Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests [*]

Viktoria Spaiser[†], Thomas Chadefaux[‡], Karsten Donnay[§], Fabian Russmann[¶]
Dirk Helbing[‖]

In 2011-2012 Russia experienced a wave of mass protests surrounding the Duma and presidential elections. The protests, however, faded shortly after the second election. We study the Russian political discourse on Twitter during this period and the main actors involved: the pro-government camp, the opposition and the general public. We analyse around 700.000 Twitter messages and investigate the social networks of the most active Twitter users. Our analysis shows that pro-government users employed a variety of communication strategies to shift the political discourse and marginalise oppositional voices on Twitter. This demonstrates how authorities can disempower regime critics and successfully manipulate public opinion on social media.

Keywords: Communication Power, Social Media, Twitter, Political Discourse, Russia, Natural Language Processing, Protest

Social media has played an increasing role in domestic and international politics, in particular in the context of social movements, demonstrations and protests (Howard & Parks, 2012). The Arab Spring, for example, is often referred to as the "Twitter Revolution," in that social media contributed to the political debate and the dissemination of the movements' message

[*]We would like to thank Michael Mäs for very helpful feedback and advice and Irina Vartunova for extremely valuable insights on Russian politics. This work has benefited from the ERC Advanced Investigator Grant "Momentum" (Grant number 324247), and the ETH project "Systemic Risks, Systemic Solutions" (CHIRP II project ETH 48 12-1)

[†]Address correspondence to: Viktoria Spaiser, School of Politics and International Studies, University of Leeds, LS2 9JT Leeds, UK (E-mail: v.spaiser@leeds.ac.uk). Viktoria Spaiser is University Academic Fellow in Political Science Informatics at the School of Politics and International Studies, University of Leeds, UK. Her research explores computational social science approaches in the field of political science with a focus on democratisation, sustainable development, and political participation.

[‡]Thomas Chadefaux is Assistant Professor of Political Science at Trinity College Dublin, Ireland. His research explores the causes of interstate conflicts and their prediction making use of large amounts of fine-grained spatial and temporal data.

[§]Karsten Donnay is Juniorprofessor of Computational Social Science in the Department of Politics and Public Administration at the University of Konstanz, Germany. In his research he develops and refines advanced quantitative methodologies to study urban violence, crime, conflict and terrorism.

[¶]Fabian Russman was a research assistant at the Chair of Computational Social Science, ETH Zurich, Switzerland. Currently he works as Data Analytics Consultant at DIONE, Zurich, Switzerland.

[‖]Dirk Helbing is Professor of Computational Social Science at the ETH Zurich, Switzerland. He is internationally known for his work on pedestrian crowds, vehicle traffic, agent-based models of social systems and big data approaches.

across the world, and helped participants coordinate and share information (Cottle, 2011; Howard et al., 2011; Lotan et al., 2011; Tufekci & Wilson, 2012). What sets these new media apart from more traditional media is that they enable private citizens to communicate on a large scale and in real time and may therefore especially benefit oppositional actors without strong institutional support and backing by traditional media outlets (Shirky, 2011; Lynch, 2011; Diamond, 2010). In autocracies in particular, social media is often perceived as a means by which the disenfranchised can express their discontent, since they are considered to be one of the few uncensored public spaces in which reliable information sharing and free political communication can take place. In other words, social media is often perceived as liberative.

Yet much less attention has been paid to the idea that social media may also be used as an instrument of oppression. As a tool that allows actors to widely disseminate information, it may not be that different from traditional media such as TV or radio, which have long been recognised as potential instruments of control and coercion (Herman, 1985; Silitski, 2005; Thompson, 2007; Enikolopov et al., 2011). Governments—not only opposition movements—can use these technologies to their advantage to spread their message, influence audiences, and change the perception of those who might be tempted to challenge their legitimacy. Indeed, oppositional challenges not only need to emerge, but also to remain strong and united over time. Social media can help achieving that goal, as the Arab Spring made clear (Cottle, 2011; Howard et al., 2011; Lotan et al., 2011; Tufekci & Wilson, 2012). But at the same time, social media can also be used by the governing elite against the opposition, through defamation, discrediting and counter-mobilisation. In this study, we focus on political communication strategies that Russian pro-government and oppositional groups used to advance their causes, mobilise their supporters and discredit their opponents on Twitter. We in particular investigate whether and how the pro-government camp employed a variety of communication strategies to shift the political discourse, marginalise oppositional voices and successfully manipulate public opinion on Twitter.

Social media, and Twitter in particular, indeed played a prominent role during the last two Russian elections—the Duma (lower legislative house of the Russian Federation parliament) elections of December 4, 2011 and the presidential elections of March 4, 2012—as well as the protests that took place during that period (Greene, 2013). People tweeted election results from their local polling stations; posted links to videos and pictures documenting electoral fraud and arrests of prominent oppositional figures such as Alexey Navalny (see Supplementary Information S4 for explanations of terms and names); and linked information about upcoming and past protest events. Twitter was particularly important because many prominent oppositional websites were taken down or hacked during and after the elections of December 2011 (Roberts & Etling, 2011). This left Twitter as one of the few platforms that was not targeted by Distributed Denial of Service attacks, although oppositional hashtags were flooded with pro-regime spam (Krebs, 2011; Kelly et al., 2012).

Twitter is certainly just a part of a larger media system that intersects with the wider political system (Chadwick, 2013). Indeed, it would go beyond the scope of this paper to try to take into account the full Russian media ecology (see Becker (2004), Lipman (2005) and Arutunyan (2009) for further information on the Russian media system). However, analysing Twitter communication as an important part of the larger media system is not only relevant for understanding political discourse in social media but also provides insights for the broader

Russian political communication context. Digital social spheres, such as the Twittersphere, mirror 'real-world' events and traditional media discourses, and hence can serve as a basis for studying the communication and interaction mechanisms between different political fractions and the wider media discourse—especially when information would otherwise be unavailable.

For social science research the popularity of social media for political communication and discourse is extremely useful, as it creates new opportunities to analyse real-time social network and political opinion formation on a large scale (Conover et al., 2011b; Tumasjan et al., 2010). Here, we examine the discourse in the Russian Twittersphere during the two Russian elections and mass protests in 2011–2012 by analysing nearly 700,000 public tweets. The uniquely fine-grained data on political discourse and affiliations over time collected from Twitter provide a unique and powerful case study for political communication on social media channels. The vast amount of text provided by Twitter was analysed with a new mixed-method approach for dynamic discourse analysis, combining methods of statistical natural language processing with context- and theory-based interpretation and social network analysis. We rely on ngrams to systematically analyse communication strategies used by both the pro-government and oppositional camp. Using a sentiment-based classification procedure we then identify pro-Putin and oppositional Twitter users/tweets. This allows us to study both the social networks of the political camps on Twitter and to follow the evolution of the political discourse within each camp over time to uncover their respective communication strategies.

Our analysis shows that an active pro-Putin campaign between the two elections decisively contributed to changing the momentum of the discourse on Twitter with the initially large and strong political opposition rapidly losing control of the discourse by the time of the March 2012 presidential elections. Our results thus cast doubt on the assertion that traditional powers are necessarily disadvantaged in an increasingly networked and digitalised society. As governments use these new technologies as means for mobilisation of their supporters and repression of oppositional voices, the balance of power on social media need not necessarily favour the opposition. In fact, our results suggest that the pro-Putin camp was very successful in regaining control over a means of communication that initially seemed particularly favourable to the opposition. These results confirm recent, more critical analyses of social media in autocratic regimes, which show that autocratic governments have increasingly adopted strategies of proactively subverting and co-opting social media for pro-regime purposes (Rød & Weidmann, 2015; Gunitsky, 2015).

Mobilization, Perceptions and the Success of Political Movements

Mass collective actions such as protests or rebellions take place when the discontent population sees a window of opportunity. Activism typically originates from a small number of radicals, then extends to a wider circle of motivated individuals, before spreading through the rest of the population (Tilly, 1978). The process can be understood as a series of crossed threshold. First the radicals mobilise. So-far inactive individuals with a higher threshold for mobilisation observe them and also mobilise as a result. In turn, their mobilisation reaches a threshold sufficient to engage others that are motivated by the size of the existing movement, and so on and so forth. Models by Granovetter (1978) and Schelling (1978) formalized this intuition, later extended by Kuran (1989), Gould (1993), Lohmann (1994), and Siegel (2009). Individual radical instigators sometimes succeed in starting a "prairie fire" (Kuran, 1989), which progressively leads others with more conservative risk preferences to follow suit.

Whether a cascade occurs, therefore, critically depends on beliefs about the probability of success, and hence about existing levels of mobilisation. Without knowledge that the radicals have mobilised, the wider circle would not mobilise by itself. And the general population needs to be informed that a wide number of individuals have already joined. This sequence is critical and explains why demonstration leaders often overstate their numbers, whereas governments seek to downplay them. Crossing certain mobilisation thresholds—and making it clear that these thresholds have been crossed—is essential to further recruitment and hence to the ultimate success of the movement.

Affecting the perception of the turnout level is therefore essential. Information on the mobilisation level is usually gathered from the media. Yet, in authoritarian regimes such as Russia, where the media is highly controlled by the government (Becker, 2004; Lipman, 2005; Arutunyan, 2009), people have learned not to rely on that information. A growing alternative source of information is social media. People in social media belong to a network and learn about the popularity of the movement from the network nodes they are connected to: friends, colleagues, peers, persons of interest, public figures but also institutions and established and alternative media who have social media accounts.

Since the perception of a political movement's success is key for a sustained and expanding mobilisation, the government's and opposition's ability to shape that perception on social media such as Twitter can be of great importance in determining the course of events. Here we show that both sides strategically used different political communication strategies on Twitter. Our analysis suggests that, in particular, the Russian government successfully used Twitter to affect population's perception of the oppositional movement's success and legitimacy.

Effectively challenging an opposition movement is a critical prerequisite to preventing any revolutionary spark from starting a 'prairie fire', or at least to prevent any further expansion and/or consolidation of the movement. By shifting the perceived balance of popular support and legitimacy towards the government and away from the opposition movement, the central government can shape the perception of success and legitimacy, and hence affect mobilisation levels. Indeed, if the balance of power and popular support is seen to be favouring the government, then only those with a relatively high level of political conviction and commitment will mobilise. In turn, this can start a downward cascade until only the most radical elements are mobilised. In short, affecting the *perception* of the movement's success can lead to an endogenously generated effect. In that sense, new media can enhance state capacity.

How mass communication technology (TV, Radio, Newspaper, Internet) can strengthen the state's capacity to disseminate political messages and as a result prevent large-scale oppositional mobilisation has been shown by Warren (2014) and Weidmann et al. (2016). Whoever controls the media and more generally the diffusion of information also influences opinions and contributes to setting political agendas. Our paper contributes to this line of works in two ways. First, we focus on social media (i.e. Twitter) and analyse to what extent they may contribute to strengthening state's ability to affect public perceptions. New Internet-based media have significantly affected traditional communication mechanisms (Bennett & Segerberg, 2013; Chadwick, 2013). In particular, social media such as Facebook and Twitter have the ability to quickly distribute information, enabling communication on a large scale and in real time, potentially sparking information cascades and the diffusing and scaling up of local protests. Therefore, social media increasingly become platforms and channels for both government and opposition campaigns (Lynch, 2011; Rød & Weidmann, 2015). Our data—who 'tweets' what and when—allow us to study the actions and reactions of all parties over time and in response to one another, with great accuracy. This enables us to track attempts to affect popular perceptions and their relative success.

Second, Warren (2014) argues that his findings about the centralised systems of mass communication may not apply to "the internet, cell phones, and other forms of 'social' media, which instead facilitate decentralised horizontal connections between individuals" (Warren, 2014, p. 136). Though this proposition has been challenged very recently by Weidmann et al. (2016), much of the interest in policy-making circles and in academia has been in the potentially liberating effect of these new forms of decentralised communication. In contrast, our analysis illustrates the ability of governments to harness these technologies. Embracing decentralisation they at the same time attempt, at least to some extent, to centralise those new media activities supporting the state.

In particular, the government may manipulate social media in a number of ways to influence the perception of an oppositional movement's dynamics and probability of success, which are critical for the movement's evolution, promoting downward spirals in mobilisations. Castells (2007, 2009) distinguishes four main ways in which Internet communication can act on people's minds and thus be used as a strategic tool in struggles for power. First, the Internet facilitates the manipulation of emotions and perceptions (*framing*) (Kramer et al., 2014). This can include diminishing, discrediting but also exaggerating, enthusing, and claiming broad public support. Indeed, the spread of manipulative information was probably never as rapid and easy as in the age of the Internet (Castells, 2009; Slove, 2007).

Second, the Internet facilitates propaganda campaigns: affecting the way in which individuals evaluate political concepts and ideas but also political figures (*priming*). This can include priming the criteria, agendas and images on which citizens base their political decisions, for instance in elections (Druckman, 2004; Domke, 2001; Roskos-Ewoldsen et al., 2011).

Third, social media change the set of people who can contribute to setting the political agenda (*agenda-setting*) and the terms of the debate. This may range from publishing certain information that would otherwise not be revealed or offering counterarguments to publicising certain political events. Social media like Twitter enable even marginalised political actors to define agendas (Drezner & Farrell, 2004; Benkler, 2006).

Finally, censorship (*indexing*) limits the range of political opinions and agendas (Castells, 2009). Censorship may go as far as cutting all access to communication networks, as witnessed for instance in Egypt (Williams, 2011). Hacking opponents' websites and disrupting their communication channels is an even more common way of censorship and was used in Russia during the protest events in the wake of the two last elections (Roberts & Etling, 2011). Online surveillance may also result in self-censorship, as people lose control over who has access to their online communication or to their private data collected on the Internet

(Castells, 2009; Bitso et al., 2012).

Data

Our analysis is based on data from the Twitter Streaming API collected between November 17, 2011 and March 12, 2012.[1] This encompasses the two Russian elections: the Duma election of December 4, 2011, and the presidential election of March 4, 2012. The collected tweets were filtered, first for Russian language, and second for political content by using various Russian keywords that broadly refer to political issues, such as 'news', 'protest', 'politics', or 'elections' (see full list of keywords in Supplementary Information S1.2). The subset of Twitter data used in our analysis then comprised 690,297 Russian language tweets with political content.

With the rising attention that social media have received in social and political research as noted in the previous section, social media data and in particular Twitter data has been increasingly used to understand various social and political phenomena.(Miller, 2011; Golder & Macy, 2011; Tonkin et al., 2012). Twitter data was for instance used to understand and predict election outcomes (Tumasjan et al., 2010; Wu et al., 2011; Larsson & Moe, 2011), political alignment (Conover et al., 2011a; Hanna et al., 2011) or shed light on the communication and recruitment strategies of political groups (Gaffney, 2010; Yardi & Boyd, 2010; Ratkiewicz et al., 2011; Conover et al., 2011b; Gonzáles-Bailón et al., 2011).

There is, however, little topic- or region-specific research on the Russian 'Twittersphere', even though by 2011 Twitter had become an increasingly important means of public communication in Russia (Kelly et al., 2012).[2] From only about 1,000 Russian Twitter users in 2007, their numbers had soared to over 3.8 millions in April 2012 (Oates, 2013).[3] While other popular Russian Social Media such as Vkontakte (Russian version of Facebook) existed in our period of analysis, they did not exhibit the same publicness in debates and are therefore less suitable for studying public debates.

Any analysis of Twitter data faces a number of well-known difficulties (Ruths & Pfeffer, 2014). First, the sample only includes public tweets from public Twitter accounts. This does not pose a problem in the context of our study though since we are interested in the use of Twitter as an instrument of communication in the public sphere. Potentially more problematic is the fact that Twitter has implemented a quality filter that filters out a small amount of tweets if they are considered to be spam or of too low quality. Unfortunately, neither the frequency of this filtering nor its exact criteria are entirely transparent (see also Supplementary Information S1.1). Despite this filtering practice, inspection of our extracted data revealed that at least 18% of the tweets were 'spam' such as automatically generated advertisements. To minimise biases in our results, we applied an additional filter to detect and remove messages using keywords related to advertisements and spam (see Supplementary Information S1.2). Note that the filtered data—now comprising 601,138 tweets—still contains some spam and advertisements that were not picked up by the filtering algorithm, but with a significantly reduced prevalence (about 5-7%).

Finally, discourse analysis faces specific difficulties when working with Twitter data. Tweets are short and thus contain only limited information. In fact, because they are limited to 140 characters, users tend to convey only part of the information directly—on average, 19% of all tweets contain links to webpages with further information (Zarrella, 2009). Despite these limitations, the short Twitter messages still allow for political discourse. And how this discourse is framed or what the actors' overall agendas and aspirations are, develops alongside the broader societal discourse. Moreover, even though Twitter users are generally not a representative sample of the overall population, almost all political groups were represented (with their respective supporters) in the Russian Twittersphere in our period of analysis.[4]

Not surprisingly, the amount of political tweets per day in our sample varies strongly—between 2,204 tweets on November 18, 2011, and 12,428 tweets on the day of the presidential election, March 4, 2012 (Median = 5,031; Mean = 5,118, SD = 1,504). In fact, the two elections are responsible for the two major peaks in the number of daily political tweets in the time period analysed. But as we show further below the relative activity of different factions on Twitter remains comparably stable over time throughout our period of analysis (see also Supplementary Information Figure S5). To relate the analysis of tweets to the timeline of the protest movement, we also collected detailed information on the election and protest events for the period of time represented in the sample. Data on political events were retrieved from various online sources[5] and compiled in a political events dataset, with information on political event type (e.g. rally, political action), time, place, involved political groups, size (e.g. number of demonstrators) and repression extent if any (e.g. number of arrests).

Methodology

Twitter data have to date only rarely been used for discourse analysis,[6] despite Twitter's potentially rich and authentic coverage of the political discourse. In fact, only few studies have analysed Twitter data beyond word counts or binary sentiment analysis. A notable exception is Wu et al. (2011), who uses a semantic network approach applied to political discourse to understand its social impact on the formation of political attitudes. Sentiment analyses are often criticised for failing to account for the complexity and contextuality of human communication, which would require, for example, to take into account the ambiguity of sentiment terms (Weichselbraun et al., 2010; Wilson et al., 2009). Moreover, Twitter users often communicate their messages through irony, sarcasm or symbols – communication means that are hard to detect by automated text processing.

In this study, we used two main text mining techniques: word counts and their temporal evolution (see Supplementary Information Figure S3 and S4), and dynamic "meme" or ngram analyses based on bi- and trigram collocation (see Supplementary Information S2.1 and Figure S2). We detected collocations of words using the association and scoring function Student's t-test (Manning & Schuetze, 1999; Perkins, 2010). The Student's t-test assesses whether two or three words co-occur more frequently than by chance. The null hypothesis is the absence of association between the two or three words beyond coincidental co-occurrence, i.e., that the words are independent. p is the corresponding probability for the non-systematic co-occurrence of two or three words. The null hypothesis is thus rejected if p is very small ($p < 0.01$ or $p < 0.05$). Maximum likelihood estimation was used to compute the likelihood that word A and word B (and word C) co-occur in the analysed text (see Supplementary Information S2.1 for further details). The Student's t-test statistic was used as bigram

(BAS) or trigram (TAS) association score (Perkins, 2010). These scores reflect the frequency of the collocations. The t-test is particularly useful to rank collocations to identify the most dominant collocations in the discourse. Significance testing is less reliable due to the normality assumption of the t-test, which is violated for natural language (Manning & Schuetze, 1999, 156). Generally speaking, the association score should be at least around 2.5, which corresponds to a confidence level of $\alpha = 0.05$ (Manning & Schuetze, 1999, 153). Only scores similar or larger than this value were considered for ranking. Association scores in our analysis then ranged from 2.45 to 13.27. Note that trigrams generally have a lower association score within this spectrum.

In order to understand the potentially distinct dynamics underlying the discourse in each of the two main political camps—the opposition camp and the pro-Putin camp—it is first necessary to identify these two camps in our dataset. This is a difficult task, since we can only rely on what users write, as typically no official affiliation information is available. We therefore proceeded as follows: first, we identified the Twitter users in our Twitter data based on the value of their "screen_name." We then used keywords (see full list of keywords in Supplementary Information S2.2) in combination with the sentiment analyser SentiStrength and scored the tweets of identified users on a scale between -3 and 3, with negative scores indicating a pro-Putin tweet, positive scores an oppositional tweet and a 0 score a neutral tweet. We classified users by the average score of all their tweets as either belonging to the pro-Putin or opposition camp by invoking that users would express positive sentiments about terms associated with their own camp and/or negative sentiments towards terms associated with the other camp (see Supplementary Information S2.2 for further details). Thus, the combination of keywords and sentiment analysis allowed us to understand the framing of the keywords used, since the keywords on their own do not indicate a political affiliation. If the keyword is Putin for instance and it appears with negative sentiment words, we can derive that the user posting this tweet, is critical of Putin, if it appears on the other hand with positive sentiment words, then the user is rather likely to be a Putin supporter. Note that we focused on and classified only the 1,000 most active users among the more than 140,000 unique Twitter users in our data. The 1,000 most active users accounted for 51% of all tweets in our dataset. Thus, these users were the most influential ones in the debate and since our focus is on the political debate and the communication strategies used in the debate to affect popular perceptions, it is sensible to focus on the most active and influential users, who are more like to affect popular perceptions. The focus allows us on the other hand to investigate the Twitter users involved in the political discourse more closely, that is, to examine who they are and how they are connected with each other.

The classification was particularly challenging because the Russian oppositional camp is highly fragmented and the often harsh criticism voiced in tweets is not only directed against Putin and his supporters but also sometimes against other oppositional groups and figures. For this reason the automatic classification may from time to time misclassify Twitter users as pro-Putin because it detects emotionally negative tweets targeted towards the "other" opposition. We therefore extended the classification procedure to include weights and additional "context" information (e.g. retweet-information) (see Supplementary Information S2.2 for further details). We estimated the quality of this classification method by selecting a sample of 100 users and manually assessing their political orientation based on their user profiles and tweet activities. By comparing this manual categorisation with the result of the automatic classification we found that around 70% of political orientations were classified correctly by our automatic routine. This accuracy level is comparable to classification accuracy achieved by common machine learning text-based classification methods (Bensusan & Kalousis, 2001; Bird et al., 2009). Note moreover that the results of our subsequent discourse analyses for the two camps, which reveal clearly pro-Putin and oppositional discourses, further lend credibility to our classification.

To get a better understanding of who the most active users are and how they are connected, we extracted and analysed the full names and profile descriptions of the Twitter users and the timing when they set up their Twitter accounts from our Twitter data. Moreover, we studied their social networks based on whom they are following within the 1,000 most active Twitter users.[7] The follower network structures allowed us to understand the communication flows and thus to what extent messages from certain camps are also noticed by the other political camps. A retweet-based social network would underestimate the links between different political camps, since oppositional Twitter users may for instance follow pro-Putin followers to stay informed about their plans and actions, but they are rather unlikely to retweet pro-Putin messages, particularly since the commenting retweet function was not available in Twitter between November 2011 and March 2012. Furthermore, the social network analysis reveals who are the most influential Twitter users in the respective camps in terms of the number of their followers and how they are linked to the other Twitter users in their own but also in the other political camps.

Finally, to analyse the political communication strategies used by the different political camps between November 2011 and March 2012, we adopted a qualitative research method approach (Saldana, 2013) and coded manually the main extracted ngrams, i.e., those with significantly high association scores, in each camp according to the four communication strategies described in the theory of communication power by Castells (2009): framing, priming, agenda-setting and indexing (see second section) and used five additional ad hoc codes (Flick, 2006) for ngrams that did not fit in either of the four categories but are important with respect to how the population perceived oppositional mobllisation: *fact*, when a ngram merely reported a fact; *demand*, when a ngram expressed a political demand such as "fair election"; *self-criticism*, when a ngram expressed an in-camp criticism; *hijacking*, when a core demand or idea from the adversary political camp was hijacked and misused by a political camp in a ngram and *mobilisation*, when a ngram informed about an upcoming or ongoing political action.[8]

Results

We first describe briefly the evolution of the protest movement and discourse following the Duma elections of December 4, 2011, showing its rise and decline in the overall Twitter discourse. We then analyse the different political camps, their most active and influential members, their social networks and their respective political discourse to understand how the communication strategies and reactions of each side contributed to the disintegration of the oppositional movement on Twitter shortly after the second elections in March 2012.

Rise and Fall of the Russian protest movement on Twitter

The election and protest events in 2011–12 were all mirrored and reflected on Twitter (see Supplementary Information Figure S3 and S4). The December election was officially an overwhelming victory for the governing party 'United Russia'. This victory was reflected in the Russian Twittersphere in the number of mentions of each party and in the number of statements referring to the parties for which people had voted, for instance, 'for Jabloko' (liberals) (BAS = 5.08), 'for KPRF' (communists) (BAS = 4.65) or 'for United Russia' (BAS = 9.24).

The Twittersphere discourse, however, also shows that the Duma elections were generally perceived as having been manipulated. The bigram 'fraud elections' (BAS = 6.63) was one of the most common bigrams for the December 4, 2011 discourse. People reported voting against United Russia in an attempt to demonstrate the inaccuracy of the allegedly manipulated official results, and demanded to 'cancel elections results' (TAS = 3.51) and to 'conduct new elections' (TAS = 4.11). Major protests followed, attended by tens of thousands of Russians on December 6, 10 and 24; on February 26; and on March 5 and 10. Here, Twitter was used as a tool for mobilisation. For example, specific protest mobilisation hashtags (e.g., #6Dec, #Triumfalnaya) were used to spread information on the timing and location of protests. Furthermore, new prominent oppositional figures emerged during the first days of protest, for example, unaligned oppositional figures such as Alexey Navalny.

The political discourse on Twitter in December 2011 was largely dominated by critical, oppositional voices. Putin was portrayed as a thief of votes ('Putin thief', BAS=3.00), and United Russia as a 'party (of) thieves' (BAS=5.14). At the same time the discourse reflects the euphoria and appeal associated with revolutionary sentiments. Tweets such as the ones on December 18, 2011, referring to a 'new level (of) evolution (of) Russian political culture' (combined TAS = 3.71, see Supplementary Information S2.1. for explanation of combined TAS and BAS) were posted frequently. A strong identification with the protest movement was shown by statements such as 'you are (the) movement' (TAS = 3.97), 'Balotnaya (Square) we come' (TAS = 4.78) or 'be one white-ribbon' (TAS = 3.54). The largest protest event on December 24, 2011, was accompanied by enthusiastic feelings among supporters of the oppositional movement: 'demonstration (was) great, thanks' (TAS = 3.44).

However, support for the protest movement began to weaken on Twitter in January 2012, despite continuing demands for fair elections and worries about the declining Russian democracy (for instance 'end (of) era (of) democratic governing' with combined TAS of 3.04). Sympathy with Putin was now expressed more frequently (e.g., 'God save Putin' with an TAS score of 3.58). Moreover, already in the wake of the first oppositional protest, the pro-Putin forces organised rallies supporting Putin and United Russia. While these rallies were initially small, attempts to delegitimise them as fake protest events appeared on Twitter immediately (e.g., 'The so called excursion turned out to be an excursion to a rally pro United Russia', combined TAS = 2.75).

At the same time and in line with a widening split in the protest movement (Sakwa, 2014), the divisions between various political opposition fractions also became visible on Twitter (e.g., 'Prokhorov against Ziuganov' TAS = 2.92, or 'LDPR gives Ziuganov Stalin mask' combined TAS = 2.52). These internal disputes created the impression of a dissolving opposition incapable of seriously challenging Putin. Increasingly, people began to express

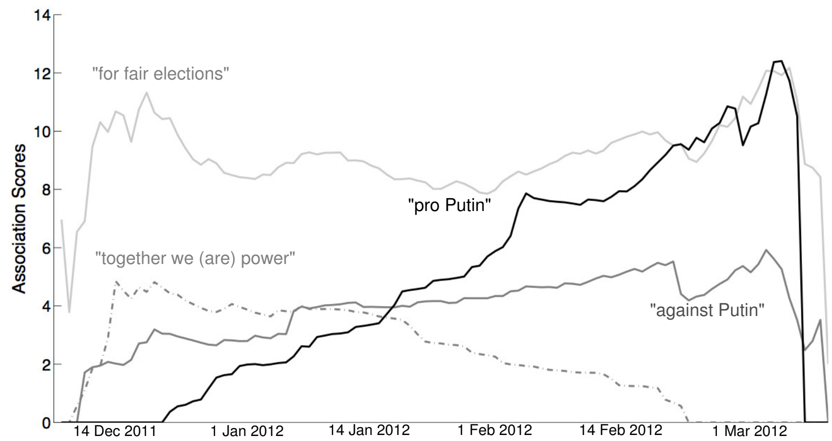


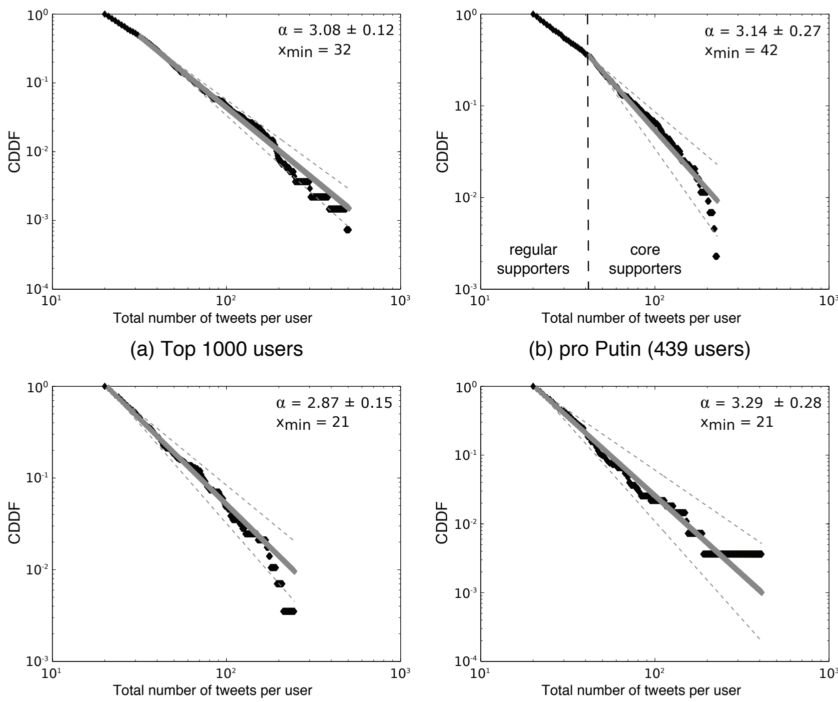
Figure 1: Smoothed trend lines for four important bi- and trigram collocations starting with the Duma elections on December 4, 2011. We controlled for linguistic heterogeneity and found that our estimated scores for the pro Putin bigram may be even underestimated (see endnote 8).

their discontent with the opposition, for instance 'Our so-called opposition, unsatisfied with elections, (but) nobody resigned' (combined TAS = 2.45).

By end-January, tweets expressing support for Putin increasingly dominated the political Twittersphere and became more frequent than tweets expressing opposition to Putin (see Figure 1).[9] In February, an increasing number of pro-Putin protest events were organised yet support for the protest movement on Twitter was still very visible. The new slogan 'Putin go-home' (BAS = 7.48) was frequently used and statements of re-assurance such as 'welcome political spring' (TAS = 3.80) as well as identification statements such as 'I took part in the protest event' (combined BAS = 4.69) were tweeted frequently. Moreover, attempts to delegitimise the pro-Putin demonstrations were intensified with users spreading statements such as 'How I started (to) love Putin for 500 Rubel' (combined TAS = 3.00), suggesting that supporters for the pro-Putin demonstrations were bribed.

At the same time, however, attempts to delegitimise the oppositional protest—the opposition was accused of having been paid and directed by the US—were also spread on Twitter, as expressed for instance in the statement 'we believe Putin, against US' revolution' (combined TAS = 3.65). The pro-Putin camp instigated popular fear of chaos and revolution, suggesting that only Putin will ensure peace and order. This resonated with an apparently growing feeling of futility and disillusionment on the side of the protest supporters. Protests were even deemed increasingly senseless at a time when the political momentum appeared to have shifted towards the pro-Putin side (for instance 'pointless protest' with BAS = 4.34 on February 26, 2012).

Despite accusations of election irregularities after the presidential election on March 4, 2012, it then seemed indisputable that Putin enjoyed broad support among Russians and the protest movement began to dissolve quickly. This is also visible in the decline in the frequency of protest-related and mobilisation keywords on Twitter following the second



(a) Top 1000 users

(b) pro Putin (439 users)

(c) pro Opposition (285 users)

(d) not assigned to factions (276 users)

Figure 2: Distribution of tweets per user (a) for the 1,000 most active users, (b) pro-Putin users, (c) pro Opposition users, and (d) users assigned to neither camp.

elections (see Supplementary Information Figure S4). At the same time the anger of those who had supported the movement turned against the oppositional leaders who were blamed to have failed: 'opposition incompetent, failed to take up people's discontent' (combined TAS of 2.58 on March 10, 2012).

Power struggle between different political camps on Twitter

The analysis of the overall political discourse on Twitter already suggests that communication power was indeed used to instigate a discursive shift in favour of Putin and to weaken support for the opposition on Twitter. Critical voices were discredited and political elites were represented as legitimate. We now turn to a more specific analysis of each political camp (pro-Putin and opposition) and their discourse. We further contrast these against the unclassified camp in our Twitter data, which may be regarded as the general public. We will in particular focus on pro-Putin camp's efforts to affect people's perception with respect to the oppositional movement to discourage further mobilisation.

We first investigated the tweeting activities in the three camps (pro-Putin, pro Opposition, unclassified). The distributions of tweets per user across the full period covered by our data—overall and in all three camps separately—consistently show relatively similar heavy-tailed signatures (Figure 2). Gray lines mark the best fit of the heavy (or power law) tail of the distribution with 95% confidence intervals. Fits were calculated using maximum likelihood estimation. The corresponding power law exponent α and cutoff x_{min} at which the tail begins are provided in the figures. This implies two important empirical characteristics of user activity in the Russian Twittersphere: first, the number of very active Twitter users is much larger than one would expect under the assumption of a normal distribution of tweets per user, for example. Second, there is no typical or mean number of tweets per user. For the full sample of the 1,000 most active users this implies that, while 90% of users contributed less than about 70 tweets over the full period considered, some users in the remaining 10% contributed over 400 tweets (Figure 2a). Consequently, these 10% most active users account for more than 46% of all tweets.

It is important to emphasise here that among the 1,000 most active users the pro-Putin, opposition and unclassified camp are not equally represented. In fact, the pro-Putin camp is by far the largest, with 439 of the 1,000 most active Twitter users classified. In comparison, the opposition camp makes up only 285 Twitter users and the unclassified camp 276. This relative difference in the size of the three camps varied little throughout the whole period analysed and in fact already suggests a communication power disbalance in favour of the pro-Putin camp.

Furthermore, we found that there is a marked statistical difference between the distribution of tweets per user in the pro-Putin camp and both the opposition and unclassified camp: the statistic for the pro-Putin camp visibly deviates from the others in that the heavy-tailed signature only statistically holds true for users with 42 tweets or more. In other words, there is a systematic difference between the activity of very active and less active users in this camp (Figure 2b). In contrast, the distribution of tweets per user follows the same regularity across all levels of individual user activity in the oppositional and unclassified camp (Figure 2c and d). This suggests that there were two distinct sub-categories of pro-Putin users: the most active users (n=157) in the tail of the distribution who contributed at least 42 tweets over the full period analysed, and the remaining less active pro-Putin supporters (n=282). Note that throughout the whole period analysed the most active users—the core Putin camp— contributed relatively more tweets to the Twitter discourse than any of the other camps, thus effectively dominating the Russian Twittersphere (see also Supplementary Information Figure S5).[10]

We can identify a notable effect of the core pro-Putin camp on the political discourse. Figure 3 shows that the "pro Putin" sentiment is almost exclusively carried by the core pro-Putin camp throughout January. The fact that the share of tweets tweeted by the different camps is comparably stable over time ensures that the effect of the core pro-Putin camp on the bigram "pro Putin" is not an artefact of activity: the camp indeed began to express pro Putin sentiment weeks before this was visible in the overall Twitter discourse.

A closer inspection of the core Putin supporters reveals that the camp is dominated by professional Twitter users, i.e., United Russia party, official governmental information outlets and major pro-government media outlets, such as Russia Today (see Table 1).

Through loyal party, institutional and media officials, the government thus seems to have had the ability to influence the discourse on Twitter more effectively than the opposition. These Twitter users have sufficient resources and leverage for flooding Twitter with dedicated messages. Among the regular Putin supporters there are also media outlets, but not the major ones. Instead, we see more single individuals supporting Putin (see Table 1). These users have lower capabilities (available time, support by a team of operators) to massively spread their views across Twitter.

Table 1: The most influential Twitter users in each political camp (In-Degree shows the number of followers).

User Name	Full Name	Description	Political Camp	In-Degree
GazetaRu_Lenta	Chronic of Daily News, gazeta.ru	Own information coverage as well as reports from major Russian and international news agencies (Gazeta.ru is the most popular Russian language news website)	Core pro-Putin	135
interfax_news	Interfax	News from Interfax (Interfax is the major Russian news agency)	Core pro-Putin	110
RU_Today	Russia Today	Peace to the World (Russia Today is seen as the main propaganda channel of the Russian Government)	Core pro-Putin	109
rgrus	Russian Newspaper	Russian Newspaper – outlet of the Russian Federation Government. Published since 11 November 1996. RG and RG.RU publish official documents and operational news	Core pro-Putin	97
radio_kp	Komsomolskaya Pravda	Informative-talkative radio station, 24 hours, format story channel. Radio of real people and non-fiction stories (Komsomolskaya Pravda, a newspaper, used to be the official organ of the Communist Union of Youth, Komsomol, since 1990 it became a daily Russian tabloid. The radio station is the radio channel of the newspaper.)	Core pro-Putin	96
er_novosti	United Russia	Official Twitter account of the United Russia party	Core pro-Putin	93
topoprf	Tribuna OP	TOP - Public Chamber Tribune - search organisations and persons, news, interviews, blogs, discussions (News website)	Core pro-Putin	71
mashina_s	Michail Kovalev	no description available	Core pro-Putin	64
VRSoloviev	Vladimir Soloviev	no description available (journalist on Rossiya 1 TV Channel)	Regular pro-Putin	65
izvestia_ru	Izvestia	Official microblog of the newspaper Izvestia. From news we create insights.(long-running, high-circulation daily broadsheet newspaper in Russia, previously official Soviet Union government newspaper)	Regular pro-Putin	42
burmatoff	Vladimir Burmatoff	First Deputy Chairman of the Education Committee of the State Duma.	Regular pro-Putin	41
ntvru	NTV	Official Twitter account of NTV and NTV.ru site (TV channel, controlled by Gazprom Media)	Regular pro-Putin	34
KFM936	Kommersant FM	Official Twitter account of the radio station Kommersant FM	Regular pro-Putin	24
AdvokatKubany	Victor Mikhaylov	Foundation of legal support for compatriots in the United States. Only proven lawyers, immigration consultants, notaries	Regular pro-Putin	23
kurginyanRU	Time Will Show	Club “Sut Vremeni” (Time Will Show). This is the Twitter account of the club members. (Russian, left, conservative political movement supporting the Putin government)	Regular pro-Putin	15
PP_Russia	Pavel Pyatnitsky	Personal views (opinionated judgements), no claim to truth. (political figure in Russia)	Regular pro-Putin	14
tvrain	TV Channel RAIN	The independent Russian TV Channel. News RAIN. (Most Popular Oppositional TV Channel in Russia)	Opposition	112
KSHN	Kashin	We have kondopoga, we have khokhloma. Russian journalist and novelist	Opposition	98
kmrsFM	Kommersant FM 93.6	Non-official Twitter account of Kommersant FM (Oppositional pendant to KFM936)	Opposition	62
lentaruofficial	Lenta.ru	Daily News (Lenta.ru is an online newspaper and the second most popular Russian language news website)	Opposition	61
GolosAmeriki	Voice Of America	Welcome to the official Twitter community service of the Russian VOA (Voice of America). (Voice of America is the official external broadcast institution of the United States federal government)	Opposition	54
korobkov	Korobkov Zemljanskij	no description available (Russian political activist, journalist and blogger)	Opposition	36
Moscow_advokat	Nikolaj Polozov	Everything you did not want to know about the Russian justice and feared to hear. Infamous farce. (Pussy Riot lawyer)	Opposition	34
maglov	maglov	no description available	Opposition	29
ddb777	Different News	A journalist, not a blogger. This account has no relation to the program ‘Vesti’ and does not reflect the information policy of VGTRK.	Unclassified	26
san4izz	Baturin	Medicine, politics, West Caucasus. Middle Volga (blogger)	Unclassified	23
crimerussia	Crime Russia	Notes on organised crime and on shadow and legal economic activities with corrupt links to Russian governing bodies	Unclassified	23
Toporintv	Toporin Aleksander	24/7, Editor-in-Chief (journalist)	Unclassified	22
bicotender	bicotender.ru	Bicotender - search system of tendering and procurement of Russia in CIS. All for success in tendering.	Unclassified	14
b111org	b111org	Service of entertaining blogs	Unclassified	14
arl_spb	Romik(18-)	Patriotism - the last refuge of a scoundrel. It’s better to be a fool, but smart rather than being a smart fool ... wife@Elisavetatheone	Unclassified	13
justmedia_ru	Justmedia.ru	News portal JustMedia - independent news on a information server. Politics, economics, culture, sports, crime	Unclassified	13

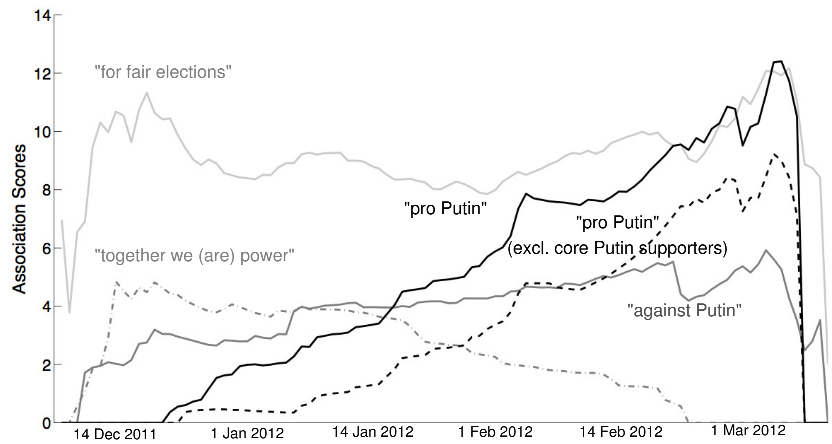


Figure 3: Smoothed trend lines for four important bi- and trigram collocations disaggregating the effect of core Putin supporters on the “pro Putin” bigram.

The most influential Twitter users from the opposition camp on the other hand combine major oppositional media outlets, notably the popular TV Channel RAIN, but also individual activists, journalists, or bloggers. We may assume that their resources are again rather limited comparing to the main media and governmental outlets on Twitter. The appearance of Voice of America in the list of most influential oppositional Twitter users shows the strong foreign support of the Russian oppositional movement (see Table 1). Expectedly, the list of most influential unclassified Twitter users contains news and individual accounts, that are rather unknown and that do not display a clear political alignment. The fact that major and minor “traditional” media outlets are among the most influential Twitter users in all camps moreover shows how strongly interlinked social media like Twitter still are with more traditional media outlets such as TV or newspapers (Chadwick, 2013).

Figure 4 shows the social ties between the most active Twitter users in our data. Additionally official government accounts such as Medvedev Russia (light violet blue) and central oppositional figures Twitter accounts, such as Alexey Navalny (orange) were added, please note these Twitter users were not in the original data among the 1,000 most active Twitter users. They were added to show the influence of these official government or central oppositional figures accounts on other Twitter users. Major hubs (nodes with highest in-degree) in each political camp are named. Interestingly we see that the two main rival political camps, the pro-Putin and the opposition camps, are well interlinked (see Supplementary Information Figure S6).

We can thus conclude that topics or issues raised by the pro-Putin camp reached the opposition and their supporters and vice versa. And respectively, it is therefore also realistic to assume that any political communication strategy adopted by any of the political camps would have indeed had a direct effect on the respective political opponent. Figure 4 (see subgraphs Supplementary Information Figures S6 and S7) moreover shows that regular Putin supporters are closely following the Twitter users in the core pro-Putin camp. This enabled the core pro-Putin camp to issue targeted political messages that are subsequently taken

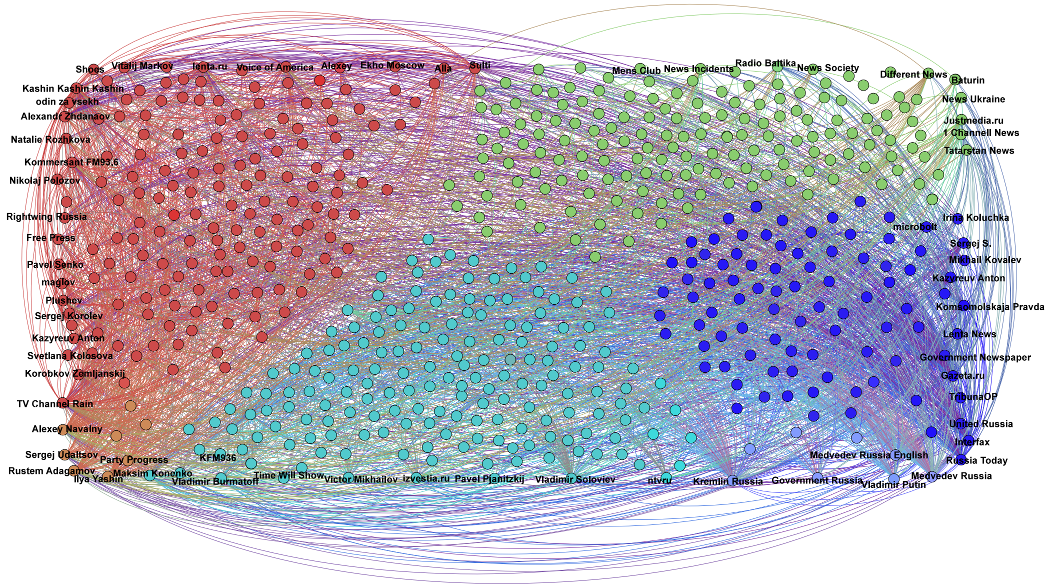


Figure 4: Social network based on whom the respective Twitter users followed. Nodes are Twitter users, directed edges are based on follower relations thus spanning the whole social network with all political camps and their ties. Red nodes are depicting the oppositional Twitter users, green nodes the unclassified and blue (core) and turquoise (regular) the pro-Putin camp.

up, echoed and further spread by the regular pro-Putin Twitter users, reinforcing the overall pro-Putin communication (Barberá et al., 2015).

Figure 4 (see also Supplementary Information Figure S7 and S8) shows also that the unclassified Twitter users, which we interpret as the general public, follow the pro-Putin and the oppositional camp. We can thus assume that pro-Putin and oppositional messages reached the general public and could potentially influence perceptions of the general public.

The analysis of the discourses in the different camps (Table 2, see extended Table S1 in Supplementary Information with BAS and TAS scores) shows the various communication strategies employed by the two pro-Putin camps and by the opposition. Table 2 shows the evolution of the political discourse described in the previous section, but additionally highlights how the different political groups contributed to the evolution of this discourse.

Initially, the opposition *set the agenda* by challenging the Duma election results. Table 2 shows that in the beginning the unclassified camp also expressed strong sympathy for the protest movement and similar indignation over election irregularities. Thus at this stage there was a high mobilisation potentiality, particularly since the number of protest participants was growing with each oppositional protest, which the opposition also tried to highlight in their Twitter communication (see Table 2, 24th December). In reaction to the strong oppositional move in the public political discourse, the pro-Putin camp increased its efforts to shift the political discourse in its favour. Putin supporters used predominantly a *framing* communication strategy, belittling, e.g. reporting lower participants numbers, (see Table 2, 24th December) and delegitimising the protest movement. A key event here seems to have been the meeting of opposition leaders with the US ambassador on January 18, 2012 (see

Table 2). Both pro-Putin camps immediately took advantage of the unique possibility to discredit the oppositional movement as being steered and financed by the US. The pro-Putin camp went beyond its *framing* strategy in this case and managed to *set an anti-opposition agenda* by revealing the secret meeting and questioning the independence of prominent oppositional leaders.

The two Putin supporter sub-camps adopted slightly different *framing* discursive means to delegitimise the opposition and challenging its quantitative (i.e. reporting lower numbers, e.g. Table 2, 6th March) and qualitative (i.e. reporting bribed participants, e.g. Table 2, 5th March) mobilisation success. The core Putin supporters seem to have acted more strategically, reporting alleged facts about discords and splits within the opposition, about the failures of oppositional leaders and about the decreasing support for the opposition or even firm rejection of the opposition in the population. On the other hand they continuously stressed the strong support for Putin in the general population. The regular Putin supporters seem to have communicated without a systematic strategy. Their tweets generally appeared to be more spontaneous reactions and were also more likely to attack the opposition directly and offensively instead of reporting matter-of-factly (Table 2).

Besides delegitimising the opposition, the pro-Putin camp seemingly adopted also a *priming* strategy, using Twitter to propagate a political program of stable progress, modernisation and innovation (see Table 2, 18th February). The opposition on the other hand failed to communicate a political program beyond demands for fair elections. The demand "for fair elections" prominently appeared also in the pro-Putin camp (Table 2). This points to a key strategic move by the pro-Putin camp: it appears as if they took over the demand for fair (presidential) elections and presented it as a genuine goal that they themselves were to pursue in the upcoming elections. This *hijacking* communication strategy deprived the opposition of one of its core political demand—a demand that, in fact, formed the basis for the union of different oppositional forces.

Meanwhile, the unclassified camp too underwent changes in its political discourse that are worth noting. After initial support for the opposition, it rather quickly began to lose interest in the political events and discourses: clear support for the oppositional camp was no longer expressed after December 2011, and after the presidential elections the unclassified camp showed even tiredness of the constant political upheaval, apparently preferring a return to normality (see Table 2). The intense discrediting campaign by the pro-Putin camp, which became more and more prominent on Twitter over time, thus seems to have not only contributed to increasing disillusion within the wider protest movement itself (for instance "Corrupted idea of first Bolotnaya protest" on February 4, 2012, Table 2), but also decisively to the weakening sympathy for the oppositional movement among unaligned Twitter users, thus, contributing to a failure of further oppositional mobilisation.

Table 2: The table shows the time evolution of the discourse based on bi- and trigrams in the three camps, opposition, pro-Putin (core and regular) and unclassified.

Time line	Core pro-Putin camp	Regular pro-Putin camp	Opposition camp	Unclassified camp
4-Dec-2011	For United Russia (*framing*)	For United Russia (*framing*)	We vote against United Russia (*framing*)	In Moscow journalists observed ballots thrown in (*agenda-setting & framing*)
	KPRF refuses to allocate votes to Jabloko (*agenda-setting & framing*)	LDPR buys votes with vodka (*agenda-setting & framing*)	Mafia throws in ballots (*agenda-setting & framing*)	Duma elections (*fact*)
5-Dec-2011	LDPR considers coalition (*agenda-setting & framing*)	For United Russia (*framing*)	Putin's criminal gang totally forged elections (*agenda-setting & framing*)	People stopped being silent (*framing*)
	United Russia meets in Moscow (*fact*)	Putin is better (*framing*)	5Dec ChP against forged elections (*agenda-setting & mobilisation*)	Demonstrators shouted 'freedom', well done (*framing*)
6-Dec-2011	Demonstration Putin supporters (*agenda-setting & framing*)	Demonstrations split country (*framing*)	6Dec Triumfalnaya rally for fair elections (*agenda-setting & mobilisation*)	6Dec Triumfalnaya rally for fair elections (*agenda-setting & mobilisation*)
		Navalny's arrest political mistake (*self-criticism*)	Navalny blogger anticorruption project (*agenda-setting & priming*)	Union of democratic forces (*framing*)
9-Dec-2011	For fair elections (*hijacking & demand*)	Damned White Ribbon, keep children away (*framing*)	White Ribbon Snow Revolution (*framing*)	Honesty best policy (*priming*)
	Udaltsov released (*agenda-setting*)	No revolution, thanks (*framing*)	United Opposition demonstration, on Bolotnaya they have to see masses (*framing & mobilisation*)	Tomorrow provocation against protesters planned (*agenda-setting*)
10-Dec-2011	Demonstration Medvedev supporters in Moscow (*agenda-setting & mobilisation*)	Our democratic bastards sully our country (*framing*)	10Dec Demonstration Revolution Square (*agenda-setting & mobilisation*)	Demonstrations in Moscow (*fact*)
		You demonstrated, those in power understood (*framing*)	Shouted 'Putin is a thief, against Putin' (*framing*)	KPRF says illegitimate elections (*agenda-setting & framing*)
23-Dec-2011	Thousand resolute Nashi members (*framing*)	Modernisation supporters, Yes Medvedev Russia (*priming*)	24Dec demonstration for fair elections (*agenda-setting & mobilisation*)	For fair elections (*demand*)
	God save Putin (*framing*)	For fair elections (*hijacking & demand*)	Revolution creative class, support political reform (*framing & priming*)	
24-Dec-2011	Burned white ribbon (*agenda-setting & framing*)	25,000 demonstrate on Bolotnaya for fair elections (*framing*)	Multiple tens of thousand people came (*framing*)	Demonstration for fair elections (*fact*)
	Huge Putin portrait launched (*agenda-setting & framing*)	Opposition overstates numbers of protesters (*framing*)	Highest level of dignity (*framing*)	Honesty best policy (*priming*)
4-Jan-2012	Political action pro-Putin (*agenda-setting & framing*)	God save Putin (*framing*)	Udaltsov was arrested (*agenda-setting*)	
18-Jan-2012	Meeting opposition leaders with US ambassador (*agenda-setting & framing*)	Meeting opposition leaders with US ambassador (*agenda-setting & framing*)	Opposition unsatisfied, but nobody resigned (*self-criticism*)	
	Support Russia, support Putin (*framing*)			
4-Feb-2012	Demonstration against fraud elections on Bolotnaya (*fact*)	Demonstration, Navalny promised a million will come (*framing*)	Navalny calls to Bolotnaya 4Feb (*agenda-setting & mobilisation*)	Honesty best police (*priming*)
	For fair elections (*hijacking & demand*)	4Feb Bolotnaya demonstration bought (*framing*)	Corrupted idea of first Bolotnaya protest (*self-criticism*)	US happy with Putin, who benefits from protest? (*framing*)
18-Feb-2012	Medvedev Modernisation Innovation, support stable progress (*priming*)	Demonstration in support of Putin (*agenda-setting & mobilisation*)	Whom Putin needs against revolution (*agenda-setting & framing*)	
	believe Putin, against US' revolution (*framing*)	believe Putin, against US' revolution (*framing*)	Putin go home (*framing*)	
23-Feb-2012	Demonstration pro-Putin 23Feb (*agenda-setting & framing*)	Demonstration pro-Putin 23Feb (*agenda-setting & framing*)	Demonstration pro-Putin not many people, about 1000-2000 (*framing*)	Zhirinovskii ranting retweet (*agenda-setting & framing*)
4-Mar-2012	FEMEN provocation (*framing*)	Elections Russian president (*fact*)	For Russia's future (*framing*)	Polling stations opened (*fact*)
	Emotional election (*framing*)	pro-Putin demonstration thousands (*agenda-setting & framing*)	We invite to come to Pushkinskaya (*mobilisation*	Fake election observers exposed (*agenda-setting & framing*)
5-Mar-2012	On Pushkinskaya they pay money (*agenda-setting & framing*)	Police bashes people (*agenda-setting & framing*)	Election observers say correct elections (*fact*)	OMON forces arrested protesters, dissolved demonstration (*agenda-setting & framing*)
	World leaders congratulate Putin (*framing*)	Kasparov was welcomed by thousandfold Boo (*framing*)	Demonstration for fair elections (*agenda-setting & mobilisation*)	Election observers say correct elections (*fact*)
6-Mar-2012	Opposition demonstration so far hardly attended (*framing*)	Majority voted for Putin (*agenda-setting & framing*)	Press conference of Electoral Association (*fact*)	Overview regions love Putin (*agenda-setting*)
	Protest blown away (*framing*)	Political columnist beaten up (*agenda-setting*)	Home Office, police state (*framing*)	Every fucking day demonstrations (*framing*)
9-Mar-2012	Navalny is dead, proclaimed (*framing*)	Obama congratulates Putin (*framing*)	Putin insulted Russian people (*framing*)	And the next protest (*framing*)
10-Mar-2012	Million protesters promised (*framing*)	Election results approved (*agenda-setting & framing*)	10Mar demonstration central on Rostov (*agenda-setting & mobilisation*)	Attempt of non-authorised demonstration (*agenda-setting*)
	Nationalists leave opposition demonstration (*agenda-setting & framing*)		Thugs are afraid of an orange revolution (*framing*)	Opposition speakers insulted (*agenda-setting*)

Conclusion

In this study we have analysed the political discourse in the Russian Twittersphere from November 17, 2011 to March 12, 2012. We demonstrated that the discourse on Twitter mirrors major political events and developments quite accurately: all that happened between November 2011 and March 2012 was communicated on Twitter and all that was communicated on Twitter had an actual "real-world" reference. The fact that we find evidence for strategic communication on Twitter that coincides initially with a broad support for the opposition and later with an increasing support for Putin on the one hand and a decline in oppositional mobilisation on the other hand additionally underlines the importance of social media as a forum of political dispute. Can we then draw direct inferences from our analysis of Twitter on the fate of the protest movement more broadly? A direct causal analysis is certainly not possible. For example, we could not predict the frequency or timing of protest events from the Twitter discourse. But understanding how the discourse on Twitter shifted in favour of the government can certainly inform our understanding of the rise and decline of the protest movement more broadly.

Our study in particular shows that while both pro- and anti-Putin Twitter users tried to influence the political discourse on Twitter, over time the balance of communication power visibly shifted towards the pro-Putin factions. The strategic communication of Putin supporters in the weeks leading up to the presidential election evidently shifted the perceptions of the protest movement on Twitter to the movement's detriment. This may thus have significantly weakened the oppositional voice on Twitter at a time the movement was already struggling to regain momentum, further mobilise and overcome internal divisions.

Our analysis highlights that the growing feeling of futility and disillusionment affecting the oppositional movement more broadly (Sakwa, 2014) was clearly reflected on Twitter in the weeks leading up to the presidential election. With the political discourse on Twitter beginning to noticeably shift in favour of the Putin supporters, oppositionally minded people on Twitter may have started to slide into a so called 'spiral of silence' (Noelle-Neumann, 1974, 1993). They perceived their political view to be in a shrinking minority, finding insufficient resonance in the general public on Twitter and gradually stopped to speak up, turning rather inwards in growing self-doubts and disillusion. The weakening sympathy and increasing indifference of the general public—as represented by the unclassified camp in our analysis—presumably contributed to this escalating de-mobilisation process. At the same time the opposition movement was increasingly confronted with discrediting allegations against its leaders, aggressively reproached by the pro-Putin camp on Twitter (and certainly on other media channels), which invoked merely disappointment among the protesters and scepticism among the general public represented on Twitter.

The pro-Putin faction's communication strategies on Twitter seem to have been more successful than the communication strategies of the opposition. However, it is important here to re-emphasise the importance of the 'institutionalised' pro-Putin support on Twitter led by loyal core supporters, which was likely instrumental in shifting the discursive power to the government aligned camps. We could clarify, in particular, that already a relatively small camp of very active and loyal core Putin supporters seems to have effectively enabled the government to decisively influence the discourse on Twitter in its favour (see Figure 3). Short of open technical manipulation the activity of the core Putin supporters thus amounts to a clear manipulation of public perceptions on Twitter in favour of those in power.[11]

It is not possible from our analysis to conclusively establish to what extent the government used paid 'Internet trolls' to spread pro-governmental propaganda, as reports revealed later with reference to the Russian-Ukrainian conflict 2014–15 (Walker, 2015). Given that the protests 2011–12 seemingly took the Russian government by surprise, their political communication strategy would in all likelihood have been a reaction to these protests. Thus, if the government would have hired massively Internet trolls to drive its communication strategy on Twitter, we would expect that many of the pro-Putin users joined Twitter after the first protests sparked in December. We checked this for the 1,000 most active Twitter users and did not find an unusual increase of newly created Twitter accounts in the pro-Putin camps following the December 2011 protests (see Supplementary Information S3). This does, of course, not exclude the possibility that existing users were directed and/or paid to support the government on Twitter.

On digital communication channels like Twitter it is generally difficult to obtain reliable proof for whether the support for established powers is real or just 'simulated'. Researchers have observed "campaigns disguised as spontaneous, popular 'grassroots' behaviour that are in reality carried out by a single person or organisation (...) to establish a false sense of group consensus about a particular idea" (Ratkiewicz et al., 2011, p.297) on the Internet. Castells (2009) pointed out that although there is no domination by one group on the Internet, those actors who have resources are capable of manipulating the discourse in their favour. The resource asymmetry between the two main camps—pro-Putin and opposition—seems to have decisively contributed to the advantage of the pro-Putin side.

This is particularly visible in the activity patterns and discursive behaviour of the core Putin supporters, who sent massive amounts of pro-Putin tweets. But there was clearly also genuine support for Putin on Twitter, as represented by the regular pro-Putin camp. Note though that our analysis also suggests that this camp of 'regular' Putin supporters was generally much less active than the opposition camp on Twitter (see Supplementary Information Figure S5).

In the end, no matter how much 'real' support Putin had, our analysis of the political discourse suggests that the perceived support had a real effect on the opposition and general public on Twitter. This shows that regardless of the promises that new digital technologies hold in terms of empowerment of marginalised or weaker (political) actors, these technologies are still part of the overall system of power—in particular, uneven resource distributions—and may therefore be still utilised by governments in their favour. In other words our study empirically confirms that indeed "whoever has enough money, including political leaders, will have a better chance of operating the switch in its favour" (Castells, 2009, 52). And this applies not only to the specific case study of the Russian political discourse during the 2011–2012 elections and protests. A study on Chinese government's massive propaganda activities on social media (King et al., 2016) for instance or reports on Erdogan's social media strategy to mobilise the population against the military coup in Turkey in 2016 (Srivastava, 2016) show clearly that the patterns found here generalize to other countries as well. The question of whether social media are at the end of the day liberative or oppressive is relevant in every political context.

Finally, our study demonstrates how Twitter data may be used for informative political science. In this paper we conducted a new kind of computational dynamic discourse analysis that is based on quantitative time-series measures (word counts, ngram association scores) but also on theory-guided and contextually embedded coding and interpretation of these measures. In the future this method could be refined for even more precise and elaborate analysis of Internet data.

Notes

1 We used the freely available Twitter Streaming API Spritzer Sample, which collects 1% of all public tweets in real-time, https://dev.twitter.com/streaming/overview. The retrieved data is in JSON format (see Supplementary Information S1.1).

2 The Berkman Center's Report "Mapping Russian Twitter" (Kelly et al., 2012) is a notable exception providing ground-breaking insights into the structure of the Russian Twittersphere. Please see Supplementary Information S1.3 for details.

3 While the Russian Twitter space extends beyond the Russian Federation and includes former Soviet states as well as Russian immigrants in other countries, the overwhelming majority of Russian language Twitter messages originates from Russia. Also note that the use of Twitter is not limited to large cities such as Moscow or St. Petersburg but rather also includes more rural and remote areas (Kelly et al., 2012).

4 The only underrepresented political groups were Russian right-wing extremists (Kelly et al., 2012).

5 In particular this included online news sites such as http://www.bbc.co.uk, http://www.ria.ru, http://en.wikipedia.org/wiki/2011--2013Russianprotests, http://tvrain.ru, http://lenta.ru/rubrics/russia

6 The term discourse analysis is often associated with a specific qualitative methodological approach advanced by Foucault, Laclau, Mouffe and others (Laclau, 1993; Weiss & Wodak, 2003), in this paper we use the term more generally to describe our analysis of written language use on Twitter in the context of political communication.

7 Since a belated follow-up extraction of followers of Twitter users is not facilitated by the Twitter API, the social network analyses are based on follower relations in 2015. See Supplementary Information S3 for further details.

8 No intercoder reliability can be provided for the collocation labelling or the manual classification of the sample of 100 active users in order to validate the automatic classification results, because only Viktoria Spaiser was a Russian speaker in the research group and thus only she could read and understand the Russian tweets and collocations.

9 We tested whether the association scores are distorted by the difference in linguistic heterogeneity between different groups, i.e., we wanted to check whether the pattern we see in the data are a results of pro-government users being more coordinated in the hashtags and phrases they use, and because of this consistency the collocations they use are more likely to be prevalent in the dataset. The opposition could still be dominating the Twittersphere in terms of number of tweets, but they could have expressed their regime criticism using more heterogeneous language, with less coordination, leading to fewer collocations showing up in the analysis. We therefore calculated the linguistic heterogeneity (lexical diversity) for each camp by diving the number of all words from the number of unique words (Bird et al., 2009). We found that the pro-Putin camp had in fact the highest linguistic heterogeneity with a score of 4.7366, while the oppositional camp had a lexical diversity score of 4.3156 and the neutral camp the lowest with 4.1295. Overall however, the scores are rather comparable.

10 The overall distributions of tweets per user in the camps are overall quite similar. Hence the pro-Putin fraction can be expected for any given day to represent the largest share of both active users and tweets posted. Since this advantage exists throughout the whole period analysed we can be sure that any shift in the political discourse is not simply an artefact of a change in the relative number of pro-Putin versus opposition users.

11 We examined our data for evidence of direct (technical) manipulation of the political Twitter discourse, particularly by the pro-Putin camp, but did not find any clear evidence for bot-produced and disseminated

pro-Putin messages. The Berkman Center researchers found that after applying a filter to the Russian Twitter data to clear the data from spam, the filtering also eliminated a number of pro-government thematic clusters (Kelly et al., 2012), i.e., especially pro-government political initiatives may have adopted aggressive online marketing strategies on Twitter. Such tweets may thus have been removed by filtering heuristics such as those applied by Twitter's Streaming API.

References

Arutunyan, A. (2009). *The Media in Russia*. Berkshire: Open University Press.

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & Gonzáles-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLoS ONE*, *10*(1), e0143611.

Becker, J. (2004). Lessons from Russia. A Neo-Authoritarian Media System. *European Journal of Communication*, *19*(2), 139–163.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Bennett, W. L., & Segerberg, A. (2013). *The Logic of Connective Action. Digital Media and the Personalization of Contentious Politics*. Cambridge: Cambridge University Press.

Bensusan, H., & Kalousis, A. (2001). Estimating the Predictive Accuracy of a Classifier. *Machine Learning: ECML 2001*, *2167*, 25–36.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol: O'Reilly.

Bitso, C., Fourie, I., & Bothma, T. J. D. (2012). Trends in transition from classical censorship to Internet censorship: selected country overviews. Published as a *FAIFE Spotlight* on the IFLA website. Retrieved from http://www.ifla.org/en/faife/spotlight.

Castells, M. (2007). Communication, Power and Counter-power in the Network Society. *International Journal of Communication*, *1*, 238–266.

Castells, M. (2009). *Communication Power*. Oxford: Oxford University Press.

Chadwick, A. (2013). *The Hybrid Media System: Politics and Power*. Oxford: Oxford University Press.

Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011a). Predicting the Political Alignment of Twitter Users. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, (pp. 192–199).

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011b). Political Polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, (pp. 89–96).

Cottle, S. (2011). Media and the Arab uprisings of 2011: Research notes. *Journalism*, *12(5)*, 647–659.

Diamond, L. (2010). Liberation technology. *Journal of Democracy*, *21(3)*, 69–83.

Domke, D. (2001). Racial Cues and Political Ideology. An Examination of Associative Priming. *Communication Research*, *28(6)*, 722–801.

Drezner, D. D., & Farrell, H. (2004). Web of Influence. *Foreign Policy*, *145*, 32–40.

Druckman, J. N. (2004). Priming the Vote: Campaign Effects in a U.S. Senate Election. *Political Psychology*, *25*(4), 577–594.

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *The American Economic Review*, *101*(7), 3253–3285.

Flick, U. (2006). *An Introduction to Qualitative Research. 3rd Edition.*. London: SAGE Publications.

Gaffney, D. (2010). #iranElection: Quantifying online activism. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line. Raleigh, NC*.

Golder, S. A., & Macy, M. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, *333 (6051)*, 1878–1881.

Gonzáles-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The Dynamics of Protest Recruitment through an Online Network. *Scientific Reports*, *1*(197), DOI: 10.1038/srep00197.

Gould, R. V. (1993). Collective action and network structure. *American Sociological Review*, *58*(2), 182–196.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420–1443.

Greene, S. A. (2013). Beyond Bolotnaia. Bridging Old and New in Russia's Election Protest Movement. *Problems of Post-Communism*, *60(2)*, 40–52.

Gunitsky, S. (2015). Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability. *Perspectives on Politics*, *13(1)*, 42–54.

Hanna, A., Sayre, B., Bode, L., Yang, J., & Shah, D. (2011). Mapping the Political Twitterverse: Candidates and Their Followers in the Midterms. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, (pp. 510–513). Association for the Advancement of Artificial Intelligence.

Herman, E. S. (1985). Diversity of News: 'Marginalizing' the Opposition. *Journal of Communication*, *35(3)*, 135–146.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M., Mari, W., & Mazaid, M. (2011). Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring? *Project on Information Technology & Political Islam (pIT-PI)*. Retrieved from http://pitpi.org/index.php/2011/09/11/opening-closed-regimes-what-was-the-role-of-social-media-during-the-arab-spring/.

Howard, P. N., & Parks, M. R. (2012). Social Media and Political Change: Capacity, Constraint, and Consequence. *Journal of Communication*, *62*(2), 359–362.

Kelly, J., Barash, V., Alexanyan, K., Etling, B., Faris, R., Gasser, U., & Palfrey, J. (2012). Mapping Russian Twitter. *The Berkman Center for Internet & Society Research Publication Series, Publication No. 2012-3*. Retrieved from http://cyber.law.harvard.edu/publications.

King, G., Pan, J., & Roberts, M. E. (2016). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. Harvard University, Department of Government Working Paper. Retrieved from http://gking.harvard.edu/files/gking/files/50c.pdf.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, *111*(29), 8788–8790.

Krebs, B. (2011). Twitter bots drown out anti-Kremlin tweets. *Krebs on Security Blog*. Retrieved from http://krebsonsecurity.com/2011/12/twitter-bots-drown-out-anti-kremlin-tweets/.

Kuran, T. (1989). Sparks and prairie fires: A theory of unanticipated political revolution. *Public choice*, *61*(1), 41–74.

Laclau, E. (1993). Discourse. In R. E. Goodin, & P. Pettit (Eds.) *The Blackwell Companion to Contemporary Political Philosophy*. Oxford: Blackwell.

Larsson, A. O., & Moe, H. (2011). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, *14(5)*, 729–747.

Lipman, M. (2005). Constrained or Irrelevant: The Media in Putin's Russia. *Current History*, *104.684*, 319–324.

Lohmann, S. (1994). The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91. *World Politics*, *47*(1), 42–101.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication*, *5*, 1375?1405.

Lynch, M. (2011). After Egypt: The Limits and Promise of Online Challenges to the Authoritarian Arab State. *Perspectives on Politics*, *9*(2), 301–310.

Manning, C. D., & Schuetze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Miller, G. (2011). Social Scientists Wade Into The Tweet Stream. *Science*, *333*, 1814–1815.

Noelle-Neumann, E. (1974). The spiral of silence: a theory of public opinion. *Journal of Communication*, *24(2)*, 287–298.

Noelle-Neumann, E. (1993). *The Spiral of Silence: Public Opinion - Our Social Skin*. Chicago: University of Chicago Press.

Oates, S. (2013). *Revolution Stalled. The Political Limits of the Internet in the Post-Soviet Sphere*. Oxford: Oxford University Press.

Perkins, J. (2010). *Python Text Processing with NLTK 2.0 Cookbook*. Birmingham: PACKT.

Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Detecting and tracking the spread of astroturf memes in microblog streams. In *Proceedings of 20th International World Wide Web Conference*, (pp. 249–252).

Roberts, H., & Etling, B. (2011). Coordinated attacks during Russian Duma elections. *Internet & Democracy Blog*. Retrieved from http://blogs.law.harvard.edu/idblog/2011/12/08/coordinated-ddos-attack-during-russian-duma-elections/.

Rød, E. G., & Weidmann, N. B. (2015). Empowering activists or autocrats? The Internet in authoritarian regimes. *Journal of Peace Research*, *52*(3), 338–351.

Roskos-Ewoldsen, D. R., Klinger, M. R., & Roskos-Ewoldsen, B. (2011). Media Priming: A Meta-Analysis. In R. W. Preiss, B. M. Gayle, N. Burrell, M. Allen, & J. Bryant (Eds.) *Mass Media Effects Research. Advances Through Meta-Analysis*. New York and London: Routledge.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346(6213)*, 1063–1064.

Sakwa, R. (2014). Whatever Happened to the Russian Opposition? Chatham House. The Royal Institute of International Affairs. Russia and Eurasia Programme. Research Paper.

Saldana, J. (2013). *The Coding Manual for Qualitative Researchers*. London: SAGE.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York & London: WW Norton & Company.

Shirky, C. (2011). The Political Power of Social Media. Technology, the Public Sphere, and Political Change. *Foreign Affairs*, *January/February*, http://www.foreignaffairs.com/articles/67038/clay--shirky/the--political--power--of--social--media.

Siegel, D. A. (2009). Social networks and collective action. *American Journal of Political Science*, *53*(1), 122–138.

Silitski, V. (2005). Preempting Democracy: The Case of Belarus. *Journal of Democracy, 16(4)*, 83–97.

Slove, D. J. (2007). *The future of reputation: gossip, rumor, and privacy on the internet*. New Haven: Yale University Press.

Srivastava, M. (2016). How Erdogan turned to social media to help foil coup in Turkey. Financial Times. Retrieved from https://next.ft.com/content/3ab2a66c-4b59-11e6-88c5-db83e98a590a.

Thompson, A. (2007). *The Media and the Rwanda Genocide*. London: Pluto Press.

Tilly, C. (1978). *From mobilization to revolution*. New York: Mcgraw-Hill College.

Tonkin, E., Pfeiffer, H. D., & Tourte, G. (2012). Twitter, Information sharing and the London riots? *American Society for Information Science and Technology Bulletin*, *38*(2).

Tufekci, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, *62*(2), 363–379.

Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*, (pp. 178–185).

Walker, S. (2015). Salutin' Putin: inside a Russian troll house. The Guardian. Retrieved from http://www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house.

Warren, T. C. (2014). Not by the sword alone: Soft power, mass media, and the production of state sovereignty. *International Organization*, *68*(1), 111–141.

Weichselbraun, A., Gindl, S., & Scharl, A. (2010). A context-dependent supervised learning approach to sentiment detection in large textual databases. *Journal of Information and Data Management*, *1*(3), 329–342.

Weidmann, N. B., Benitez-Baleato, S., Hunziker, P., Glatz, E., & Dimitropoulos, X. (2016). Digital discrimination: Political bias in Internet service provision across ethnic groups. *Science*, *353*(6304), 1151–1155.

Weiss, G., & Wodak, R. (2003). *Critical Discourse Analysis*. New York: Palgrave Macmillan.

Williams, C. (2011). How Egypt shut down the internet. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/worldnews/africaandindianocean/egypt/8288163/How-Egypt-shut-down-the-internet.html.

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, *35*(3), 399–433.

Wu, Y., Wong, J., Deng, Y., & Chang, K. (2011). An Exploration of Social Media in Public Opinion Convergence: Elaboration Likelihood and Semantic Networks on Political Events. In *Proceedings of the IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, (pp. 903–910).

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327.

Zarrella, D. (2009). State of the Twittersphere. Retrieved from http://blog.hubspot.com/Portals/249/sotwitter09.pdf.