authors participated in the research design, editing, and revision of this article. Sanovich is the corresponding author (sanovich@nyu.edu). Sanovich wrote the first draft of what is now Sections 1–4, and Stukal wrote the first draft of Section 5, as well as the code for the computational analysis in Section 5. Stukal recruited and Sanovich trained coders, whose coding is presented in Section 5. Both Sanovich and Stukal supervised the coders. We are extremely grateful to Duncan Penfold-Brown and Richard Bonneau, who have collaborated with us on the larger project of which this article is a part. We are also grateful to our volunteer coders from the Higher School of Economics (Moscow, Russia): Ivan Aleksandrov, Valeria Babayan, Maret Bochaeva, Daria Bushina, Viktoria Dimova, Yulia Gavrilova, Anastasia Gergel, Tatyana Glushkova, Alexandra Goncharova, Egor Ilyin, Christina Ilina, Aleksandra Izyumova, Artem Kolganov, Nikita Konyukhov, Yulia Korneeva, Maria Kuz, Alena Kuznetsova, Nikita Lata, Alina Lyutikova, Maria Makarova, Kamila Malikova,

Elena Malysheva, Polina Malyutina, Ekaterina Mikhaylova, Maria Milosh, Dmitry Muravyov, Mikhail Murzin, Pavel Myslovskiy, Timur Naushirvanov, Maxim Novokreschenov, Veronika Pankina, Anastasia Parshina, Dmitry Prodanov, Anastasia Rodygina, Zlata Sergeeva, Rais Shaidullin, Viktor Sinitsyn, Anna Skosyreva, Timur Slavgorodskiy-Kazanets, Anna Sokol, Elizaveta Sokovnina, Georgy Tarasenko, Oksana Tyulpinova, Azizbek Tulaganov, Natalia Vasilenok, Anna Velikanova, Alena Volodkina, Alexey Volokhovich, Anna Zaychik, and Kirill Ziborov. The data were collected by the NYU Social Media and Political Participation (SMaPP) laboratory (https://wp.nyu.edu/smapp/), of which Tucker is one of the co-Directors along with Richard Bonneau, John T. Jost, and Jonathan Nagler. The SMaPP lab is supported by the INSPIRE program of the National Science Foundation (Award SES-1248077), the New York University Global Institute for Advanced Study, the John S. and James L. Knight Foundation, the Moore-Sloan Data Science Environment, and Dean Thomas Carew's Research Investment Fund at New York University. Stukal and Sanovich gratefully acknowledge funding support from the Jordan Center for the Advanced Study of Russia at New York University.

[1] Bruce Etling, Karina Alexanyan, John Kelly, Robert Faris, John G. Palfrey, and Urs Gasser, "Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization," Berkman Center Research Publication No. 11, October 19, 2010, 33, available at http://papers.ssrn.com/abstract=1698344.

[2] John Kelly, Vladimir Barash, Karina Alexanyan, Bruce Etling, Robert Faris, Urs Gasser, and John G. Palfrey, "Mapping Russian Twitter," Berkman Center Research Publication No. 3, March 20, 2012, available at http://papers.ssrn.com/abstract=2028158.

[3] Adrian Chen, "The Agency," *New York Times*, Jun. 2, 2015, available at https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

[4] Due to space constraints, the Appendix is not in the print version of this article. It can be viewed in the online version, at www.ingentaconnect.com/cuny/cp.

[5] We draw upon a collection of 28 million tweets collected using key-word filtering over a year and a half period (November 25, 2013-July 13, 2015) during a particularly tumultuous period in recent Russian history including the Sochi Olympics, the Euromaidan uprising in Ukraine and the subsequent annexation of Crimea, and Russian involvement in the war in Eastern Ukraine.

[6] We further develop our bot-detection methods in Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker, "Detecting Bots on Russian Political Twitter," *Big Data*, 5 (December 2017), 310–24.

[7] Due to space limitations, we provide short summaries of both the case study and the bot-detection methods and findings in the text; much more information is available in the appendices.

[8] Giampiero Giacomello, *National Governments and Control of the Internet: A Digital Challenge* (London: Routledge, 2008); Hannibal Travis, ed., *Cyberspace Law: Censorship and Regulation of the Internet* (London: Routledge, 2013).

[9] Gary King, Jennifer Pan, and Margaret Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 107 (May 2013), 326–43; Kyungmin Ko, Heejin Lee, and Seungkwon Jang, "The Internet Dilemma and Control Policy: Political and Economic Implications of the Internet in North Korea," *Korean Journal of Defense Analysis*, 21 (September 2009), 279–95; Rebecca MacKinnon, "China's 'Networked Authoritarianism,'" *Journal of Democracy*, 22 (April 2011), 32–46; Zubair Nabi, "The Anatomy of Web Censorship in Pakistan," Paper presented at the 3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI '13), Washington DC, August 2013, available at http://arxiv.org/abs/1307.1144.

[10] Vladimir Barash and John Kelly, "Salience vs. Commitment: Dynamics of Political Hashtags in Russian Twitter," Berkman Center Research Publication No. 9, April 10, 2012, available at http://papers.ssrn.com/abstract=2034506; Katy Pearce, "Two Can Play at That Game: Social Media Opportunities in Azerbaijan for Government and Opposition," *Demokratizatsiya: The Journal of Post-Soviet Democratization*, 22 (January 2014), 39–66.

[11] Freedom House, *Freedom on the Net 2011–* (Washington, DC: Freedom House, 2011–), available at https://freedomhouse.org/report-types/freedom-net.

[12] Ronald Deibert, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain, eds., *Access Contested: Security, Identity, and Resistance in Asian Cyberspace* (MIT Press, 2011); Evgeny Morozov, "Whither Internet Control?," *Journal of Democracy*, 22 (April 2011), 62–74; Margaret E. Roberts, *Censored: Distraction and Diversion Inside China's Great Firewall* (Princeton University Press, forthcoming).

[13] Thus, we see our classification as a compliment, rather than alternative, to Roberts' framework.

[14] Technological responses include Internet-filtering, which spans from targeted bans of particular websites and keywords to larger national-level schemes to block entire segments of the Internet (China) or—in

the extreme—any outside Internet access outside the country (North Korea). Sociopolitical responses range even more widely from distributed denial-of-service (DDoS) attacks to employing both automated bots and paid trolls to destroy online communities' social capital to physical attacks on bloggers.

15. Larry Diamond, "Liberation Technology," *Journal of Democracy*, 21 (July 2010), 69–83.

16. Morozov, 63.

17. Deibert et al.; Bruce Etling, Hal Roberts, and Robert Faris, "Blogs as an Alternative Public Sphere: The Role of Blogs, Mainstream Media, and TV in Russia's Media Ecology," Berkman Center Research Publication No. 8, April, 2014, available at http://papers.ssrn.com/abstract=2430786; Pearce; Sarah Lange, "The End of Social Media Revolutions," *The Fletcher Forum of World Affairs*, 38 (Winter 2014), 47–68; Seva Gunitsky, "Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability," *Perspectives on Politics*, 13 (March 2015), 42–54; Maxim Ananyev and Anton Sobolev, "Fantastic Beasts and Whether They Matter: Do Internet 'Trolls' Influence Political Conversations in Russia?," Paper presented at the Midwest Political Science Association Annual Meeting, Chicago, IL, April 6–9, 2017; Kevin Munger, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker, "Elites Tweet to Get Feet off the Streets: Measuring Regime Social Media Strategies during Protest," forthcoming at *Political Science Research & Methods* (2018), available at http://kmunger.github.io/pdfs/PSRM_revised.pdf.

18. See a detailed discussion in Robert Ackland, *Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age* (SAGE Publications, 2013).

19. Neil Macfarquhar, "Russia Quietly Tightens Reins on Web With 'Bloggers Law,'" *New York Times*, May 6, 2014, available at http://www.nytimes.com/2014/05/07/world/europe/ russia-quietly-tightens-reins-on-web-with-bloggers-law.html. See also Appendix B.3.

20. Ackland, 143, 145–46.

21. Sergej Sumlenny, "Bad News: What Does the Closure of RIA Novosti Mean for Media in Russia?," *Calvert Journal*, December 12, 2013, available at http://calvertjournal.com/comment/show/1837/ RIA-novosti-putin-russian-media-kiselyov.

22. Gabrielle Tetrault-Farber, "RIA Novosti Begins Cutting 1/3 of Staff," *Moscow Times*, Mar. 12, 2014, available at http://www.themoscowtimes.com/news/article/ria-novosti-begins-cutting-13-of-staff/495980.html.

23. Yandex is incorporated in the Netherlands as Yandex N.V.—a fact that was publicly condemned by Vladimir Putin at his meeting with the All-Russia People's Front in 2014. See Christopher Brennan, "Putin Says CIA Created the Internet, Cites Foreign Influence at Yandex," *Moscow Times*, Apr. 24, 2014, available at http://www.themoscowtimes.com/news/article/ putin-says-cia-created-the-internet-cites-foreign-influence-at-yandex/498903.html.

24. Nikolay Grishin, "Yandexed Everything," *Kommersant—Trade Secret*, Mar. 12, 2012, available at http://www.kommersant.ru/doc/2065978.

25. Nickolay Kononov, "The Kremlin's Social Media Takeover," *New York Times*, Mar. 10, 2014, available at http://www.nytimes.com/2014/03/11/opinion/the-kremlins-social-media-takeover.html; Joshua Yaffa, "Is Pavel Durov, Russia's Zuckerberg, a Kremlin Target?," *Bloomberg Businessweek*, Aug. 1, 2013, available at http://www.businessweek.com/articles/2013-08-01/is-pavel-durov-russias-zuckerberg-a-kremlin-target.

26. Ingrid Lunden, "Durov, Out for Good from VK.com, Plans a Mobile Social Network Outside Russia," *Techcrunch*, Apr. 22, 2014, available at http://techcrunch.com/2014/04/22/durov-out-for-good-from-vk-com-plans-a-mobile-social-network-outside-russia/.

27. In the Russian case, these pleas were especially suspicious given that Russia already had a "national" search engine, social network, and email service, all created without any government aid. Yandex and VKontakte have larger market shares in Russia than Google and Facebook, respectively, an impressive result in the absence of any protectionist measures against foreign competitors.

28. Morozov.

29. *The Economist*, "The Art of Concealment," Apr. 4, 2013, available at http://www.economist.com/news/special-report/21574631-chinese-screening-online-material-abroad-becoming-ever-more-sophisticated. See as well Roberts.

30. King, Pan, and Roberts.

31. Philip N. Howard and Muzammil Hussain, *Democracy's Fourth Wave? Digital Media and the Arab Spring* (London: Oxford University Press, 2013), 71–72.

32. Morozov.

33. Helmi Noman, "The Emergence of Open and Organized Pro-Government Cyber Attacks in the Middle East: The Case of the Syrian Electronic Army," *Information Warfare Monitor*, May 30, 2011, available at http://www.infowar-monitor.net/2011/05/7349/.

34. However, hacking and publishing bloggers' personal communications (such as emails, instant messages, etc.) could allow the government to expose and implicate the opposition and shape the conversation that way. A typical example of the latter is the case of Russian blogger and hacker known online as *Torquemada Hell*, a Russian-speaking person allegedly living in Germany. In 2010–2011, he successfully hacked the email accounts of multiple Russian opposition politicians and released potentially damning information to the public. See Alexey Sidorenko, "Russia: Analysis of Hacker Attacks on Bloggers," *Global Voices*, Jun. 20, 2010, available at http://globalvoicesonline.org/2010/06/20/russia-analysis-of-hacker-attacks-on-bloggers/.

35. Roberts.

36. Alexey Sidorenko, "Russia: Major Search Engine Closes Its Blog Rating," *Global Voices*, Nov. 6, 2009, available at http://globalvoicesonline.org/2009/11/06/russia-major-search-engine-closes-its-blog-rating/. The more pressing concern for Yandex, though, was government outrage in the cases when, instead, anti-government posts got traction in the ratings, see Alexandra Odynova, "Yandex to Close List That Annoyed State," *Moscow Times*, Nov. 6, 2009, available at http://www.themoscowtimes.com/news/article/yandex-to-close-list-that-annoyed-state/388969.html.

37. Sarah Cook, "China's Growing Army of Paid Internet Commentators," *Freedom at Issue Blog*, Oct. 11, 2011, available at http://www.freedomhouse.org/blog/china%E2%80%99s-growing-army-paid-internet-commentators.

38. Anton Nossik, "11 Rubles and 80 Kopecks per Comment," *Echo of Moscow*, Sep. 10, 2013, available at http://www.echo.msk.ru/blog/nossik/1154616-echo/.

39. Miriam Elder, "Hacked Emails Allege Russian Youth Group Nashi Paying Bloggers," *The Guardian*, Feb. 7, 2012, available at http://www.theguardian.com/world/2012/feb/07/hacked-emails-nashi-putin-bloggers; Miriam Elder, "Polishing Putin: Hacked Emails Suggest Dirty Tricks by Russian Youth Group," *The Guardian*, Feb. 7, 2012, available at http://www.theguardian.com/world/2012/feb/07/putin-hacked-emails-russian-nashi.

40. Miriam Elder, "Emails Give Insight into Kremlin Youth Group's Priorities, Means and Concerns," *The Guardian*, Feb. 7, 2012, available at http://www.theguardian.com/world/2012/feb/07/nashi-emails-insight-kremlin-groups-priorities.

41. However, again in line with the Soviet blueprint, an elaborate system of digital surveillance called SORM was established. See Ivan Zasursky, *Media and Power in Post-Soviet Russia* (New York: M.E. Sharpe, 2004), 181–83.

42. That it became one of Medvedev government's defining policies suggests how limited was the change overall. See on limits of Medvedev's modernization Lena Jonson and Stephen White, eds., *Waiting for Reform Under Putin and Medvedev* (Basingstoke: Palgrave Macmillan, 2012).

43. Miriam Elder, "Emails Give Insight into Kremlin Youth Group's Priorities, Means and Concerns."

44. According to most observers, protest, triggered by the alleged major irregularities in vote count during Duma elections, did not just coincide with Putin's return to Kremlin, but was largely caused by the perceived undemocratic nature of the deal between Putin and Medvedev, which was announced just a few weeks before elections at the ruling party convention, and was kept secret until the last minute even from the convention delegates. See Richard Sakwa, *Putin Redux: Power and Contradiction in Contemporary Russia* (London: Routledge, 2014), 111–34.

45. Lunden.

46. Human Rights Watch, *Online and On All Fronts. Russia's Assault on Freedom of Expression* (July 18, 2017), available at https://www.hrw.org/report/2017/07/18/online-and-all-fronts/russias-assault-freedom-expression.

47. Ibid.

48. *Human Rights Watch*, "Russia: Halt Orders to Block Online Media," Mar. 23, 2014, available at https://www.hrw.org/news/2014/03/23/russia-halt-orders-block-online-media.

49. "Grani.ru v. Office of Prosecutor General," Global Freedom of Expression, Columbia University, September 2, 2014, available at https://globalfreedomofexpression.columbia.edu/cases/grani-ru-vs-office-of-prosecutor-general/. Alexey Navalny was able to successfully set up a free-standing blog that is not blocked, but he had to regain the audience he lost on LiveJournal before being able to expand it.

50. Sergei Guriev, "Facebook Faces down Putin," *Project Syndicate*, January 9, 2015, available at http://www.project-syndicate.org/commentary/facebook-versus-putin-by-sergei-guriev-2015-01.

51. We do, however, intend to greatly expand this focus in future research.

52. This could include using a set of sources, such as news feeds, other Twitter accounts, search queries, or even simply a list of predefined texts in a spreadsheet. Note as well that a loose understanding of this definition would also encompass AI (Artificial Intelligence) bots that change what they produce based on what they

encounter online, but at the same time we recognize that as bots become increasingly guided by AI our classification scheme will likely need to be revisited.

53. Politicians' names include Putin, Medvedev, Navalny, Khodorkovsky, Udaltsov, etc. The events feature Sochi Olympics, opposition rallies at Bolotnaya Square in Moscow, "Direct Line with Vladimir Putin," etc. We also searched for slogans like "Party of thieves and crooks," "Sobyanin is our mayor," "Stop feeding the Caucasus," etc. The full list of keywords and hashtags is provided in Appendix A.

54. We have data on 483 days between those dates, with the exception of two periods: October 2–29, 2014 and November 6, 2014-January 29, 2015; we stopped collecting tweets between those days for technological reasons. Our collection also includes additional 14 million tweets from 3.3 million users who do not list Russian as their account language. We chose not to include these tweets in the current analysis because our goal here was not to characterize the entirety of bot behavior on Russian Twitter, but rather to demonstrate the feasibility of bot-detection techniques on Russian Twitter.

55. Around 37% of all tweets in our collection are retweets, which means the user in question shared a tweet posted by another user. However, even among the remaining 63% of the collection, we were able to identify about 6.7 million tweets that were identical to at least one other tweet in the collection. Most of these are short and repeat just once, which might be a pure coincidence, whereas others repeat numerous times.

56. Bot detection is a relatively new topic in political science emerging from the burgeoning research in political communication on social media platforms. The scant literature that exists on the subject mostly borrows methods developed in computer science to detect email spam. Bot detection in social media is a different (but closely related) task that can be achieved with a wider range of techniques that make use of both textual and non-textual account information (see Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia, "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?," *IEEE Transactions on Dependable and Secure Computing*, 9 (November 2012), 811–24). Nevertheless, bot detection in social media is still regarded as a challenging task within the computer science community, making Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu claim that 80% of bots are undetectable (Yazan Boshmaf et al., "The Socialbot Network: When Bots Socialize for Fame and Money," *Proceedings of the 27th Annual Computer Security Applications Conference* (ACSAC '11), 93–102, available at https://doi.org/10.1145/2076732.2076746). Here, we combine some of the existing automated bot-detection techniques with domain specific knowledge and human coding. For more details, see Stukal et al., 2017.

57. In retrospect, it should not be surprising how easy it is to setup simple accounts like this. Adding diverse streams of content, videos, pictures, etc., on the other hand, requires more sophisticated programming. Further discussion of the results in Table 2 can be found in Appendix C.

58. Thus, for example, we didn't examine the political orientation of the news stories a bot was spreading, although this would be an interesting subject for future research.

59. We again adopted stringent inter-coder reliability requirements to ensure a high-level of reliability for any positive classification of a political orientation. To this end we imposed the following rule: if even just 2 of 5 coders put the account into two different partisan categories (for instance, pro-Kiev and pro-opposition), we considered this account as coded unreliably (the "Unclear" category in Table 3). If there was no disagreement between coders regarding the partisanship of the account and coders only differed if it belongs to a particular group or is neutral, we opted for neutral if three or more coders categorized it as such.

60. We also cannot rule out the possibility that at least some of the anti-Kremlin bots, particularly pro-Kiev bots with the most vicious content, were created as a provocation against the Ukrainian cause. A similar logic could also potentially imply that some of the most fervent pro-Kremlin bots were set up by regime opponents. However, the methods presented here are not suited to test these hypotheses, but an analysis of network structure in the future could shed light on this matter.

61. Diamond, 2010; Larry Diamond and Marc F. Plattner, *Liberation Technology: Social Media and the Struggle for Democracy* (JHU Press, 2012); Howard and Hussain; Joshua A. Tucker, Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá, "From Liberation to Turmoil: Social Media and Democracy," *Journal of Democracy*, 28 (October, 2017), 46–59;

62. While this is clearly a subject far beyond the current manuscript, "Social Media 3.0" may very well turn out to be the story of how anti-systemic forces in existing democracies harnessed social media to fuel the rise of populist political movements. For a much more detailed elaboration of this argument, see Tucker et al., 2017.

63. We address the former of these tasks in Stukal et al., 2017 and are currently at work on the latter.

APPENDIX

Appendix A Keywords and Hashtags Used for Collecting Twitter Data

1. #сочи
2. #Erdogan
3. #ExpelTurkeyFromNATO
4. #FreeSavchenko
5. #Latakia
6. #Nemtsov
7. #Putinkiller
8. #putinsgames
9. #RussianJet
10. #Russianplane
11. #sochi
12. #sochi2014
13. #sochi2014problems
14. #sochifail
15. #sochiproblems
16. #витишко -
17. #МинутаНеМолчания
18. #Немцов
19. #олимпиада
20. #олимпийскаязачистка
21. #ПутинУмер
22. #самолет
23. #считаемвместе
24. 37ґодвернулся
25. 6may
26. 6мая
27. PussyRiot
28. Su24
29. бирюлево
30. болотная
31. болотноедело
32. всезаодноґо
33. высурковскаяпропаґанда
34. ґолодовка
35. ґорожанепротив
36. ДМП
37. духовныескрепы

38. едро
39. жалкий
40. занавальноґо
41. зачестныевыборы
42. кировлес
43. команданавальноґо
44. кровавыйрежим
45. майданер
46. майданутый
47. майдаун
48. маршмиллионов
49. медведев
50. МинутаНеМолчания
51. навальный
52. народныйсход
53. немцов
54. Одесса
55. одинзавсех
56. оппозиция
57. партияжуликовиворов
58. пжив
59. привет37ґод
60. прямаялиния
61. ПуссиРайот
62. путин
63. путинах
64. путинвор
65. рассерженные
66. росузник
67. русскиймарш
68. савченко
69. свободуполитзаключенным
70. свободуузникам6мая
71. сднемпобеды
72. собяниннашмэр
73. Сочи2014
74. спасибопутинузаэто
75. Стратеґия31
76. Су-24
77. Су24
78. судвкирове
79. сурковскаяпропаґанда
80. толокно

81. толоконникова
82. триумфальная
83. удальцов
84. хватитвинитькавказ
85. хватиткормитькавказ
86. честныевыборы

Appendix B The Russian Government Online: A Long Way from Putin to Putin

B.1 *2000–2008: Putin I: Laissez-faire Regime for Online Media*

The many Russian political and economic reforms of the 1990s got a mixed assessment from both outside observers[1] and reformers themselves[2], and their reception among Russians remains ambiguous at best.[3] However, a wide consensus holds that if anything worked well during the painful transition from Communism, it was the introduction of media freedom.[4] Diverse, influential and competitive news outlets emerged almost immediately, and by the end of 1990s several powerful media conglomerates were operating alongside a large network of independent federal and regional media, usually free of any government control.[5] This is proved by the enormous role media played in political fights throughout 1990s[6] and, above all, during the so-called war for Yeltsin's succession,[7] when high-powered media was mobilized by both Putin and his opponents. The role of the media in Putin's rise to power is well documented by rigorous quantitative studies,[8] Putin's biographers,[9] Western observers,[10] and Russian political memorialists alike.[11] This experience allowed Putin to fully appreciate the power of the media to change public opinion and reverse political fortunes. Putin's media policy in the next 15 years demonstrates that he took this lesson extremely seriously and worked tirelessly to put the media under his control.[12] However, as we shall see, this policy was not universally applied to all types of media. To the contrary, online media enjoyed the *laissez-faire* regime that was in many respects on par with most advanced democracies. To uncover the reason why Putin drew such a sharp line between traditional and online media, we will start by examining the Soviet experience of media control. The seemingly inconsistent strategy Putin adopted becomes much less surprising when observed in light of the strategy once devised by Putin's (former KGB lieutenant colonel) employers at the Central Committee and Lubyanka. Next, we will discuss the main features of Putin's dual strategy and assess the changes in political and media environment that ultimately rendered it unworkable.

The, Soviet Union, as a relatively long-lived 20th century dictatorship,[13] survived several waves of technological advancement. Already in 1917, the Bolsheviks famously recognized the role of modern technologies and communications by seizing (along with the Winter Palace, railway stations, bridges and army headquarters) Petrograd's Telegraph and Telephone Exchanges. This recognition entailed two different strategies – for mass and personal communications – which were implemented fairly persistently throughout Soviet history. The government maintained a complete monopoly over the mass communication and fiercely prosecuted those who tried to challenge this monopoly. The most famous examples include the jamming of Western radio broadcasters (Radio Liberty, Voice of America, BBC, etc.) and strict regulations over using printers and photocopiers after they were installed at various Soviet administrative departments and institutions.[14]

Personal communications were a different matter. Instead of monopolizing their usage, the government allowed Soviet citizens to use them promiscuously and then used it to identify and prosecute those who were disloyal. Phones, for example, spread rapidly in the USSR, approaching roughly 37 million, or about 13 per 100 inhabitants in 1990.[15] This was still six times as small as in the U.S. at the time, but the difference was due to technological and economic reasons, not political restrictions. However, the government used every opportunity to spy on its citizens using wiretapping. Under Stalin serious efforts were made to study voice identification, which was famously depicted by Aleksandr Solzhenitsyn in the autobiographical novel *In the First Circle*.[16] Later the elaborate system of surveillance and the spread of personal, but publicly registered, home phones eliminated the need for voice identification. Similarly, typewriters were allowed for personal use, however their printout had to be handed to the *First Department* (local KGB affiliates at any Soviet enterprise or institution) and could be cross-verified to identify the exact typewriter used in printing "inappropriate" materials.

After assuming power in 1999, Putin gradually implemented a similar strategy of complete monopolization of mass media paired with a more liberal policy on personal communications. While the latter was virtually left free from interference (but not from surveillance[17]), national television networks were returned to government ownership and Soviet-style management with weekly instructions delivered from the Kremlin to the news executives.[18] Print media and radio remained more diverse, with some pro-government and some relatively independent outlets competing with each other.[19] The process of putting traditional media under government control in the early 2000s included such colorful episodes as the imprisonment of media mogul and oligarch Vladimir Gusinsky (in order to be released he signed a secret protocol with Russian Minister of Communications Mikhail Lesin and handed over his media assets to the state natural gas monopoly Gazprom) and stunning reversal of fortunes for the architect of both Yeltsin's electoral victory in 1996 and Putin's in 2000 Boris Berezovsky (who lost control of main Russian TV channel and was already in exile by 2001).[20] This and other episodes and their consequences for the media landscape are well documented in the literature.[21] Putin's approach to the internet, on the other hand, is less well documented. For example, observers have simply noted that "the Russian blogosphere is a space that appears to be largely free of government control"[22] or "the absence of Internet filtering is notable. Based on tests run through the OpenNet Initiative, we continue to find no evidence of significant technical filtering of the Russian Internet"[23], etc.

A recent account by the leading Russian internet news producer Anton Nossik suggests that this was no accident. Instead, already in 1999, the then still prime-minister Vladimir Putin had a clear preference for non-interference in the internet space:

> . . . in December 1999, three days before he became acting president of Russia, Vladimir Putin [. . .] summoned all the heads of Russia's nascent Internet industry for a meeting, including me. [. . .]. In his brief but passionate speech that day, Putin made special mention of Chinese and Vietnamese models of Internet regulation, stating that he viewed them as unacceptable. "Whenever we'll have to choose between excessive regulation and protection of online freedom, we'll definitely opt for freedom," he concluded to the puzzlement and disbelief of everyone in the room.[24]

Under the auspices of such a benevolent government policy, Russian online media flourished, becoming a vibrant sector of the economy and a reliable source of information for millions of Russians. Russia is one of the few countries where Google is not the most popular search engine and Facebook is not the most popular social network. Remarkably, both occurred without restrictions on American competitors. Unlike the Chinese Baidu and Weibo, the Russian platforms Yandex, Odnoklassniki and VKontakte won virtually fair competition with their American counterparts.[25] Successful development of local services did not mean that foreign ones were not actively used by Russian bloggers and readers. LiveJournal, the most popular Russian social network in 2001-2011, while being originally American and predominantly English-speaking, developed a Russian community so large that it was eventually overtaken by a Russian media holding and became dominated by Russian users.[26] And as of April, 2014 Facebook had 24 million users from Russia[27] and Twitter had more than 8 million[28], which makes Russian one of 10 most popular languages on Twitter.[29]

Again, in stark contrast with most other countries, the most popular Russian news websites do not represent traditional media such as newspapers, radio and TV broadcasters.[30] Instead, *Gazeta.Ru*, *Lenta.Ru*, *NewsRu.com*, *Polit.ru* and alike were built from scratch and became major news outlets in their own right (i.e. their staff does original reporting, often as an eyewitness, rather than just digitizing others' content).

As a result, Russia developed a strong, powerful and independent internet media sphere, which was a remarkable achievement for any non-democratic country, but especially for one where traditional media are so tightly controlled. As Karina Alexanyan and co-authors noted "Russia is unusual in the degree of freedom found online compared to offline media and political restrictions".[31] Such imbalance, however, proved to be unsustainable. In the late 2000s Internet media increasingly supplemented and eventually supplanted TV as the main news source at least for educated Russians.[32] One of the leading Russian TV anchors Leonid Parfenov, who has been banned from the air since 2004, aptly summarized this process in a 2010 speech, which went viral on YouTube[33]:

> These evergreen tricks are known to everyone who has witnessed the Central Television of the USSR. Reports are replaced by protocol shootings like "Meeting at the Kremlin"; reporter's intonations support the officials in the picture; broadcasting models are implemented to show "the leader receiving a minister or a governor", "the leader campaigns among the masses", "the leader holding a summit with his foreign colleague", etc. These are not news; this is old record that repeats the already established patterns of broadcasting. Even a news hook isn't a must. In the emasculated

> media environment any small fry will pass for a big shot just because of getting some airtime.
>
> [. . .]
>
> It hurts twice as much to speak about television journalism, given the evident success of the large-scale TV shows and Russian school of television series. Russian TV is getting more and more sophisticated in exciting, fascinating, entertaining and amusing people, but it hardly could be called civic social and political institution. I am convinced it is the reason for the dramatic decline in TV viewership among the most active part of the population. People of our type say: "Why bother turning on the box? It's not intended for us."

However, as Nossik notes, this dual strategy – tight control of traditional media and almost complete nonintervention in the web – was devised when Russian internet penetration was almost negligible.[34] Even three years after Putin came to power, in 2002, Russia had 2.1 million people (2% of the adult population) who used the Internet daily.[35] By 2008 this share increased to 14 million (16% of the adult population), and by 2013 to 52.2 million people (46% of the adult population).[36] Needless to say, the quality of access changed dramatically after wide access to broadband connection replaced slow dialup. These circumstances diminished the value of the monopoly in TV broadcasting and strong influence in other traditional media which the Kremlin enjoyed,[37] and simultaneously made the online communities sufficiently large and well-structured to become politically significant. The dual nature of social media, which is simultaneously mass and personal communication, presented a particular challenge for the government.

These changes coincided with the constitutionally required transition of power from Putin (who served two consecutive terms) to Medvedev in 2008. While Putin was appointed Prime Minister of Russia immediately after elections and Medvedev was widely considered a weak leader who never freed himself from Putin's oversight, Medvedev had his own agenda and probably nowhere else it was more visible than in his approach towards information technologies and the Internet in particular.

B.2 *2008–2012: Medvedev: The Blogger-in-Chief and his Followers*

Dmitry Medvedev's approach towards the Internet was an integral part of his general agenda. Laid out in an article "Russia, Forward!", which was published in the liberal (and online-only) newspaper *Gazeta.ru*, his *modernization* plan aimed to preserve the basic parameters of the political system built by Putin, but make it more efficient and friendlier towards businesses and citizens.[38] This included, for example, establishing Moscow as an international financial center, police reform, boosting the international competitiveness of higher education and the creation of a functioning e-government.[39] Medvedev's signature project was Skolkovo, a publicly funded, but semi-independently, managed high-tech incubator near Moscow. Obviously, the success of these projects was dependent on the creative class in major population centers, and IT professionals in particular. Thus, channels with these people, who were largely ignored by the blatant Soviet-style TV propaganda, was the first order of business for Medvedev. And unlike in many other areas, he did not hesitate to break with Putin's legacy, and put the traditionally solemn and unquestioned presidential speech in the caustic domain of the social networks.

Less than a year after assuming office, in early 2009 Medvedev started a video blog which quickly moved to LiveJournal – by then the main Russian social network and blogging platform. In 2010 he visited Silicon Valley, met Steve Jobs. and opened a twitter account at Twitter headquarters in San Francisco. Notably, his account began to follow (in addition to foreign heads of states and Russian officials) several bloggers known for their criticism of the government and newsfeed from the radio station *Echo of Moscow* – perhaps the most critical of the government among major media outlets in Russia. In 2011 he even set up his own Facebook page, which he occasionally used to communicate with its readers on matters not covered or ill-covered by the official media (such as 2011 protests) using a different, more frank tone. In all social networks, he built a large readership, which is typical for heads of states, but still notable since the environment was completely different from the general media environment Medvedev was used to: here he could not get his message through simply by eliminating competition and controlling the platform and the agenda.[40] In addition, in a rare occasion in 2011 he visited a small private TV channel *Rain (Dozhd)*, which then (and now) was mainly accessible online. As a result, Medvedev got permanently associated with blogging and social networks, and even was called both in Russia and abroad the "Blogger-in-Chief"[41], which simultaneously gave him credit for being up-to-date with the internet age and suggested that his rhetoric translated to little action.

Medvedev was not embarking on social media platforms alone. While it still remained an exception for high-level public officials at the time, several of his aids established a significant presence on social media as well. In particular, his close aid and economic adviser Arkady Dvorkovich maintained one of the most popular Russian Twitter accounts at the time, with close to half a million followers; he also has a Facebook page, as does Medvedev's press-secretary Natalya Timakova (who as a former *Kommersant* journalist is a Facebook friend of many prominent liberal reporters).

However, probably even more important was the establishment of a large-scale and permanent operation to push a pro-government agenda on the web and in social media in particular. Until then the Russian government presence on social media had been very limited. A report by the Berkman Klein Center for Internet and Society at Harvard University, which was published in late 2010 and covered the Russian blogosphere – concentrated in LiveJournal at the time – found that "pro-government bloggers are not especially prominent and do not constitute their own cluster".[42] Moreover, those affiliated with the government "are not central nodes in any of the political or social clusters [. . .] investigated".[43]

Following a long-standing Russian tradition, government action came late, but swiftly. Pro-Kremlin youth movements, created as part of efforts to prevent a "colored revolution" on Moscow streets and squares,[44] were partially repurposed to push a pro-government agenda online. Its leaders (in the case of *Nashi,* they were called *commissars*) became active bloggers, but they never relied on the persuasive capacity of their messages. Instead they gradually created a network of online support.

The network of support started with technological rather than human effort. Networks of bots were frequently employed first to flood opposition blogs with meaningless or assaultive content. Later they began to push alternative, pro-government messages to top charts and help pro-government bloggers to attract new followers. That same Berkman Center report mentioned previously also noted that "there is a concentration of bloggers affiliated with pro-government youth groups among the Instrumental bloggers [i.e. bots]".[45] However, real bloggers soon followed. In less than a year – which also witnessed the transition of the discussion core of the Russian blogosphere from LiveJournal to Twitter – pro-government bloggers emerged as a distinct and larger cluster on Russian political topics.[46] This result holds even after filtering out bots and other instrumental accounts, which remained numerous in the pro-government segment.

Continuous monitoring of the Russian blogosphere, undertaken by the *Internet in Russian Society* program at the Berkman Klein Center for Internet and Society at Harvard University from 2010 – 2014, reveals several distinctive characteristics of the pro-government segment in Russian social networks, as compared both to oppositional and "uncommitted" users. First, due to the general weakness and high fragmentation of the Russian opposition, "many active Russian bloggers [...] engage on political topics without 'choosing a team'. [...] most Russian bloggers prefer to declare an independent intellectual posture, and eschew group affiliations".[47] In contrast, pro-government bloggers tend to declare their political preferences and affiliation. Moreover, the usage of predominantly pro-government hashtags in Twitter was highly concentrated among pro-government users, at least compared to predominately oppositional hashtags, which were more widely used in different clusters. Finally, while pro-government users demonstrated high commitment in terms of the number of hashtag mentions (after the first one), they usually did it in a short time period, producing sharply peaked distribution of hashtag popularity.[48]

As the blogosphere remains the most ideologically diverse media environment in Russia, pro-government users experience pressures absent in other media. A comparative study of the Russian blogosphere and TV in the year before the Duma elections of 2011 reveals that this competitive environment forced pro-government bloggers to engage with their adversaries in cases when TV and even newspapers could largely ignore oppositional activity. Etling et al. give an example of the oppositional youth retreat in the outskirts of Moscow, which was intended to countervail the large government-sponsored youth camp "Seliger". Largely overlooked by the traditional media, it became the subject of the heated discussion between leading oppositional and pro-government bloggers on Twitter.[49]

The online response to hostile (or perceived as such) internet activity through direct engagement with users remained the "strategy of choice" during the Medvedev presidency, but certainly it wasn't the only one. Both offline responses and attempts to go through the online infrastructure to limit access to content did take place, but the latter were relatively rare and quite limited in their scope and the former was not a part of any systematic internet policy, and as such could not (and wasn't intended to) change the digital media landscape.

Up until the end of Medvedev's presidential term the only type of internet infrastructure infringement known in Russia were relatively brief (lasting up to several days) DDoS attacks on particular web resources.[50] The first major attack was launched on August 6, 2009, the first anniversary of the Russo-Georgian 5-days war. The target was the pro-Georgian blogger *cyxymu*. The attack was strong enough to significantly disrupt Facebook and completely shut down Twitter and LiveJournal.[51] The series of smaller attacks on various LiveJournal blogs and independent media culminated on the weekend of the Russian Duma elections of 2011, when two dozen of the most prominent independent media (including *The New Times*, *Kommersant*, *Echo of Moscow*, *Novaya Gazeta*, *Slon*, etc.), blogs (including the entire LiveJournal platform) and, most crucially, election monitors' coordinating portals (including the largest one, GOLOS) were shut down for hours.[52] Later many of the very same resources were attacked during opposition rallies after the elections and in the early 2012.

Importantly, DDoS attacks, unlike filtering (and offline responses), could be used not only by the government, but also by the opposition. For example, in early 2012, the Russian branch of the international cyber activist group *Anonymous* blocked the web sites of the Russian government, the Kremlin and several major state media, such as *Vesti* and *RIA Novosti*.[53] These attacks, however, lasted only several hours (compared to several days in the case of LiveJournal), and obviously could not impede the offline state response to demonstrations.

Finally, the offline response by the Russian government to unfriendly internet activity was not yet separated from general anti-opposition activity and was not legally or organizationally institutionalized. Market regulation and government entrepreneurship was still targeted at traditional media: for example, in 2011 the newspaper *Moskovskiye Novosti* was relaunched by the state news agency *RIA Novosti*. As it was widely assumed, the project was aimed to provide moderate competition to the privately owned (and quite critical) *Vedomosti*, simultaneously being friendlier to Medvedev than most state media and still loyal to Putin.[54] Later that year Medvedev announced the establishment of the Public Television of Russia (*OTR*), which was supposed to compete with TV Rain (*Dozhd*) and shared the second goal with *Moskovskiye Novosti* as well.[55]

Violence and legal action against bloggers was relatively rare and mostly took place in the North Caucuses. Legal restrictions, if any, were imposed under the auspices of the general anti-terrorist laws and orders, mostly having to do with combating Chechen and Dagestani insurgencies. While the anti-terrorist rational was often abused for the sake of winning over political enemies in the respective republics, these cases were rarely consequential at the federal level.[56] In a few cases outside the Caucuses prosecutions were largely a regional matter or the result of the local security apparatus initiatives rather than implementation of any national strategy. Prominent cases from that time included the blogger Savva Terentyev from the Komi Republic, who in 2008 was convicted of defamation of the "social group 'law enforcement personnel'" and sentenced to one year of imprisonment with a probation period of one year after an anti-police comment at LiveJournal. Another prominent case took place in 2009 in the Republic of Tatarstan, where a former government official turned opposition blogger posted a false rumor that the governor of the republic had died. He was convicted of libel and defamation of the "social group 'government officials'" and sentenced for 2 years in prison.[57]

Institutionalization of the offline response, as well as the means of control over the online infrastructure, happened only after Dmitry Medvedev handed his office back to Vladimir Putin in 2012. However, the process was so quick that already by 2014 the relative importance of different types of government responses were reversed: the sheer force of offline response and establishment of a comprehensive system of internet filtering rendered the online engagement with users, created by Medvedev, almost irrelevant.

B.3 *2012–2014: Putin II: Cracking Down and Giving Up*

Compared to the transition from Putin to Medvedev in 2008, the reverse transition in 2012 was much less smooth. Announced on September 24, 2011 and immediately nicknamed as "castling", it was met with resentment by both Medvedev supporters and those in opposition to both Medvedev and Putin.[58] This resentment manifested itself in large-scale street protests after the December, 2011 parliamentary elections, which were widely considered rigged.[59] As mentioned above, the close relationship between Putin and Medvedev did not mean that Medvedev lacked his own agenda. In this case too his response was a program of moderate, but significant, political reforms, announced in the Address to the Federal Assembly (the Russian equivalent of the State of the Union) in late December of 2011, three weeks after the Duma elections, and just after major protests had started. This program included, most importantly, the reinstatement of popular elections of Russian governors and the election of MPs in single-member districts (switching back from pure proportional to a mixed electoral system).[60] These reforms, however, were either striped of any substance (like changes in party registration rules) or explicitly reversed (like decriminalization of libel).[61] Protest activity, on the other hand, was severely restricted after on May 6, 2012 (one day before Putin's inauguration) when an opposition rally was dispersed by force (hundreds of people were arrested and several dozens of them were subsequently prosecuted for inciting riots and assaulting police).[62]

It is in this context that the freedom of the Russian internet from filtering came to an end.[63] Already in July of 2012, despite vocal protests, including a temporary voluntary shutdown of Russian Wikipedia, the Russian State Duma adopted (and Vladimir Putin signed into law) the so-called Internet Restriction Bill (Federal law of the Russian Federation no. 139-FZ), which created a continuously updated Russian Internet Blacklist.[64] The list, maintained by the Russian Federal Service for Supervision of Communications, Information Technology, and Mass Media (*Roskomnadzor*), contains domain names which any Russian ISP has to permanently block on the grounds of containing pornography, copyright infringement or "extremist content". Initially, items were to be included in the list per a court order and only if the hosting website failed to remove the content in 24 hours after receiving the notification. However, in December of 2013 new amendments to the Law on Information, Information Technology, and Information Protection provided the Office of the Prosecutor General with the authority to block websites without any court order. Moreover, the procedure was changed, so the web page was to be blocked first, and allowed to be accessible again only after it removes the content deemed as "calling for mass disorders, extremist activity, and participation in mass public events, which fail to follow appropriate regulations".[65]

However, when at the height of the Russian-Ukrainian conflict in March of 2014 several oppositional news web sites were blocked, even these loose rules were not followed. On March 13, 2014 *Grani.ru*, *Kasparov.ru* and *EJ.ru*, as well as popular opposition politician Alexey Navalny's LiveJournal blog, were blocked by all ISPs per government order. Since then several suits have been brought to courts demanding the reason for the blocking. Journalists and Alexey Navalny have asked the authorities to identify the specific materials on these websites that triggered the blocking, so that the materials could be removed and access reestablished. Throughout 2014 the authorities repeatedly denied that they were under any obligation to provide such information and courts repeatedly dismissed the cases.

Still, the Russian government's incomplete control over online infrastructure significantly impedes its ability to crack down on opposition activity simply by blocking web pages. The greatest threats are the large foreign social media platforms, i.e. Facebook and Twitter. First, unlike most other web resources, Facebook's and Twitter's individual pages (say, a particular post or user profile) could not be blocked by the filtering software currently available to the Russian authorities.[66] Blocking the entire platforms, on the other hand, is still considered undesirable: it would further hurt the Putin's regime reputation abroad and simultaneously hurt and potentially antagonize a large number of politically indifferent (and regime-friendly) users in Russia. In a rare event, a public official, *Roskomnadzor* deputy head Maxim Ksenzov, who speculated about such a possibility was publicly rebuked by Prime Minister Dmitry Medvedev in a Facebook post and later was formally reprimanded.[67]

Of course, instead of blocking these platforms, the Russian government could ask them to police themselves and remove access to certain pages at least for users inside Russia. However, unlike VKontakte, foreign social networks can easily ignore such orders. For example, in December of 2014 authorities requested Facebook and VKontakte to block access to pages, allowing supporters of Alexey Navalny to register for a rally protesting his looming criminal conviction and receive updates about the place and time of the event. VKontakte blocked the page and all subsequent attempts to create a copy, posting a warning that "This page is blocked upon receiving a *Roskomnadzor* notification of restricting access to information, which contains calls to participate in mass public events, which fail to follow appropriate regulations, as per the request of the Office of the Prosecutor General of Russia."[68] Facebook also blocked access to a similar page inside Russia,[69] but after a huge outcry in the Western media, refused to block any other pages. Moreover, some Russian media outlets, which were afraid to report the scheduling of the event itself, covered the *Roskomnadzor* order and social networks response. As a result, more people learned about the event and the new event page opened on Facebook attracted even more people.[70]

Given that second page attracted more than 33 thousand people, who stated that they are “going to the rally” (plus almost 6 thousand, who stated that they were “likely going”), it is not surprising that the authorities resorted to an offline response: they simply changed the day of the court proceedings to two weeks earlier. The new date was the day before the largest Russian holiday (The New Year’s Eve) and Navalny was informed less than 24 hours in advance. While the third event also attracted a considerable number of supporters, the combination of suddenness, cold weather and pre-holidays preparation likely reduced the turnout.

Offline responses were certainly not limited to the types of ad hoc solutions just described. Instead, government complete control over the law enforcement apparatus and law making was actively used to augment its limited ability to censor social media platforms. Criminalization of online activity was first implemented through targeted amendments to existing criminal law, but was soon institutionalized in dedicated laws. Using media to spread information deemed extremist was always an aggravating circumstance in Russian criminal law. Laws missing such provisions were sooner or later corrected: for instance, when in 2011 the punishment for Article 280 of the Criminal Code was severed, using mass media for “extremism propaganda” became an aggravating circumstance. However, when just two years later, in 2013, a new extremism crime appeared in the Criminal Code (Article 280.1, Public Appeals to the Violation of the Territorial Integrity of the Russian Federation), using “mass media, including telecommunication networks (including ‘Internet’)” was added as an aggravating circumstance.[71]

In May, 2014 Vladimir Putin signed into law a requirement for any blogger with a daily readership in excess of 3000 people to register with the government and reveal her true identity and email address.[72] In addition, bloggers will be held accountable for failure to verify the information they “spread”, have to keep archives of their postings, and follow laws which regulate news production during electoral campaigns. However, institutionalized regulations – as might be expected – are much less effective then targeted actions: in half a year after the law came into force just 369 people got registered with *Roskomnadzor*[73] and the only known real consequence is the shutdown of Intel’s forum for developers – hardly a platform of political significance, which was closed by Intel voluntarily out of an abundance of caution.[74] Among the reasons are unclear definitions of “readership”: *Roskomnadzor* guidelines on the subject call for the use of rigorous “page views” count (rather than hits, number of friends or followers or any other metric), but not all platforms generate such statistics, and it is especially hard to accomplish this using social network platforms.[75]

Using loyal business groups to restructure the online media market proved to be a much more reliable tool to ensure that at least Russian major platforms fall in line. Hitherto mostly focused on traditional media (TV and press), power brokers in the Presidential Administration, Ministry of Communications and the largest media conglomerates have been increasingly preoccupied with online news outlets and platforms. The methods they used were not much different. Two cases are particularly revealing. In 2014 billionaire Alexander Mamut fired the editor-in-chief of the most popular Russian online news portal *Lenta.ru*, allegedly on the grounds of insufficiently "pro-Russian" coverage of the Ukrainian revolution of 2013-2014. The complete lock out of the entire editorial staff was strikingly similar to the one at the NTV channel in 2001 and countless others since then. However, in contrast to their colleagues from TV, this fired team of journalists was able to relaunch their media. The insurance of their independence and security from outside pressure was a physical relocation of most of the editorial staff to the neighboring Baltic country of Latvia and opening the website in an *.io* domain zone, which belongs to the British Indian Ocean Territory and is administered by a UK company. The new name of the company, Meduza (Russian for jellyfish), matches the geographical location of its domain.

The hostile takeover of VKontakte in 2013-2014 by Kremlin-affiliated businessmen also followed the approach which earlier successfully secured the loyalty of various media outlets (such as *Izvestia* and *Kommersant*): involuntary ownership transfer, usually compensated at the market rate if the former owner cooperates. This transfer usually came after former owners and/or managers refused to cooperate in politically sensitive matters for too long. VKontakte for years received requests from the FSB similar to the ones described (and followed) in the case of the pro-Navalny rally in late 2014. Specifically, requests to remove pro-Navalny groups came first in the wake of large-scale protests after Duma elections in 2011[76], but VKontakte owner and CEO, libertarian internet-guru Pavel Durov, refused to comply. However, when in early 2014 VKontakte was served with a request to disclose the personal data of administrators of Euromaidan-related pages in VKontakte,[77] the government did not take no for an answer. Durov had to sell what was left of his share, resigned, and left the country.[78]

The lesson of VKontakte was taken seriously not only by Russian media managers and owners who wanted to keep their positions and businesses; foreign companies that wanted to be able to refuse involuntary cooperation with Russian government had to assess if they had any vulnerable assets in Russia. For instance, Facebook's ability to change its response and refuse to block any more groups in the late 2014 episode of the pro-Navalny rally was secured by the company's lack of any significant assets within Russia. Google, which had a development office in Russia, closed it, and the entire engineering team was invited (and, for the most part, accepted the offer) to move to Google offices in Europe and elsewhere. While reasons for this move were not disclosed, observers assumed the company was concerned with the potential access of the Russian government to the Google code and, especially, with the possibility of the government employing coercive methods to gain this access through pressure on individual engineers.[79]

What in this context happened with online responses? Did the offline response and through the online infrastructure supplant any engagement with users on behalf of the government? If by engagement we mean an attempt to persuade and generate support, then the answer is in the affirmative: distortion, confusion and discouragement emerged as the new goals of the Russian government's online propaganda strategy.

Another key change was in the target audience of the government online effort. If Medvedev was trying to build a coalition around values of modernization and reformist policy agenda, "Putin Redux" entailed rather dramatic change in regime ideology, not just compared to the Medvedev years, but also compared to Putin's first two terms.[80] A reorientation towards conservative, even traditionalist values in domestic policy was paired with expansionist, revanchist foreign policy.[81] Among the consequences of this change was the reorientation of the online propaganda machine from winning over neutral or even already opposition-inclined users towards protecting the wider public (those receiving most of their news from TV, but starting to use the internet for entertainment or consumption) from the dangerous influence of the dissident voices by spreading division and mistrust.

A typical example from the same episode with the pro-Navalny rally in late 2014 was the production of YouTube videos with prominent Navalny supporters, who were showing on air the web address of the supposedly pro-Navalny website with information about the rally. In reality these were fake and the address lead to a page full of anti-Navalny videos. These videos and the apparent endorsement of them by famous artists and journalist were then promoted on social media.

Such provocations obviously could not build the reputation that Medvedev had been seeking to build online. The new goal was not to engage and invite discussion it is to disorganize, discourage and mute opposition. And this goal is much better served by filtering technologies and targeted prosecution of influential bloggers. Extensive online debates between oppositional politicians and the pro-Putin "Nashi" youth movement, which occasionally happened before 2012, were no longer in demand. With the gradual, but persistent, political retreat of government officials of liberal inclinations (which in many cases included leaving public service or even the country for good), the government presence online did not vanish completely. Government did not go offline, but it was no longer trying to respond to anybody online, much less waiting for responses from anyone. Its communication became a monologue, and its propaganda a jamming device. The experience in more artful digital propaganda was not wasted, though, but put for a productive use abroad to support Russian increasingly expansionist foreign policy.

Appendix C Online Engagement: Bots in Russian Political Twitter

C.1 Bot-Detection Methods

Entropy of inter-tweeting time intervals. Our first technique is predicated on the idea that bots show a much higher regularity in their activity on Twitter than human beings. This is also true for cyborgs, at least to the extent that they rely on automated sourcing of content. In the simplest case, for instance, bots may be programmed to send tweets every *k* seconds. On the contrary, humans' tweeting activity is much more sporadic. These differences in predictability may be captured by entropy, which is a measure of uncertainty popular in computer science and information theory. In order to compute entropy, we created a list of all accounts that have at least three tweets in our collection.[82] Then, for each of these accounts, we computed the length of time intervals between consecutive tweets, and used those time intervals to compute average entropy as follows:

$$Av\text{·}Entropy_i = \frac{1}{T_i} \sum_{t=1}^{T_i} p_t^{(i)} \times log_2\left(p_t^{(i)}\right),$$

where p_t is the probability of interval t for account i; T_i is the total number of time intervals for account i. The higher the value, the more unpredictable an account is. We expect that accounts with low entropy are either bots or cyborgs.

Followers/friends ratio. One concern with an entropy-based approach to bot detection in the case of a dataset acquired by selecting tweets containing particular keywords is that we could be missing activity by accounts in our collection that also tweeted on other topics. Thus, we use another method of detection that does not depend on tweeting activity, but instead is based on the idea that bots should have fewer followers than normal human users. Indeed, most humans would probably refuse to follow a bot that does not show signs of a normal human online activity. At the same time, bots would tend to follow lots of users (in Twitter parlance, have many friends) in the hope that some of them will accidentally follow them back. Thus, we expect that some bots will tend to have a very small followers/friends ratio defined simply as:

$$ratio_i = \frac{|\{followers_i\}|}{|\{friends_i\}|},$$

where $|\{followers_i\}|$ denotes the number of accounts that follow account i, and $|\{friends_i\}|$ stands for the number of accounts that are followed by that account.

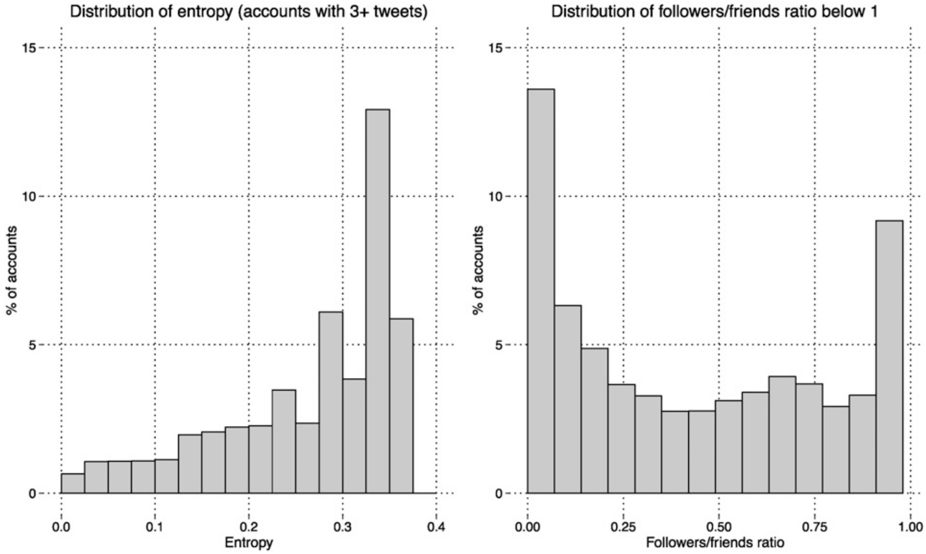
Another interesting case includes accounts that have no friends (i.e. do not follow anybody). Obviously, the followers/friends ratio is undefined for these accounts due to division by zero, and we code them separately.

Identical tweets. Our two final bot-detection techniques involve identifying accounts that send identical tweets. There are two subtypes of identical tweets an account could send out: *intra-* and *inter-account* identical tweets. The first subtype refers to the case when an account is repeatedly sending the same tweet. We doubt that a human being would engage in such an activity, whereas a pre-programmed primitive bot could easily do that. The second subtype refers to the situation when a group of accounts are sending out identical tweets. This strategy can be employed by bots to maximize the spread of specific information over the network.

C.2 Empirical Assessment of Bot Detection Methods

We use all the methods outlined above to identify bot accounts for the subsequent verification using human coding. Doing so required setting a set of thresholds that

Figure C1 Distribution of Twitter Accounts in the Data Set by the Entropy of Inter-Tweeting Time Intervals (left panel) and Followers/Friends Ratio (right panel)



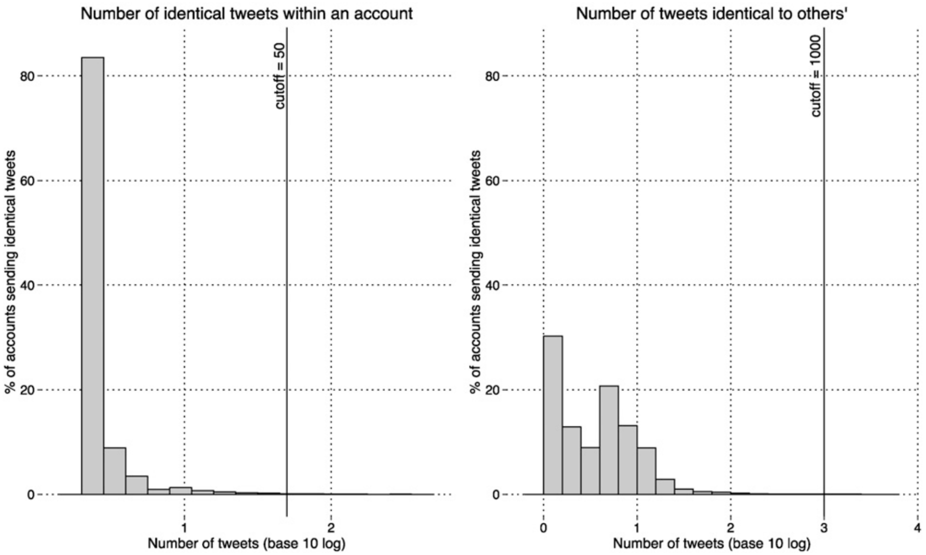
determine how many accounts each method recovers for us to code by hand. Similar to most cases of threshold selection, this is not a straightforward exercise since no theory has so far been developed to justify the choice. Given the lack of theory-driven guidance, we followed an empirical approach with two main considerations in mind. On the one hand, thresholds should be sufficiently loose to produce enough accounts for a meaningful verification. On the other hand, the thresholds needed to be sufficiently stringent to keep hand-coding of the recovered accounts feasible.[83]

For the *entropy* measure, we restricted our attention to those accounts that produced at least 300 tweets throughout the period under study. As there were more than 5,500 such accounts, we selected 100 accounts with the lowest entropy values.

A similar threshold (300+ tweets in our collection) was applied to 36,500 accounts with no friends, yielding 99 accounts.

To set the threshold for the *followers/friends ratio*, in the right panel of Figure C1, we plotted the distribution of all accounts in our dataset that have a ratio below 1 (i.e. follow more people than they are followed by). There are more than 900,000 accounts like that, or about 70% of our accounts. Among these, a disproportionally large number of accounts have a ratio around 0 (when nobody follows them) and 1 (when they are followed by the same number of accounts that they follow themselves). We expect bots to have a very low ratio and have chosen 0.01 as our threshold, thus selecting accounts that follow (at least) 100 times more accounts than they are followed by. Together with

Figure C2 Distribution of Accounts Sending Repeated Tweets



a 50+ tweets (in the collection) activity requirement, this method yielded 135 accounts for verification.

Lastly, 43,000 accounts in our collection tweeted the same text several times and another 600,000 tweeted text that some other accounts in our collection tweeted (distributions shown in Figure C2). We expect a large proportion of accounts engaging in such behavior to be bots. For the former category, we have set the threshold at 50 repetitions, yielding 90 accounts to verify. For the latter category, we set the threshold at 1,000 repetitions, which left us with 102 accounts to verify.

Thus, we end up with five sets of "suspicious" Twitter accounts (a total of 526) that we identified using different bot-detection techniques. Table C1 shows intersections of these sets. There are in total 14 duplicates, therefore we proceed with an analysis of 512 accounts. The low number of accounts these sets have in common suggests that we probably managed to identify different kinds of bots that might be used for different purposes, or at the very least were created using different techniques.

In order to assess the reliability of our bot-detection algorithms, we enlisted 20 coders (native Russians, undergraduate students in political science, and familiar with Twitter) and tasked them with classifying 512 accounts into five categories: in addition to humans, bots, and cyborgs, described above, we have two miscellaneous ones, spam and official accounts. Spam includes accounts that feature no meaningful content, and consist mostly of gibberish or consumer advertisement. Official accounts refer to the accounts run by organizations (such as media and government bodies) whose tweeting

Table C1 Intersection of Sets With Suspicious Accounts

	No friends	Low ratio	Entropy	Repeat themselves	Repeat others
No friends	**99**	0[a]	4	2	1
Low ratio		**135**	0	0	0
Entropy			**100**	0	5
Repeat itself				**90**	2
Repeat others					**102**

Note: Entries are numbers of Twitter accounts that are common to a pair of sets.

[a] "Low ratio" and "no friends" sets are mutually exclusive by definition, as ratio of followers to friends is undefined for accounts with no friends.

patterns are expected to be different from personal accounts. Some of these five categories were further split into subcategories. This was primarily done to reduce noise in human coding (more narrow categories are easier to define precisely for coders), and they were reaggregated at the analysis stage. With respect to bots, however, we present some interesting findings about their subtypes below. For details on the coding schema, see Appendix D.

Our coders received detailed coding instructions in Russian (Appendix D contains a brief summary), and were provided with a list of 20 Twitter accounts (not drawn from 512 accounts of interest) to code as an exercise. Next, each of them went through a 90-minutes Skype session with one of the co-authors, to ensure their clear understanding of the coding schema. We then randomly split coders into 4 groups of 5 people, and randomly assigned different Twitter accounts of interest to different groups. All coders were instructed to work independently and did not know the names of other coders in their groups. Thus, every account was classified independently by 5 coders.

It is important to note that coders did not work with live Twitter accounts. Instead, we used tweets and other data from our collection to re-create Twitter accounts *based on the data collected* at the time our collection was running. Obviously, there is no chance to replicate accounts exactly (for example, we do not include non-political tweets because they did not contain any of our keywords or hashtags and therefore were ignored by our collection filters). However, in addition to featuring up to 100 tweets from the account,[84] we made use of account metadata to reproduce the number of followers, friends, tweets, as well as account description, geo-location, user and background pictures, date of account creation, and other information typically available on a Twitter web page. This makes our entire approach completely replicable; for example, another team of scholars could look at the exact same collection of tweets using the same visual display to try to replicate our coding.

In the text of the article we report the accuracy of these bot detection methods (Table 1), as well as disaggregate identified bots by detection method and sub-type (Table 2, reprinted here for the reader's convenience).

Several interesting patterns emerge from breaking down the types of bots by their source of detection (first five columns in Table 2). For example, more accounts that do

Table 2 Verified Bots by Type and Method of Identification

	No friends	Low ratio	Entropy	Repeat themselves	Repeat others	Total
Retweets only	3	8	26	3	3	**9**
Videos only	2	0	0	0	0	**< 1**
Pictures only	2	0	0	0	1	**1**
Text only:						
– News headlines only:						
News headlines with links	38	74	40	41	15	**40**
News headlines without links	48	15	26	36	73	**42**
– Other text	3	2	2	5	1	**3**
Diverse content	3	2	7	15	7	**6**

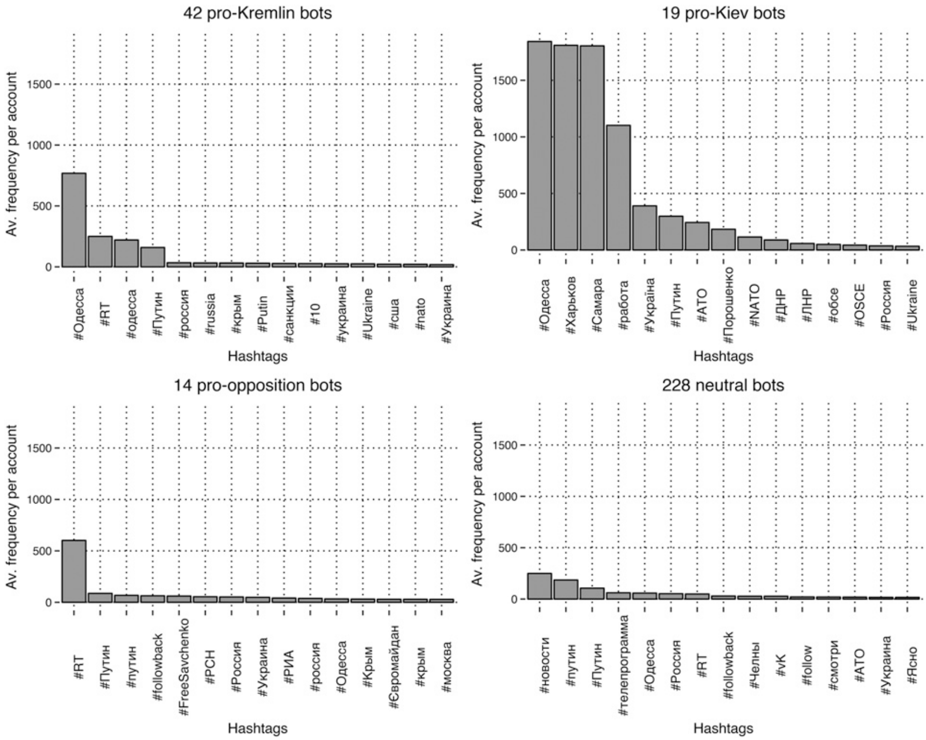
Note: Entries are column percentages (may not sum up to 100 due to rounding).

not follow any other accounts (first column) feature tweets *without* links than *with* links to the news story referenced by the headline; for accounts that do follow other accounts (second column) this proportion is reversed, with share of tweets with links jumping from 38% to 74%. While the former most likely are simply aimed at promoting specific news in search engine rankings, the latter seem to be designed in the hope that at least a few people would follow them back and then click on the news link. Accounts repeating others are similar to accounts with no friends: 73% of them tweet news headlines with no link to the story. Accounts that repeat themselves are different; they more often feature diverse content (quotes from famous people, beauty and character advice, etc.). It is this content that tends to be repeated to attract followers, who are then exposed to fresh news stories, often with a link to the website. The low entropy method as well catches a different kind of bots: those that retweet content from other accounts.[85]

C.3 Pattern of Bot Activity by Political Orientation

In addition to coding each account as bot, human, cyborg, spam or official account, we also coded their political orientation. In the text of the article, we explain the categories we employ, as well as present an initial distribution of bots by ideological orientation in our set of identified bots. Here, we delve a bit deeper into the question of whether the behavior of bots with different ideological orientation vary. We explore this issue in two ways by focusing both on content and the dynamics of tweeting activity. In terms of the former, Figure C3 presents bar plots illustrating the popularity of the 15 most common hashtags within every orientation group, including neutral bots. One can see from the graph that all types of bots discussed political developments in Ukraine and include a variety of hashtags about the Euromaidan protests, as well as further tragic events in Odessa and Eastern Ukraine. Despite these similarities, though, there are also important differences across orientation groups. For instance, the Organization for Security and Co-operation in Europe (OSCE), which plays a peace-keeping role in

Figure C3 Most Common Hashtags



Eastern Ukraine, is more popular among pro-Kiev bots, whereas the U.S. and NATO appear among common hashtags for pro-Kremlin bots. Hashtags in support of a Ukrainian officer Nadezhda Savchenko, who was captured and put on trial in Russia (allegedly in relation to the death of Russian journalists who covered the conflict in the Eastern Ukraine), are among the most popular hashtags among pro-opposition, but not pro-Kiev, bots.

Table C2 Jaccard Similarities of 50 Most Popular Hashtags

	pro-Kremlin	**pro-opposition**	**pro-Kiev**	**Neutral**
pro-Kremlin	1.00	0.27	0.30	0.25
pro-opposition		1.00	0.39	0.19
pro-Kiev			1.00	0.23
Neutral				1.00

Note: Entries are Jaccard similarities between sets of 50 most popular hashtags in every group of bots. Jaccard similarities range from 0 to 1, where 0 refers to a pair of sets that have no common element, whereas 1 refers to a pair of sets that are identical.

Figure C4 Political Bots' Tweeting Activity over Time

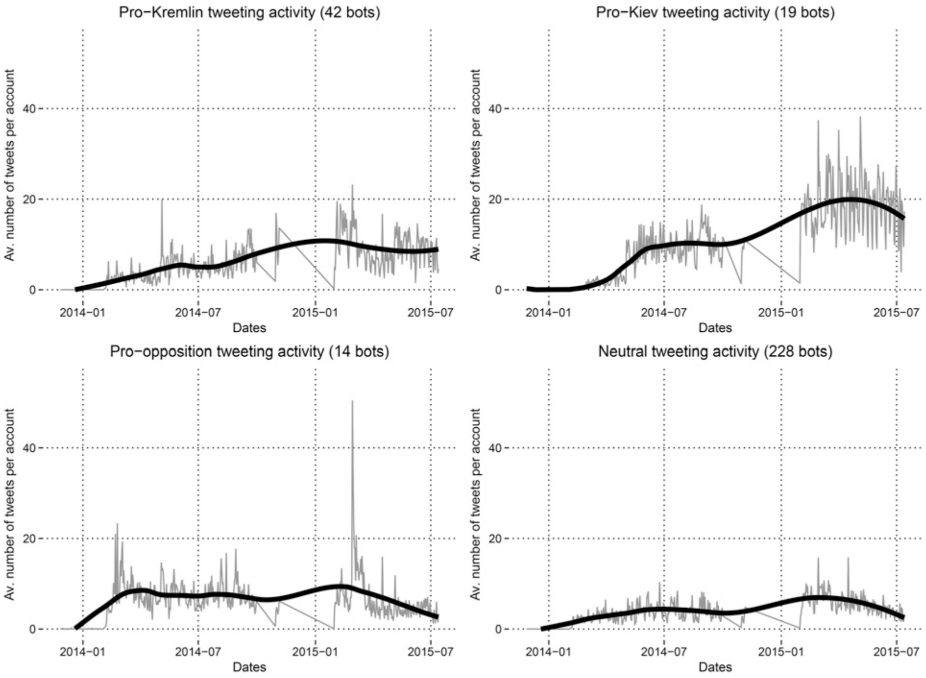


Table C2 reports Jaccard similarity scores, a popular measure of the similarity between two sets that shows how many elements the two sets have in common in comparison to the total number of elements in the sets.[86] Thus we find that pro-opposition and pro-Kiev bots are the closest in terms of their used hashtags, whereas neutral bots are furthest apart from the rest, as Table C2 demonstrates.

Another way to examine the similarity of bots with different political orientation is to look at when and how much they tweet. Figure C4 reveals that the pro-opposition bots we identified were more active than the pro-Kremlin bots we examined at the very beginning of 2014, and pro-Kiev bot activity was virtually non-existent. With the annexation of Crimea, the conflict escalation, and Russian involvement becoming more assertive, pro-Kremlin bot activity quickly caught up with the pro-opposition bot activity by the late Spring of 2014. By summer, the pro-Kiev bots we identified were in full swing too.

All three sets of bots produced roughly the same levels of activity through the end of 2014, but then began to diverge again. Pro-opposition activity was gradually dwindling, but this trend was interrupted when late at night on February 27th, 2015 one of the most prominent leaders of the Russian opposition, Boris Nemtsov, was shot dead in front of the Kremlin. His tragic death generated one of the largest waves of identical tweets in our collection that involved more than 1,800 of the users for whom we have data. Meanwhile, pro-Kremlin bots' activity remained relatively constant and did not match the marked increase in activity by pro-Kiev bots in the beginning of 2015, when the war in Eastern Ukraine escalated again.[87]

One final point is worth reiterating: these hand-coded accounts should not in any way be taken as a representative sample of all bot activity. As noted in the first section of this appendix, we selected these accounts according to peculiar criteria, including the frequency with which they appeared in our collections. Future research that applies machine learning to code the entire data collection in terms of bot presence and bot orientation will be needed before we can make more general claims about bot behavior in Russian political Twitter; the previous examples are intended to demonstrate the exciting potential of such research.

Appendix D Classifying Twitter Accounts

For all accounts, the first order of business for coders was to check if it features some meaningful content or if instead it contains exclusively advertisements and/or gibberish. The latter category of accounts mostly features various kinds of **spam** links (sales coupons, lottery, etc.), pictures of consumer products and instant earnings ads. Sometimes, they switch to feeds of spam links in other languages. If an account features meaningful content, coders first checked whether it is an **official** account for an organization. We put these accounts into a separate category, for institutional accounts are qualitatively different from the ones individual people run (for example, they have larger content volume). Both public and private account holders belong to this category. In our collection, these accounts usually belong to news media. Note that we put only official accounts[88] in this category: if a bot features the same news feed as a news agency, it would be classified as a bot, not as an official account.

If an account belongs neither to spam nor to the category of official institutional accounts, the goal was to classify it most reliably as a human or a bot. We proceeded by developing a comprehensive and empirically motivated classification scheme that would allow assigning unambiguously each account of interest to one of these categories.

First, we checked how diverse the account content is. Some accounts feature only one specific type of content, thus indicating bot activity. It could be just few hundred or millions of tweets, but all of them are identical in their form. For example, this could be easily identifiable **news headlines** ("President to visit China Tuesday"), posted one after another without tweets of any other kind. Alternatively, it could be only **retweets** from other Twitter accounts, or just **pictures**, or just **videos**. In this case, the substantive content of the tweets does not matter: pictures could be from tennis events, videos from theater performances in Saint-Petersburg and news headlines about agriculture in the U.S. What matters here is that these are not tweets written by a human being, but posted by a robot, likely querying the web or feeding content from a pre-defined RSS feed. The only exception to this rule are retweets: they can feature different content, produced by both bots and humans. Here the decisive factor is the absence of content actually posted by the user in question, as well as the consistency and volume of retweets: few humans maintain Twitter accounts featuring exclusively retweets from others, but even those who do could rarely demonstrate the dedication to the task exhibited by bots.

News headlines are further distinguished into those **containing a link** to a news website (it could be both the real source of news, like a major newspaper, or a makeshift website that simply republishes content taken from elsewhere) and those **without any links**.

The last group that belongs to this category are accounts posting **other text** that does not come from the news headlines, but clearly was not generated by the account holder either. For instance, feeds entirely consisting of famous quotes from historical figures belong to this category.

The most challenging classification task arises in cases when the account features **diverse** content. The first group in this category are accounts that may contain retweets, pictures, videos and text, but clearly do not contain anything that is not available elsewhere on the web and thus is not personally produced by the account holder. These could be pictures from a news agency, alternating with retweets, alternating with links to news stories or unidentified quotes from other blogs and Twitter feeds. If in doubt about the latter, coders were instructed to google a chunk of text from a randomly chosen tweet to check if it was available elsewhere on the Internet.

All of the account types we mentioned above usually exhibit a level of content consistency so high that one can spot bot activity even without paying close attention to the content of the tweets: multiple links that lead to the same news website; pictures of the identical size posted one after another and never featuring a person unfamiliar to the general public, etc. Naturally, the first encounter with human content comes when the consistent pattern is broken.

For instance, take a Twitter feed consisting of many retweets featuring similar content, such as pictures of Ukrainian soldiers allegedly killing civilians in the most brutal way possible. But occasionally something different is popping up: a reply to another user, wishing her a nice morning, or lamenting about bad weather. Googling this content confirms that it could not be found, at least easily, elsewhere. If this kind of content is sparsely dispersed between many similar tweets clearly lifted from elsewhere, we are likely dealing with a bot manually maintained by (a team of) supervisors, who occasionally tweet a real reply to another user (who could be both a real person and another bot). We classify this account as a **cyborg**. We believe that this tactic is used to avoid Twitter spam filters as well as to increase the so-called account reputation (a computed characteristic of a Twitter account that might increase its visibility inside and outside the Twitter network).

Importantly, we clearly distinguish between trolls and cyborgs. Trolls are human beings dedicated (because of their persuasion, for money or out of fun) to spreading a particular kind of message online. We classify them, together with perfectly "normal" individual users into **personal accounts**. A different kind of human account is a feed featuring links to a user's online presence elsewhere. For example, one might want to post links to her Instagram pictures or Facebook posts. We call such accounts **transmitters**. Finally, sometimes a small group of people maintain a community, for example, with local news or dedicated to a particular artist. It is a rare case on Twitter, but for such cases we keep the **community** category.

Therefore, the coding schema for our verification effort is as follows (terminal categories are in **bold**):

1. Official institutional accounts
2. Bots
  a. Single-content
    i. Retweets
    ii. Pictures
    iii. Videos
    iv. Text
      A. News headlines
        - **Text with links**
        - **Text without links**
      B. Other text
  b. Diverse content
3. Cyborgs
4. Humans
  a. Transmitters
  b. Personal accounts
  c. **Communities**
5. Spam

APPENDIX NOTES

1. Andrei Shleifer and Daniel Treisman, *Without a Map: Political Tactics and Economic Reform in Russia* (Cambridge: MIT Press, 2001).

2. Egor Gaidar, *The Economics of Transition* (Cambridge: MIT Press, 2003).

3. Irina Denisova, Markus Eller, and Ekaterina Zhuravskaya, "What Do Russians Think About Transition?" *Economics of Transition*, 18, (April, 2010): 249–80.

4. Zasursky.

5. Maria Lipman, "Media Manipulation and Political Control in Russia," Chatham House Russia and Eurasia Programme Publication No. 09/01, January, 2009, available at http://www.academia.edu/download/31134451/300109lipman.pdf.

6. Olessia Koltsova, *News Media and Power in Russia* (London: Routledge, 2006).

7. Vladimir Gel'man, Dmitry Travin, and Otar Marganiya, *Reexamining Economic and Political Reforms in Russia, 1985–2000: Generations, Ideas, and Changes* (London: Lexington Books, 2014), 104–108.

8. Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya, "Media and Political Persuasion: Evidence from Russia," *American Economic Review*, 101 (December 2011): 3253–85.

9. Masha Gessen, *The Man Without a Face: The Unlikely Rise of Vladimir Putin* (New York: Penguin, 2012), Chapter 2.

10. Ben Judah, *Fragile Empire: How Russia Fell in and out of Love with Vladimir Putin* (New Haven: Yale University Press, 2013), Chapter 2.

11. Yelena Tregubova, *The Tales of a Kremlin Digger* (Moscow: Ad Marginem, 2003), Chapter 10. (in Russian).

12. Lipman; Tina Burrett, *Television and Presidential Power in Putin's Russia* (London: Routledge, 2010).

13. According to Adam Przeworski, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi, *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990* (Cambridge University Press, 2000) the average dictatorship which was overthrown between 1950 and 1990 lasted 27.4 years and average dictatorship still in course in 1990 was 26.2 years old.

14. Ann Komaromi, "The Material Existence of Soviet Samizdat," *Slavic Review*, 63 (September 2004): 597–618. The last among these regulations, which was not enforced anymore, was struck down by the Russian Supreme Court only in 2009.

15. Arthur S. Banks and Kenneth A. Wilson, *Cross-National Time-Series Data Archive* (Jerusalem, Israel.: Databanks International, 2013), available at https://www.cntsdata.com/.

16. Andrei Soldatov and Irina Borogan, *The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries* (PublicAffairs, 2015), 5-7, 180-185.

17. Sophisticated "deep pocket" web surveillance system known as SORM (-2,-3) was installed by the Russian government no later than in the late 1990s and has been being updated constantly ever since, see Andrei Soldatov and Irina Borogan, "Russia's Surveillance State," *World Policy Journal*, 30 (September, 2013): 23–30.

18. Scott Gehlbach, "Reflections on Putin and the Media," *Post-Soviet Affairs*, 26 (January 2010): 77–87.

19. Lipman; Helen Womack, "Making Waves: Russian Radio Station Is Last Bastion of Free Media," *Index on Censorship*, 43 (September, 2014): 39–41.

20. Jonathan Becker, "Lessons from Russia: A Neo-Authoritarian Media System," *European Journal of Communication*, 19 (June, 2004): 139–63.

21. See academic studies by Burrett, 2010; John A. Dunn, "Where Did It All Go Wrong? Russian Television in the Putin Era," in *The Post-Soviet Russian Media: Conflicting Signals*, eds. Birgit Beumers, Stephen Hutchings, and Natalia Rulyova (Routledge, 2008), 42–55; Tina Burrett, "The End of Independent Television? Elite Conflict and the Reconstructing the Russian Television Landscape," in *The Post-Soviet Russian Media: Conflicting Signals*, eds. Birgit Beumers, Stephen Hutchings, and Natalia Rulyova (Routledge, 2008), 71–86; for more informal discussion see Gessen, Chapter 7 and an illuminating personal memoir by Tregubova, Chapters 11-13.

22. Etling et al. 2010, 33.

23. Karina Alexanyan, Vladimir Barash, Bruce Etling, Robert Faris, Urs Gasser, John Kelly, John G. Palfrey, and Hal Roberts, "Exploring Russian Cyberspace: Digitally-Mediated Collective Action and the Networked Public Sphere," Berkman Center Research Publication No. 2, March 2, 2012, available at http://papers.ssrn.com/abstract=2014998, 10–11.

24. Anton Nossik, "I Helped Build Russia's Internet. Now Putin Wants to Destroy It," *New Republic*, May 15, 2014, available at http://www.newrepublic.com/article/117771/putins-internet-crackdown-russias-first-blogger-reacts.

25. Still, VKontakte (but not Yandex or Odnoklassniki) had significantly benefited from the lax enforcement of the property rights. However, this doesn't make comparison with China less impressive, given that China is also famous for wide-spread piracy.

26. Robert Greenall, "LiveJournal: Russia's Unlikely Internet Giant," *BBC News*, February 29, 2012, available at http://www.bbc.co.uk/news/magazine-17177053.

27. "Facebook has more than a billion users now," *RIA Novosti*, October 4, 2012, available at http://ria.ru/technology/20121004/766127348.html.

28. "Number of Russians on Twitter doubled in the last six months," *Digit.ru*, October 31, 2014, available at https://web.archive.org/web/20150608005425/http://digit.ru/internet/20131031/407481403.html#.

29. Ashwin Seshagiri, "The Languages of Twitter Users," New York Times Bits Blog, March 9, 2014, available at http://bits.blogs.nytimes.com/2014/03/09/the-languages-of-twitter-users/.

30. Nossik, 2014.

31. Alexanyan et al., 10.

32. Charles Clover, "Internet Subverts Russian TV's Message," *Financial Times*, December 1, 2011, available at https://www.ft.com/content/85dd8e96-1c2d-11e1-9631-00144feabdc0.

33. David Remnick, "Putin's Television," *The New Yorker Blogs*, December 9, 2011, available at http://www.newyorker.com/online/blogs/newsdesk/2011/12/putins-television.html.

34. Nossik, 2014.

35. FOM ‘Internet in Russia’ bulletin No. 33, Spring, 2011, available at http://bd.fom.ru/report/map/projects/internet/internet1133/vesna2011.

36. “Internet in Russia; Penetration Dynamic, Winter 2014-14,” FOM.RU, April 1, 2014, available at http://fom.ru/SMI-i-internet/11417.

37. Sarah Oates and Tetyana Lokot, “Twilight of the Gods?: How the Internet Challenged Russian Television News Frames in the Winter Protests of 2011-12,” SSRN Working Paper, July 2013, available at http://papers.ssrn.com/abstract=2286727.

38. Sakwa, Chapters 3-5.

39. See and overview of Medvedev’s reforms in J. L. Black, *The Russian Presidency of Dimitri Medvedev, 2008-2012: The Next Step Forward or Merely a Time Out?* (London: Routledge, 2014); J. L. Black and Michael Johns, *Russia after 2012: From Putin to Medvedev to Putin – Continuity, Change, or Revolution?* (London: Routledge, 2013).

40. Dmitry Yagodin, “Blog Medvedev: Aiming for Public Consent,” *Europe-Asia Studies*, 64 (October, 2012): 1415–34.

41. See, for example, Darrell West, “President Dmitry Medvedev: Russia’s Blogger-in-Chief,” *The Brookings Institution*, April 14, 2010, available at http://www.brookings.edu/research/opinions/2010/04/14-medvedev-west.

42. Etling et al. 2010, 3.

43. Etling et al. 2010, 33.

44. Henry Hale, “Democracy or Autocracy on the March? The Colored Revolutions as Normal Dynamics of Patronal Presidentialism,” *Communist and Post-Communist Studies*, 39 (September 2006): 305–29.

45. Etling et al. 2010, 3.

46. Kelly et al. 2012, 11.

47. Etling et al. 2010, 19.

48. Barash and Kelly 2012.

49. Etling et al. 2014, 37-45.

50. Freedom House, 2011; Agora, *Internet Freedom in Russia 2011* (Kazan, Russia: Association of Human Rights Organizations “Agora,” 2011), available at http://openinform.ru/fs/j_photos/openinform_353.pdf.

51. Elinor Mills, “Twitter, Facebook Attack Targeted One User,” *CNET*, August 6, 2009, available at http://www.cnet.com/news/twitter-facebook-attack-targeted-one-user/.

52. Hal Roberts and Bruce Etling, “Coordinated DDoS Attack During Russian Duma Elections,” *Internet & Democracy Blog*, December 8, 2011, available at http://blogs.law.harvard.edu/idblog/2011/12/08/coordinated-ddos-attack-during-russian-duma-elections/.

53. “Kremlin web site under DDoS attack,” *Habrahabr.ru*, May 9, 2012, available at http://habrahabr.ru/post/143501/; “RIA-Novosti website under DDoS attack,” *Lenta.ru*, May 10, 2012, available at: http://lenta.ru/news/2012/05/10/attack/.

54. Alexander Morozov, “‘Moskovskie Novosti’s Space,” *OpenSpace.Ru*, March 28, 2011, available at http://os.colta.ru/media/projects/18065/details/21397/; Olga Barykova and Maria Zotova, “Ushlo li vremya ‘Moskovskih Novostey’?,” *BBC Russian*, April 1, 2011, available at http://www.bbc.co.uk/russian/russia/2011/04/110331_moscow_news_scandal.shtml; Henry Meyer, “Putin Revives Gorbachev Glasnost Paper to Widen Election Appeal,” *Bloomberg*, March 30, 2011, available at http://www.bloomberg.com/news/2011-03-30/putin-revives-gorbachev-glasnost-paper-to-widen-election-appeal.html.

55. As mentioned, execution of government projects in media suffers from general government inefficiency and the TV channel went on air only in 2013, long after Medvedev had switched offices with Putin.

56. Dr Greg Simons, *Mass Media and Modern Warfare: Reporting on the Russian War on Terrorism* (Farnham: Ashgate Publishing, 2013).

57. Natalia Yudina, “RuNet, Hate Crime and Soft Targets: How Russia Enforces Its Anti-Extremism Law,” *Open Democracy*, October 30, 2012, available at http://www.opendemocracy.net/od-russia/natalia-yudina/runet-hate-crime-and-soft-targets-how-russia-enforces-its-anti-extremism-la.

58. Judah.

59. Post-election analysis revealed that suspicion was well-grounded, see Ruben Enikolopov, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov, “Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections,” *Proceedings of the National Academy of Sciences*, 110 (January, 2013): 448–52; Dmitry Kobak, Sergey Shpilkin, and Maxim S. Pshenichnikov, “Statistical Anomalies in 2011-2012 Russian Elections Revealed by 2D Correlation Analysis,” ArXiv preprint, May, 2012, available at http://arxiv.org/abs/1205.0741.

60. Sakwa, 129-132.

61. Sakwa, Chapters 7-8.

62. Julia Ioffe, "Vladimir the Unstable," *Foreign Policy*, May 7, 2012, available at https://foreignpolicy.com/2012/05/07/vladimir-the-unstable/.

63. Freedom House, 2012, 2013.

64. Available at http://eais.rkn.gov.ru/.

65. Text of the law is available at http://www.rg.ru/2013/12/30/extrem-site-dok.html. See analysis by Human Rights Watch 2014.

66. Alena Sivkova, "We Don't See Much Risk in Blocking Twitter in Russia," *Izvestia*, May 16, 2014, available at http://izvestia.ru/news/570863.

67. Available at https://www.facebook.com/Dmitry.Medvedev/posts/10152047885946851.

68. Available at https://vk.com/blank.php?rkn=32274605. It should be noted that according to those "appropriate regulations" authorities could not be notified about the upcoming rally earlier than 15 days in advance. The page was blocked 26 days before the event it announced was scheduled to take place.

69. Available at https://www.facebook.com/events/417200101767938.

70. Available at https://www.facebook.com/events/406603402849183.

71. Agora, *Internet Freedom in Russia 2012* (Kazan, Russia: Association of Human Rights Organizations "Agora," 2012), available at http://www.hro.org/node/15685; Agora, *Internet Freedom in Russia 2013* (Kazan, Russia: Association of Human Rights Organizations "Agora," 2013), available at http://eliberator.ru/files/Internet_2013.pdf.

72. According to the survey reported by Alexanyan et al., even without any legal requirement Russian bloggers rarely conceal their identity. They do use pseudonyms (following internet tradition), but usually alongside, not instead of their real names. This is particularly true for politically-engaged bloggers.

73. Kevin Rothrock, "Meet Russia's 369 Kremlin-Registered Bloggers," *Global Voices*, January 8, 2015, available at http://globalvoicesonline.org/2015/01/08/meet-russias-369-kremlin-registered-bloggers/.

74. Ingrid Lunden, "Intel Shuts Down Russian Developer Forums to Comply with Russia's 'Blogger Law'," *TechCrunch*, January 5, 2015, available at http://techcrunch.com/2015/01/05/intel-shuts-down-russian-developer-forums-to-comply-with-russias-blogger-law/.

75. Available at http://rkn.gov.ru/docs/prikaz_Roskomnadzora_ot_09.07.2014_N_99.pdf.

76. Olga Razumovskaya, "Russian Social Network: FSB Asked It to Block Kremlin Protesters," *Wall Street Journal*, December 8, 2011, available at http://blogs.wsj.com/emergingeurope/2011/12/08/russian-social-network-fsb-asked-it-to-block-kremlin-protesters/.

77. Available at https://vk.com/wall145621.

78. Brian Ries, "Founder of 'Russia's Facebook' Says Government Demanded Ukraine Protestors' Data," *Mashable*, April 16, 2014, available at http://mashable.com/2014/04/16/vkontakte-founder-fsb-euromaidan/; Kononov; Lunden 2014.

79. Leonid Bershidsky, "Google's Retreat from Moscow," *BloombergView.Com*, December 12, 2014, available at http://www.bloombergview.com/articles/2014-12-12/googles-retreat-from-moscow.

80. Sakwa.

81. Regina Smyth and Irina Soboleva, "Looking beyond the Economy: Pussy Riot and the Kremlin's Voting Coalition," *Post-Soviet Affairs*, 30 (July, 2014): 257–75; Timothy Snyder, "Fascism, Russia, and Ukraine," *The New York Review of Books*, March 20, 2014, available at http://www.nybooks.com/articles/archives/2014/mar/20/fascism-russia-and-ukraine/.

82. Since we are computing the entropy of *inter*-tweeting time intervals, we need at least two intervals to compute a meaningful entropy value, or, in other words, three tweets.

83. Thus, these thresholds should be considered simply a first cut in developing bot-detection methods: we check whether we can reliably identify bots if we set the thresholds at their most extreme levels keeping the sample size for coding reasonably large but feasible for hand-coding (roughly 100 accounts per method). In future work, we intend to use machine learning tools to empirically identify both the threshold levels that distinguish bots from humans and those account characteristics that are most informative for bot detection.

84. If an account had more than 100 tweets in our collection, we used the most recent 100 tweets.

85. We suspect that there could be two explanations for this pattern. First, it is possible that tweeting in this case might not be triggered by the news headline appearing in the news agency feed, and therefore the bot designer would have to specify the frequency of posting as part of the code governing the bot's behavior. Alternatively, accounts that retweet everything from many other accounts may simply get to tweet more often than a typical news agency comes up with a new story.

86. Technically, the Jaccard similarity coefficient J(A,B) between sets A and B is $J(A,B) = |A \cap B| / |A \cup B|$ where $|A \cap B|$ is the number of elements that A and B have in common, and $|A \cup B|$ is the total number of elements in the two sets. Table C2 presents Jaccard similarities for sets of 50 most common hashtags in different groups of bots by political orientation.

87. This disparity in cyber-war effort could be said to mirror the changes on the ground, as Kremlin was no longer interested in the conflict and, instead of taking advantage of escalation to stage an offensive, put just enough resources to coerce Ukrainians into the Minsk agreement, which essentially preserved the status-quo and frozen the conflict based on the Transnistria model. See Cristian Ghinea,"Transnistria and Eastern Ukraine – Any Similarities?," *Emerging Europe*, February 11, 2015, available at http://emerging-europe.com/voices/voices-economy/ghinea-transnistria-and-eastern-ukraine-any-similarities/.)

88. But not necessarily verified as such by Twitter itself.