Network Structure, Brokerage, and Framing: How the Internet and Social Media Facilitate High-Risk Collective Action

Bruce E. Etling
St. Antony's College

Trinity Term 2016

73,349 Words

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information, Communication, and the Social Sciences at the Oxford Internet Institute, University of Oxford

Abstract

Network Structure, Brokerage, and Framing: How the Internet and Social Media Facilitate High-Risk Collective Action

Bruce E. Etling, St. Antony's College
Thesis submitted for Doctor of Philosophy in Information, Communication, and the Social Sciences
Trinity Term 2016

This thesis investigates the role of network structure, brokerage, and framing in high-risk collective action. I use the protest movement that emerged in Russia following falsified national elections in 2011 and 2012 as an empirical case study. I draw on a unique dataset of nearly 30,000 online documents and the linking structure of over 3,500 Russian Web sites. I employ a range of computational social science methods, including Exponential Random Graph Modeling, an advanced statistical model for social networks, social network analysis, machine learning, and latent semantic analysis.

I address three research questions in this thesis. The first asks if a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if that structure changes over time. Polycentric networks are conducive to high-risk collective action and are robust to the targeted removal of key nodes, while hierarchical networks can more easily mobilize protesters and spread information. I find that the Russian protest network has a polycentric structure only at the beginning of the protests, and moves towards a less effective hierarchical structure as the movement loses popular support. The second research question seeks to understand if brokered text is actually novel, and if that text is more novel in polycentric networks than in hierarchical ones. Brokers are the individuals or nodes in a network that connect disparate groups through weak ties and close structural holes. Brokers are advantageous because they have access to and spread novel information. I find that the text among nodes in brokered relationships is indeed novel, but that information novelty decreases when networks have a hierarchical structure. The last research question asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames, and within such a movement, whether moderate or extremist framing is more prevalent. I find that the opposition is far more effective than the government in spreading its frames, even when the government organizes massive counter protests. Within the movement, moderates are more likely to have their framing adopted online than extremists, unless violence occurs at protests.

The findings suggest that movements should build flatter, more diffuse networks by ensuring that brokers tie together diverse protest constituencies. The findings also provide evidence against those who claim that authoritarian governments are more effective in shaping online discourse than oppositional movements, and also suggest that movements should advance moderate framing in order to attract a wider base of support among the general population.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I have had two intellectual homes while writing this thesis: the Berkman Center for Internet and Society at Harvard University (recently renamed the Berkman Klein Center), and the Oxford Internet Institute (OII). I benefited from the intellectual curiosity and support of the faculty, students, fellows, staff, and affiliates at both centers. Most of the data used in this thesis was drawn from the Media Cloud platform, so I owe special thanks to the entire Media Cloud team at the Berkman Center, and especially Hal Roberts for setting up the platform, keeping it running smoothly, and providing the protest data. Rob Faris has also been a constant source of inspiration and support since I joined the Berkman Center, a great sounding board for ideas good and bad, and a wonderful research partner. I am grateful to the Berkman Center for financial support during nearly all my time as a doctoral student. Thanks to Berkman Center's current and former leadership for creating a stimulating, diverse, and welcoming environment, especially John Palfrey, Yochai Benkler, Colin Maclay, Urs Gasser, and Jonathan Zittrain. This thesis grew from the work I did as part of the Internet & Democracy project at Berkman, and I benefited from all those that worked on the project over the years, especially those who worked on the Russia project, including John Kelly, Karina Alexanyan, Vladimir Barash, and Gregory Asmolov. I owe a special debt of gratitude to John Kelly for introducing me to social network analysis, providing the LiveJournal blogs that form the backbone of the Russian data in Media Cloud, and for encouraging me to pursue a doctorate. I owe thanks as well to Ethan Zuckerman at MIT's Center for Civic

Media for conceiving of the Media Cloud platform, and for his critical approach to digital activism. Thanks to his whole team at MIT for nurturing and building out the Media Cloud platform. I also benefited from conversations with several individuals affiliated with the Center for New Media & Society in Moscow, including Floriana Fossato, Sam Greene, Ivan Klimov, and Olga Kravtsova. Sam Greene's work on the Russian election protests was an important source of ideas as I planned the thesis. Besides the data provided by the Berkman Center, the work in this thesis is entirely my own.

I found the culture and community at OII to be equally open, supportive, and inspiring. I also appreciated OII's emphasis on serious social science research. I thank the entire OII community, but particularly Vicki Nash, Eric Meyer, Grant Blank, Bill Dutton, Vili Lehdonvirta, and Jonathan Bright for support while pursing the DPhil or feedback at various stages of the thesis. I am particularly grateful to OII for financial support through the OII scholarship. Thanks in particular go to my supervisors at Oxford, Bernie Hogan and Paul Chaisty. Both were willing to take on supervision of a thesis that was not perfectly aligned with their research interests and provided exceptional guidance. I am especially grateful to Bernie for his sense of humor and for sharing his expertise in social networks, coding, and computational social science methods. Paul was a great resource on Russian politics, but also helped me envision the thesis as a whole, provided excellent, timely feedback on individual chapters, and helped prepare me for each assessment phase. Sandra Gonzalez-Bailon provided excellent supervision the first year of the thesis, and her work remains a touchstone in the field. I also thank all of my fellow DPhil students for constructive feedback throughout the development of the thesis.

Finally, special thanks to my wife Sierra for her encouragement and support, for taking on additional childcare duties, and for putting up with all the missed weekends and holidays so I could write, and finally finish, the thesis.

Chapter 1 Introduction

You cannot stop people any longer. You cannot control them any longer. They can bypass your established media; they can broadcast to one another; they can organize as never before. … [The Millennial generation] will use technology to displace old modes and orders.

--Andrew Sullivan, "The Revolution Will be Twittered"

Facebook and the like are tools for building networks, which are opposite, in structure and character, of hierarchies. Because networks don't have a centralized leadership structure and clear lines of authority, they have real difficulty reaching consensus and setting goals. They can't think strategically; they are chronically prone to conflict and error. … If you're taking on a powerful and organized establishment you have to be a hierarchy.

--Malcolm Gladwell, "Small Change: Why The Revolution Will Not Be Tweeted"

A wave of recent protests, from the Arab Spring to Occupy Wall Street, has marked an increasing shift in politics away from formal institutions and towards the streets. One of the most heated debates surrounding these protest movements is whether the Internet and social media have played any role in them. The overarching question examined in this thesis is how networked communication technologies facilitate high-risk collective action in hybrid regimes. Hybrid regimes are those that have some democratic attributes, but that are largely authoritarian in nature (Diamond, 2002; Levitsky & Way, 2002, 2010). To narrow this otherwise impossibly broad question, I focus on three general themes that can be empirically tested and that have a solid theoretical base in the social sciences: the role of network structure, brokerage, and framing in high-risk collective action. Instead of analyzing a particular online tool or platform, I focus on the sociological processes at play within protest movements. This approach allows for greater generalizability of the results, and also makes it possible to take on several claims made by skeptics of social media's potential to facilitate risky collective action. These include claims that the Internet creates flat networks, which are poorly suited to high-risk collective action; that governments are better able to shape online discourse than protest movements; that the Internet emboldens extremists instead of moderates; and that the Internet creates echo chambers that make it difficult for novel ideas and information to flow between clusters of homophilous, strongly-connected groups.

I begin this thesis by developing a theoretical framework for understanding the role of networked communication technologies in high-risk collective action. I start by examining the importance of network structure and tie strength for high-risk collective action. One common belief is that networked communication platforms create flat networks instead of hierarchical ones (Gladwell, 2010; Shirky, 2008, 2011). Conventional wisdom holds that hierarchical networks are better than flat networks for high-risk collective action because hierarchical networks have leaders at the top and clear lines of authority. This facilitates decision-making, strategy development, and reduces conflict (Gladwell, 2010; McAdam, 1988). Theoretically, hierarchical networks are also able to diffuse information more quickly and to turn out protesters more easily, because the lines of communication are short and most people are connected directly to key nodes in the center of the network.

However, the collective action and network science literatures have also shown that flat, or in network terminology, polycentric or small-world networks, are actually more conducive to high-risk collective action than hierarchical ones (Baldassari & Diani, 2007; Gerlach, 2001). Polycentric networks are made up of both strong and weak ties. These networks have clusters of strongly tied nodes that are connected by brokers through weak ties. Polycentric networks are conducive to high-risk collective action because they are more robust than hierarchical networks to targeted removal of nodes or forced defections, and are therefore more resistant to repression. Social movements with this type structure tend to have multiple centers of power and are made up of diverse groups who operate independently of any central authority. This structure allows movements to penetrate and recruit from different segments of society, and has also been found to encourage social innovation and problem solving (Baldassari & Diani, 2007; Gerlach, 2001).

It is also argued that the Internet and social media are more likely to facilitate the creation of weak ties instead of strong ties (D. Faris, 2008; Gladwell, 2010; Shirky, 2008). The conventional view is that strong ties are necessary for high-risk collective action. Studies of the civil rights movement in the United States, for example, show that those most likely to engage in high-risk collective action had strong tie connections to others in the movement (Gladwell, 2010; McAdam, 1986, 1988). Strong ties also help to build internal group solidarities, are more likely to influence behavior than weak ties, allow protest participants to receive repeated affirmation of protest information, and reinforce the normative importance of participation (Centola, 2010; Centola & Macy, 2007; McAdam, 1988; McAdam & Paulsen, 1993). However, weak ties allow groups of friends with strong tie bonds to receive novel information from groups they would not otherwise be exposed to, and have also been shown to aid in community mobilization and diffusion processes (Borgatti & Lopez-Kidwell, 2011; Granovetter, 1973, 1983). Weak ties also allow disparate social groups that rarely interact with each other to come together under a common protest banner. While weak ties and strong ties are often portrayed as in competition with one another and leading to binary outcomes for movements, it is better to view them as complementary (Tufekci, 2010). Polycentric networks, being made up of both strong and weak ties, can draw on the benefits of both type of tie, and also correct for some of the weaknesses of each.

Another concern, more generally, is that networked communication is leading people increasingly to isolate themselves into echo chambers of the likeminded, filtering out information they disagree with, and surrounding themselves with people who are similar to them and who hold similar opinions (Pariser, 2011; Sunstein, 2007). It is believed that this process is contributing to increased polarization in politics, which can exacerbate racial and ethnic tensions. In Western democratic contexts, it is argued that this is undermining democracy, as the collapse of civic institutions has removed one of the key ways that we connect with, understand, and empathize with those who are different than us (Putnam, 2000). However, when networks have a polycentric structure, brokers are able to bridge otherwise isolated clusters of strongly connected individuals. Empirical evidence from social movements show that brokers are able to tie together diverse political groups and social classes, aid in the diffusion of revolutionary ideas, and contribute to the emergence of broad-based, geographically dispersed movements instead of isolated, local ones (D. Faris, 2008; Han, 2009; Hedstrom, Sandell, & Stern, 2000; Kim & Pfaff, 2012). The main reason that brokers are advantageous is because they provide access to novel ideas and information that we would not be exposed to from our close family and friends (Burt, 1992, 2004; Granovetter, 1973, 1974; Putnam, 2000). This is a bedrock assumption of the brokerage literature, but it is rarely actually tested.

Another common argument among those skeptical of the ability of the Internet and social media to aid protest movements, especially in authoritarian settings, is that these tools are actually more beneficial to governments than they are to protesters (Gunitsky, 2015; Kalathil & Boas, 2003). These skeptics contend that authoritarian and semi-authoritarian regimes are becoming increasingly sophisticated in their control of online speech, especially as it relates to criticism of the regime and organization of collective action. Beyond just blocking offensive content, critics maintain that governments are now able to use paid armies of supporters and other techniques to shape online conversations to their own advantage (Deibert, Palfrey, Rohozinski, & Zittrain, 2010; Fossato & Lloyd, 2008; Gunitsky, 2015). Russia is frequently cited as an example of this type of tactic. There is also a concern that the Internet and social media are leading to a rise in extremism (Copsey, 2003; Perry & Olsson, 2009; Wojcieszak, 2010). Those who previously held views outside the mainstream can now more easily find each other online and form communities. There are concerns that this community formation and online mutual support for extremist views may lead to offline acts of aggression. There are also concerns that individuals who would never previously align themselves publicly with extremist views can do so anonymously online. For collective action, this is a major concern; since if a movement is seen as dominated by extremists, it is less likely to win the support of moderates and gain a wide following (Gitlin, 1980; Lohmann, 1994). Authoritarian governments also frequently frame popular protest movements as led by extremists in order to undermine them.

Research questions

These debates lead to the three core research questions that I investigate in this thesis. The first asks if a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if that structure changes over time. Again, critics argue that the Internet and social media facilitate the creation of movements that are flat, which in network science are known as polycentric networks. There are potential benefits for both type of network structure. Polycentric networks are conducive to high-risk collective action and are more robust to defections or the removal of key nodes than hierarchical networks, while hierarchical networks are able to turn out protesters and diffuse information quickly. The second research question seeks to understand if brokered information is actually novel, and if brokered text is more novel in polycentric networks than in hierarchical ones. Brokers are the individuals or nodes in a network that, via weak ties, connect disparate groups and close structural holes. They are believed to be essential to protest movement emergence and success. The ability to access, share, and translate heterogeneous information between different groups is the primary explanation for the many benefits that accrue to brokers and the groups they connect, but this is rarely actually tested. The last research question asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames, and within such a movement, whether moderate or extremist framing is more prevalent.

Collective action in hybrid regimes

Research on the impact of networked communication technologies on collective action focuses in large part on protests and social movements in advanced democracies. This is logical given that Internet penetration is greater in those countries with more advanced, richer economies and, thus, social media tools were more readily available for social movements and protesters in Western democracies. The initial focus on the United States and Western Europe has begun to shift, especially following the protests connected to the Arab Spring and interest in social media's role in those events (Hussain & Howard, 2013; Tufekci & Wilson, 2012). And among autocracies, studies of how social media contribute to an alternative public sphere and facilitate collective action in China are increasingly frequent, as are studies of government attempts to censor online speech and the organization of collective action (Huang & Yip, 2012; King, Pan, & Roberts, 2013; Yang, 2009). However, studies that investigate how networked communication technologies in hybrid regimes facilitate collective action are still relatively rare.

Hybrid regimes fill an important space between liberal democracies and authoritarian regimes. After democracies, hybrid regimes are the most frequent type of regime in the world (Diamond, 2002). Hybrid regimes are broadly defined as political systems that have some democratic attributes but which are largely authoritarian in nature. Electoral authoritarianism is one common type of hybrid regime, and describes states where elections are seen as the only legitimate source of political authority, but incumbents so frequently violate basic tenants of democracy that the playing field and rules of the game are heavily tilted in favor of incumbents. Comparative political scientists also call these states competitive authoritarian regimes or pseudo-democracies (Diamond, 2002; Levitsky & Way, 2002, 2010). In these states routine elections occur but are largely falsified, opposition parties exist but are prevented from participating in elections, and online criticism of the regime and even organized collective action may be allowed, but the government often tightly controls mainstream media. Further, tax authorities, the judiciary, and other compliant state institutions are used to harass, jail, and otherwise neutralize the government's political opponents. In these types of regimes, barriers to political competition for the opposition are so high that governing elites have little chance of being removed from power (Diamond, 2002; Levitsky & Way, 2002, 2010).

This thesis focuses on an important case of political protest in Russia, which is a prime example of a hybrid regime. There are two general views on the impact of the Internet on civil society in Russia that mirror broader debates about the impact of networked communication technologies on high-risk collective action. The more dystopian school doubts the Internet will be able to overcome the power of the state or bring much real political change to an increasingly authoritarian regime. For example, in the first major study of the Internet and civic activity in Russia, Fossato and Lloyd (2008) argue that the qualitative level of Internet discussion is low and that lack of trust online is widespread and skillfully manipulated by authorities. Further, online networks seem to be generally rather closed and intolerant, leaders of Internet sites can often be co-opted, compromised or frightened; and Russian Internet users do not respond actively to political campaigning on the Web. Deibert et al. (2010) argue that even though they find no technical filtering in Russia, as is typical in most semi-authoritarian states, the Russian government instead engages in more subtle 'third generation' controls designed to shape and affect when and how information is received rather than denying access to it outright. The methods the Russian government employs include support for pro-Kremlin bloggers to push the government's messages and discredit opposition voices online, and support for groups that carry out Distributed Denial of Service (DDoS) and other hacking attacks against perceived enemies of the state (Etling et al., 2010).

The other school of thought has found evidence of a more active Russian online civil society, defined as individuals undertaking activities that benefit society broadly, but do so outside the bounds of the state (Diamond, 2008). Through social network analysis methods, Etling et al. (2010) find clusters, or communities, of bloggers that actively discuss politics from a 'non-aligned' perspective, a small cluster of bloggers affiliated with formal opposition political parties, as well as a large number of nationalists. Contrary to concerns among the more dystopian school, pro-government bloggers are not prominently linked to and do not form their own cluster within the blog network. They also find many examples of online mobilization, including online and offline actions related to drivers' rights, the environment, nationalism, and human rights. Summarizing a three-year study of the Russian Internet, Alexanyan et al. (2012) find the emergence of the Russian Internet as an alternative public sphere, the growing use of the digital platform in social mobilization and political action, and that the

Internet in Russia remains generally open and free. Through case studies of small-scale protests related to health and children's issues, Oates (2013) also shows that Russians are using the Internet to demand their rights and air individual grievances against local officials. Oates also argues that the government has failed to control the online space as effectively as traditional media in Russia.

Russia is worthy of study for a number of reasons. Through analysis of the Russia case, we can gain a better understanding of how the Internet facilitates political participation in hybrid regimes. In other words, we learn how those who are prohibited from engaging in formal electoral politics engage in 'politics by other means.' Russia is also a useful case because until recently, unlike China, Russia took a very hands off approach to Internet regulation. For example, Russia blocks a small number of Web sites compared to China and has a much less sophisticated infrastructure for blocking content. This light regulatory touch allowed the Russian Internet to emerge as an alternative public sphere where oppositional, critical voices could emerge and a civil society develop. It is believed this is one of the factors that contributed to the emergence of the election protest movement that is the focus of this thesis (Litvinenko, 2012; Shvetsova, 2012).

Empirical case

The Russian election protest movement of 2011-2012 is used as an empirical case study in this thesis. This movement emerged spontaneously after falsified parliamentary election in December 2011 (OSCE, 2012). Over the next six months, despite government countermeasures, freezing temperatures, and the arrest and jailing of protest leaders, the movement was able to organize the largest, most successful protests in Russia since the fall of the Soviet Union. The movement consisted of a loose coalition of oppositional civic groups, formal political parties, and individual activists with a range of different political affiliations. The movement included democrats, nationalists, leftists, independent journalists, writers, musicians, and activists associated with anti-corruption, environmental, drivers' rights, and gay rights causes (Greene, 2013; Ioffe, 2011c; Popescu, 2012; Smyth, 2013; Taylor, 2011). These movement elements drew heavily on the non-systemic opposition in Russia, or those individuals, civic organizations, and political parties that are either informally or legally barred from competing in elections. Large, public demonstrations were the main method the movement used to publicize election fraud, make claims against the government, and gain support among the broader Russian population. While the largest protests took place in major urban centers, such as Moscow and St. Petersburg, protests took place across the country. At one point, the focus of the national movement even moved to the region of Astrakhan, in support of a mayoral candidate who believed local elections there were falsified.

The Internet played a significant role in the Russian protests. Networked communication technologies were used to collect and distribute evidence of election fraud, to publicize protest events, to mobilize supporters, and to coordinate the protests (Greene, 2012, 2013; Koesel & Bunce, 2012; Litvinenko, 2012; Shvetsova, 2012; S. White & McAllister, 2014). A number of collaborative online platforms, including Google docs and Facebook groups, also allowed users to vote on who should speak at the protest events, providing a more interactive and bottom-up form of protest organization and participation. Several of the movement's leaders, including Alexei Navalny and Evgenia Chirikova, also built their reputations online in the years leading up to the protests, and used the Internet to promote their causes and build supportive online communities. The election protests also built off of many instances of successful civic action in which citizens used the Internet and social media to fight against the government over specific or very localized grievances (Oates, 2013). The Russian election protest movement is a good example of a contemporary, networked social movement.

Understanding the affordances of the Internet for protests in reaction to election falsification is important since this type of protest is likely to continue to occur. This is because leaders in hybrid regimes must hold routine elections in hopes of gaining domestic and international legitimacy, but they must also increasingly falsify them to ensure their own victory. Russia is also a useful case to study because the protest movement was not ultimately successful in overturning the election results and forcing the government to hold new elections, although the movement did win some concessions from the government resulting in changes to elections and political party laws (Koesel & Bunce, 2012). So for some this is a case of a failed movement. However, social movements are long-term, often multi-year efforts. All movements go through periods when they score major victories, and others when they suffer significant losses. Tarrow (1998) calls these cycles of contention. By looking at the Russian protest movement over its entire life cycle, we can gain a more nuanced appreciation of how movements change over time, and how they operate during periods of both success and failure.

Data and methods

This thesis draws on a unique dataset of nearly 30,000 Russian-language documents and the linking structure of over 3,500 Web sites that discussed the Russian election protests between December 1, 2011 and May 31, 2012. These dates include the six months of peak protest activity by the movement, as well as the parliamentary and presidential elections. The dataset was created by first using a regular expression to mine a large database with text and link content from tens of thousands of Russian-language blogs, news sites, and government Web sites to identify those that discussed the protests. A Web crawler was then used to identify additional Russian-language Web sites, blogs, forums, and social networking sites that also discussed the protests. This data set is unique because it includes data from multiple platforms, a rarity among studies of online collective behavior. It is also significant that I use data that covers the entire period of the protests, instead of just one month or one major protest event. This makes it possible to study how the protest communication network changes over time, as well as the success of framing by different groups as time passes.

I use a variety of computational social science methods to analyze this data. To understand whether the network has a polycentric or hierarchical structure I build and fit a series of Exponential Random Graph Models (ERGMs). ERGMs are advanced statistical models for social networks that allow researchers to account for the presence or absence of ties in a network, and therefore a way to understand the micro-processes that account for macro structure in an observed network of interest. An ERGM models a given network in terms of a set of theoretically relevant parameters, or terms, selected by the researcher that account for local tie-based structures. These parameters consist of processes that are exogenous and independent of network ties (such as node attributes, like location or media type) as well as endogenous or structural terms, where observations are dependent (such as reciprocity or transitivity). ERGMs therefore allow social scientists to understand how these endogenous and exogenous parameters interact to create and maintain network structure (Lusher, Koskinen, & Robins, 2013).

Supervised learning, or machine learning, methods are used to investigate whether the government or the opposition is more effective at spreading their preferred framing of the protests, as well as the prevalence of extremist versus mainstream frames. Machine learning requires the creation of a small training set of documents that represent categories of interest to the researcher. I create a training set that includes sources and text that represent the framing of each of the major protest constituencies, as well pro-government sources and the frames that they use to describe the protests. I then use a machine-learning algorithm created by Hopkins and King (2010) to identify the proportion of text in the larger corpus that matches the framing of each of the categories of interest over each week of the protests.

I use a combination of both social network analysis and automated text analysis techniques to investigate if brokered text is novel. These techniques are also used to understand if local information novelty is affected by global network structure. I use a community detection algorithm to partition the network, identify brokers as defined by Gould and Fernandez (1989), extract the ego networks for each broker, and then use Latent Semantic Analysis (LSA) to measure the novelty of text among nodes in the brokered relationship. Latent Semantic Analysis (LSA) is a method used to identify latent topics in documents through the co-occurrence of words (Landauer & Dumais, 1996, 1997; Landauer, Foltz, & Laham, 1998).

Structure of the thesis

This thesis proceeds as follows. In the next chapter, Chapter 2, I develop the analytical framework used in the thesis. In that chapter, I review the relevant literature on networks, tie strength, collective action, and framing. I also lay out the specific research questions asked in the thesis, and develop the hypotheses tested in each of the substantive chapters.

Chapter 3 provides a detailed description of the empirical case. As noted above, I use the Russian election protest movement that emerged following falsified elections as a case study in this thesis. I begin with a description of the falsified legislative elections and the general conditions that led to the movement's emergence. I then describe the different constituencies that made up this movement. I list the movement's major demands and also describe the importance of the Internet to the protests. The government's reaction to the protests is also discussed. A chronology of major protest events is also provided.

Chapter 4 describes the data used to answer this thesis's main research questions, the methods used to collect that data, and the limitations of the data. I discuss in detail the specific Russian-language data sources that served as the seed set for the protest data and the Media Cloud platform they were drawn from (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2015). I then describe the Web crawling method used to create the larger Russian protest data set that is used throughout this thesis, and provide some descriptive statistics for that data set. In this chapter, I also introduce the methods at a general level. More detailed discussion of the methods and specific slices of the data used to answer specific as are left for each of the substantive chapters.

Chapter 5 is the first substantive chapter. Chapter 5 addresses the first research question on the importance of network structure. Specifically, I investigate whether a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if this structure changes over time.

Chapter 6 seeks to understand why brokers matter, and investigates the second research question. I test if the text among nodes in brokered relationships is novel. I also investigate whether brokered text is more novel when networks have a polycentric structure.

Chapter 7 addresses the third research question. This chapter investigates the framing battles that took place during the protests, and asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames. I also investigate whether moderate or extremist framing is more prevalent within such a movement.

I conclude the thesis in Chapter 8. In that chapter, I discuss the major themes that emerge from the thesis's substantive findings. I discuss the importance of polycentric networks and how network structure impacts information novelty. I also reflect on the implications of the framing battles between the opposition and the government, and within the movement between moderates and extremists. I discuss the implication of the Russia-specific case as well. I highlight the methodological contribution and the benefits of computational social science methods for the study of collective action. Finally, I also address the study's main limitations and make suggestions for future research.

Chapter 2 Theoretical Framework

Understanding the impact of the Internet and social media on protest movements is one of the most important political science questions of our time. The Internet has arguably had an impact on nearly every major recent protest or revolutionary movement against authoritarian regimes. These movements span the globe. From the Iranian Green Revolution to the successful Arab uprisings in Tunisia, Libya, Yemen, and Egypt in the Middle East, to the 2011 Russian election protests and the successful Ukrainian Euromaidan protests in Eurasia, and the 2014 Hong Kong protests and the numerous smaller, primarily online campaigns against corruption and abuse of power in China, digitally enabled protests have struck and even overturned authoritarian regimes, many once thought invincible. And yet, despite this increased frequency of political protests and overthrow of long-standing dictators, no one theory or scholarly field has been able to explain the impact of the Internet on these protest movements and their success or failure.

In this chapter, I develop a multi-disciplinary analytical framework to explain how the Internet facilitates collective action in authoritarian and hybrid regimes. In doing so, I draw on theories and recent scholarship from network theory, political communication, sociology, political science, and the social movements literature. Where possible, I attempt to highlight causal processes and underlying social mechanisms that can be empirically tested. While much of the collective action and social movements literature I draw on is focused on collective action in democratic regimes, my focus in this thesis is on digitally enabled protests in authoritarian and hybrid regimes.

Differences in regime types have important implications for risks to protesters and the opportunity structures available to them, and so are worth discussing briefly. It is clear that protest movements in democratic regimes are subject to lower levels of risk and have more avenues for collective action available to them than in authoritarian regimes. However, there are also important differences between closed authoritarian systems and hybrid regimes, as well as differences among hybrid regimes themselves. Hybrid regimes are broadly defined as political systems that have some democratic attributes but which are largely authoritarian in nature. Electoral authoritarianism is one common type of hybrid regime, and describes states where elections are seen as the only legitimate source of political authority, but incumbents so frequently violate basic tenants of democracy that the playing field and rules of the game are heavily tilted in favor of incumbents. Comparative political scientists also call these states competitive authoritarian regimes or pseudo-democracies (Diamond, 2002; Levitsky & Way, 2002, 2010).

Whatever the label chosen, Russia before the protests is a good example of this common type of hybrid regime. Routine elections occurred but were largely falsified, opposition parties existed but were either 'in the pocket' of the Kremlin or prevented from participating in elections, and online criticism of the regime and even organized collective action was allowed, but TV channels were tightly controlled by the government. Further, tax authorities, the judiciary, and other compliant state institutions were frequently used to harass, jail, and otherwise neutralize the government's political opponents. In these types of regimes, barriers to political competition for the opposition are so high that governing elites have little chance of being removed from power, although it is possible if the opposition is extraordinarily unified or if the regime underestimates the level of discontent in society before elections. Generally, the more restrictive the hybrid regime, the more difficult it is for protest movements to organize collective action and the greater the risks to them. The upshot of operating in electoral authoritarian regimes for protest movements is that those states need to hold routine elections to maintain domestic and international legitimacy. This creates routine opportunities for large-scale protests if elections are seen as falsified. Election fraud thus becomes the primary mobilizing event for protesters and also provides a unifying, common grievance for the broader public to rally against (Diamond, 2002; Levitsky & Way, 2002, 2010; Tucker, 2007; Wegren & Herspring, 2010).

The framework developed in this chapter contributes to three research questions. The first asks if the Russian protest network has a polycentric or hierarchical structure, and if this changes over time. I hypothesize that the Russian protest network will have a polycentric structure only at the beginning of the protests, and will move towards a less effective hierarchical structure as the movement loses popular support. The second research question seeks to understand if brokered information is actually novel, and if brokered text is more novel in polycentric networks than in hierarchical ones. I hypothesize that the text of nodes from different communities that are connected by brokers will be novel, since information and ideas shared between groups tends to be novel compared to that shared within groups. I further hypothesize that the text of brokers will be the most novel when the network has a polycentric structure, since such structures will have many communities of different types discussing a diverse range of topics, more structural holes, and more brokers and weak ties connecting disparate communities than a hierarchical network. The last research question asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames, and within such a movement, whether moderate or extremist framing is more prevalent. I hypothesize that the Russian protest election movement will be more effective than the government at spreading its preferred frames online, except when the government organizes large pro-government rallies. I also hypothesize that within the movement moderates will be more successful than extremists at spreading their frames, unless violence occurs at protest events.

I begin this chapter by drawing on network theory to explain how both strong ties and weak ties aid social movements. Weak ties allow individuals to receive novel information not available from their immediate social circles, and are also believed to aid in community mobilization and diffusion processes (Borgatti & Lopez-Kidwell, 2011; Granovetter, 1973). Strong ties help to build internal group solidarity and exert more influence on individuals than weak ties. Strong ties also allow for social reinforcement of the credibility of protest information and the normative importance of participating in risky collective action (Centola & Macy, 2007). Polycentric networks have both strong and weak tie structures. They are more robust than hierarchical networks, more resistant to repression, can penetrate and recruit from different segments of society, and encourage social innovation and problem solving (Baldassari & Diani, 2007; Gerlach, 2001). Drawing further on network and social movement theory, the importance of bridges and brokers is examined next. Bridging ties and brokers connect disparate constituencies in protest movements into a larger whole and aid in the spread of political movements. Finally, I tap the political communication and the social movements literature to explain the importance of framing, which, among other attributes, allow social movements to gain new adherents and spread their message, in their own terms, to the larger public.

The relevance of tie strength and network structure

Both strong and weak ties are necessary for risky collective action. Strong ties build internal group solidarity and are more likely to influence individual behavior than weak ties. Strong ties also allow for both repeated affirmation of protest information and repeated emphasis on the normative importance of participation in risky collective action. Weak ties allow for individuals, or groups, to receive novel information that they are unlikely to receive from their close friends or kin. Weak ties can also help to tie together disparate constituencies that can develop into a larger, more integrated movement with a common purpose.

Granovetter (1973) originally proposed that tie strength could be measured by the frequency of interaction between two individuals, along with the emotional intensity, intimacy, and reciprocity of a relationship. Intuitively, strong ties are generally those that occur between close friends and family. Those ties are also likely to be affectively charged and highly salient to both individuals. There is also often a correlation between strong ties and homophily – that is, strong ties are more likely between individuals who are similar to each other. It has also been suggested that there is a geographically spatial element to strong ties. Those who live or work in close proximity to each other are more likely to form strong tie type relationships than those who live far from one another. Weak ties are more likely to connect acquaintances that interact less frequently, are less emotionally invested in one another, and are less likely to influence each other than strong ties (Centola & Macy, 2007; Granovetter, 1973). Empirical evidence has shown that a measure of 'closeness,' or the emotional intensity of a relationship, is the best indicator of tie strength compared to others initially proposed by Granovetter (Marsden & Campbell, 1984).

According to Granovetter's (1973, 1974, 1983) strength of weak ties theory, weak ties are more important than strong ties for information diffusion and the spread of novel information or new ideas. This is because cliques, or more densely connected clusters of individuals, are able to receive novel information that they would not be exposed to from their strong ties, or friends. Granovetter (1973) writes, "Weak ties are more likely to link members of different small groups than are strong ones, which tend to be concentrated within particular groups" (p. 1376). Granovetter showed how this counterintuitive finding occurs in the real world through his study of job seekers, uncovering the now widely-cited finding that weak ties are more important in helping job seekers find work, even though one's close friends would theoretically be more invested in helping one find employment.

Granovetter also suggests that weak ties can play a role in community mobilization. Granovetter (1973, 1983) drew on a case study by Gans (1961) of community mobilization by two neighborhoods, the West End and Charlestown, against assimilation into the city of Boston. Granovetter argues that the West

End community was assimilated into Boston while Charlestown remained independent because Charlestown possessed a greater number of weak ties that facilitated community-level organizing against assimilation. The West End, as a bedroom community where people slept while tending to work in Boston or elsewhere had a social structure of distinct clusters with many dense strong ties, but a dearth of bridging weak ties. It also lacked social organizations, which are often sources of weak ties. This lack of bridging ties between different groups in the West End inhibited organization against the assimilation threat because it led to fragmentation of the community and distrust of leaders (Granovetter, 1983). In contrast, Charlestown residents worked in the community and had more opportunities to interact with their neighbors during their everyday work and social lives. Borgatti and Lopez-Kidwell (2011) extend Granovetter's theory as it pertains to community organizing by arguing that Charlestown's diffuse weak-tie structure created group-level social capital that allowed it to unify around a common goal, in this case mobilizing resources and organizing community action in response to the assimilation threat.

The notion of weak and strong ties and their role in strengthening communities and supporting democracy are very similar to Putnam's (2000) notion of bridging and bonding social capital. He developed this idea further with Goss, writing that:

> Bridging social capital refers to social networks that bring together people of different sorts, and bonding capital bring together people of a similar sort. This is an important distinction because the externalities of groups that are bridging are likely to be positive, while networks that are bonding (limited within particular social niches) are at greater risk of producing externalities that are negative. (Putnam & Goss, 2002, p. 11)

In this view, there is a danger to societies with strong ethnic and religious cleavages that do not have both bonding and bridging social capital. Civil strife, such as the ethnically and religiously-fuelled war in Bosnia, can result from communities made up predominately of strong ties between homogeneous individuals but lacking bridging social capital that, according to Putnam, is provided by civic associations.

There are some problems with weak ties, though. Since we interact primarily with our strong ties, it is very hard to access people that are more than one hop away from our immediate social network (Watts, 2003). Since we have little knowledge of people that we are weakly connected to, it is extremely difficult to know what information they have that could be useful to us. To access information from weak ties we must go through an intermediary. That intermediary may not necessarily share information as easily as Granovetter and other advocates of weak ties hypothesize. For example, Marin (2012) finds that information holders with the opportunity to mention job openings with those in their network share that information only 27% of the time. This is because people hesitate to share information with others unless they know that information is being sought. Since we are more likely to know the career goals of our strong ties, information about jobs flows more frequently to them. The information that comes from weak ties may also not be as novel as one would expect. Aral and Van Alstyne (2011), for example, show that since interactions with weak ties are by their nature rare, the information that flows through them is slower and of lower volume, or bandwidth, than that from strong ties. They argue that high-bandwidth channels can provide access to more diverse and non-redundant information. This is because information shared through high bandwidth channels is likely to be detailed, cover multiple topics, and address complex and interdependent ideas. Brokers can therefore be at a disadvantage because lower bandwidth communication from spatially distant nodes actually reduces the amount of novel information received. Finally, it is not obvious that weak ties operate the same way in different political contexts. For example, Bian (1997) finds that in China government-assigned jobs are acquired more frequently through strong ties than weak ties. More importantly for this thesis, a weak tie relationship will typically not provide access to sensitive information because there is less trust between individuals connected by weak ties (Aral & Van Alstyne, 2011). In an authoritarian country, where risks for dissident activity are high, weak ties would appear to be poor conduits of sensitive protest information.

Strong ties can help to overcome some of the problems affiliated with weak ties, especially those related to risky collective action. Strong ties serve to reinforce group solidarities, and individuals are more likely to be influenced by their strong ties than weak ones. As Centola and Macy (2007) argue, "Strong ties increase the trust we place in close informants, the exposure we incur from contagious intimates, and the influence of close friends" (p. 703). Centola and Macy further argue that a supposed weakness of strong ties, that information from them is likely to be redundant, can actually be an advantage for risky collective action. This is because the willingness to participate in such risky behavior may require independent affirmation and reinforcement from multiple sources. In their view, weak ties may spread simple information, but strong ties are needed to turn that information into action. Multiple contacts in risky political contexts serve to both verify the information's credibility and emphasize the normative importance of taking action. In an online experiment testing this hypothesis, Centola (2010) found that individual adoption of a behavior is more likely in clustered networks where individuals receive social reinforcement from multiple neighbors. Centola also found that this behavior spread more quickly and reached more people in clustered networks made up of strong ties, than in random ones.

The social movement literature also provides evidence that strong ties are important for recruitment to high-risk activism. McAdam (1986, 1988), as well as many Internet skeptics (Gladwell, 2010), conclude that strong ties are more important than weak ties for social movements. As evidence for this claim, these authors point to the significant role that strong ties played in recruitment of student activists during the high-risk activism of the U.S. civil rights movement's Freedom Summer. Even when accounting for a range of independent variables such as gender, race, age, college major, education level, and home and college regions, the single greatest predictor of participation in the Freedom Summer was the existence of a pre-existing strong tie to another participant (McAdam, 1986; McAdam & Paulsen, 1993). Diani (2010) also argues that pre-existing bonds and solidarities are more likely to lead to effective mobilization, and doubts that social media will be able to create new social ties that did not previously exist. In this view, weak ties are both a blessing and curse, since they are conduits for new ideas, but are poor at transmitting behaviors that are in some way risky or costly to adopt. According to much of the diffusion research, risky behaviors such as joining a protest in an authoritarian country require higher thresholds of neighbors engaging in that activity before an undecided individual will also choose to adopt that behavior (Granovetter, 1978; E. Rogers, 2003; Suk-Young Chwe, 1999).

While much of the literature tends to focus on the benefits of one type of tie over the other, a combination of weak ties and strong ties are probably most conducive for risky collective action, with each serving specific positive roles for social movements while also frequently correcting for the limitations of the other. In a review of the social movement literature on the relevance of tie strength for collective action, McAdam and Paulsen (1993) sum up the relative importance of strong and weak ties for collective action and describe what an ideal mix of ties might look like. At the meso level, they note that a number of studies find that diffusion of collective action is enabled by weak bridging ties, and that movements can fail if they lack such ties. In their view, this implies that communication is the essential function being served by weak ties at the meso level. However, at the individual, micro level, the evidence of the influence of strong ties on participation suggest that ties are less important as conduits of information but more important as sources of social influence. Further, the stronger the tie, the stronger the influence exerted on a potential new recruit. McAdam and Paulsen (1993) conclude that, "the ultimate network structure for a movement would be one in which dense networks of weak bridging ties linked numerous local groups bound together by means of strong interpersonal bonds" (p. 655). While McAdam and Paulsen hypothesize that this type of structure is ideal, they do not test this hypothesis or seek out empirical evidence to support that claim.

While Granovetter and others focused on tie-strength at the dyadic level, we can also look at the whole network to understand how well the actors within a given network are connected. A network is said to be dense if the number of ties in a network are close to total number of possible ties. A network that has all nodes connected is also known as a complete network. On the other hand, a sparse network has a relatively low number of connections compared to the total number of possible links. At a substantive level, a dense network can be thought of as cohesive in the sense that it is tightly knit together, the information flowing through it is likely to be redundant, and, due to homophily, the actors within the network to be similar to one another. Dense networks are also resilient since they will not be affected much if one node is removed. Social movements with dense social networks are linked both theoretically and empirically with greater solidarity, mutual support, and incentives for self-sacrifice, and participants in dense social networks are also more likely to engage in high-risk collective action (Krinsky & Crossley, 2014). A sparse network, on the other hand, is more likely to have more diverse information flowing through it, and to have more structural holes between communities. A sparse network is also more easily disrupted through the strategic removal of nodes.

Whole networks can also be classified based on the strength of ties within them. Watts and Strogatz (1998) describe what they call small world networks. In their study of network evolution, they find that it only takes a small number of randomly rewired ties to turn a highly clustered network made up of predominately strong ties, also called lattice or regular networks, into a network that resembles a small world network. In small world networks, information can be diffused much more easily than in a lattice network since weak ties now connect those previously isolated clusters. In their view, this has a large potential impact on diffusion of information and disease, since only a few weak ties are necessary for the introduction of information or for a disease to spread even among remote, isolated communities. These types of small world networks are marked by high average clustering coefficients, which accounts for clusters made up of strong ties ('isolated caves' in the terminology of Watts (2003)), and low average path lengths, which mean that an isolated cluster can be reached from any other in just a few steps.

Although Watts and Strogatz (1998) used simplified network models to make this argument about small worlds, they provide three real world examples: a collaboration graph of actors in the Internet Movie Database (IMDB), the neural network of the earth worm *C. elegans*, and the U.S. power grid. Each of these real world networks demonstrate a higher clustering coefficient and shorter path lengths than random networks with the same number of nodes and average number of edges. In the real world network of actors in the IMDB dataset, which might be closest of the three networks to a real world social network, the average geodesic was just 3.65 while the clustering coefficient was quite high, at 0.79 (Watts, 2003). Most useful for this thesis, then, is that Watts and Strogatz provide a way to measure small world networks. A criticism of their approach is that nearly every real world network falls somewhere between the extremes of a lattice network and a random graph, and Watts and Strogatz are not clear what the ideal combination of ties might look like. If most emergent social networks are small world networks, how can we differentiate between those that are conducive to collective action and those that are not? They also focus more on information and disease spread than social movement emergence and success.

Baldassari and Diani (2007) provide direct evidence of the importance of network structure for political networks engaged in collective action. They investigated the structure of civic networks in Bristol and Glasgow, two UK cities with different socio-economic contexts. They asked if each city's network had a hierarchical, centralized form or a more horizontal, polycentric form compared to random networks. Hierarchical networks have a core-periphery structure with one or a few central nodes in the center of a network surrounded by many peripheral nodes in the edges of the network. Those peripheral nodes link to the central nodes, but not to each other. Multiple clusters of strong ties connected by a few weak ties, on the other hand, characterize polycentric networks. The polycentric network are similar to what Watts and Strogatz call a small world network, and are identified in the same manner; by finding networks with a high clustering coefficient and short average path lengths compared to random graphs (Baldassari & Diani, 2007; Watts, 1999). Baldassari and Diani found that the civic networks in both cities were more horizontal and polycentric than random networks despite their socio-economic differences.

Baldassari and Diani (2007) also argue that their two civic networks resemble what Gerlach (2001) called SPIN models of social movement organization, because they are segmented, polycentric and integrated networks. These types of movements are segmented because they are composed of many diverse groups that come and go, and divide and fuse over time; polycentric because they have many leaders or centers of leadership that are not organized hierarchically and also evolve over time; and networked because they form loose, integrated networks with multiple linkages, facilitated in part by the Internet, and created through shared memberships, joint activities, common reading material, and shared ideals (Gerlach, 2001). Movement leaders in SPIN models of organization are more likely to be evangelical and inspirational than experienced bureaucratic managers sitting on top of a formal hierarchical organization. Gerlach (2001) further notes that one way to become and remain a leader in a SPIN organization is to become a node that connects many groups. These types of networks consist of dense clusters connected by sparse bridging ties, and are powerful and very effective forms of organization (Baldassari & Diani, 2007). They are especially effective in repressive environments because they are robust, with the defection or removal of nodes mattering little. These SPIN-type movements can also penetrate into different social factions or niches of society, and are also found to promote innovation (Baldassari & Diani, 2007; Gerlach, 2001). Hierarchical models, on the other hand, can theoretically mobilize the network for collective action more easily than polycentric networks, but are more susceptible to the removal of key nodes.

A criticism of Baldassari and Diani is that it is not clear how generalizable their findings are given that they look at networks in just two cities. While those cities are different in their socio-economic makeup, both are based in the same country and in an advanced Western democracy. They conclude by proposing that their findings should be tested in other political contexts, including semi-authoritarian countries, to see if these types of horizontal, polycentric civic networks are also present (Baldassari & Diani, 2007).

This section has shown that, although weak ties are often seen as more important than strong ties for social movements, both are necessary for risky collective action. Weak ties allow for the introduction of novel information and can tie together otherwise isolated clusters into a larger movement with a common purpose. However, individuals connected by weak ties are not likely to exert a great deal of influence on one another compared to those with strong ties.

Strong ties also reinforce group solidarity, allow for repeated confirmation and affirmation of protest information in risky political settings, and help signal the normative importance of participation. However, without weak ties, it would be difficult for protests and information about them to diffuse beyond small cliques. Weak ties can also help movements overcome the risk of fragmentation and allow them to develop common goals. Polycentric networks have clusters of strongly connected actors that are connected to each other via weak ties. Polycentric networks are more robust than hierarchical networks, more resistant to repression, can penetrate and recruit from different segments of society, and encourage social innovation and problem solving (Gerlach, 2001). This literature leads to the first research question and related hypothesis.

RQ 1: Will a protest network challenging a hybrid regime have a polycentric or hierarchical structure, and does this structure change over time?

<u>*Hypothesis 1:*</u> *The Russian protest network will have a polycentric structure only at the beginning of the protests, and will move towards a less effective hierarchical structure as the movement loses popular support.*

Building on theories of weak ties and their role in connecting clusters of densely connected individuals, the next section of this chapter looks at the role of brokers. Brokers are those individuals who sit at the ends of weak ties and close structural holes in networks, and therefore play special roles in a range of different networks.

Brokerage and structural holes

Brokers play critical roles in social movement emergence and success. While I focus in this thesis on the role of brokers in protest movements, much of the theory and empirical evidence on brokerage comes from organizational studies. Burt (1992) based his classic formulation of brokerage on an analysis of workers in firms who connect otherwise isolated clusters of employees. Burt argues that brokers can exert influence and gain prestige within the firm by having access to and translating information between otherwise disconnected clusters. Burt calls these gaps in the network structural holes because there is an absence of ties between the various clusters of workers. Brokers gain a number of advantages from their structural position. For example, brokers have been shown to have greater levels of influence within organizations and are more likely to be promoted compared to those who are not situated in such boundary spanning positions.

Brokers are not necessarily the most prestigious individuals in an organization. Brokers can include people such as the CEO's secretary, janitors, or other lower level employees who happen to share smoke breaks with employees from different departments. While Burt does not discuss the importance of his work for protest movements, his work suggests that individuals who are able to close structural holes will be able to influence the flow of knowledge in the network, and thereby exert influence over the movement (Benkler et al., 2015; Burt, 1992).

The social movement literature also sees an important role for weak ties, bridges, and brokers. Network analyses of social movements often study the links between formal movement organizations (Ansell, 2003; Baldassari & Diani, 2007; Diani, 2003; Osa, 2003). However, in a particularly relevant study for protest movements in authoritarian countries, where formal civil society groups are prohibited or tightly controlled, Goldstone and Useem (1999) find that informal networks and bridging ties between neighborhood or friendship networks are more conducive to movement success than formal organizations. They cite as an example informal networks in East Germany and Czechoslovakia from 1989-1991 that were able to spread the belief of the efficacy of protests, and were ultimately successful in bringing down authoritarian regimes. Goldstone and Useem (1999) also argue that prison riots show how informal networks facilitate collective action "even from a position of close supervision and near-zero autonomy" (p. 998). The fact that prison riots occur in an environment where formal organizations are often prohibited undercuts the argument put forward by McAdam and others that formal organizations are required for movement success. In the view of Goldstone and Useem, the risky collective action such as prison riots are enabled by informal networks, including those created by inmates who share cell blocks, dining, recreation, and work facilities, which are used to communicate, build shared grievances, and form common beliefs (Goldstone & Useem, 1999).

Gould (1991) finds that the interaction of formal and informal organizational network structures are critical to the mobilization process and to group solidarity. This point is echoed by Goldstone and Useem (1999), who observe that informal networks can also support movements by allowing them to continue to live on through their individual members following setbacks, such as the arrest or jailing of movement leaders by the government. Gould also shows that mobilization does not just depend on pre-existing ties, but that protesters can also form new ties through their participation in collective action, which also strengthens movements.

Specifically, Gould (1991) analyzed how informal networks interacted with the formal organization of the National Guard in the Paris Commune of 1871. Gould argues that this type of interaction was critical for mobilization in the final weeks of fighting because it linked informal social networks defined by neighborhoods to the formal organizational network of the National Guard. Battalions were typically formed in a specific neighborhood and were often composed primarily of fighters who lived in that neighborhood, but there were often cross-district enrollments. Using network autocorrelation models, Gould shows that overlapping enlistments influenced the degree of commitment to the insurrection. Gould writes, "High levels of commitment in one area enhanced commitment elsewhere when enlistment patterns provided a conduit for communication and interaction" (p. 726). This is an effect that cannot be explained by spatial diffusion, meaning that neither social similarities among neighborhoods nor a shared border between neighborhoods explain joint commitment to the fight, but instead overlapping enlistments. This effect was only found in only one direction. That is, levels of resistance in each district were affected by its own residents serving in other districts in the National Guard, but not by the experiences of Guardsmen who lived in other districts. In other words, insurgents serving away from home positively influenced the behavior of their neighbors serving at home, but these insurgents did not affect the behavior of the Guardsmen with whom they served. Gould concludes from his study that the greatest opportunity for formal and informal networks to influence each other is when social movements are mobilized mostly along the lines of indigenous social structure, but in a way that also allows some interaction across pre-existing social boundaries (Gould, 1991).

Han (2009) provides additional evidence of the importance of weak ties for movements through his analysis of the American revolution, showing how individuals who served as brokers were able to tie together distinct revolutionary organizations and social classes into a broader social movement. Specifically, Han describes how the overlapping memberships of Paul Revere and Joseph Warren in various revolutionary organizations, and their professions, allowed them to play key brokerage roles in the American revolution. In Han's analysis, the social structure of New England during the revolution was very fragile and had many political, economic, and social fault lines that could easily have undermined the revolutionary movement. However, Revere and Warren were able to bridge these fault lines through extensively overlapping memberships in revolutionary organizations as well as ties that spanned various social classes. For example, as a silversmith Paul Revere stood in the middle of the social hierarchy in revolutionary America that allowed him to operate between the 'bully boys' of the Boston waterfront and the Harvard-educated gentleman that served as the revolution's leaders. Warren, as a doctor, occupied a similar structural position due to a patient list that included members from all social classes, including his experience having treated a typhoid outbreak in Boston, a disease that tended to disproportionately strike the lower classes. Both also belonged to multiple revolutionary organizations that spanned social classes and political allegiances. Han argues that it is unlikely that the American revolutionary movement would have emerged or been successful without these types of actors, and the same could be said of other broad-based social movements. This is probably an overstatement given the many factors contributing to the successful revolution, but he nonetheless makes a compelling case about the importance of brokers to social movements.

Kim and Pfaff (2012) provide additional evidence for the importance of brokerage and the closure of structural holes for the spread of protests and the diffusion of revolutionary ideas during the protestant reformation, showing that ideologically mobilized students were able to move ideas across structural holes from critical university communities back to their hometowns. Hedstrom et al. (2000) also argue that evangelical labor leaders who travelled between disconnected working communities in Sweden created meso-level networks and closed structural holes, enabling the creation of a national level labor movement. More recently, D. Faris (2008) argues that blogs and social networking sites have an important role to play in amplifying weak ties by making those ties transparent and usable, and simplifying the process of activating them. He writes that, "Facebook takes dormant social ties and makes them active, takes musty acquaintances and wipes the cobwebs from them, and can potentially plug you into social networks you never even knew you wanted to be a part of" (D. Faris, 2008, p. 7). In his analysis of labor strikes in Egypt before the Arab Spring, Faris found that Facebook served as a bridge between labor activists and college-educated elites in Cairo, magnifying and multiplying weak links between these groups who together contributed to successful labor strikes.

One of the major reasons that brokers are found to occupy such an advantageous position in both protest and organizational networks is because they introduce or have access to novel information from actors in the communities they connect. This is because information and opinions are more homogenous within groups than between them (Burt, 1992, 2004; Granovetter, 1973, 1974; Putnam, 2000). Information within groups tends to be homogenous in large part because of homophily; the tendency for individuals with similar characteristic to form ties and cluster together in groups in networks (McPherson, Smith-Lovin, & Cook, 2001). This general principle of tie formation results in homogenous personal networks and can limit access to information, lead to polarization, and negatively impact the attitudes people have towards other political, social, and ethnic groups (Granovetter, 1973; McPherson et al., 2001; Putnam, 2000). The internal focus on the information within one's own group creates gaps in the information flow between groups, which, again, Burt calls structural holes. Brokers have unique access to novel information found in the diverse groups they connect.

It is broadly believed that there is value to being exposed to novel information from individuals who come from different economic and social backgrounds, belong to different professional or scholarly communities, and those that hold opposing political views. Novel information is believed to be valuable because it scarce locally within one's social network (Aral, 2016; Aral & Van Alstyne, 2011). Those with access to novel information are better positioned to broker opportunities, make better decisions, and apply that knowledge to solve problems that are otherwise intractable. Access to novel information can also increase an individual's breadth of knowledge, help them share information on a wide range of topics to diverse audiences, and to be more persuasive. Since there is greater heterogeneity of information between groups, brokers have early and unique access to novel information and can translate that information between actors in different groups. This is known as the social capital of brokerage (Burt, 2004). Unique access to the information flow between heterogeneous groups allows brokers to, "see early, see more broadly, and translate information across groups. Like over the over-the-horizon radar in an airplane, or an MRI in a medical procedure, brokerage across structural holes between groups provides a vision of options otherwise unseen (Burt, 2004, p. 354)."

However, most organizational and firm-based brokerage studies focus on the benefits that accrue to brokers from their position, without testing whether the information that brokers have access to is actually novel. These studies are based on the premise that brokered information is novel, but are surprisingly 'content agnostic' (Aral & Van Alstyne, 2011; Hansen, 1999). And studies of the role of brokers within social movements are too frequently descriptive, identifying who brokers are within a movement and the differences among the groups they bridge, taking as a given that brokers are beneficial to the movement and have access to novel information (D. Faris, 2008; Han, 2009). Others focus on the outcomes made possible by brokers, such as the spread of revolutionary movements or the commitment of revolutionary fighters to the cause (Gould, 1991; Kim & Pfaff, 2012). Brokerage studies within both the social movement and organizational studies fields rarely test whether the information that brokers access from the different groups they connect is actually novel, and how network structure might affect their access to novel information.

As these historical and more contemporary examples show, weak ties, bridges and brokers play especially important roles in movement emergence and success. These examples also cover a range of different countries and types of movements, including political, religious, labor, and revolutionary movements. The main reason that brokers are advantageous to social movements is because they access and help diffuse novel information. This literature informs the second set of research questions and hypotheses.

RQ2a: Is brokered information in a protest communication network discussing high-risk collective action actually novel?

Hypothesis 2a.1: *The text of nodes from different Russian protest communities that are connected by brokers will be novel, since information and ideas shared between groups tends to be novel compared to that shared within groups.*

Hypothesis 2a.2: *The text of brokers in the Russian protest network should be novel compared to other nodes in their own communities, since brokers are linked to and have access to novel information from outside groups.*

RQ 2b: Is brokered information more novel when a network has a polycentric structure than when it has a hierarchical structure?

Hypothesis 2b: *The text of brokers will be the most novel when the Russian protest network has a polycentric structure, since such structures will have many communities of different types discussing a diverse range of topics, more structural holes, and more brokers and weak ties connecting disparate communities than a hierarchical network.*

In the next section, I discuss how framing helps social movements to spread their messages to the larger public and win new supporters, thereby helping to overcome preference falsification and pluralistic ignorance. I also show how media framing of protesters can have important impacts on public perceptions of movements, since the general public is much more likely to support moderates over radicals. Finally, I look at personal action frames and the logic of connective action before the final set of research questions is provided.

Framing

Framing is a contested but useful theory for understanding the origins and success of social movements. At a theoretical level, framing is helpful because it links the psychological and structural views that seek to explain movement participation, and also reintroduces the notion of grievances into theories of social movement emergence and success (Snow, Rochford, Worden, & Benford, 1986). Framing benefits social movements by diagnosing social problems, identifying targets of action, and mobilizing adherents (Halfmann & Young, 2010). A frequent criticism of the framing paradigm, in both the social movement and political communication literature, is that the concept is poorly defined. Framing is therefore often stretched beyond its useful original meaning and applied inappropriately. Still, it serves as a useful concept for understanding how grievances, beliefs, and emotion play a role in the emergence and growth of social movements, including protest movements in authoritarian and hybrid regimes. I first discuss the sociological roots of the framing paradigm before moving to a discussion of how framing is defined and employed in the political communication literature, especially by Entman (1993).

Framing and frame alignment in the social movement literature

Perhaps the most consistent criticism of framing as a theoretical concept is the lack of clarity in its definition and meaning. Snow et al.'s (1986) touchstone article on framing, however, serves as a useful starting point. Building off Goffman's (1974) original notion of the term 'frame,' Snow et al. define a frame as a "'schemata of interpretation' that enable individuals to 'locate, perceive, identify, and label' occurrences within their life spaces and the world at large. By rendering events or occurrences meaningful, frames function to organize experience and guide action, whether individual or collective" (p. 464). Put another way, frames are mental models created over time through an individual's life experiences that allow one to interpret events in a certain meaningful way, attribute blame, and assign responsibility (Crossley, 2002; Opp, 2009; Snow et al., 1986).

Social movements have three core framing tasks, which are diagnostic, prognostic, and motivational (Benford & Snow, 2000). The goal of diagnostic framing is to identify the cause or source of social problems and assign blame to those responsible. Prognostic framing refers to the articulation of a plan to solve the problem and tactics for carrying out that plan. And motivational framing represents calls to action that mobilize both core supporters and casual bystanders to participate in collective action. The product of these processes are often referred to in the social movement literature as collective action frames (Benford & Snow, 2000).

Snow et al. (1986) also argue that for social movements to mobilize support, they must connect their frames with those of their potential recruits through a process of 'frame alignment.' They describe four such frame alignment processes: frame bridging, frame amplification, frame extension, and frame transformation. 'Frame bridging' occurs when two or more frames of a social movement are ideologically congruent but structurally unconnected with their potential supporters. In this case, sympathetic 'sentiment pools,' or groups of uncoordinated individuals with common grievances and attributes, are sympathetic to the movement but lack the organizational means to mobilize around common claims. With these types of supporters, a social movement's primary task is outreach and information diffusion, which is achieved through interpersonal and intergroup networks, the media, direct mail, and today, one would assume, social media and the Internet. Frame amplification concerns the clarification or invigoration of existing beliefs in order to mobilize supporters. Frame extension on the other hand concerns the process of a social movement broadening their own framing of an issue in order to attract new recruits. This is done, for example, when movements align themselves with popular musicians, celebrities, or other causes, sometimes quite removed from their own original values, in hopes of winning additional supporters. Finally, frame transformation entails the situation in which a social movement's frames stand in contrast to a society's existing beliefs or lifestyles. Movements in this situation must systematically change how potential members interpret and experience the issue at hand. This frame transformation process is similar in some ways to religious awakenings, in which adherents see the world in a new light and change their views on who is to blame for the issue of concern to the movement (Snow et al., 1986).

Framing can also be contentious, as various constituencies engage in arguments over which grievances to focus on, who should be held accountable for social problems, and what tactical responses are necessary and appropriate (Benford & Snow, 2000). Some argue that frame alignment processes allow disparate movement actors and organizations to identify common grievances and targets, and to develop unified strategies to fight them (Snow et al., 1986). However, it is debatable whether such frame alignment processes actually occur, and whether frame alignment is necessary for movement emergence and success (Opp, 2009). Framing battles often also emerge between movements and those they are making claims against. For example, counter-framing by authoritarian governments is a common strategy for undercutting oppositional protest movements. Specifically, governments often frame popular protest movements against them as led by radicals and extremists in hopes of scaring away mainstream supporters.

A major problem with framing theory as laid out by Snow et al. (1986) is their assertion that frame alignment is a necessary condition for social movement formation, implying that in every case of social movement emergence, frame alignment will have already occurred. This seems unlikely. Emergent protest movements that come together quickly, such as those that form in reaction to a falsified election, are made up of different constituencies. They may be in agreement that they would like to see a dictator removed or new elections held, but each group will often couch their opposition in their own terms and based on their own experiences and political proclivities. However, as Opp

(2009) points out, it is possible that Snow and his colleagues may have actually meant that frame alignment is a sufficient condition. That is, when present and combined with other factors, frame alignment can lead to social movement emergence, but is not required or necessarily the only cause of social movement emergence.

Framing in the political communication field

Instead of focusing on how movements frame themselves, communication scholars tend to focus on framing as an act carried out by the media. Along with agenda setting, framing is a dominant theory of media effects in the political communication field. If agenda setting can be thought of as 'what' the media talks about, then framing is 'how' the media talk about an issue. I adopt Entman's (1993) definition of framing in this thesis. Like Goffman and Snow et al., Entman proposes that to frame an issue is to select some aspect of a perceived reality and make it more salient for an audience, though he expands the original notion to the communication field by arguing that framing occurs through a communicating text. Frames matter, according to Entman, because they serve to define problems, diagnose causes of those problems, make moral judgments, and suggest remedies. To clarify the notion of framing, Entman offers an example of the Cold War frame that previously dominated U.S. media coverage of foreign affairs. When the media in the United States employed the Cold War frame, they identified certain events such as civil wars in the developing world as problematic, identified communist rebels as the source of that problem, made moral judgments about the atheist aggression of the rebels and their Soviet sponsors, and proposed solutions, such as U.S. support for the other side (Entman, 1993; Etling, Roberts, & Faris, 2014).

Entman's focus on framing by the media is too limited. Certainly, the media retain significant influence in determining what issues the public comes to view as salient and how those issues are talked about, but individuals and movements clearly also can now play a role in how the public views issues of concern to movements. Benkler's (2006) notion of the networked public sphere is useful here, since he explains how the Internet enables bottom-up agenda setting and framing processes that networked social movements and activists can take advantage of today. Social movements, therefore, are no longer required to go through the mainstream media filter to gain attention, and they are able to use the Internet and social media to engage in framing battles with those they are making claims against. Movements can also use social media to directly counter governments or other power holders that often frame movements in a negative light.

Research explicitly focused on the success of moderate versus extremist framing among constituencies *within* movements is limited. Instead, studies of extremism and framing tend to focus on how the media often choose to highlight extremist or potentially violent elements within movements, especially if they are new or perceived to be out of the mainstream, such as leftists during the 1960s in the United States (Gitlin, 1980). Similar recent work has investigated how the contemporary global justice movement is framed, with empirical evidence showing that the media often frame social justice protesters as violent, disruptive, or 'freaks' (Boykoff, 2006). This results in a situation where protesters are actually ignored if their actions are not seen as extreme enough to deserve coverage. This then requires movements to radicalize their rhetoric and tactics in order to gain the attention of the mass media.

In a rare case in which the framing of moderate and more extremist protest constituencies are compared to one another, Bennett and Segerberg (2013) show that the personal action frames and protest events against the G20 meeting in London in 2009 that were advanced by a moderate group were more likely to gain support than that put forward by a more radical one. Bennett and Segerberg argue that this difference is due at least in part to the moderate group having more open participation practices and easier modes of affiliation with the movement, but they also acknowledge that the radical framing employed by the more extreme group (who advocated for 'crushing the banks' and 'overthrowing the capitalist system') were also part of the explanation for the smaller and less publicized event held by that group. Bennett and Segerberg also note that another 2009 G20 protests in Pittsburg, where extremists played a larger role and that included violence, resulted in a focus by the media on that violence. Clashes at the protests thus drowned out the protesters core global justice frames and messages in the media. Implicit in all of these studies is the notion that framing by the media of protests as radical, extreme, or violent undermines social movements, detracts from their preferred frames, leads to less effective protest actions, and can even lead to the death of protest movements, or at least the end of protest cycles.

Such media centered analysis drawn from Western democratic contexts is not especially useful in analysis of social movements in hybrid regimes, where governments often formally or informally control the media. However, this work does help explain why governments in hybrid regimes choose so often to use violent or extremist frames to describe popular protests against their rule. As already noted in this chapter, leaders in authoritarian or hybrid regimes frequently frame protest leaders as extremists or terrorists. In the 2013 Euromaidan protests in Ukraine, the government and its supporters often referred to the protesters as 'terrorists' and 'political radicals' and often amplified the importance of the radical Right Sector group in the organization of protests, drawing attention to the group's Nazi and white nationalist connections (Etling, 2014; Miller, 2014). Even in Ukraine, where the Euromaidan protests removed the pro-Russian government from power, periods of violence between protesters and the police led to a rise in pro-government sentiment online (Etling, 2014). As I discuss in Chapter 3 and Chapter 7, the Russian government and its supporters also framed election protest leaders, such as Alexei Navalny, as extremists. For example, they circulated a cartoon video of Navalny leading protesters in a heil Hitler salute. Similar to the Ukraine case, the Russian government and conservative media frequently focused on more extreme elements of the movement as the primary organizers, especially the far-left firebrand Sergei Udaltsov, while minimizing the role of more moderate leaders and groups.

These extremist elements cannot be ignored, though, as they often play important roles in opposition protests in hybrid regimes. For example, soccer 'hooligans' helped fight back against government forces during the first Arab Uprising protests at Tahrir Square in Egypt, and more radical groups in Ukraine were key defenders of the protest camps against the Berkut riot police. Russian nationalists are also a large constituency in Russia, and formed an important part of the 2011-2012 election protest movement. Since Putin often claims he has the ardent support of Russian nationalists, their public criticism of him and clear support for the election protests raised the question in the minds of many of who exactly did support Putin. The finding that Russian protest participants were more likely to support ethno-nationalism and authoritarianism instead of democracy (Chaisty & Whitefield, 2013), also shows that the movement had a large base of nationalist support. Given this literature, what may be even more interesting than how the government or the media frame protesters, is how successful extremists are compared to moderate elements *within* the same movement in advancing their frames online. Since violence and radical actions tend to lead to less effective protests and distract the media and the public from social movements' core framing and goals, it is also interesting to understand if extremist framing increases when violence occurs at protest events or during less successful protest periods.

While framing clearly has it critics and shortcomings, the concept remains incredibly popular, and understandably so. Framing helps to explain how social movements build solidarities among existing and potentially new adherents, identify common problems, and propose solutions to those problems. In combination with structural factors, framing processes help to explain how emotion, ideas, group identity, and culture play important roles in individual decisions to participate in collective action, especially risky collective action. Framing therefore serves as an important addition to purely structural or rational explanations of social movement emergence and success and remains a bedrock theory of media effects in the political communication field. The framing literature leads to the final research questions and hypotheses.

RQ 3a: Can a protest movement in a high-risk political setting be more successful than the government at spreading its preferred frames?

<u>*Hypothesis 3a:*</u> *The Russian protest election movement will be more effective than the government at spreading its preferred frames online, except when the government organizes large pro-government rallies.*

RQ 3b: Are moderate or extremist elements within a protest movement in a hybrid regime more effective at spreading their preferred frames?

<u>*Hypothesis 3b*</u>*: Moderates will be more successful than extremists in the Russian election protest movement at spreading their frames, unless violence occurs at protest events.*

Before concluding this chapter, I offer some thoughts on other relevant theories and research on the role of social media and the Internet in collective action, and reflect on ways that traditional theory may not perfectly fit with online communication networks and activities. Next, I describe the recently developed theory of connective action (Bennett & Segerberg, 2012, 2013). While this is clearly the most coherent and comprehensive theory in this space to date, I also explain its shortcomings and why I do not rely on it more heavily in this thesis.

The logic of connective action

Bennett and Segerberg (2012, 2013) nicely tie together different strains of the framing, social movements, and network literatures in their logic of connective action theory. They argue that traditional rational actor and framing theories are not able to explain how the Internet and social media are changing collective action. Building off of Benkler (2006), Bennett and Segerberg argue that a peer-produced, sharing-based networked economic logic leads to fundamentally different forms of social organization and technologies of association than predicted by traditional microeconomic models of participation built around the rational actor. They also describe how collective action has become more personalized, with young people in particular placing less emphasis on membership in formal political groups and showing less trust in institutions. Instead, individuals increasingly participate in collective action by making changes in their daily habits in ways that address issues they care about, which they then share broadly through their social networks. This personalization theme draws on the observations of Bimber, Flanagin, and Stohl (2012), who argue that personal engagement with causes is increasingly pushing aside traditional notions of participation tied to formal membership in social movement organizations. This requires that organizations, which Bimber et al. (2012) still see as essential for social movement success, now have to create spaces for individual participation.

Bennett and Segerberg propose a typology of networks that explains the two logics of contentious action at play due to the trends of decreasing power of institutions and increased personalization of participation. At one extreme are organizationally brokered networks that are best explained by Olson's (1965) traditional logic of collective action. These networks are led by formal social movement organizations, and have rigid collective action frames that are negotiated by those formal organizations. At the other extreme are crowd-enabled networks, such as Occupy Wall Street, that operate under the logic of connective action. These networks have little or no formal organizational coordination, allow large-scale personal access to multi-layered social technologies, have communication that is driven by emergent, inclusive personal action frames, and are resistant to the involvement of formal organizations. In the middle is a hybrid, organizationally enabled connective action network. This type consists of loose networks of formal organizations that may be involved in more traditional collective action networks, but that adopt tactics and technology that allow for personalized engagement with publics through organizationally-generated personal action frames. These organizations may also lightly moderate personal expression. Bennett and Segerberg draw on Chadwick's (2007, 2013) notion of hybridity, in which the repertoires, practices, and norms of the Internet have forced traditional organizations and political institutions to adopt tactics more typical of social movements. Similar hybrid forms are also emerging in Western democratic media systems, with older, traditional media adopting newer online practices common with bloggers, for example, while bloggers adopt practices of traditional journalists to gain authority.

Bennett and Segerberg argue that personal action frames help to tie together both the crowd-enabled networks and the hybrid, organizationally enabled network types. They propose that instead of collective action frames, which they argue are often negotiated with great difficulty among SMOs and require a shared identity or rigid interpretation, that personal (as in easy to personalize) action frames are more important for tying together looser connective action networks. Personal action frames share two components. First, they are inclusive of different motivations for individual participation in a movement, and therefore require little in the way of persuasion or reframing to bridge differences among participants from different backgrounds or classes. Second, they are shared widely online through open, personal communication technologies. Methodologically, they use basic social network analysis and tracking of twitter hashtags, and create inventories of interactive elements made available on protest organizations' Web sites or special purpose campaign sites. They draw on a number of national cases of protests related to the global justice and environmental movements in advanced Western democracies. The most relevant methodological contribution to this thesis is their finding that connective action networks (both hybrid organizationally-enabled and crowd-enabled networks) have less steep power law curves than the typical power law distribution found online.

I find the notions of increasing attention to loosely connected, hybrid organizations and their success over more traditional, hierarchical institutions and organizations to be the most useful part of Bennett and Segerberg's theory. This is a common theme identified among scholars that study the Internet and collective action, with others also noting ways that social media are disrupting existing hierarchies and institutions, be they state institutions connected to international affairs (Owen, 2015), political parties (Bimber et al., 2012), social movements (Earl, 2015; Earl & Kimport, 2011), or the media Chadwick (2013).

I chose not to lean more heavily on this theory for a number of reasons that I enumerate here. First, even though Bennett and Segerberg draw on network theory and use basic social network analysis to support their theory, they lean too heavily on the idea of 'networks' as a metaphor and as a heuristic device. This is a common issue with other theorists who write about social media and collective action (Castells, 2012). This causes them to also play a bit fast and loose with network terms that have specific meanings to network scientists, such as describing hierarchical networks as 'brokered networks', when brokers are more likely to exist and play more important roles in flat, polycentric networks instead of hierarchical ones. Bennett and Segerberg also do a poor job of describing the crowd-enabled network in a network theoretic or more empirically grounded way. This makes it difficult for them to measure the existence of such a network and compare it to more hierarchical networks in their typology. For example, Bennett and Segerberg describe the crowd-enabled, 'Occupy Wall Street' type of network as a 'network of networks' but fail to describe it as a polycentric network in the way that Baldassari and Diani (2007) do. This means that they are left describing this type of network and its influence in a largely ethnographic and descriptive way. Even in more recent work they explicitly choose not to investigate network structure as part of further theory building on crowd-enabled networks (Bennett, Segerberg, & Walker, 2014).

Second, Bennett and Segerberg focus their analysis and empirical work too heavily on formal organizations, even though personalized participation is an important aspect their theory. For example, Bennett and Segerberg catalog the ways that formal organizations supply opportunities for personal engagement on their Web sites instead of tracking and more systematically analyzing the way that individual participants talk about protests and frame their own actions online. This approach of cataloging the affordances of organizations' Web site offerings for personal engagement seems to be influenced by Earl and Kimport (2011), which seems a bit dated given the wealth of personalized, individual data available online today. Also, given the many ways that framing is already so frequently modified and redefined by scholars, it is not clear that yet another definition of framing, specifically personal action frames, is entirely helpful. It is similarly not clear that collective action frames and framing battles within movements have entirely disappeared just because we see the existence of personalized frames and more individualized forms of participation. This is a major issue with personal action frames because Bennett and Segerberg never prove whether these types of frames are actually more prevalent or have replaced traditional collective action frames within the cases they analyze.

Bennett and Segerberg, along with many others in this space, including but not limited to Chadwick (2013), Bimber et al. (2012), Baldassari and Diani (2007), Gonzalez-Bailon, Borge-Holthoefer, and Moreno (2013), and Karpf (2012, 2016), look only at examples of movements, organizations, or media systems in Western democratic contexts. This limits their use as a starting point for movements operating in the much more restrictive and higher-risk environments of hybrid regimes. The political opportunity structures and levels of risk in these liberal democratic contexts are likely to be much different than those in hybrid regimes.

Finally, Bennett and Segerberg ascribe too much agency to networks and to digital media. Drawing on Latour (2005), they argue that digital media can be thought of as non-human organizing agents (or actants) in social movements. They also claim that communication can be seen as a form of protest organization, as the digital traces of past actions are archived for future use and adaptation. I believe that agency continues to lie with the human actors in movements, and not the technology they choose to use. While Bennett and Segerberg do a very nice job of tying together quite distinct literatures, they treat networks too often as a metaphor, do not draw heavily enough on network theory, and do not provide a convincing case for why collective action frames should be abandoned in favor of personal action frames. They also fail to provide evidence that personal action frames have actually replaced collective action frames in connective action networks.

Online versus offline theory and practice

Much of the theory I draw on in this thesis was written well before the emergence of the Internet. This leads to a less than perfect fit between digital data and theory. For example, in terms of network ties, some online connections are easier to make and reciprocate than offline ones, as I note in Chapter 5 when discussing network parameters used in ERGM models. However, we now know there is often overlap between online and offline friends, and friends we frequently communicate with online or via a mobile device are also likely to be close 'real world' friends. Levels of trust between nodes in a network may also be different in an online network compared to offline. Notions of strong tie bonds as they relate to high-risk offline collective may not be equivalent in an online communication network. Talking about a protest or sharing a link is probably less risky than attending a protest in a hybrid regime, although many people who are talking about protests online also participated in them, and use the online space to share their experiences. At the same time we should not assume that there is no risk from online speech or organizing given government surveillance of online activity, especially in places like Russia. Throughout this thesis, I try to highlight when a traditional theory based on offline interactions or face-to-face communication might not fit as cleanly as with online interactions or data. So some caution is necessary in adapting more traditional theories discussed in this chapter. That said, I would discourage the reader from dismissing online activity as cheap or unworthy of study, thinking of digital data as exclusively 'virtual', or envisioning 'cyberspace' as completely detached and separate from the 'real world' (R. Rogers, 2013).

Conclusion

Despite the wave of Internet-enabled protests that have swept across the globe in the last decade, few have put forward a coherent framework to explain the impact of the Internet on protest movements in authoritarian regimes. The question of whether the Internet is better for dictators or democracy activists has not led to much illumination since, as Farrell (2012) notes, it is impossible to answer that question as posed. Others, especially in the media, that focus on the benefits of a given technology (e.g., a Twitter or Facebook revolution) have been equally unhelpful since technological platforms come and go, and the media narrative of their importance have gone along with them. Analysts also often fail to focus on the underlying sociological causal processes at play in these protest movements.

In this chapter, I attempted to rectify these problems by creating a multi-disciplinary framework to explain how the Internet and social media facilitate high-risk collective action in hybrid regimes. I focused, when possible, on causal processes that can be examined and tested empirically. I also concentrated on sociological processes within movements instead of the attributes of a particular tool. Special attention was paid to the importance of strong ties and weak ties for high-risk collective action. Strong ties help build group solidarities, are more influential than weak ties in inducing individual participation, allow for repeated affirmation of protest information, and reinforce the normative importance of participation. Weak ties facilitate the introduction of novel information not available from ones strong ties. They can also help prevent social fragmentation and allow a movement's disparate constituencies to form common goals. Polycentric networks have both strong and weak tie structures. These networks are a powerful form of social organization in repressive environments because they are robust, can penetrate and recruit from different niches of society, and also promote innovation. Brokers, or individuals who span structural holes in networks, were shown to be important in the emergence and success of movements by tying together distinct revolutionary organizations and social classes into broader movements during the American revolution and in Egypt before the Arab Spring, diffusing ideas beyond universities during the Protestant reformation, and enabling the creation of a national-level labor movement in Sweden. I also discussed the benefits of framing as described in the political communication and social movements literature, and how the Internet enables bottom-up agenda setting and framing processes. Before beginning to test the hypotheses generated in this chapter, I first provide a detailed description of the Russian protest movement in the next chapter.

Chapter 3 The Russian Election Protest Movement

On December 4, 2011, Russians went to the polls to elect a new parliament. Nobody predicted what would happen next. Throughout Election Day, stories and video evidence of election falsification exploded across the Internet. In the days, weeks and months that followed, online rage turned to mobilization. Protests began on December 5 when 6,000 rallied at Chistye Prudy in Moscow, a huge number for an oppositional protest in Russia (Barakovskaia, 2011; Ioffe, 2011a). That event was quickly dwarfed when as many as 50,000 protesters attended the December 10 protest at Moscow's Bolotnaya Square and in cities across Russia, followed by more than 100,000 at Sakharov Prospekt on December 24 despite sub-zero temperatures, then again with 20,000 attending protests on February 4 and 50,000 on May 6, 2012 (Barry, 2011; Elder & Parfit, 2011; Greene, 2013). These events represented the largest protests in Russia since the fall of the Soviet Union, and shattered the widely held belief that the Russian electorate is apathetic and that Russian civil society is weak.

In this chapter, I provide essential background on the Russian election protest movement, which is used as an empirical case study in this thesis. I begin with an overview of the movement and the falsified Duma election that led to the movement's emergence. I next describe the protesters, with a focus on the different constituencies that made up the movement. Next, I highlight the movement's major demands. I also describe the importance of the Internet to the protests. I then discuss the government's reaction to the protests and the counter movement that supported the government. This is followed by a chronology and description of major protest events as they unfolded over the most intense six months of protest activity.

What was the Russian election protest movement?

The Russian election protest movement of 2011-2012 was a spontaneous protest movement that emerged as the direct result of falsified parliamentary (Duma) elections that were held on December 4, 2011 (Greene, 2013). In that election, the pro-government party United Russia barely secured a majority of the votes for the Duma. This majority would allow United Russia to continue to rubber stamp the Kremlin's legislative agenda. That the Central Election Commission and its pro-Putin chairman Vladimir Churov declared United Russia the victor was not surprising, but seemed improbable to many given the low levels of support for the party. The party had suffered an onslaught of negative publicity by activists such as protest leader Alexei Navalny, who had labeled it 'a party of crooks and thieves' and had encouraged people to 'vote for anyone but United Russia.' While it was rare for attacks of this sort on Putin, it was much easier for Russians to see local bureaucrats and party functionaries as corrupt and to blame for many of Russia's problems. However, there had also been signs in the months before the protests of growing weariness with the government, and even Putin. For example, there was significant outrage over the 'castling' move by Putin and Medvedev, in which it was announced that the two had long beforehand agreed to swap positions, with Putin running for president and Medvedev returning to his previous role as prime minister. To many Russians the brazenness of this move proved that Russian elections were indeed a sham. Putin was also widely ridiculed in the largely free and unregulated Russian blogosphere, probably to an even greater extent than United Russia (Etling et al., 2010; Etling et al., 2014). However, before the Duma elections a young, male, and macho audience that was supposed to be an important base of his support also publicly booed Putin at an ultimate fighting match. The event was even briefly broadcast on television, but was edited out of later newscasts. In short, public discontent was on the rise and the government's carefully orchestrated image of invincibility had begun to crumble even before the elections took place (Koesel & Bunce, 2012).

International observers found that the elections had numerous serious violations, including frequent procedural violations, apparent manipulation of results, and ballot box stuffing (OSCE, 2012). However, the evidence collected by Russians themselves was probably much more damning. The Internet and social media played an important part in allowing Russians to collect and share personal stories and video evidence of election fraud. Journalists and independent media outlets also uncovered and highlighted election fraud. Investigative journalists even surreptitiously participated in organized ballot stuffing and election falsification rings, or carousels ("I was a participant in the falsification of the 2011 elections," 2011). Several outlets, such as the main independent radio station Echo of Moscow and major independent newspapers, also helped to promote an online tool created by the election-monitoring group Golos for reporting of election violations. These independent traditional media outlets supported the narrative of those citizens who had observed fraud, but also helped to magnify those citizen voices to the broader public. The many stories and accompanying video evidence of electoral fraud shared online and via independent media outlets led many voters to conclude that the elections had been falsified. Independent experts found that United Russia would not have more than 35% of the vote without engaging in election fraud, instead of the 54% the party claimed to have won (Shvetsova, 2012).

Electoral fraud thus became a key mobilizing event for protesters and also provided a unifying, common grievance around which the broader Russian public could rally. This is significant because scholars believe mass protests in opposition to electoral fraud both lowers the risks and increases the expected benefits of anti-regime activity compared to a citizen confronting the regime alone over daily abuses, such as bribery by the police (Tucker, 2007). The calculation of individual risk changes because the entire country experiences the abuse of electoral fraud at the same time, giving protesters a common grievance, or focal point of contention. Risk is also reduced as the size of protests grows, thereby lowering the chance of individual punishment. Finally, protests against electoral fraud also hold out the possibility of actually removing those who hold political power in the country, a potentially major payoff to consider in combination with the already lowered risk of harassment or arrest. This may explain, in part, why these protests were so large and widespread.

Who were the protesters?

The Russian protest movement consisted of a loose coalition of oppositional civic groups, formal political parties, and individual activists with a range of different political affiliations. The movement included democrats, nationalists, leftists, independent journalists, writers, musicians, and activists associated with anti-corruption, environmental, and gay rights causes (Greene, 2013; Ioffe, 2011c; Popescu, 2012; Smyth, 2013; Taylor, 2011). These movement elements drew heavily on the non-systemic opposition in Russia, or those individuals, civic organizations, and political parties that are either informally or legally barred from competing in elections. Leaders of issue-specific civic groups played an especially important role in the leadership of the election protests. These individuals led groups focused on corruption, ecological destruction, housing and development fraud, repression of artists and journalists, and other violations of the core needs of specific groups (Greene, 2013; Robertson, 2013). This included anti-corruption activists and blogger Alexei Navalny, environmental activist Evgenia Chirikova, and automobile activist Sergei Kanaev (Greene, 2013). Civic organizations that work on concrete social issues, such as Chirikova's group that attempted to save the Khimki forest near Moscow, have proven to be far more popular with Russians than opposition political parties or abstract political movements (Volkov, 2012). Examples of independent journalists and writers sympathetic to the movement included Alexei Venediktov of Echo of Moscow radio, Yevgenia Albatz of The New Times, journalist Oleg Kashin, satirist Viktor Shenderovich, fiction author Boris Akunin, writer and journalist Dmitri Bykov, and socialite and TV host Kseniya Sobchak, who was is also the daughter of Putin's mentor, former St. Petersburg mayor Anatoly Sobchak (Greene, 2013; D. White, 2015).

Among more traditional democratic parties and politicians, the umbrella political group Solidarity played a key role in the logistics and organization of protests. Solidarity represents individuals and parties affiliated with the democratic opposition, such as RPR-PARNAS and Yabloko. The first protest event at Chistye Prudy on December 5 was planned by Solidarity and approved before the Duma elections took place, and the group continued to help organize and publicize future protest events (D. White, 2015). Individuals associated with Solidarity include longtime democratic activists and politicians such as Ilya Yashin, Boris Nemtsov, Vladimir Ryzhkov, and Grigorii Yavlinskii. Former world chess champion and democratic opposition politician Garry Kasparov also participated in the protests. Kasparov founded and leads the United Civil Front movement, and participated in Other Russia, another umbrella oppositional group of political and human rights activists that organized a series of 'dissenters' marches' before the electoral protest began (Gel'man, 2007). According to Levada Center polling conducted at several protest events, 60 to 70% of protest participants identified themselves as democrats or liberals (Volkov, 2015). This mix of independent civic groups, parties, journalists, writers, and artists form what I call the mainstream core of the protest movement.

More radical elements of the protest movement included ultra-nationalists and far-leftists. On the far left, Sergei Udaltsov played a prominent role in the protest movement. Udaltsov is the leader of the radical socialist group Left Front and is a self-proclaimed revolutionary who calls for a social-democratic revolution in Russia. He was frequently arrested before and during the election protests. He also went on hunger strikes while in prison as a form of protest and to draw attention to his imprisonment (Balmforth, 2012a; Parfit, 2011; Sakwa, 2015; Schwirtz, 2012a). On the far right, Russian ultra-nationalists who participated in and sometimes spoke at the protests included Eduard

Limonov, founder of the banned National Bolshevik Party and also a leader of the broader Strategy 31 protest movement, Vladlen Kralin (a.k.a. Vladimir Tor) of the Movement Against Illegal Immigration (DPNI), Ilya Lazarenko, former head of the fascist party National Front, Dmitry Dyomushkin of Slavic Union, and Konstantin Krylov of the Nationalist Democratic Party, who was also member of the Opposition Coordinating Council (Umland, 2012). These ultranationalist leaders and affiliated groups tend to focus on illegal immigration, argue that the Kremlin should 'stop feeding the Caucasus' and promote the idea of 'Russia for Russians.' These nationalist groups formed an uneasy alliance with the democrats and leftists during the election protests (Popescu, 2012). These individuals and groups make up the core of the extremist constituency in the protests movement. Chaisty and Whitefield (2013) find that supporters of the protests were more likely to support authoritarian leadership and ethno-nationalism than a democratic transition, hinting at the influence of nationalists within the movement.

There is some debate about whether protest leader Alexei Navalny should be considered part of the extremist Russian nationalist camp. Pro-government sympathizers certainly tried to frame Navalny as an ultranationalist, and compared him to Hitler in online videos released after the first protest events (Mackey & Kates, 2012). Navalny certainly does not hide his nationalist sympathies. He frequently appeared at the annual 'Russian March' organized by many of the nationalists who participated in the election protests, was kicked out of the Yabloko political party for his nationalist views, and is generally more supportive of nationalist causes than other democrats and liberals in Russia. However, it is difficult to know Navalny's true feelings on the nationalism issue, and whether his language and positions are heartfelt or used to gain the support of an important oppositional constituency in Russia (Popescu, 2012). As he has become a more public figure, Navalny has moderated his more extreme positions, stopped attending the Russian March, and has apologized for the more racist comments he has made in the past (Coalson, 2013). Navalny is also not nearly as extreme as Russian ultranationalists such as Vladimir Tor, or far right European politicians like Jean-Marie Le Pen of the Nationalist Front in France or the late Jorg Haider of the Austrian Freedom Party (Popescu, 2012; Umland, 2012). He is also much more focused on anti-corruption issues than nationalism. For these reasons, I consider him to be more of a mainstream leader, especially when compared to the much more radical ultranationalists who made up the core of the extremist protest constituency.

The systemic opposition (the three 'opposition' parties besides United Russia that are allowed to compete in elections and hold seats in the Duma) played a more limited role in the protests. These parties, the Communist Party of the Russian Federation (KPRF), Just Cause (Spravedlivaya Rossiya), and the Liberal Democratic Party of Russia (LDPR), are often also called the 'pocket opposition,' since they are seen as 'in the pocket' of the Kremlin. March (2012) argues that the systemic opposition could not tap into discontent in society after the falsified elections because of the informal political rules of the game in Russia. These rules require the systemic opposition to show deference to the federal government in public. The Communist Party of the Russian Federation (KPRF) and Just Cause initially backed the protest movement, while the Liberal Democratic Party of Russia (LDPR) never waivered in its support of Putin. According to Sakwa (2015), the KPRF was opposed to mass mobilization and, like the government, did not want an 'Orange Revolution.' KPRF support for the protests faded after violence between the protesters and police in May 2012. Just Cause's Gennady Gudkov, the parliamentarian most supportive of the protesters, was also stripped of his seat as punishment for his endorsement of the protests, and as a warning to others in the systemic opposition (March, 2012; Sakwa, 2015).

In Chapter 7, the substantive framing chapter, I focus on five different elements of the protest movement: the Mainstream, Environmental, Extremist, LGBT, and KPRF constituencies. Texts from Web sites and blogs associated with each constituency, as well as pro-government supporters, are used to identify the framing employed by each group. That chapter also includes qualitative summaries of the framing used by each constituency when discussing the protests.

What were the protesters' demands?

While the Russian protest movement was based on largely informal organization of different civic groups, it had a unified and clear set of demands. Specifically, the movement demanded that:

1. The results of the fraudulent election be nullified and new elections for the Duma held;
2. The Central Election Commission chairman Churov, who protesters believed was not independent and working at the behest of the Kremlin, must step down and be replaced by a truly independent

election commissioner. One of the heads of the Russian search giant Yandex was often mentioned as a good candidate to replace Churov;

3. Election laws must be rewritten so that all parties are allowed to participate in elections;
4. Political reforms must be made to allow for a more democratic political system, including changes to the elections and political party laws;
5. Peaceful protests, such as the election protests, must be allowed to take place;
6. The release of political prisoners.

The movement was able to win some concessions from the government. For example, Medvedev announced the return of direct elections for regional governors, new rules for federal parliamentary elections, and simplified rules for the registration of political parties (Koesel & Bunce, 2012). Protests were also generally allowed to take place, but only when they were preapproved by the government and typically with restrictions on location, size, and duration (Barry, 2011; Batty, 2011).

What role did the Internet play in the movement?

The Internet played significant information dissemination, mobilization, and coordination roles in the largest demonstrations in Russia since the fall of the Soviet Union (Greene, 2012, 2013; Koesel & Bunce, 2012; Litvinenko, 2012; Shvetsova, 2012; S. White & McAllister, 2014). As Litvinenko (2012) writes, “Protesters were mobilized mainly via social networks sites such as Vkontakte and Facebook – a fact that finally demolished the argument that the Russian Web was ineffective at mobilizing citizens." Facebook, Vkontakte, Twitter, blogs and other online tools were used to inform Russians about the protests and to encourage users to invite their friends to attend. According to the Levada Center, at least one third of those who attended the Sakharov Prospect protest may have been recruited via online media (Greene, 2013; Levada Center, 2011). White and McAllister (2014) also found, using a nationally representative survey, that being a Facebook user increased the likelihood that respondents agreed with the protesters' demands. They also found that the more frequently someone used the Internet the more likely they were to see the elections as fraudulent, while those who watched TV more frequently were more likely to see them as fair. A number of collaborative online platforms, including Google docs and Facebook groups, also allowed users to vote on who should speak at the protest events, providing a more interactive and bottom-up form of protest organization and participation. Before the protests, many of the eventual leaders of the movement, such as Alexei Navalny and Evgenia Chirikova, also built their reputations online, where they also promoted their causes and gained dedicated groups of supporters. The protests also built off of many instances of successful civic action in which citizens used the Internet and social media to fight against the government over specific or very localized grievances (Oates, 2013).

An important option that the Facebook and Vkontakte protest event pages included was the ability to publicly declare if a user planned to attend, and to see that their friends would attend as well. For the Bolotnaya protest, nearly 40,000 indicated that they would attend, while at least 10,000 more actually participated in person (http://www.facebook.com/events/198328520252594/). For our purposes, the exact number of protesters is not as important as the fact that tens of thousands signalled beforehand that they would participate. This may have encouraged even more individuals to attend since the risk of harassment or arrest by the police is reduced when protesters number in the tens of thousands instead of hundreds, as was often the case for Russian opposition protests in the preceding years (Tucker, 2007).

Organization of election observers is another example of online mobilization. Alexei Navalny informally promoted election observation online before the Duma election, but efforts became much more organized for the Presidential election (Litvinenko, 2012). Specifically, Navalny created the Web platform RosVybori (http://www.rosvybory.org) to help coordinate election-monitoring efforts. The site contains training materials, training videos, and online forms that allowing users to sign up to become election observers. The independent Russian election monitoring organization Golos (Voice or Vote in Russian) also created an online tool (http://www.kartanarusheniy.org) that allowed Internet users to report and aggregate cases of election violations for formal submission to the Russian Central Election Commission (CEC). This platform was similar to the very successful Russian "Help Map" (http://russian-fires.ru) that was created in the summer before the elections to coordinate volunteer efforts, collect supplies and direct resources to the areas of greatest need during the major forest fires outside Moscow in 2010 (Asmolov, 2014a). This also contributed to transparency around deaths that increased due to the smog that enveloped Moscow during the forest fires. Similar volunteer efforts that were aided by ICTs also took place in 20012 when floods struck the southern Russia. Asmolov (2014b) argues that these types Web-facilitated mutual aid projects represent a new type of activism in Russia. As Asmolov (2014a) writes, "While in the past the coordination of large-scale collective action was expensive and available primarily to hierarchical institutions, event-mapping makes it accessible to individuals and non-institutional entities and empowers these."

Government reaction and counter-movement

The Russian government and its proxies formed the core of the counter movement that mobilized against the anti-government election protests. They began by trying to ignore the protests. Neither Putin nor Medvedev spoke about the protests and possible election fraud for weeks. The major TV news channels were also silent on the protests until the day of the first major protest on December 10, when the topic was simply impossible to ignore (Litvinenko, 2012). Fearing that youth would be active participants, as they had in the color revolutions in Ukraine and elsewhere in the region, the Ministry of Education changed the date at the last minute for the national entrance exam required of all students hoping to attend university for the same date as the protests. The government also threatened to conscript any young men who were found to have avoided military service. The Minister of Health warned people would get the Flu or even SARS (Barry & Schwirtz, 2012; Elder, 2011b).

The government and its supporters also engaged in negative framing of the protest movement and its leaders as part of an overt strategy aimed at undermining the protests. The government used a mix of negative frames to describe election protesters and the movement. This included highlighting extremist, nationalist elements and slogans in the protest movement instead of moderates. The government also framed the protesters as a minority, and as part of a spoiled and poorly behaved, cosmopolitan 'creative class' in major cities. The government also employed anti-Western rhetoric, claiming that the protests were orchestrated and paid for by the United States, and that student protesters were like cattle herded around by their paid masters. Protest leaders and participants were called traitors like Judas Iscariot who were part of a fifth column bent on U.S.-supported regime change. Anti-Putin protesters were also described as unpatriotic and anti-Russian, adopting Western European values and rejecting Russian culture and tradition (Koesel & Bunce, 2012; Smyth, Sobolev, & Soboleva, 2013; Volkov, 2015).

In the online space, a coordinated distributed denial of service (DDoS) attack took down approximately twenty independent news sites, including the Web sites of the radio station Echo of Moscow, the oppositional newspaper The New Times, the business newspaper Kommersant, and the Map of Election Violations created by the election monitoring group Golos (Soldatov & Borogan, 2015). DDoS attacks on individual sites in Russia, or against its foreign rivals, are frequent. However, this DDoS attack was rare because just two bot nets took down such a large group of related sites, and because the attacks were timed almost exactly with the opening and closing of the polls on election day. We can assume the attacks were intended to prevent citizens from reporting on election fraud on those Web sites (Roberts & Etling, 2011). Using similar automated tactics to the DDoS attacks, thousands of Twitter accounts were used to spam protest-related hashtags. These accounts were successful in drowning out the protest conversation with garbage text or the same short, simple messages. Twitter was able to delete many of these accounts, but there was little that protest leaders could do in response besides switching to new hashtags, but those were also quickly overwhelmed with spam (Krebs, 2011). Finally, the Federal Security Service, or FSB, asked Russia's largest social networking site, Vkontakte, to take down protest pages because, it argued, some commenters were inciting violence. However, the creator of the site refused to comply with the request so the site stayed active and went on to play an important part in online mobilization (Razumovskaya, 2011; Soldatov & Borogan, 2015).

The government and its proxies also orchestrated a series of pro-government, 'Anti-Orange' rallies that also coopted the opposition's demand for 'honest elections.' The rallies were labeled 'Anti-Orange' because organizers believed the opposition protesters were trying to replicate the 2004 Ukrainian Orange Revolution, which Putin and his supporters argue was orchestrated by the United States. The implication was that such a revolution in Russia would lead to economic decline, political instability, and civil war (Koesel & Bunce, 2012). These rallies were also intended to demonstrate overwhelming support for Putin, especially among 'average Russians.' The pro-government events took place at symbolically important sites, such as Manezh Square in the shadow of the Kremlin walls, and at Reverence Hill (Poklonnaya Gori). Opposition protesters, on the other hand, were denied permits for events at sites such as Red Square and Revolution Square, and instead forced to protest at places like Swampy Square (Bolotnaya Ploshad), on an island in the Moscow River. Some government rallies also took place on symbolically important dates that emphasized Russian war victories and sacrifice, such as the Day of the Defenders of the Motherland (Koesel & Bunce, 2012; Smyth et al., 2013). I discuss the major protest events of both the opposition and the government next.

Major protest events

Large, public protests were the primary vehicle by which the Russian protest movement sought to publicize election fraud, make demands of the government, and win support among the broader Russian public. As noted above, public protests were also a tool of the pro-government counter movement, which sought to demonstrate that most 'average' Russians supported the government and were opposed to an 'Orange Revolution' in Russia. However, it is widely believed that most of the participants at pro-government protests were paid or public sector employees who were coerced to participate (Smyth, 2013; Smyth et al., 2013; Weir, 2012). In this section, I describe the major protest events that took place, on both the anti-government and pro-government sides, but with a primary focus on the oppositional events. Figure 3.1 provides a timeline of major protest events and federal elections.

The anti-government protests began in early December following the Duma election on December 4, 2011 that was widely regarded as falsified and that led to the emergence of the protest movement that would remain active over the next six months. The first significant anti-government protest occurred the day after the election, on December 5, at Chistye Prudy in Moscow with approximately 6,000 taking part, a huge number for an opposition protest (Barakovskaia, 2011; Ioffe, 2011a). The oppositional Solidarity group organized this event before the elections took place. Blogger and anti-corruption activist Alexei Navalny as well as more traditional activists such as Ilya Yashin headlined the event. Navalny called for the protesters to march to the Kremlin, but this was not approved ahead of time by the government and the police quickly arrested the approximately 300 other protesters. This number included Alexei Navalny and Ilya Yashin, who were both subsequently sentenced to 15 days in jail for failing to follow the instructions of a police officer. While the jail term kept these leaders off the streets immediately after the election, it seems to have backfired and made Navalny and other activists even more popular (Ioffe, 2011b).

The month of December also saw the first major anti-government protest at Bolotnaya Square. This event took place on December 10 and drew 50,000 to 60,000 participants (Barry, 2011; Greene, 2013). Major TV channels, which are under the editorial control of the government, did not publicize plans for the protests at all before they took place. Instead, they were promoted online thorough social media and through smaller independent or oppositional media outlets. However, on the day of these major protests even pro-government channels gave largely neutral reports of the protests in Moscow and across the country. There are reports that journalists at one of the government owned stations, NTV, would quit if the protests were not covered (Batty, 2011). Simultaneous regional protests also took place across Russia on December 10, although none were as large as the Bolotnaya protest in Moscow. A second major protest took place on December 24 on Sakharov Avenue in Moscow, with as many as 120,000 taking to the streets despite freezing temperatures (Elder & Parfit, 2011).

There were no major protests in the month of January. January is typically a very quiet month in Russia due to the New Year and Orthodox Christmas holidays. Most commercial and government offices are closed for much of this time, and the protesters largely took a breather during this month.

The protest movement reemerged in February. The next major protest event took place again at Bolotnaya Square on February 4. There were again tens of thousands of participants at the February Bolotnaya protest (Parfit, 2012). On that same date, a large pro-government 'Anti-Orange' counter protest took place at Poklonnaya Hill in Moscow. This pro-government rally rivaled the anti-government protests in size with tens of thousands in attendance. However, there is evidence that the 'Anti-Orange' protests were orchestrated by the government and that many participants were either paid or otherwise coerced to participate (Ioffe, 2012; Koesel & Bunce, 2012; Smyth et al., 2013). A large pro-Putin rally took place at Luzhniki stadium in Moscow on February 23. However, reports by the popular press argued that the crowd at this event were unenthusiastic, largely unwilling to speak to reporters, and again appeared to have been paid or coerced to participate (Weir, 2012). An anti-government protest action took place on February 26 when as many as 30,000 protesters formed a human-chain along the Garden Ring in Moscow (a major street that encircles central Moscow) (Schwirtz, 2012b).

The month of March included the March 4 presidential election and also saw two large anti-Putin protests in Moscow. The first took place at Pushkin Square on March 5, the day after the election, and another on March 10 on Novy Arbat street (L. Kelly & De Carbonnel, 2012). While still drawing tens of thousands of participants, these March protests were smaller than earlier protests, as the energy behind the movement began to wane following Putin's victory in the presidential election (Lankina & Voznaya, 2015; Rosenberg, 2012). The reelection of Putin, and the lack of strong evidence of election fraud that many anticipated, was a major blow to protest movement, and the energy and momentum around the movement began to decrease at this time.

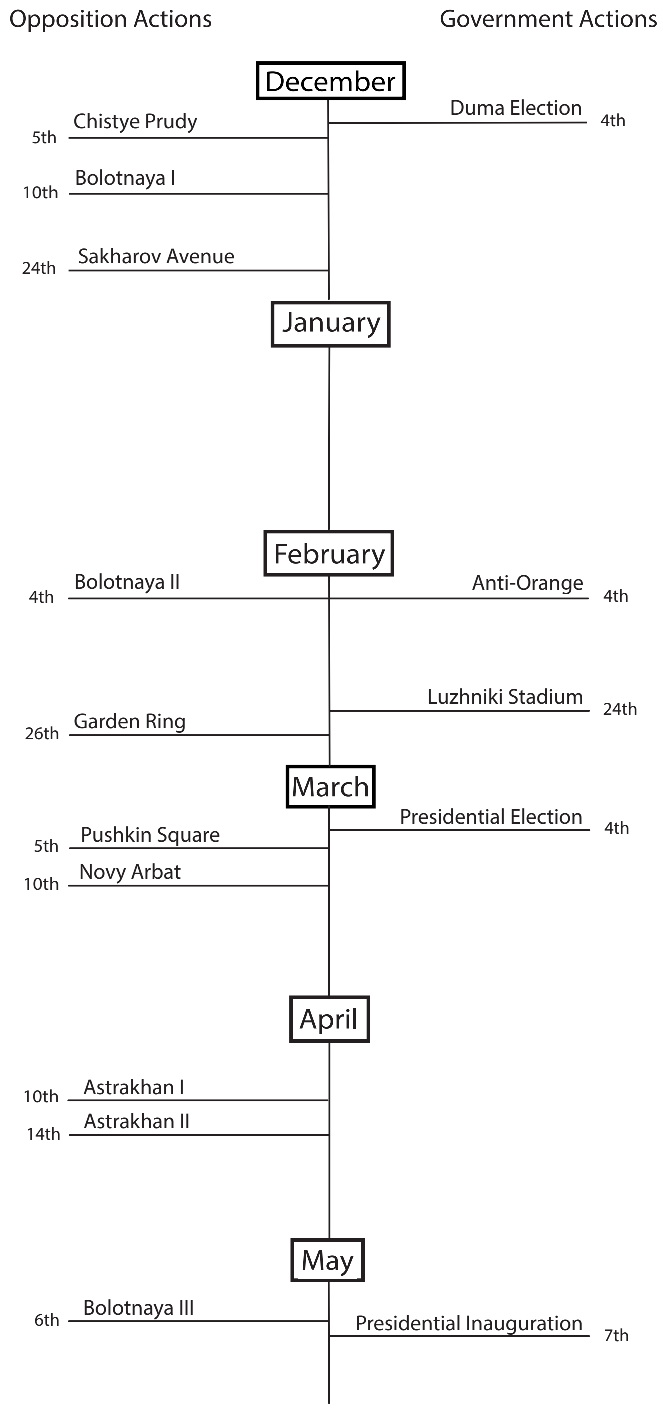


Figure 3.1: Timeline of major protest events and federal elections.

In the month of April, the attention of the entire movement moved to the regional city of Astrakhan, a city located approximately 800 miles to the Southeast of Moscow. The falsification of the Astrakhan elections led to a hunger strike by Oleg Shein, a leftist politician in the region with a national profile.

Shein is part of the Just Russia party, and had previously served two terms in the Duma. Shein argued that the March mayoral election for the city of Astrakhan, in which the pro-government United Russia candidate was named the winner, was falsified. Shein also believed that election officials did not seriously investigate the fraud complaints. He engaged in a hunger strike to raise awareness of the election falsification in the region. National election protest leaders turned their attention to Shein and the Astrakhan election fallout this month. Many traveled to the region to take part in a local protests and also provided moral support on social media (Balmforth, 2012b). Supporters included Alexei Navalny, and two opposition members of parliament, Dmitry Gudkov and Ilya Ponomaryov, among others. Two protests took place in the second week of April in Astrakhan (on April 10 and 14) (Balmforth, 2012b; Herszenhorn, 2012a). These protests were small compared to those in Moscow, with just a few hundred in attendance, but significant for a regional protest, where the numbers of protesters are typically much smaller than those in the major cities (Balmforth, 2012b; Herszenhorn, 2012a).

May 2012 was the final month of major protests connected with election falsification. This month included the final major protest event on May 6 at Bolotnaya Square, which was the day before the presidential inauguration. The protest was meant as another peaceful protest against the election of Putin. The protest drew at least 20,000 participants (Barry & Schwirtz, 2012). Critically, the May 6 protest included violence between protestors and the police. The government used that violence as justification for a crackdown on the protest movement, and May effectively marked the end of the election protests and the movement behind it (Barry & Schwirtz, 2012). It is believed that pro-government actors actually instigated this violence in order to justify the strong government response. Scores of protest leaders and everyday participants were arrested, tried, and convicted as part of the fallout from this last major protest event (Elder, 2013).

Conclusion

The Russian election protest movement emerged spontaneously as a result of falsified parliamentary elections in early December 2011. Over the next six months, despite government countermeasures, freezing temperatures, and the arrest and jailing of protest leaders, the movement was able to organize the largest, most successful protests in Russia since the fall of the Soviet Union. The movement was also the most public challenge to Vladimir Putin since he came to power early in the new millennium, and the most significant political event in Russia in decades. The movement consisted of a broad coalition of groups and individuals, including liberal democrats, nationalists, leftists, informal civic movements, and formal political parties. Their main demand was to nullify the results of the falsified election and to hold new, free and fair elections. As in a number of other protest movements that have emerged around the world in the last several years, the Internet and social media were used to help spread information about election fraud, mobilize participants, and help organize the protests and election monitoring efforts. The protest movement was not ultimately successful in achieving its main goal of forcing the government to hold new elections, but it was able to puncture the perception of Putin's invincibility and significantly damaged the brand of Putin's political party, United Russia. It also showed the power of Russian civil society to mobilize at a national scale, and that the source of real political reform in Russia will come not from the top, but instead bottom-up from its citizens (Shvetsova, 2012).

In this chapter, I provided essential background on the Russian election protest movement that is used as an empirical case study in this thesis. In the next chapter, I explain the data and the methods I use to answer the thesis's main research questions. I use a novel data set of nearly 30,000 online documents and over 3,500 Russian-language Web sites that discussed the election protests between December 1, 2011 and May 31, 2012. I also use a mix of computational social science methods to analyze and extract meaning from this large data set, including supervised learning methods, Exponential Random Graph modeling, social network analysis, and an unsupervised text analysis method, Latent Semantic Analysis.

Chapter 4 Data and Methods

This is an amazing time to be a social scientist. The Internet and social media have radically transformed the amount and type of data available to scholars of politics, collective action, and social movements. This data also allows scholars to ask different questions than in the past. Instead of just studying protest leaders and formal movement organizations, scholars of social movements can now investigate how individuals within movements discuss protests, debate tactics, and enact strategy for the causes they care about. We can now also much more easily analyze change over time in movements, and the way that people communicate within them. In other words, we can more fully investigate cycles of contention, the rise and fall of movements as they make claims against the powerful and attempt to attract new adherents. Critically, we now have digital traces of individual behavior, and are no longer limited to surveys and self-reporting of respondents describing how they feel about issues, or how they think they might act in the future. This is especially important in studies analyzing movements in hybrid regimes, where it can be dangerous for study participants to speak out against the state, and where it can also be dangerous for scholars to study sensitive political topics. The barrage of data available online has also led to challenges. Online data is almost never representative of the larger population under study, and is rarely randomly drawn. New computational social science methods that are not yet part of the standard toolkit for social scientists are necessary to find patterns in the flood of data available online, to identify communities of interest around certain topics, and to test specific hypotheses. I use a large Russian-language data set drawn from thousands of online sources in this thesis, and employ a number of computational social science methods.

In this short chapter, I describe the data used to answer this thesis's main research questions, the methods used to collect that data, and the limitations of the data. I also introduce the methods at a general level. More detailed discussion of the methods and different slices of the data used to answer specific research questions are left for each of the substantive chapters that follow. I begin this chapter with a review of the research questions asked in this thesis and provide an overview of the data and methods. I next discuss the core Media Cloud tool and database, and then discuss in detail the specific Russian-language data sources in Media Cloud that the protest data was initially drawn from. I then describe the method used to create the larger Russian protest data set that is used throughout this thesis, and provide some descriptive statistics for that data set. Finally, I introduce the methods used to investigate each research question.

Overview

I address three research questions in this thesis. The first asks if a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if that structure changes over time. Polycentric networks are conducive to high-risk collective action and are more robust to defections or the removal of key nodes than hierarchical networks. The second research question seeks to understand if brokered information is actually novel, and if brokered text is more novel in polycentric networks than in hierarchical ones. The last research question asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames, and within such a movement, whether moderate or extremist framing is more prevalent. I use the Russian election protests that occurred after falsified national elections in 2011 and 2012 as an empirical case study.

To answer these questions, I draw on a unique dataset of nearly 30,000 Russian documents and the linking structure of over 3,500 Web sites that discussed the Russian election protests between December 1, 2011 and May 31, 2012. These dates include the peak of protest activity, as well as the parliamentary and presidential elections. The protests data set was created by first identifying all the Russian-language documents in the Media Cloud database that discussed the Russian election protests. Links were extracted from all of those documents and used to feed a spider that identified thousands of additional sources and documents that also discussed the protests. I apply a number of computational social science methods to this data set. This includes Exponential Random Graph Modeling, an advanced statistical model for social networks that I use to determine if networks are polycentric or hierarchical. An unsupervised method, latent semantic analysis (LSA), is used in combination with social network analysis to determine whether brokers have access to novel information. Finally, supervised learning techniques are used to investigate the framing research question. Next, I describe the Media Cloud system and its Russian-language sources.

Media Cloud platform and data

The data used in this thesis was drawn initially from Media Cloud, an open source platform for the discovery, crawling, extraction, and storage of text from online sources (mediacloud.org). Media Cloud was created as part of a joint project of Harvard's Berkman Center for Internet & Society and MIT's Center for Civic Media to aid researchers in the quantitative study of media. The Media Cloud platform collects online content several times a day from tens of thousands of sources, primarily online news and blogs, and stores that data in a large PostgreSQL database. Data is primarily collected from each site's RSS feeds, but some sites are manually scraped if RSS feeds are not available. There are often multiple RSS feeds for each Web site. For example, the New York Times and Washington Post each have over 120 feeds, while a blog will typical have just one. As of June 2016, the Media Cloud platform crawls 65,000 feeds from 25,000 online sources. The database includes over 425 million text documents (which the Media Cloud team refers to as stories, whether they are traditional news stories, blog posts, or other online content). Every day 300,000 documents are added to the database. The text of those documents has been split into 5.3 billion sentences that are stored in a Lucene/Solr database for quick searching and retrieval. The text data is accessible via an API (Benkler et al., 2015; Etling et al., 2014).

The Media Cloud platform collects primarily English-language content, but also includes foreign language data. In total, Media Cloud currently includes language support for and data from media sources in sixteen languages: Danish, German, English, Spanish, Finnish, French, Hungarian, Italian, Lithuanian, Dutch, Norwegian, Portuguese, Romanian, Russian, Swedish, and Turkish (mediacloud.org). The Russian-language collection is one of the largest foreign language sets in Media Cloud. Data collection for most of the Russian-language sources began in January of 2011, approximately eleven months before the election protests. The Russian-language collection includes text and links drawn from thousands of blogs, mainstream media sources, national TV Web sites, and Russian government Web sites (Etling et al., 2014). Individual sources are grouped into media sets that are described in detail next. The sources in these media sets were queried to identify all the documents in the Media Cloud database that discussed the election protests.

Russian top 25 mainstream media

This set includes the top 25 most popular news Web sites in Russia according to Google Ad Planner's data on unique Russian speaking users as of January 2011. This excludes sites affiliated with TV stations as they are captured in a separate TV news media set described below. This set consists of online news sites, sites affiliated with major traditional newspapers, and Russian news agencies. Table 4.1 lists the name and URL of the top 25 Russian mainstream media. At the time of the protests, newspapers and online media generally enjoyed more editorial freedom than national TV channels. This is at least in part because these outlets had smaller audience than national TV channels, which reached nearly 90% of the population (Etling et al., 2010; Lipmann & McFaul, 2010).

Russian TV news channels

This set includes text from nine major TV news channels in Russia. Russian news channels are firmly under the editorial control of the Kremlin, and are an important tool of political control in Russia (Lipmann & McFaul, 2010). All major national TV channels are either under direct control of the government or state-owned enterprises, such as the gas monopoly Gazprom (Etling et al., 2014; Lipmann & McFaul, 2010). A sense of how biased Russian TV news is in favor of the government can be seen from the results of a study that analyzed content from a random sample of TV newscasts in 2011 and 2012. That study by the Central Intelligence Agency's Open Source Center found that on average 96% of the airtime in a 30 minute newscast is devoted to coverage of either the Russian President or the Prime Minister (*Lapse in airtime targets suggests effort to improve Putin's image*, 2013). The sites included in this data set are Channel 1, NTV, Vesti, REN TV, TV Tsentra, Channel 5, Mir, Zvezda, and TV Stolitsa. Zvezda is a Russian military channel, and TV Stolitsa is focused on the city of Moscow. The others are mainstream news channels with national reach and geared towards a general audience. Transcripts of these news sites would have been preferred to content scrapped from the Web sites of TV channels, but the cost of transcripts was found to be prohibitively expensive.

Russian government Web sites

This media set aims to include major parts of the federal government's official Web presence. It includes text and links from President Putin's official site, Prime Minister Medvedev's site, the Russian government portal government.ru, and the Web sites of the Ministry of Emergency Situations, Ministry of Justice, Ministry of Defense and the Ministry of Foreign Affairs.

Table 4.1: Russian Top 25 Mainstream Media Set

Rank	Name	URL	Description
1	RIA Novosti	rian.ru	News agency
2	Komsomolskaya Pravda	kp.ru	Newspaper
3	Lenta	lenta.ru	Online news site
4	Gazeta	gazeta.ru	Online news site
5	3D News	3dnews.ru	Technology news and links
6	Regnum	regnum.ru	News agency
7	Vzglad	vz.ru	Conservative news site
8	Newsru	newsru.ru	Online news site
9	Svobodnaya Pressa	svpressa.ru	Online news site
10	Inosmi	inosmi.ru	Translation of foreign news
11	Vedomosti	vedomosti.ru	Business newspaper
12	Argumenti i Fakti	aif.ru	Newspaper
13	Rossiskaya Gazeta	rg.ru	Russian Government Gazette
14	Pravda	pravda.ru	Newspaper
15	Cnews	cnews.ru	Technology news
16	Dni.ru	dni.ru	Online news site
17	Rosbalt	rosbalt.ru	Online news site
18	Interfax	interfax.ru	News agency
19	Kommersant	kommersant.ru	Online news site
20	Moskovskii Komsomolets	mk.ru	Newspaper
21	Expert Online	expert.ru	Business news
22	Izvestiya	izvestiya.ru	Conservative newspaper
23	bfm.ru	bfm	Business news portal
24	Trud	trud.ru	Newspaper
25	fontanka.ru	fontanka.ru	St. Petersburg news

Note. Rank determined by Google Ad Planner for unique Russian speaking users as of January 2011. The Web site of TV news channel Vesti ranked #3 but is not included in this media set. Instead it is included in the Russian TV media set. Many other popular news sites, such as the blog of Echo of Moscow, are also included as part of the Yandex top blog set.

Russian-language Blogs

There are three sets of Russian-language blogs that are included in the Media Cloud database. I discuss each next. It is worth pointing out first, however, that what I refer to as blogs are actually hybrid platforms that include elements of both social networking sites and traditional blogs (Etling et al., 2010). LiveJournal and other popular blogging platforms in Russia allow users to post content in a traditional blog format, with user-created content listed in reverse chronological order, but also provide social networking features, such as friends and followers lists and public profiles. Indeed, LiveJournal, which constitutes the majority of blogs in the Media Cloud and protest data sets, meets all the definitional requirements of a social network. Social networking sites are defined as:

> Web-based services that (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. (Boyd & Ellison, 2008, p. 211)

LiveJournal provides all of these functions. Boyd and Ellison include LiveJournal among a list of representative social networking sites, and also show that it was one of the first ones created. The fact that LiveJournal and other blog platforms in Russia have social networking features may explain, at least in part, why they remained so popular in Russia for so long, while more traditional blog platforms popular in other countries lost users to social networking sites, or were absorbed into more traditional mainstream media outlets.

Berkman top blogs

This set includes 10,734 of the most linked to Russian blogs as identified as part of a large-scale study of the Russian-language blogosphere by Harvard's Berkman Center and Morningside Analytics (Etling et al., 2010). This set was created initially from a Yandex list of over five million Russian-language blogs that was collected in May 2009. This list was used to feed a crawler that identified several thousand additional blogs. However, it was found that many of these blogs were inactive or had very low levels of posting activity. The Etling et al. study focused on data collected from approximately one million active blogs from May 2009 through September 2010. From this set of one million, the most linked to 11,792 were identified for social network and content analysis. An updated, slightly smaller list of 10,734 blogs from the original study was provided to the Berkman center for inclusion in the Media Cloud database. Data collection began in January of 2011. The Berkman blog study focused in part on identification and description of the major communities that discuss politics and public affairs, although there are a fair number of blogs more concerned with culture than politics. The major politics and public affairs clusters cover a broad range of attitudes and agendas, but the study found that most oppositional bloggers discuss politics from a non-aligned perspective. That is, they are typically opposed to the government, but they do not do so as part of a political party or from a clearly identifiable partisan political perspective. This is unsurprising to most Russia area scholars given the generally low levels of support for Russian political parties (Volkov, 2015). However clusters that represented well-known leaders and parties that are part of the democratic opposition in Russia exist. Nationalists also form their own large community in the network. There are also communities focused on finance and economics, and Russian and international political news. Pro-government bloggers were not especially prominent in the network, although to the extent they exist, they revolve around pro-government youth groups (Etling et al., 2010).

Yandex top blogs

This media set includes 1,140 blogs from the 'top blogs' list generated by the Russian search engine Yandex. Yandex is the most popular search engine in Russia, and is often referred to as the 'Google of Russia.' This top blogs list is a ranked list of the most linked to Russian-language blogs according to Yandex. This list includes the blogs as they were ranked on January 3, 2011. Yandex stopped ranking and curating top blogs and top blog posts in 2014 when Russian legislation was introduced that would treat any blog with a daily audience in excess of three thousand readers as a mainstream media outlet, forcing their registration with the government and making bloggers responsible for the accuracy of information published (Denisova, 2016; MacFarquhar, 2014). Therefore, it is not possible to find a similar list today. This list was used to identify the most popular bloggers in Russia using a metric separate from the Etling et al. study.

Russian long-tail blogs

In an effort to include a set of blogs that are not extremely popular or highly linked to, this final set consists of a random sample of 1,184 Russian-language blogs that were identified by a spider built at the Berkman center (separate from the Morningside Analytics blog set). That spider found approximately 1 million blogs. Links to blogs typically follow a power law distribution, in which a few sites receive the vast majority of the links while the majority of sites receive just a few links, or none at all. While this media set could by chance include popular blogs, the aim of creating this set was to have a collection of blogs from the long tail of the indegree distribution. That is, blogs that are not frequently linked to and are more likely to represent an average, everyday Russian blogger without a huge audience.

This core Russian-language data set was created as part of a MacArthur Foundation funded study of the impact of the Internet on Russian politics and society at the Berkman Center. One of the objectives in the creation of this data set was to include both oppositional and pro-government sources. Too often scholars of social movements study only movements whose goals they agree with politically, or fail to include data from the counter-movements. Including data from pro-government sources helps to counter this trend. It is also necessary for testing whether the government is better able to spread its preferred framing of the protests compared to the opposition, which I explore in Chapter 7. Now that the core Media Cloud data sources have been described, the specific method used to create the protest data set is described next.

Russian protest data

To create the Russian protest data set that is used throughout this thesis, a regular expression was first used to identify all the Russian-language documents in the Media Cloud database, drawn from the media sets described above, that include terms associated with the protests. Specifically, the regular expression searched for documents that included the Russian terms for 'election' plus 'protest,' or 'rally,' or 'march' (Russian language data is used throughout this thesis. The author, who has general professional proficiency in Russian, did any translations shown throughout the thesis text). I reviewed 1000 randomly selected stories that resulted from the use of this regular expression to ensure that these documents were indeed relevant to the election protests. Less than 10 were off topic. A broader search pattern that just looked for the word 'protest' was also tried, but this returned a large number of stories that were completely off topic, for example finding a story about a wrestling match in St. Petersburg where the coach 'protested' against the results. A spider was then used to identify additional sources and documents that discussed the protests. The link data from stories in the Media Cloud data set was then used to feed a spider that discovered additional links and stories that also matched the regular expression from the initial data-mining step. Links are drawn from the main body of a document, so the story in mainstream news article or the links in a post by a blogger. This is particularly important for blogs since this does not take account of friends on LiveJournal or a blog roll on a more traditional blog. This approach allows for a standard way to mine links from multiple online sources, which is an important methodological step in a general move away from single-platform studies and towards multi-platform research. Any relevant stories or media sources that did not already exist in the database were then added to the protest data. The crawler ran for 94 iterations, until only one or two new documents were being discovered in each run. This method resulted in a data set consisting of 29,544 time-stamped documents, 3,583 Web sites, and 23,041 links. This data collection method follows that of Benkler et al. (2015) in their study of mobilization around Internet-related legislation in the United States. This data collection method has also been used to study a range of different substantive topics, including the Trayvon Martin killing in Florida (Graeff, Stempeck, & Zuckerman, 2014), the debate over net neutrality legislation in the United States

(R. Faris, Roberts, Etling, Othman, & Benkler, 2016), and a case study of the quality of information sources that were cited during the Ebola outbreak (Roberts, Seymour, & Fish, 2015).

Next, I provide some descriptive statistics from the protest data set. Table 4.2 shows the top ten sites by indegree, or the number of inlinks each site receives. The top sites are a mix of news sites, blogs, user-generated content sites, online communities, and a social network. It is worth highlighting that even though social networking sites such as Vkontakte were not included in the original data set, the spider was able to pull in data from that site, and it is one of the most linked to in the protest data. Although it is worth noting that the Media Cloud code collapses multiple feeds for the same site into one media source unless a researcher specifies that a given Web site's feeds should split into individual media sources. This is primarily because traditional newspapers have multiple feeds. Future research could split the Vkontakte data into individual media sources for each page in the data. I discuss the impact of this choice in the limitations section of this chapter.

Table 4.2: Top Sites by Indegree in Protest Data

Rank	Name	Media Type	Indegree	Documents
1	Novaya Gazeta	News	1117	117
2	Oleg Shein (LJ)	Blog	774	1103
3	Vkontakte	Social Network	485	120
4	Lenta.ru	News	363	383
5	Gazeta.ru	News	358	356
6	Picasa Web Photos	User-generated	355	71
7	YouTube	User-generated	351	205
8	Solidarity	Community	326	339
9	Rosbalt	News	302	142
10	LGBT Grani	Community	288	290

Note. Vkontakte totals are aggregates of all Vkontakte pages included in the data. Oleg Shein's page is his LiveJournal (LJ) blog.

Figure 4.1 shows the distribution of documents over time. Peaks in document frequency are the results of major oppositional protest events and federal elections. These peaks are labeled with the names of these events. As part of Chapter 5, the ERGM modeling chapter, the data was divided into five temporal networks. The dashed lines on Figure 4.1 indicate these five time periods, which roughly correspond with the months of December, February, March, April, and May. As part of that chapter, Web sites were also coded into different media-type categories. Table 4.3 shows the number of documents generated by each media type, and the number of Web sites that fall into each category. Table 4.3 shows that blogs, news sites, and communities account for the largest number of documents in the protest data set, while the largest number of sites in the data are blogs, news sites, and Web sites of NGOs. Since this data only accounts for the months that were included in the ERGM models, overall totals for the entire protest data set would be slightly higher.

Different slices of the protest data set are used in the three substantive chapters of this thesis, depending in part on the method used. In the next section I discuss the limitations of the data collection process and sampling frame. I then introduce the methods used in the thesis before concluding the chapter.

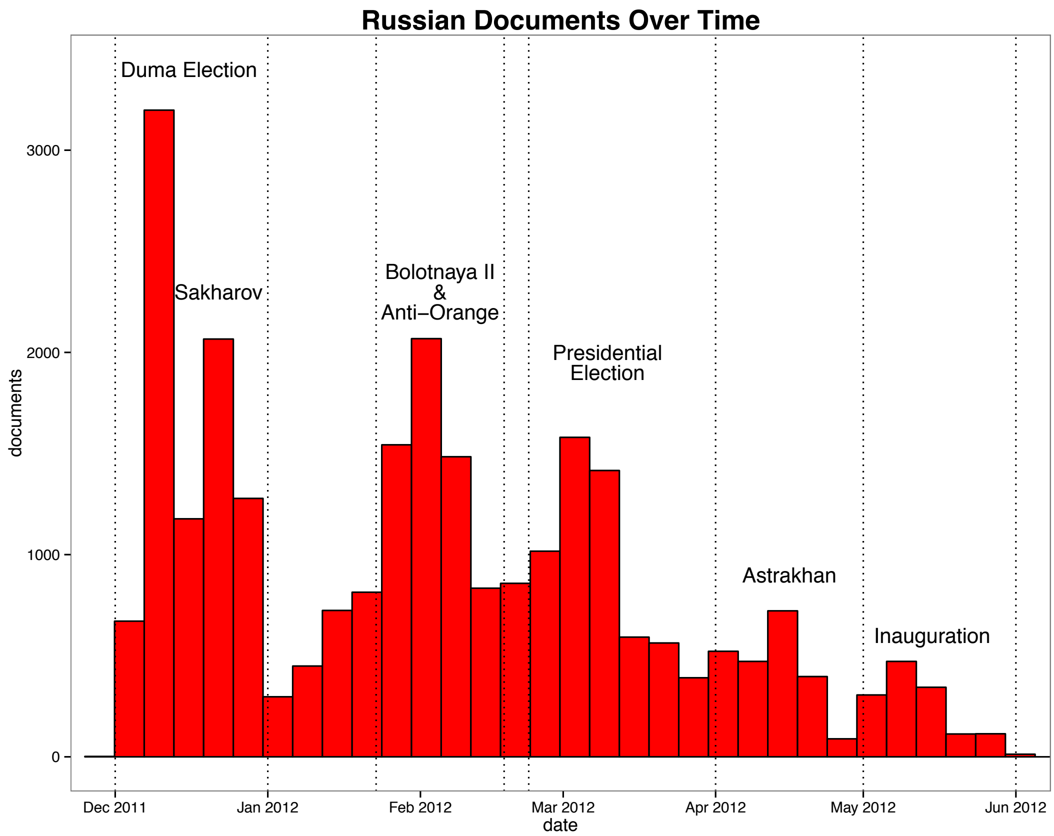


Figure 4.1: Distribution of documents over the full six months of protests. Dashed lines represent the five protest periods that are used in the Exponential Random Graph Modeling chapter (Chapter 5).

Table 4.3: Number of Documents and Web Sites by Media Type

Media type	No. Documents	No. Sites
Blog	7224	2293
News	2350	306
Community	438	35
NGO	244	112
User-generated	142	10
Online tools	85	22
Social Network	84	12
Other	52	38
Personal	48	34
Commercial	44	35
Government	15	12

Note. No. = Number. Includes only sites that were coded by media type as part of ERGM modeling chapter. The time periods modeled in that chapter correspond with major protest events and election. Those networks include nodes that sent or received at least one link in December, February, March, April, and May. Overall totals would be slightly higher.

Limitations

The main limitation of this data set, and the data collection method, lie in the initial choice of sources to include in the data set. The core Russian-language Media Cloud data set includes a large and diverse set of sources, but it is far from a universal representation of the Russian networked public sphere. The objective of creating the Media Cloud Russian data set was to have a range of sources that included traditional Russian media outlets, including text from national TV channels, Russian government Web sites, and a wide range of Russian blogs. Russian blogs, particularly LiveJournal, which was the most popular blog platform in Russia, had a very large and diverse audience and user base at the time of the protests. According to Yandex, there were over five million Russian-language blogs in the year before the protests (Etling et al., 2010). Still, other US-based social media platforms like Twitter and Facebook were growing in popularity by late 2011. A number of homegrown Russian platforms such as Vkontakte and Odnoklassniki were also rising in popularity at this time.

To attempt to identify text and links from other popular platforms and news sites that discussed the protests, but that were not included in the core Media Cloud data set, the Russian protest data was created by feeding a spider or Web crawler that mined links from all the documents in Media Cloud that discussed the protests. This process was described above. This method helped to identify a large number of additional sources and documents not originally included in core Media Cloud data set, and I believe helps to overcome some of the limitations in the choice of the original sampling frame. The Web crawler discovered about one-third of the documents in the Russian protest data set. More precisely, the spider found 13,566 documents that were not included in the original Media Cloud database, and also discovered 1,795 new media sources. This process led to the inclusion of popular protest pages on the Vkontakte social network, but resulted in relatively few Twitter accounts. It is also important to note that the nature of link mining is such that it is difficult to find links outside the core of a network. Instead, link crawlers tend to find additional links belonging to similar nodes towards the center of a network, instead of peripheral nodes or communities on the edges of a network. This means that smaller communities that are not part of the mainstream conversation may be missing from this data set. Future research could attempt to add additional social media sources.

As noted above, the Vkontakte pages in the data are collapsed into one node as part of the data collection process for social network analysis. This is the default behavior in Media Cloud, unless it is manually changed beforehand, and is not an ideal scenario in terms of trying to understand the importance of various Vkontakte pages in the protests. This was not realized until late in the writing of the thesis, and is another possible limitation. This likely results in Vkontakte appearing more important than its individual protest or participant pages in the social network data and in tabulations of top media sources by inlinks, such as in Table 4.2. Given that Vkontakte is more popular with general users within Russia, as opposed to the elites who lead the protest movement that preferred to use Facebook and LiveJournal (Soldatov & Borogan, 2015), this probably results in less individualized data on regional participants outside of the major cities. There may also be more pro-government content on Vkontakte due to its large user base. While it is always better to have more data, I do not believe too much is missed in terms of tracking major actors or groups in the protest movement.

This is born out by Sherstobitov (2014) who analyzed major oppositional and pro-government Vkontakte groups during and after the election protests. The oppositional groups he identifies on Vkontakte are very similar to those found in the protest data I use in this thesis, while the pro-government sites are generally less well connected, are more likely to be isolates, and have fewer members than the oppositional groups. Based on survey data, S. White and McAllister (2014) find that Vkontakte was less influential than Facebook in mobilizing support for anti-regime protests. They suspect that this may be due to the fact that Vkontakte is Russian owned, has its servers in Russia, and therefore would be more susceptible to Russian government pressure than Facebook. Similarly, Reuter and Szakonyi (2015) find that Twitter and Facebook were more important for increasing political awareness of electoral fraud in the 2011 protests than Vkontakte or the other popular homegrown Russian social network, Odnoklassniki. They also suggest this is due to users expectations of greater privacy on the foreign owned social networks. In sum, Vkontakte is a very popular Russian social network and more data from more platforms is always preferable in this type of research. However, I do not see strong evidence to date that Vkontakte was more important than LiveJournal, Facebook, or Twitter in mobilizing anti-government protesters, or that it included a significantly larger pro-government user base that discussed the protests. Existing research reinforces the need, however, for obtaining Twitter and Facebook data in the future.

Another possible limitation is that some sites could have become popular after the data collection began. However, the data collection process started less than one year before the protests began, generally using the most recently available ranking data. The spider used to create the protest data set should also have found popular sites that are not in the core Media Cloud data.

While an effort was made to identify pro-government content, it is possible that pro-government sources are underrepresented in the data. However, this is at least in part due to the fact that pro-government bloggers in particular are not frequently linked to in Russia. Russian Twitter is more likely to have pro-government users and content, so that is another reason to try to include data from that platform in future research (J. Kelly et al., 2012). In the end, as the framing chapter (Chapter 7) will show, there was no difficulty in finding good examples of pro-government sources or sentiment in the protest data.

To conclude this limitations discussion, the protest data set includes a large and diverse set of online Russian-language sources. The breadth of this set is significant when compared to other studies of protest movements that use online data. Those studies typically limit their analysis to short time periods and usually draw data from just one platform. Still, the main limitation of this data set and data collection method lies in the initial choice of sources and platforms in the core Media Cloud data set. To address this limitation, an automated spider was used to create the final Russian protest data set. That spider successfully identified over 10,000 additional stories and over 1,500 media sources that discussed the protests. Future research could nonetheless include additional sources from other social media platforms. With the data set and its limitations now fully described, I turn next to a brief introduction of the methods used in this thesis.

Methods

While online data offers new opportunities and significant new data sources, it also requires the use of new methods that are not part of the traditional social science tool kit. Instead, what are increasingly referred to as computational social science methods are required (Giles, 2012; Lazer et al., 2009). This set of new methods typically refers to techniques for the study of social networks, which look at the links between individuals or Web pages, as well as another class of methods that allow for automated analysis of the massive amounts of text created online. In this thesis, I use both network and text-based approaches, and describe each next based on the chapter in which those methods appear.

In the first substantive chapter (Chapter 5), I build a series of networks around five document frequency peaks, which are caused by major protest events. A limitation of social network analysis is that it has had to rely on descriptive network statistics instead of statistical inference, since the links between nodes means that the observations are not independent. The independence of observations is a strong assumption for most traditional statistical methods. However, new statistical methods for social networks have been devised and are growing in popularity. In Chapter 5, I use one these statistical methods, Exponential Random Graph Modeling. Exponential Random Graph Models (ERGMs) are part of a family of statistical models for social networks. ERGMs allow researchers to account for the presence or absence of ties in a network, and therefore a way to understand the micro-processes that account for macro structure in an observed network of interest. An ERGM models a given network in terms of a set of theoretically relevant parameters, or terms, selected by the researcher that account for local tie-based structures. These parameters consist of processes that are exogenous and independent of network ties (such as node attributes, like location or media type) as well as endogenous or structural terms, where observations are dependent (such as reciprocity or transitivity). ERGMs therefore allow social scientists to understand how these endogenous and exogenous parameters interact to create and maintain network structure (Lusher et al., 2013). I build a series of Exponential Random Graph Models to understand if the network has a polycentric or hierarchical structure. Polycentric networks are more conducive to high-risk collective action and are more resistant to targeted removal of key nodes. I use node location as one of the parameters in those models. Since the Media Cloud software does not collect that data, I wrote a Python script to pull location data from the profile pages for all LiveJournal accounts and manually coded the remaining sites in the data set.

Those who use computational social science methods tend to develop expertise in either social networks or in automated text analysis. The two methods are rarely combined in a robust fashion. In Chapter 6, however, I use both automated text analysis and social network analysis methods to investigate a bedrock assumption in the network theoretic literature. Specifically, I investigate whether brokers, nodes that connect different communities in social networks and are hypothesized to play special bridging and translating roles between groups in social and revolutionary movements, have access to and share novel information. The brokerage literature rests largely on the idea that brokers have vision advantages gained from access to novel information and that they are able to translate this information between groups that are otherwise isolated. However, this idea is rarely actually tested using node level textual data.

In broad strokes, my method in the brokerage chapter (Chapter 6) is to partition the network into groups, identify brokers, extract their ego nets, and use latent semantic analysis (LSA) to determine the similarity, or novelty, of the text of the nodes in each brokerage relationship. Specifically, I use social network analysis techniques to identify communities in the protest network, as well as the nodes that play different sorts of brokerage roles. I use the Louvain method to partition the nodes in the network into non-overlapping groups (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). I then identify brokers of various types based on the theory and method of Gould and Fernandez (1989), and focus on the brokerage relationships where a broker connects nodes from different communities. I then extracted the ego network of each broker with both in and out links at a distance of one. I then use techniques popular in the information retrieval field to determine the novelty of brokered text. Specifically, I weigh the terms in the term-document matrix by term frequency-inverse document frequency (TF-IDF), identify latent topics with Latent Semantic Analysis (LSA), and calculate the similarity of topics among documents using cosine similarity. TF-IDF identifies important words by comparing the frequency of a word within a document to its frequency within the entire corpus (Manning, Raghavan, & Schutze, 2008). Latent Semantic Analysis (LSA) is a method used to identify latent topics in documents through the co-occurrence of words (Landauer & Dumais, 1996, 1997; Landauer et al., 1998). The basic idea behind LSA is that topics or concepts can be found in patterns of words that occur together in documents. So, for example, the co-occurrence of the words

Obamacare, health, and insurance could represent a document discussing the Affordable Care Act and troop, war, Taliban and Afghanistan could indicate discussion of the war in Afghanistan (Grimmer & Stewart, 2013). Finally, cosine similarity is used to rank the similarity of the latent topics. Cosine similarity is a method used to measure the similarity of any two vectors or documents by measuring the cosine of the angle between those two vectors. This results in a score between 0 and 1, where documents closer to 1 are more similar and those closer to 0 are the more dissimilar, or in this analysis, novel. In Chapter 6, I also explore whether the novelty of text is impacted by macro network structure, by testing whether brokered text is more novel when the network has a polycentric structure.

In the final substantive chapter, Chapter 7, I use a non-parametric supervised learning, or machine learning, method developed by Hopkins and King (2010). I use this method to determine if the framing by the government of the protest movement is more prevalent than the opposition. Within the movement itself, I also test if the framing of the mainstream or extremist constituencies is more prevalent. I use all the text from all documents in the protest data set between December 1, 2011 and June 1, 2012. Machine learning requires creating a small training set of documents that represent categories of interest to the researcher. The machine-learning algorithm then identifies the proportion of text in the larger corpus that match each of the categories of interest. I create a training set that includes sources and text that represent the framing of each of the major protest constituencies, as well pro-government sources and frames. Most supervised methods classifying individual documents into categories, assess the level of accuracy of that classification process and further tune the machine for acceptable levels of true and false positives (precision and recall), and then aggregate individual classifications across the entire corpus. The Hopkins and King method skips this intermediate classification and aggregation step. Instead, it directly estimates document category proportions in the corpus based on the word profiles for each category in the training set (Hopkins & King, 2010). I discuss each of the methods introduced in this section in detail, along with their strengths and weaknesses, in each of the substantive chapters that follow.

Conclusion

In this brief chapter, I described the data used to investigate the role of polycentric networks, brokers, and framing in high-risk collective action. Since I use the Russian election protests that emerged after elections widely viewed as falsified as the empirical case study, I rely on a large dataset of Russian-language text and links drawn from the Web. I described the core Russian Media Cloud database and its sources, which consists of data from top Russian mainstream media sources and a large, diverse set of over ten thousand Russian-language blogs. I then described the method for creating the Russian protest data set used throughout this thesis. That method involved using a regular expression to identify documents in the core Media Cloud data set that discussed the protests, and extracting the links from those documents. Those links fed an automated spider that discovered over 13,000 thousand additional documents from over 1,700 new media sources. This data collection process resulted in a final dataset consisting of 29,544 time-stamped documents, 3,583 Web sites, and 23,041 links. Unlike many previous studies of collective action, this data set covers the entire six months of protest activity, not just one month or one major protest event. Finally, I also introduced the computational social science methods used in the thesis. These methods include social network analysis, Exponential Random Graph Modeling, machine learning, and Latent Semantic Analysis.

Thus far in this thesis, I have introduced the research questions, developed related hypotheses, and laid out my analytical framework. I also provided essential background information on the Russian election protests. With the data and methods now also described, I turn in the next three chapters to the substantive research undertaken as part of this thesis. I begin, in the chapter that follows, by investigating the relevance of network structure for the Russian protest movement. Specifically, I develop a series of Exponential Random Graph Models to determine whether the movement has a polycentric or hierarchical network structure, and if this structure varies over time. Both types of structure have potential advantages for high-risk collective action. Polycentric networks are made up of clusters of strongly connected actors that are tied together by brokers, and are more resistant to repression than hierarchical networks. Hierarchical networks, on the other hand, have clear leadership and decision-making structures that can more easily create and implement strategy, and can more easily diffuse protest information and mobilize protesters to turn up at events.

Chapter 5 The Importance of Tie Strength and Network Structure for High-Risk Collective Action

> *High-risk activism…is a 'strong-tie' phenomenon. … The platforms of social media are built around weak ties. This is in many ways a wonderful thing. … But weak ties seldom lead to high-risk activism.*
>
> *--Malcolm Gladwell*

> *If weak ties are critical to building bridges between different tight-knit social networks, then blogs and social networking sites like Facebook might have an incredibly important role to play in amplifying weak ties, making them transparent and usable, and simplifying the process of activating them.*
>
> *-- David Faris*

The relationship between strong and weak ties is often presented as one of opposition and tension, when in fact it should be seen as one of complementarity and mutual benefit (Tufekci, 2010). This is why networks consisting of both strong and weak ties are necessary for successful protest movements. Strong ties are especially important for risky collective action. They build internal group solidarities, are more likely to influence behavior than weak ties, allow protest participants to receive repeated affirmation of protest information, and reinforce the normative importance of participation (Centola & Macy, 2007; McAdam, 1988). Weak ties allow groups of friends with strong tie bonds to receive novel information from groups they would not otherwise be exposed to, and also aid in community mobilization and diffusion processes (Borgatti & Lopez-Kidwell, 2011; Granovetter, 1973, 1983). Weak ties also allow disparate social groups that rarely interact with each other, such as nationalists and gay rights groups in Russia, to come together under a common protest banner. Polycentric networks have both strong and weak tie structures, with clusters of strongly connected actors linked together by weak ties (Baldassari & Diani, 2007). These networks resemble Gerlach's (2001) SPIN models of successful social movement organization because they are segmented, polycentric, and innovative. These types of networks are conducive to high-risk collective action because they are more robust then hierarchical networks, more resistant to repression, can penetrate and recruit from different segments of society, and encourage social innovation and problem solving (Baldassari & Diani, 2007; Gerlach, 2001).

The specific research question addressed in this chapter is whether a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if this structure changes over time. I hypothesize that the Russian protest network will have a polycentric structure only at the beginning of the protests, and will move towards a less effective hierarchical structure as the movement loses popular support. In order to test this hypothesis, five Exponential Random Graph Models (ERGMs) are built and fitted, each representing a peak period of protest activity identified by spikes in story frequency over the six months of Russian election protests under study.

This chapter begins with a general description of the data and a brief review of the major election protest events in Russia. The data collection process for node attributes is then explained and descriptive statistics for the five protest networks are provided. In the method section, the ERGM modeling approach is justified and an overview of the modeling process is given. The parameters used in the models are then described. The results section begins with a detailed description of the April model to explain the ERGM modeling process, the choices made during modeling, and the methods used to assess goodness-of-fit. Summary results are then provided for the remaining final models, followed by a discussion of goodness-of-fit for those models. The discussion and conclusion section compares the topline results of the various models to identify general trends, including whether the models have microstructures indicative of polycentric or hierarchical networks. The chapter concludes with a summary of the findings, proposals for how the models might be improved, and suggestions for future research.

Data

Five periods of peak protest activity are modeled and compared. Peaks are identified based on the distribution of stories between December 1, 2011 and May 31, 2012 (Figure 5.1). These peaks coincide with major protest events and the legislative (Duma) and presidential elections. The month of December 2011 is the first time period modeled and includes the falsified elections on December 4 that led to the emergence of the protest movement. The first significant anti-government protest occurred on December 5 at Chistye Prudy in Moscow , and the first large protest at Bolotnaya Square on December 10 that drew as many as 50,000 participants (Barry, 2011). Smaller regional protests also took place across Russia on December 10. This time period also includes a second large protest on Sakharov Avenue in Moscow on December 24 that had up to 120,000 participants (Elder & Parfit, 2011).

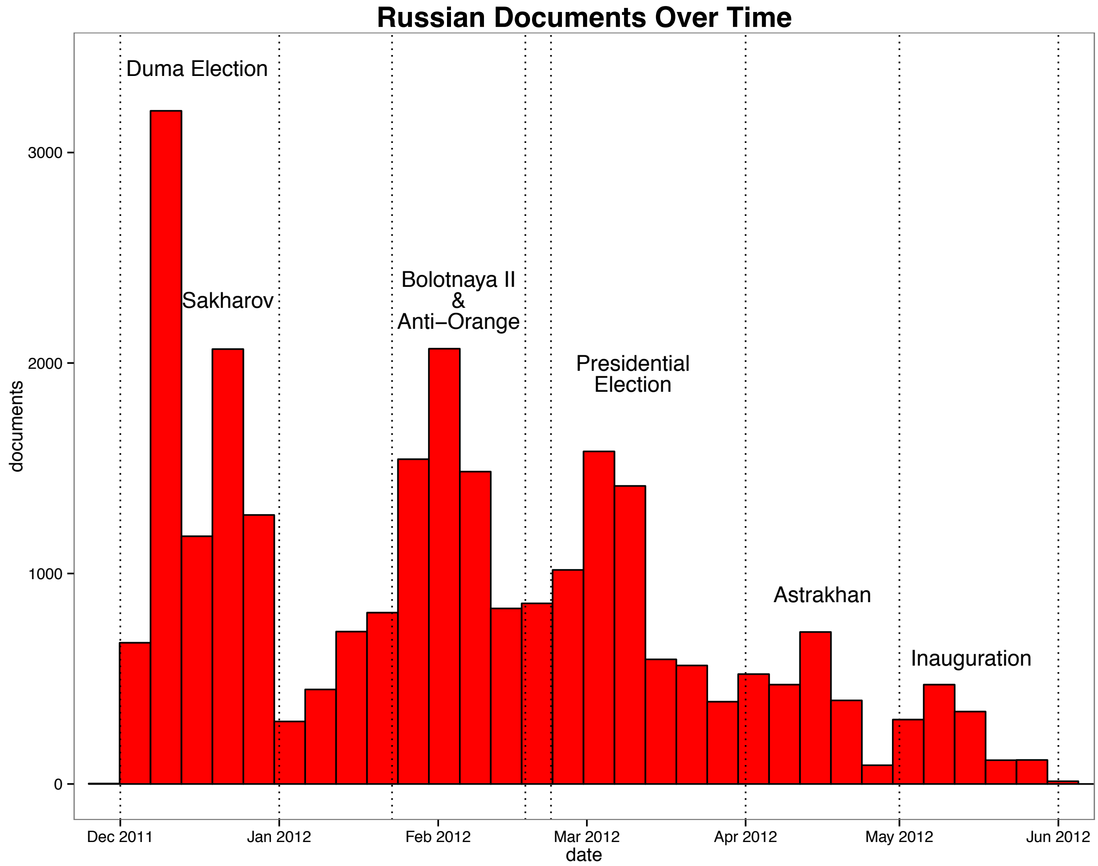


Figure 5.1: Protest peaks based on document distribution. Dashed lines separate the five protest periods.

The second time period modeled runs from the last week in January through February 21, 2012. This time period included the second significant protest at Bolotnaya Square on February 4, which again saw tens of thousands in attendance (Parfit, 2012). This date also saw the first major pro-government, or 'Anti-Orange', counter protest at Poklonnaya Hill in Moscow. The first three weeks of January are not modeled since no major protest events took place in those weeks. The level of online activity was also very low due to the New Year and Christmas holidays.

The third time period runs from February 22, 2012 through the end of March. This time period included a large pro-government rally at Luzhniki stadium in Moscow on February 23. The next anti-government action took place on February 26, when 30,000 protesters formed a human-chain along the Garden Ring in Moscow, a major street that circles the city (Schwirtz, 2012b). This time period also includes the March 4 presidential election. Two large anti-government protests took place in Moscow; one at Pushkin Square on March 5 and another on March 10 on Novy Arbat street. Although tens of thousands of participants attended these two protests, the excitement around the movement began to dissipate after Putin won the presidential election.

The month of April 2012 constitutes the fourth time period modeled and coincides with elections in the Russian region of Astrakhan, which many observers believed were falsified. Oleg Shein, a leftist politician in the region who was running in the local mayoral election, began a hunger strike and also helped organize smaller protest events. National protest leaders shifted their focus to Astrakhan and Shein during this month.

The final time period modeled is the month of May 2012. This month included the final major protest on May 6 at Bolotnaya Square. The government used the violence that broke out at that protest event as an excuse to crackdown on the broader movement. May therefore effectively marked the end of the election protests. Additional details on protest events are available in Chapter 3.

The network for each time period includes all those nodes that sent or received at least one link during the given period of interest. Each time period results in a directed, simplified network. The ERGM function in the statnet package was used to fit each model and requires that multiple edges in each network be collapsed to one (Handcock, Hunter, Butts, Goodreau, & Morris, 2003). This may cause some loss of explanatory power; although it is possible to add edge weight as a parameter to the ERGM models and this could be attempted in future research. I chose not to include an edge weight parameter in order to create more parsimonious models. Descriptive statistics for each protest period are included in Tables 5.1 and 5.2.

Table 5.1: Descriptive Statistics for Temporal Protest Networks

Network	Nodes	Edges	Degree	Density	Reciprocity	Transitivity	Triangles
December	1493	4584	6.14	0.0021	0.055	0.146	3017
February	1118	2042	3.65	0.0016	0.063	0.120	725
March	1043	2167	4.16	0.0020	0.086	0.125	1306
April	339	616	3.63	0.0054	0.214	0.055	326
May	254	259	2.04	0.0040	0.085	0.120	28

Table 5.2: MAN Dyad Census and Mutual Divided by Total Ties

Network	Mutual	Asymmetric	Null	Mutual/Total Ties
December	127	4330	1109321	0.028
February	64	1914	622425	0.032
March	93	1981	541329	0.043
April	66	484	56741	0.11
May	11	237	31883	0.043

Note. MAN = Mutual, Asymmetric, and Null.

Data collection for node attributes

Node attributes used in the ERG models include indegree, outdegree, story count, media type, and location. Here I describe how node attribute data was collected, while the specific ERGM terms for each model are discussed in the next section. Indegree and outdegree for each node was determined through use of the igraph package in R after simplifying the network by removing multiple edges and self loops, as required for use with the statnet software. All other structural parameters are determined automatically by the network and sna packages in R, which are part of the statnet software suite.

The media type was determined through a combination of automated regular expressions and manual coding of each Web site in the protest data.

Each node was assigned to one of the following media type categories: LiveJournal, blog, community, government, news, NGO, personal Web site, social networking site (SNS), general online tools, user generated content site, and other. These media types categories were created, with some modifications, from manual coding of blog outlinks in a large-scale study of Russian blogs and other research projects that draw on similar online protest data (Benkler et al., 2015; Etling et al., 2010). LiveJournal is treated as a distinct category separate from blogs and social networking sites since it is a unique hybrid platform that includes elements of both a social network and a blog platform (Etling et al., 2010). LiveJournal communities are separated from other LiveJournal pages and included in the communities category. The automated portion of the media type classification process identified LiveJournal sites, LiveJournal communities, and blogs from various popular Russian blog platforms through a series of simple regular expression that looked for relevant text patterns in URLs (e.g., URLs that matched the text pattern 'community.livejournal' were coded as LiveJournal communities). The author manually coded the remaining Web sites. The communities category consists of online forums and discussion groups, in addition to the LiveJournal communities already discussed. NGOs include Web sites of nongovernmental and nonprofit organizations, broadly defined, as well as political parties. However, sites connected to the dominant pro-government United Russia party are coded as government Web sites, along with other official government Web pages. The personal Web sites category consists of stand-alone personal Web sites that are not blogs, including individual politicians' Web sites. The tools category includes search engines and similar online 'tools' such as Yandex, Google, Yandex Money, and Scribd. Social networking sites include sites such as Vkontakte, Odnoklassniki, Facebook, and Twitter. User-generated sites are those where general Web users contribute the majority of the content, such as Russian equivalents of link-sharing sites like Reddit and citizen journalism sites such as Ridus. Any site that did not fit into one of the above categories was coded as 'other.' The media type categories are mutually exclusive.

Location data was obtained automatically for LiveJournal blogs through a Python script that automatically pulled city, region, and country data from the relevant LiveJournal profile pages. All non-LiveJournal sites were manually coded for location by the author. To maximize the number of sites with location data, and also to reduce the number of possible categories, a user's city was used to determine the region of each LiveJournal blog when a region was not listed in the journal's profile. Likewise, city and region data were used to identify the country where each site appears to be located when country data was not available. The region and country categories therefore had the most amount of data available and were used in the analysis for homophily and categorical indegree terms, while city was not. Overall, a region was identified for 77% of the sites in the data and a country for 90% of the sites.

Method

Exponential random graph models are part of a family of statistical models for social networks. ERGMs allow researchers to account for the presence or absence of ties in a network, and therefore a way to understand the micro-processes that account for macro structure in an observed network of interest. A key assumption of this method is that micro processes lead to macro network structure. An ERGM models a given network in terms of a set of theoretically relevant parameters, or terms, selected by the researcher that account for local tie-based structures. These parameters consist of processes that are exogenous and independent of network ties (such as node attributes, like location or media type) as well as endogenous or structural terms, where observations are dependent (such as reciprocity or transitivity). ERGMs therefore allow social scientists to understand how these endogenous and exogenous parameters interact to create and maintain network structure (2013).

Exponential random graph modeling is used in this chapter to determine whether the online Russian protest network demonstrates tendencies towards polycentric or hierarchical structure over time. The main hypothesis is that the Russian protest network will have a polycentric structure only at the beginning of the protests, and will move towards a less effective hierarchical structure as the movement loses popular support. A combination of structural terms, described in detail below, is used to determine if a network has polycentric or hierarchical tendencies. Specifically, a positive geometrically weighted edgewise shared partnership (GWESP) parameter estimate, negative geometrically weighted dyadwise shared partnership (GWDSP) estimate, negative geometrically weighted indegree (GWIDEGREE) estimate, and negative or small reciprocity (mutual) estimate indicates tendencies towards a polycentric network structure. Tendencies towards a core-periphery structure are indicated by positive parameter estimates for geometrically weighted indegree (GWIDEGREE) and reciprocity (mutual). A number of structural and node attribute terms are included as controls. Figure 5.2 shows ideal polycentric and hierarchical network structures.

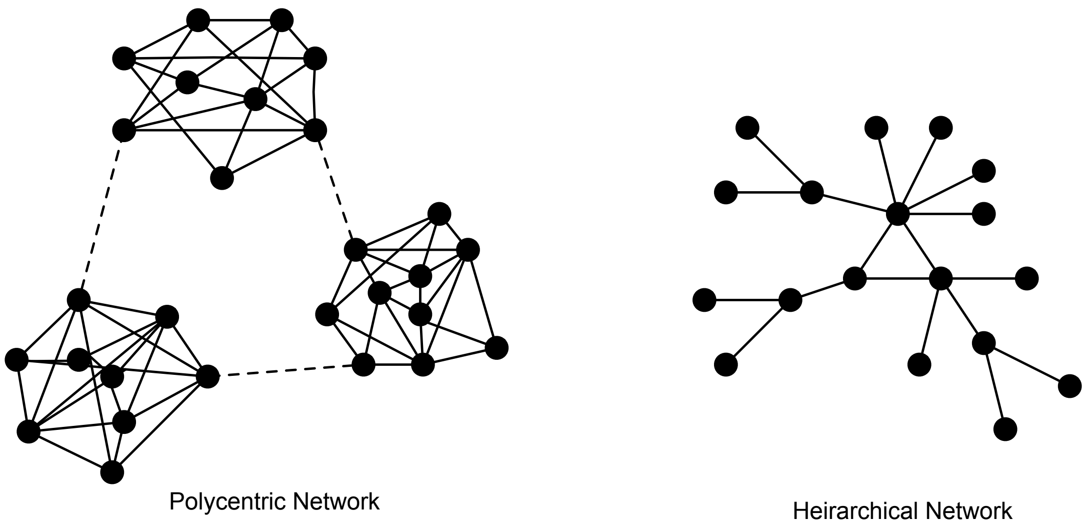


Figure 5.2: Ideal network structures. Dashed lines represent weak ties in the polycentric network. Adapted from Baldassari and Diani (2007).

A brief word on the choice of statistical model for social networks is necessary. A Stochastic Actor-Oriented Model (SOAM) model is a natural choice for modeling change over time in social networks (Snijders, van de Bunt, & Steglich, 2010). The Russian protest data was ultimately not appropriate for such a model since there is so much variation, or turn over of nodes, between each time period. This results in zero or near zero Jaccard index scores for networks between each time period, indicating that, for statistical purposes, these networks are essentially independent. Stochastic Actor-Oriented Models are not appropriate when there is this much change in network composition, and so that modeling approach was abandoned. The same assumption is true of temporal ERGMs. Instead, each time period of interest is modeled separately as a stand alone ERGM and then the results of all models are compared.

To explain the modeling process, the parameters chosen, and the decisions made during modeling, the April model is discussed in detail in the results section. First, though, the modeling process is described in broad strokes. The specific parameters used in each model are then described in detail.

Modeling overview

For each protest peak time period the following method was employed, following the general approaches recommended by Harris (2014), Goodreau, Handcock, Hunter, Butts, and Morris (2008), and Lusher et al. (2013). First, a null model is created with only the edges parameter, which models the basic propensity to create ties. The null model also is used to compare the fit of future models as it provides a Bayesian Information Criterion (BIC) statistic that can be used to judge the fit of future models as additional parameters are added. As in a standard logistic regression, a lower BIC statistic in larger models indicates that adding additional terms has increased the explanatory power of the model. A higher BIC indicates the additional terms have not improved the fit and so are removed from future models.

Next a model with the edges parameter plus all structural terms of interest is built. This is also referred to in the literature as an endogenous model. This type of model can show how important structural terms are to an eventual full model. On its own a structural model is rarely the best model. In practice, structural models are also sometimes degenerate without inclusion of exogenous terms, especially homophily terms.

An exogenous model is build next. This model is also referred to in the literature as a local, actor-relational, main effects, or node attribute model. This model is similar to a standard logistic regression, since it does not include any structural, or dependent, terms. This model includes the edges parameter from the null model, as well as continuous, categorical, and homophily terms for node attributes (Harris, 2014; Robins & Lusher, 2013). As in standard logistic regression model building, non-significant terms are removed from the local model before creating a full model (Harris, 2014).

Finally, a full model is created by combining all significant terms from the exogenous model together with the structural terms from the endogenous model. Non-significant terms are then removed to create the final model, as long as doing so does not negatively impact fit (Snijders, Pattison, Robins, & Handcock, 2006). Structural terms are generally retained, unless they lead to degeneracy in a model or otherwise have a large, negative impact on fit.

Description of parameters by model

This section describes the specific parameters used in the ERGM models and their connection to theory. Parameter names are highlighted in bold text below. Some time is spent discussing the geometrically weighted terms since they play a special role in helping prevent degeneracy in ERGMs and are rather complex compared to other terms.

Null model parameters

The ***edges parameter*** accounts for the basic tendency to create a tie in an ERGM and is typically a required control in every model. Once additional terms are added to the null model, the edges parameter is equivalent to the intercept in an OLS regression, and so it is no longer interpreted. The edges coefficient is often negative and significant. This term often has a parameter estimate very similar to if not exactly the same as the density of a graph. In real-world social networks, we should expect the number of actual ties to be less than the total number of possible ties (Robins & Lusher, 2013).

Structural or endogenous parameters

This section describes the structural parameters used in each model. I begin with all the terms recommended by Robins and Lusher (2013) for a positive affect (friendship) network. Structural terms include, first, the ***mutual parameter***, which is the statnet term that measures reciprocity, or the tendency for node pairs to form mutual links between each other. In a directed network, if node i links to node j, then j reciprocates this tie back to i. This is a very common feature of social networks and of human relationships (Blau, 1964). We should expect higher levels of reciprocity in an observed network than in a random graph (Lusher & Robins, 2013).

Network theorists propose that the proportion of reciprocated ties in a network can help to describe the amount of cohesion, trust, and social capital in a network. Some theorists also argue that reciprocal ties tend towards an equilibrium state, with tendencies towards either reciprocated or null dyads. Networks dominated by reciprocated or null dyads are often seen as more stable, balanced, or equal than those dominated by asymmetric dyadic relationships (Hanneman & Riddle, 2005). Theories of structural balance are especially relevant to interpretation of the mutual parameter. In those theories it is believed that people seek to find a state of balance with their relations by maximizing the pleasure and reducing the psychological tension in their relationships (Cartwright & Harary, 1966; Heider, 1958; Wasserman & Faust, 1994). It is also believed that reciprocity is influenced by homophily, or the propensity for ties between nodes that are similar.

In the Russian protest data, it seems reasonable to expect higher levels of reciprocity between nodes of similar media type, such as nodes that are both on LiveJournal or social networking sites, and lower reciprocity between bloggers and large traditional media outlets, where the ties will typically go in one direction. We can also likely expect that star nodes with very high indegree are less likely to reciprocate links to low degree nodes, especially if those star nodes are mainstream media sites. This is line with theorists who argue that structurally equivalent nodes and those sharing other similar attributes are believed to be more likely to reciprocate links than those that are different from one another (Lazarsfeld & Merton, 1954). Reciprocity is a common form of interaction online, following the dynamic that if I link to or follow you, you will link to or follow me. As such, reciprocity in an online communication network may be more likely than in an offline relationship, and so might say less about trust, cohesion, and social capital than in traditional social network studies where the ties under study might be determined by who lends money to whom, or who seeks advice from whom (Borgatti, Everett, & Johnson, 2013).

Transitivity is another common tendency in social networks (Simmel, 1908, 1950). In laymen's terms, transitivity accounts for the tendency for friends of friends to become friends. More formally, if i links to k, and j links to k, then i will also link to j. Transitivity is measured in statnet by the ***Geometrically Weighted Edgewise Shared Partners (GWESP) parameter***, a geometrically weighted term developed to prevent degeneracy in ERGMs. The GWESP parameter is discussed in greater detail below with the other geometric terms.

Transitivity has been shown to lead to clustering, since triangles tend to clump together in groups in networks. Transitivity is believed to come about because of three processes. One is propinquity, in which two people get to know each other because of time spent together with a third. Another cause is the cognitive process related to structural balance theory mentioned above, in which two people come to like each other because of their shared affection for or value in a common friend. Homophily can also contribute to transitivity, since the increased chances of ties between similar others also increases the chance that triads will be closed within homophilous groups (Cartwright & Harary, 1966; Goodreau, Kitts, & Morris, 2009; Heider, 1958; Wasserman & Faust, 1994). Recent research on the evolution of friendship networks finds a negative relationship between reciprocity and transitivity. When interpreted together, low reciprocity in a network and high transitivity might indicate a more cohesive network, while the opposite relationship between the two parameters might indicate a less cohesive, more locally focused network (Block, 2015). Since GWESP accounts for multiple shared partners, it is also a good indicator of the type of network structure that is conducive to risky collective action. This is because the willingness to participate in high-risk activism requires independent affirmation of protest information as well as reinforcement of the normative importance of participation from multiple sources (Centola & Macy, 2007).

Four geometrically weighted terms are added to the initial structural models. These terms were developed by Snijders et al. (2006) to help fit ERG models that include triangle terms, and in general to help fit ERGMs. Hunter and

Handcock (2006) modified these terms for the statnet software suite to deal with the common problem of degeneracy in ERGMs. Degeneracy occurs during model estimation when the simulated networks are either nearly empty (almost no ties) or nearly complete (nearly all nodes connected). Nearly empty or complete simulated distributions are not reasonable models for observed network data since these types of social networks rarely occur in the real world; therefore parameter estimates generated by software are likely to be poor reflections of the observed network under study (Harris, 2014; Koskinen & Snijders, 2013; Robins, Snijders, Wang, Handcock, & Pattison, 2007).

The four geometrically weighted terms used are ***Geometrically Weighted Edgewise Shared Partners (GWESP), Geometrically Weighted Dyadwise Shared Partners (GWDSP), Geometrically Weighted Indegree (GWIDEGREE), and Geometrically Weighted Outdegree (GWODEGREE)***. GWESP is equivalent to alternating k-triangles, GWDSP is equivalent to alternating 2-paths, and Geometrically Weighted Degree (the undirected version of GWIDEGREE and GWODEGREE) is equivalent to alternating k-stars as originally developed and defined by Snijders et al. (2006). Generally, the geometric weight in these terms anticipates exponential rather than normal distributions in network data and accounts for this skew. The weighting places declining emphasis on, for example, nodes with higher degree, since in an empirical network most nodes will have low degree while few will have high degree centrality. During modeling higher degree nodes are geometrically down-weighted, which dampens their impact and helps prevent the dramatic phase transition where simulated networks become very sparse or very dense and can lead to degeneracy (Hunter, 2007; Koskinen & Daraganova, 2013; Snijders et al., 2006).

As already discussed, GWESP accounts for transitivity or triadic closure in networks and the tendency for triangles to form together into clumps, or clusters, in a network (Robins et al., 2007). The geometric weight in this term expresses the expectation that higher order triangles, where nodes share many partners, are less likely than lower order triangles, where nodes share fewer partners (Wimmer & Lewis, 2010). So, this term might be thought of as accounting for structures in which two friends have many friends in common. Again, the existence of multiple shared partners in a network is indicative of densely connected clusters, and also indicates structures beneficial to high-risk activism (Centola & Macy, 2007; Harris, 2014).

When included without GWESP in a model, GWDSP measures the tendency for two actors to have one or more partners in common, regardless of whether the two actors are connected themselves. GWDSP can be thought of as a measure of pre-transitivity, since only one additional edge is required to create one or more triangles in the network. GWDSP indicates whether shared partners tend to clump together in the same region of the network, regardless of whether they are connected or not (Harris, 2014). A positive GWDSP is also indicative of many two paths in a network. A high GWDSP statistic can also be caused by many two paths cycling through a small number of high degree nodes, and therefore may indicate tendencies towards a core-periphery structure (Snijders et al., 2006).

GWESP and GWDSP are often both included in models and can be interpreted together, and also control for each other. Both terms are included in each Russian protest network model. When both terms are included in a model, it is common for GWESP to have a positive parameter estimate, for GWDSP to be negative, and for both terms to be significant. These joint parameter estimates (again, a positive GWESP and negative GWDSP) can be interpreted as a tendency towards a global network structure with several connected small regions of overlapping triangles, or clusters (Harris, 2014; Robins et al., 2007). A larger positive GWESP statistic indicates overlapping triangles in denser parts of the graph (Robins et al., 2007; Snijders et al., 2006). A positive GWESP statistic can also be seen as representing the closing of structural holes in a network (Robins et al., 2007). Figure 5.3 provides images of GWESP and GWDSP structures for one, two, and three shared partners in an undirected network.

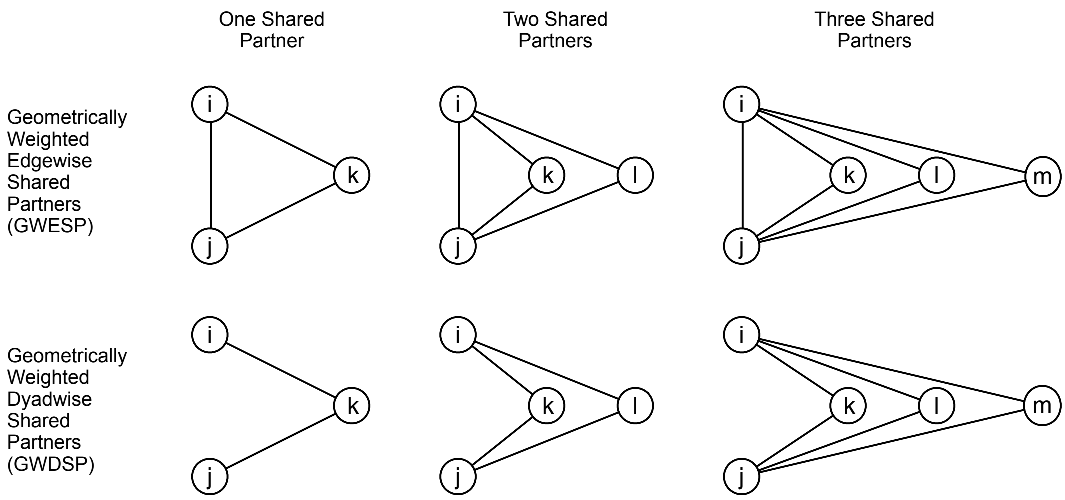


Figure 5.3: Edgewise shared partners (top) and dyadwise shared partners (bottom) for the dyad ij, with one, two, and three shared partners. Adapted from Harris (2014).

For undirected networks, a ***geometrically weighted degree (GWD)*** statistic is typically used in combination with GWESP and GWDSP. This term was developed to account for the power law, or declining degree, distribution common in observed networks and which is especially prevalent in online networks. This parameter models the spread of the degree distribution (Robins

& Lusher, 2013; Robins, Pattison, & Wang, 2009). In statnet, the GWD statistic "Multiplies the frequency for each value of degree by a weighting parameter and sums the values" (Harris, 2014, p. 27). The resulting parameter estimate for GWD is higher if there is a dispersion in the degree distribution, with some nodes having high degree and most nodes a low degree. A negative parameter indicates that high degree nodes are improbable and that the variance in degree among nodes is small (Robins et al., 2009). A positive GWD parameter can also indicate a tendency for a core-periphery network structure (Robins et al., 2007; Snijders et al., 2006). This is because a positive parameter estimate suggests a preference for connections between a larger number of low degree nodes and a smaller number of high degree nodes (Robins et al., 2009). Since the observed Russian protest network is a directed network, the relevant ***geometrically weighted indegree (GWIDEGREE) and geometrically weighted outdegree (GWODEGREE)*** terms are included in each model. A positive and significant parameter estimate for GWIDEGREE indicates a tendency towards a core-periphery network centered around a few nodes of high indegree with most other nodes having a low indegree, while a positive and significant estimate for GWODEGREE indicates a tendency towards a network centered around nodes with high outdegree (Lubell, Robins, & Wang, 2014; Lusher et al., 2013; Snijders et al., 2006). GWODEGREE is used as a control in the Russian protest models. The best indication of a hierarchical or core-periphery network using just one structural parameter is likely a positive, large, and significant GWIDEGREE estimate.

Robins et al. (2007) have shown that GWD can be interpreted together with GWESP. They show that a positive and significant GWESP (alternating k-triangle) combined with a negative and significant GWD (k-star) estimate indicates a global structure with several, usually connected, smaller regions of overlapping triangles instead of a single core of densely connected nodes. Further, in simulations with a fixed positive and significant GWESP parameter estimate, a significant and increasingly negative GWD estimate indicates a move away from centralization and towards more segmentation the further the GWD parameter estimate moves away from zero. Network closure, as indicated by the positive GWESP estimate, is occurring, but the negative GWD estimate indicates that those closure processes are distributed across a number of smaller, denser regions in the network instead of in one central region (Robins & Lusher, 2013). Therefore, a positive GWESP parameter estimate in combination with a negative GWD estimate (or GWIDEGREE in the case of a directed network) suggests a tendency towards a polycentric network structure. This is not a perfect indication of a polycentric network, since the clique-like regions of the network are not definitely connected by weak ties in the graph, but it appears to be the best indication of a polycentric network based on just two structural parameters available.

The value of the statistic for each geometrically weighted term is determined by two things: First, and primarily, the proportion of nodes with high values for that statistic, and, second, the value of the alpha weighting parameter. A fixed alpha of 0.25 was used in all models, which is suggested as an alpha value that achieves reasonable estimates in ERGMs. This approach resulted in a better fit and more stable models than using a Curved Exponential Family (CEF) model, where the alpha for each geometric term is estimated during the modeling process (Harris, 2014; Hunter & Handcock, 2006; Koskinen & Daraganova, 2013).

Returning to the remaining structural terms, ***cyclic triple***, or ***ctriple*** in statnet, measures the tendency for cyclic triples in the network, which is sometimes interpreted as a tendency towards generalized exchange or egalitarianism in the network (Robins & Lusher, 2013). This term adds one statistic to the model equal to the number of cyclic triples in the network, defined as a set of edges of the form (i,j), (j,k), (k,i) (M. Morris, Handcock, & Hunter, 2008). This term is included as a control in the models.

The ***mixed two star*** **(m2star)** parameter measures the extent to which actors who send ties also receive them, and controls for the tendency for in and out degree to be correlated. In practice this term is often negative (Robins & Lusher, 2013). A mixed two star is also known as a two path. In statnet "this term adds one statistic to the model, equal to the number of mixed 2-stars in the network, where a mixed 2-star is a pair of distinct edges (i,j), (j,k)" (M. Morris et al., 2008, p. 8). The theoretical focus of interest with the mixed two star term is on the midpoint j, instead of on the endpoints i and k (M. Morris et al., 2008). This term is included as a control in the models.

Exogenous model parameters

In addition to the edges parameter from the null model, the following node attribute terms are included in exogenous models for each protest peak time period.

The ***story count parameter*** is a continuous variable that measures if nodes that publish many stories are also more likely to attract links than chance. Theoretically, one can hypothesize that nodes that publish many stories, such as newspapers or very popular blogs, are more likely to attract links than by chance. The story count parameter is simply a count of the number of stories that a given node published in the specified time period of interest. This data was drawn from a query of the database containing all the Russian protest data.

Homophily, or the tendency for individuals with similar characteristics to form ties to one another, structures all types of human networks. This general principle of tie formation results in homogeneous personal networks, and can limit people's access to information and the attitudes they form about political or social issues (McPherson et al., 2001). Strong ties also tend to be homophilous ties. Individuals tend to have strong ties with those who are similar to them and with those who live close to them. In extreme cases, homophily can lead to social and political polarization if there are no weak ties serving as bridges between clusters of individuals with homophilous strong ties (Putnam, 2000; Putnam & Goss, 2002). As noted above, homophily can also contribute to reciprocity and transitivity. Homophily is represented by inclusion of ***nodematch parameters*** in statnet, and, again, identifies the tendency for nodes with similar characteristics to be linked to each other more frequently than by random chance (Harris, 2014). Homophily terms for media type, country, and region are included in each model.

Categorical variables for nodes are represented by ***nodefactor parameters*** in statnet. Since the Russian protest networks are directed and indegree is typically of more theoretical interest than simply degree, categorical indegree (denoted by ***nodeifactor*** in statnet) terms are included in each model. For a selected node attribute of interest, the nodeifactor statistic in statnet gives the number of times a node with that attribute appears as the terminal node of a directed tie (Handcock et al., 2003). Nodeifactor terms are included for the most popular media types and regions for each network, with the exact cut off for inclusion of sub-categories determined by indegree frequencies in each peak protest network. For example, since there are far fewer media types than there are regions or countries in the data, all but the least frequently occurring media types are initially included as nodeifactor terms (typically 'other' is excluded as a minimal starting point since at least one term is required to be used as a base of comparison), while for regions and countries only the top 5-8 regions and countries are included for each model.

Returning to the main hypothesis tested in this chapter, it is expected that the early protest periods will have tendencies toward a polycentric network structure, while later protest time periods, when the movement begins to lose popular support, will have greater tendencies towards a hierarchical structure. As noted at the beginning of this chapter and in detail in the theoretical framework of this thesis (Chapter 2), polycentric networks are conducive to high-risk collective action as they are more robust than hierarchical networks, more resistant to repression, can penetrate and recruit from different segments of society, and encourage social innovation and problem solving (Baldassari & Diani, 2007; Gerlach, 2001).

A combination of terms is used to assess whether networks are polycentric or have a core-periphery structure. The following mix of structural parameter estimates indicate a tendency towards a polycentric network structure:

- GWESP: Positive, large, and significant;
- GWDSP: Negative and significant;
- GWIDEGREE: Negative and significant;
- Mutual: Negative and significant, or positive and small.

And the following combination of parameter estimates suggest tendencies towards a global core-periphery network structure:

- GWESP: Small or insignificant;
- GWDSP: Positive and significant;
- GWIDEGREE: Positive, large, and significant;
- Mutual: Positive, large, and significant.

Figure 5.4 provides a summary of all the parameters included initially in each model. A figure for each term is also provided, as well as a description of the network tendency that each parameter measures. Now that each parameter included in the models has been described, the next section of this chapter provides results from the models. Given that these are complex models unlikely to be familiar to most social scientists, the April model building process and results are first discussed in depth before moving to summary findings for all models.

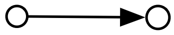
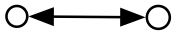
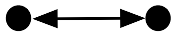
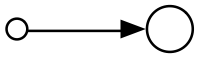
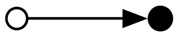
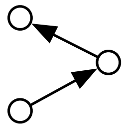
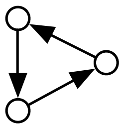
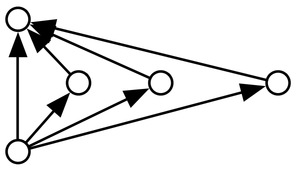
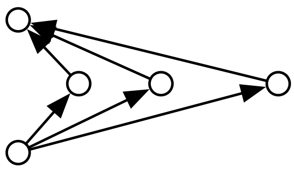
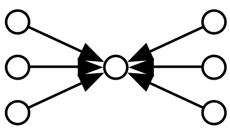
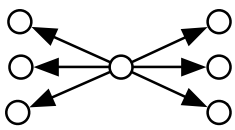
Parameter	**Figure**	**Tendency**
Edges		Basic tendency to create an edge
Reciprocity (Mutual)		Tendency for node pairs to form mutual links
Homophily (Nodematch)		Tendency for links between nodes with similar characteristics
Story Count		Tendency for nodes that write many stories to attract links
Categorical (Nodeifactor)		Tendency for nodes with a given categorical attribute to attract links
Mixed Two Star		Measures extent to which nodes that send links also receive them
Cyclic Triple		Tendency towards generalized exchange
Transitivity (GWESP)		Tendency towards path closure and clustering in the network
GWDSP		Tendency towards two paths in the network
Indegree (GWIDEGREE)		Tendency towards core-periphery network centered around high indegree nodes
Outdegree (GWODEGREE)		Tendency towards core-periphery network centered around high outdegree nodes

Figure 5.4: Model parameters, figures, and tendency measured.

Results

Detailed description of model building

To illustrate the model building process, the April time period is explained in detail. This model is chosen for detailed description because it is well fit, has unique results, and covers an interesting time period in the protests, as attention moved to falsification of the Astrakhan regional elections. Summary results are then reported for each remaining protest period. This is followed by a discussion of model fit for all the models using goodness of fit plots found in Appendix A. Looking forward, an assessment of trends that emerge from an analysis of all the final full models is included in the discussion and conclusion section of this chapter.

Descriptive statistics

First, some descriptive network statistics for the April time period. The April network consists of 339 nodes and 616 edges, making it one of the smaller networks in the Russian protest data. The density of the network is 0.0054, indicating that there are fewer edges in this network than possible. Reciprocity is 0.214 and transitivity is 0.055. As Table 5.1 in the data section shows, this network has the highest density and reciprocity of any time period, but the lowest transitivity. There are 326 triangles, 30 cyclic triples, and 5421 mixed two stars. A Holland and Leinhardt MAN dyad census shows that the network has 66 mutual dyads (a <-> b), 484 asymmetric dyads (a <- b or a ->b), and 56741 null dyads (a <-!-> b) (Holland & Leinhardt, 1970; Wasserman & Faust, 1994).

Figure 5.5 shows a histogram of the indegree distribution for the April network. This plot suggests a power law distribution common in data drawn from the Internet, with most nodes having low indegree centrality and relatively few nodes having high indegree centrality. Observing this distribution on a log-log plot (insert) we see that this is not a pure power law distribution, which would be indicated by a straight line on the plot. However, it is still highly skewed. The outdegree distribution (Figure 5.6) shows the same general power law distribution in the histogram, but again the log-log plot indicates this is not a pure power law distribution. However, the slope of the line for the outdegree log-log density plot is more uniform than the indegree log-log plot. All plots show that there are some outliers far out in the tail of each distribution, indicating there are some nodes with very high indegree and outdegree in the network.

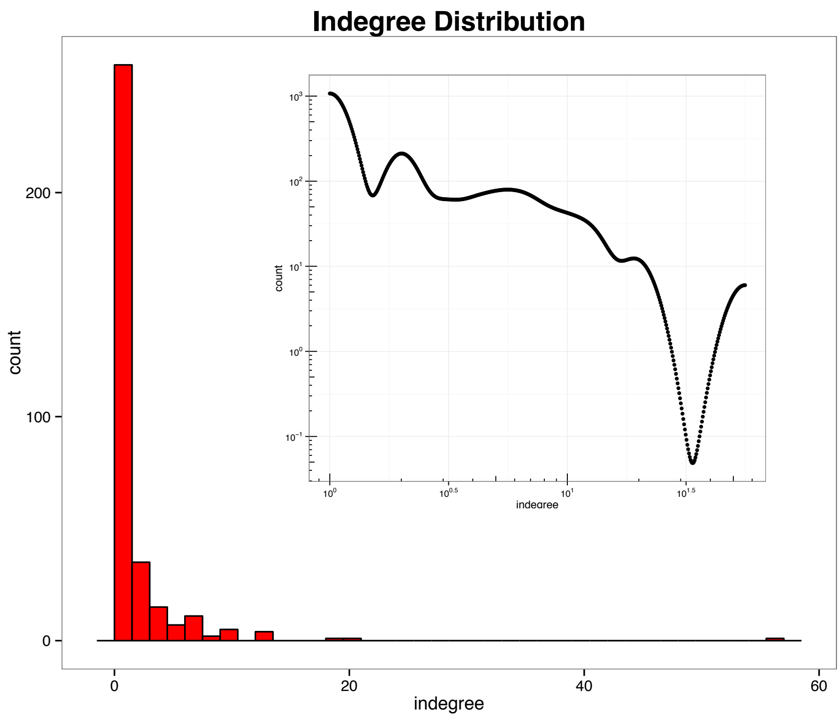


Figure 5.5: Indegree distribution of the April network (insert, log-log density plot).

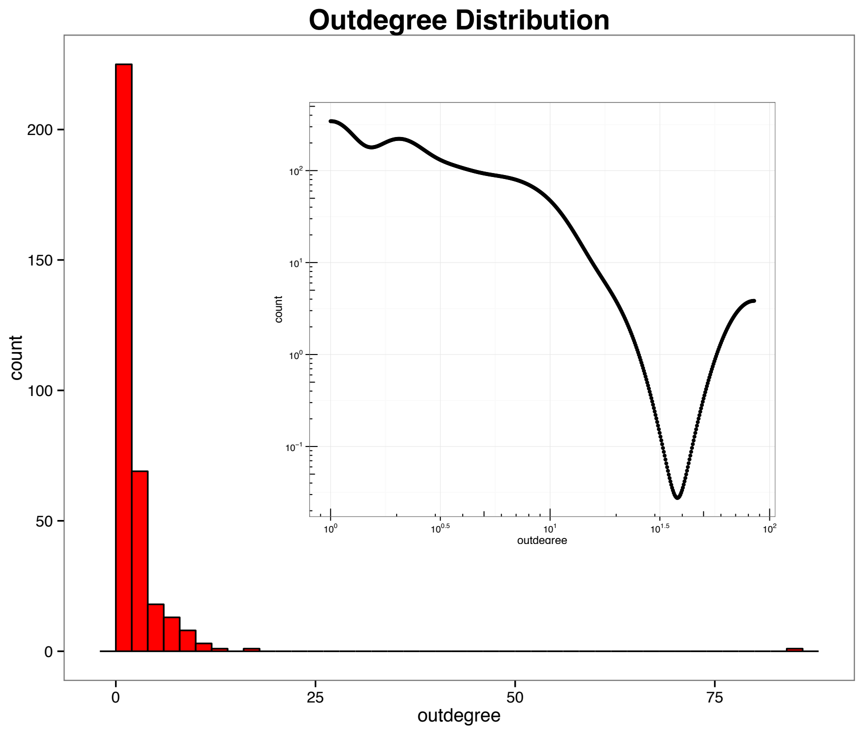


Figure 5.6: Outdegree distribution of the April network (insert, log-log density plot).

A mixing matrix is useful for showing homophily by media type (Table 5.3). Reading along the diagonal, this table shows that the highest number of matches by media type is among LiveJournal (293) nodes, but there are also matches for news (12), community (2), NGO (1), blogs (1), and user-generated content (1) categories. There must be at least one link between nodes with a common attribute to be included as part of a homophily term in an ERGM, so commercial, other, personal, and social networking sites are not included in the April model's media type homophily term.

The mixing matrix for region is too large to include here. In terms of links among nodes from the same region, there are matches for Astrakhan (20), Kurgan (1), Moscow (95), St. Petersburg (1), Ulyanovsk (1), and Volgograd (2) in the April data. The number of matches for Astrakhan, a relatively small region compared to Moscow and St. Petersburg, is indicative of the importance of the Astrakhan election and related protests during this time period.

Table 5.3: April Media Type Mixing Matrix

	TO										
FROM	Blog	Commercial	Community	LJ	News	NGO	Other	Personal	SNS	User	Total
Blog	1	0	1	5	12	2	0	0	0	1	22
Commercial	0	0	0	0	0	0	0	0	0	0	0
Community	0	0	2	50	19	7	0	0	4	0	82
LJ	8	2	10	293	128	17	8	1	4	6	477
News	0	0	0	8	12	1	1	1	2	1	26
NGO	2	0	0	0	3	1	0	0	0	0	6
Other	0	0	0	0	0	0	0	0	0	0	0
Personal	0	0	0	0	0	0	0	0	0	0	0
SNS	0	0	0	0	0	0	0	0	0	0	0
User	0	0	0	2	0	0	0	0	0	1	3
Total	11	2	13	358	174	28	9	2	10	9	616

Note. Shaded cells along the diagonal indicate the number of links between nodes of the same media type. LJ = LiveJournal. NGO = Non-Governmental Organization. SNS = Social Networking Site.

April null model

The null model includes only the edge parameter and captures the basic tendency for the creation of an edge in the network. An exponential transformation of the edges coefficient is made to convert each parameter estimate to an odds ratio (OR), which is used in all tables in this chapter to aid in interpretation of the results (Harris, 2014; Snijders et al., 2006). Generally, an odds ratio above 1 indicates positive odds for that parameter, a value of 1 indicates no association, and below 1 indicates negative odds. A confidence interval (CI) at the 95% confidence level is also reported, as a well as a p-value. If a confidence interval includes 1 (no association), the result is not statistically significant. The April null model results in Table 5.4 show the edges parameter has an odds ratio value of 0.005405. The BIC value, which is used as a baseline for comparison to future models, is 7679.

Table 5.4: April Null Model

		95% CI			
Parameter	OR	Lower	Upper	p-value	
Edges	0.005405	0.004994	0.006	< 0.0001	***

Note. BIC = 7679. OR = Odds Ratio. CI = Confidence Interval.
*Significance codes: *** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1*

April structural or endogenous model

Table 5.5 shows the results of the structural model. The results of this model show a very high, positive odds ratio of 35.05 for the reciprocity (mutual) term. This parameter is highly statistically significant. Both the GWESP and the GWDSP terms are significant and positive, with odds ratios of 3.14 and 1.16, respectively. The positive GWESP term indicates a tendency towards triadic closure and clusters of triangles in the network. The positive GWDSP term indicates a tendency for two paths (or open triangles) in the network. Neither the geometrically weighted indegree nor the geometrically weighted outdegree terms are significant, but they are retained in future models as structural controls. The cyclic triple term has a negative odds ratio of 0.23, indicating a tendency against generalized exchange and egalitarianism in the network. The mixed two star term is not significant at the 0.05 alpha level. The BIC value from the structural model is 6762, which is lower than the null model. Adding endogenous terms has improved the model.

Table 5.5: April Structural Model

		95% CI			
Parameter	OR	Lower	Upper	p-value	
Edges	0.00	0.00	0.01	< 0.0001	***
Mutual	35.05	23.16	53.02	< 0.0001	***
M2star	0.89	0.77	1.01	0.08	†
Ctriple	0.23	0.10	0.53	0.00	***
GWESP	3.14	1.54	6.41	0.00	**
GWDSP	1.16	1.02	1.33	0.03	*
GWIDEGREE	0.60	0.01	26.84	0.79	
GWODEGREE	0.29	0.01	5.74	0.41	

Note. BIC = 6762. OR = Odds Ratio. CI = Confidence Interval.
*Significance codes: *** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1*

April exogenous model

Next an exogenous model is built by adding node attribute terms to the null model. All the possible terms for an exogenous or local model discussed above were added to the null model. For the nodeifactor terms, only the terms with the most inlinks are added to the model, since many categories, especially regions, have zero or just one inlink per category.

The results of the exogenous model (Table 5.6) show that there is a tendency for nodes with a high story count to attract links. The homophily terms for media type and region both have positive odds ratios, but only homophily by media type is significant. However, since the homophily by region term is close to significant at the 0.05 alpha level, I retain this term for future models since adding structural terms or making adjustments to the MCMC algorithm can change the significance of terms. All of the categorical, or nodeifactor, terms for media type are significant, but only social networking sites and news sites have positive odds ratios. Social networking sites are 9.5 times more likely to receive an inlink, the second largest effect in the model, and news sites are 1.85 times more likely than chance to receive an inlink, controlling for other terms in the model. Finally, the categorical indegree terms for Moscow and Smolensk are significant and have positive odds ratios. Nodes from Smolensk are 10.19 times more likely to receive an inlink than chance, which is the largest effect in the model, while those from Moscow are 1.28 times more likely to receive an inlink. It is surprising that Astrakhan region is not significant since this time period is focused on protests connected to the Astrakhan election, but when controlling for other parameters, this term is not significant.

Table 5.6: April Exogenous Model

Parameter	OR	95% CI Lower	95% CI Upper	p-value	
Edges	0.00	0.00	0.01	< 0.0001	***
Story count	1.01	1.01	1.01	< 0.0001	***
Homophily					
Media type	1.28	1.03	1.60	0.03	*
Region	1.22	0.97	1.54	0.09	†
Categorical Indegree (Media Type)					
Blog	0.42	0.21	0.84	0.01	*
Community	0.45	0.24	0.84	0.01	*
LiveJournal	0.48	0.33	0.70	0.00	***
News	1.85	1.29	2.65	0.00	***
SNS	9.50	3.68	24.53	< 0.0001	***
User-generated content	1.79	0.85	3.77	0.13	
Categorical Indegree (Region)					
Astrakhan	1.31	0.95	1.82	0.10	
Moscow	1.28	1.02	1.59	0.03	*
Petersburg	0.68	0.36	1.29	0.24	
Smolensk	10.19	5.08	20.46	< 0.0001	***

Note. BIC = 157910. OR = Odds Ratio. CI = Confidence Interval.
*Significance codes: *** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1*

The BIC value for the exogenous model is 157910, which is much larger than the null model and the structural model. The addition of terms representing node attributes to the edges parameter has not improved model fit compared to the null and structural models.

Full model

Next, a full model is created by combining the exogenous model with the structural model. At this point, node attribute terms that lose their significance are removed if doing so does not negatively impact fit. MCMC controls are also increased to improve fit and prevent degeneracy. Since geometrically weighted indegree is theoretically of interest and an important control in the model, it is retained in the final model. It is generally acceptable to include non-significant terms in ERGMs if they aid fit (Robins & Lusher, 2013). Since homophily terms generally tend to improve fit, the homophily term for region is retained, even though it is not significant. The final model represents the best fit of the full model without removing the terms of greatest theoretical interest.

In the final April model (Table 5.7), reciprocity has the largest odds ratio value of any term in the model. Controlling for other terms in the model, if a link is extended to a node, it is very likely to be reciprocated (the odds increase by 39.8 times). The GWESP term odds ratio of 4.3 indicates that there is evidence of triadic closure and clustering in the network, although this term is fairly small relative to the mutual parameter and several other terms in the model. The GWDSP parameter has a positive odds ratio of 1.7 and is statistically significant, which is atypical of most ERGM models that include both GWESP and GWDSP. This indicates more two paths in the networks than chance and the clustering of those two paths together in the network. It may also indicate a number of two paths cycling through high degree nodes, so may also indicate a core-periphery network structure (Snijders et al., 2006). Since GWESP is controlled for, the positive GWDSP parameter in this model can also be thought of as an indication of structural imbalance, since it represents situations in which nodes i and j are not connected, even though both those nodes link to one or more common sites (Wimmer & Lewis, 2010). The GWIDEGREE term is positive, indicating tendencies towards a core-periphery network, although the term is not significant. The GWODEGREE parameter is negative and significant; the network is not centered on high outdegree nodes and there is not a great amount of variance in the outdegree distribution. The story count parameter odds ratio is 1.01, indicating that nodes that publish many stories are more likely to be linked to, controlling for other terms in the model. The only homophily term that remains significant in the final model is for homophily by media type. Nodes that are of the same media type are 1.34 times more likely to link to each other than chance would predict. Among the nodeifactor terms social networking sites have a reasonably large odds ratio, being 7.82 times more likely than chance to receive a link. Astrakhan, Moscow, and Smolensk have positive and significant odds ratios, with Smolensk having the largest effect of the regional categorical indegree terms, and the second largest effect overall in the model. In summary, this model has a greater tendency towards a core-periphery structure than a polycentric one based on a very large mutual term, relatively small GWESP, positive GWDSP, and positive GWIDEGREE (although this last term is not significant).

Table 5.7: Final April Model

Parameter	OR	95% CI Lower	95% CI Upper	p-value	
Edges	0.01	0.00	0.01	< 0.0001	***
Mutual	39.82	26.23	60.44	< 0.0001	***
M2star	0.56	0.47	0.68	< 0.0001	***
Ctriple	0.25	0.14	0.45	< 0.0001	***
GWESP	4.31	3.52	5.27	< 0.0001	***
GWDSP	1.70	1.40	2.06	< 0.0001	***
GWIDEGREE	1.38	0.93	2.06	0.11242	
GWODEGREE	0.23	0.16	0.31	< 0.0001	***
Story count	1.01	1.01	1.01	< 0.0001	***
Homophily					
Media type	1.29	1.08	1.55	0.00554	**
Region	1.13	0.94	1.38	0.20142	
Categorical Indegree (Media Type)					
Blog	0.40	0.20	0.79	0.00875	**
Community	0.43	0.23	0.79	0.00689	**
LiveJournal	0.44	0.31	0.62	< 0.0001	***
News	1.65	1.19	2.29	0.00293	**
SNS	7.82	3.09	19.83	< 0.0001	***
Categorical Indegree (Region)					
Astrakhan	2.28	1.68	3.08	< 0.0001	***
Moscow	1.27	1.03	1.57	0.02562	*
Petersburg	0.68	0.36	1.27	0.22762	
Smolensk	9.04	4.47	18.29	< 0.0001	***

Note. BIC = 6222. OR = Odds Ratio. CI = Confidence Interval.
*Significance codes: *** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1*

Model fit and MCMC diagnostics

Two methods are used to assess model fit and to aid in selection of the best model for each time period. First, as already described and reported, the BIC statistic is used to determine if adding terms to a model improves fit. As in a standard logistic regression, a lower BIC indicates a better fit. For the April model, the final model BIC value is 6222, which is lower than both the endogenous and exogenous models. Adding structural terms together with significant node attribute terms results in the best fit according to the BIC statistic.

Model fit can also be assessed by simulating 100 networks and comparing simulated networks statistics with those from the observed network, usually with terms that are not already included in the model (Goodreau et al., 2008). If these repeated simulations are able to reproduce structures in the observed graph that are not explicitly modeled, we have greater confidence that the mechanisms being modeled are similar to those that generated the empirically observed network (Wimmer & Lewis, 2010). The goodness of fit procedure in statnet generates a table of frequencies comparing the 100 simulated network statistics to the observed network, as well as goodness of fit plots that compare the log-odds for each term in the observed network against the log-odds for the term in the simulated networks. Table 5.8 shows the frequencies and associated p-values for triad census. This table lists the triad type by its commonly know identification code (see Appendix B for a description of the 16 possible triad types and codes), the observed value, the minimum, mean, and maximum value for the simulated networks, and the MC p-value. The MC p-value represents the proportion of simulated values of the statistic that are at least as extreme as the observed value (Harris, 2014). Therefore, large p-values (above 0.05) indicate that the statistic being assessed in the observed network is similar to values in the simulated networks (in other words, the values are significantly similar, indicating that the simulated networks are doing a good job of capturing the observed network statistic). The majority of the triads in the April model have p-values above 0.05 (those that are *not* are highlighted in grey in the table) (Harris, 2014).

Table 5.8: Goodness-of-Fit Statistics for Triad Census (April Model)

Triad Type	Observed	Minimum	Mean	Max	MC p-value
003	6257005	6238505	6258687.02	6280093	0.86
012	154619	130048	151921.14	170777	0.70
102	17576	12606	17999.4	22072	0.84
021D	1366	925	2025.69	5850	0.38
021U	1055	320	517.35	928	0.00
021C	548	354	980.82	1474	0.04
111D	350	86	544.34	1102	0.36
111U	1786	1390	2147.99	3221	0.26
030T	88	16	39.26	69	0.00
030C	0	0	0.35	2	1.00
201	1207	264	727.62	1075	0.00
120D	2	2	13.67	30	0.02
120U	60	31	59.28	92	0.92
120C	3	0	2.36	8	0.88
210	21	7	18.08	27	0.58
300	3	0	4.63	9	0.70

Note. Shaded MC p-values indicate significantly different simulated values compared to observed frequencies. MC = Monte Carlo.

The goodness of fit plots, shown in Figure 5.7, present similar information visually, and are typically used to assess and report goodness of fit for ERGMs. In the plots, the black line represents the observed network, the box plots represent the range of a statistic's value in the simulated networks, and the grey lines represent the 95% confidence interval for the simulated network statistic. When the black line falls within the grey lines, the simulated networks are capturing the parameter of interest from the observed network well. Triad census and minimum geodesic distance are used to assess the fit of the April model since those two terms are the only parameters available in the statnet GOF routine that are not already included in the model. The GOF plot for triad census shows that the simulated networks are capturing the triad types in the observed network well, while geodesic distance is not as well captured. Since there is no research question of interest related to minimum geodesic distance, the fact that the simulated networks are not capturing the observed network as well is not of great concern. Geodesic distance can also be very hard to fit in ERGMs. Also, just as no standard OLS regression can be expected to account for 100% of the variance in a model, no ERGM can be expected to account for all features of a network (Robins & Lusher, 2013). Based on the combination of BIC and goodness of fit plots and statistics, I argue the final April model is well fit, although it does not capture geodesic distance as well as it could (Harris, 2014).

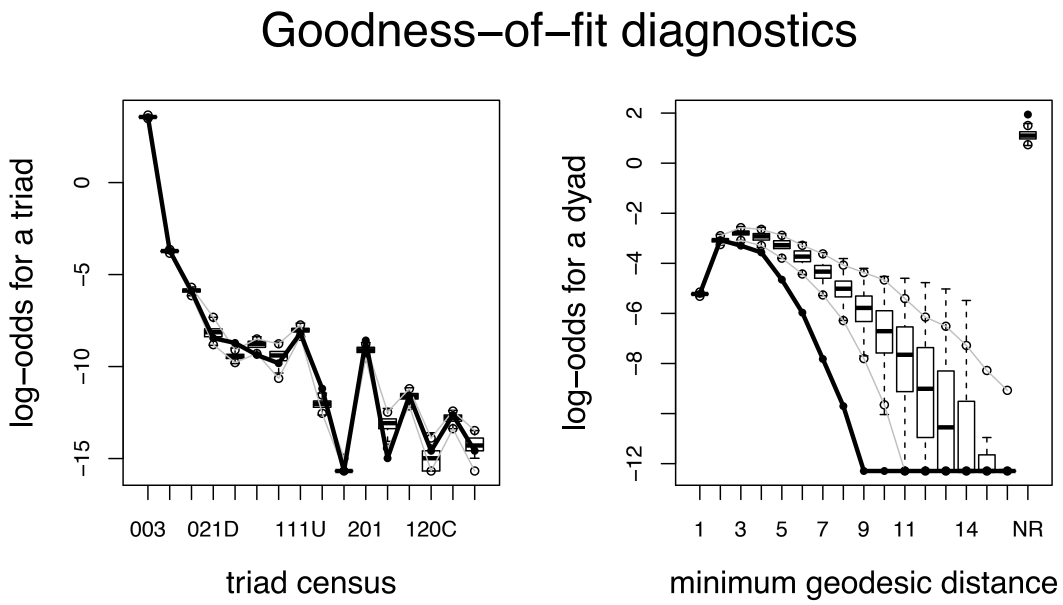


Figure 5.7: Goodness-of-fit diagnostic plot for final April model

MCMC diagnostics

In addition to the two methods available for investigating fit discussed above, MCMC diagnostics and plots are used to determine if the estimation algorithm has converged or if there are degeneracy problems. First, increases in the log-likelihood during estimation indicate how far the iterations are from starting values. Large improvements in log-likelihood indicate that starting values are off and that the model may be degenerate. Generally, increases in log-likelihood should be small and decrease at each iteration, which was the case during estimation of the April model. Graphic diagnostics within statnet are also used to show if the model has converged or if degeneracy is an issue. The built-in MCMC diagnostics routine in statnet generates a set of two plots for each term in the model based on the final iteration of the algorithm. In Figure 5.8, plots are shown for the first three parameters in the April model – edges, mutual, and mixed two star. The left plot is a trace plot that shows the MCMC chain as a time series based on the values the statistic took over time in the chain. The trace plots should show that the trajectory of the chain was consistent over time, which each of the April plots indicate. The right hand density plot is a smoothed histogram of the same process, showing the distribution of the values of the statistic in the chain. Parallel processing is used to speed up the estimation and each colored line represents one of the eight MCMC chains. Each tells a similar story. The right hand density plots should appear roughly as a normal distribution varying stochastically around a mean of zero, which represents the estimate for that term in the model. For the April data, the MCMC controls were modified to increase the burn-in (the amount of time the chain runs before starting values are used for the parameter) from 20,000 to 50,000, and the sample size was increased from 10,000 to 70,000. While the default MCMC control settings led to generally good plots for the April model, tweaking the algorithm as described above improved the MCMC diagnostic plots and statistics. The three plots in Figure 5.8 indicate a stable model, as does each plot for the remaining terms in the model (Harris, 2014).

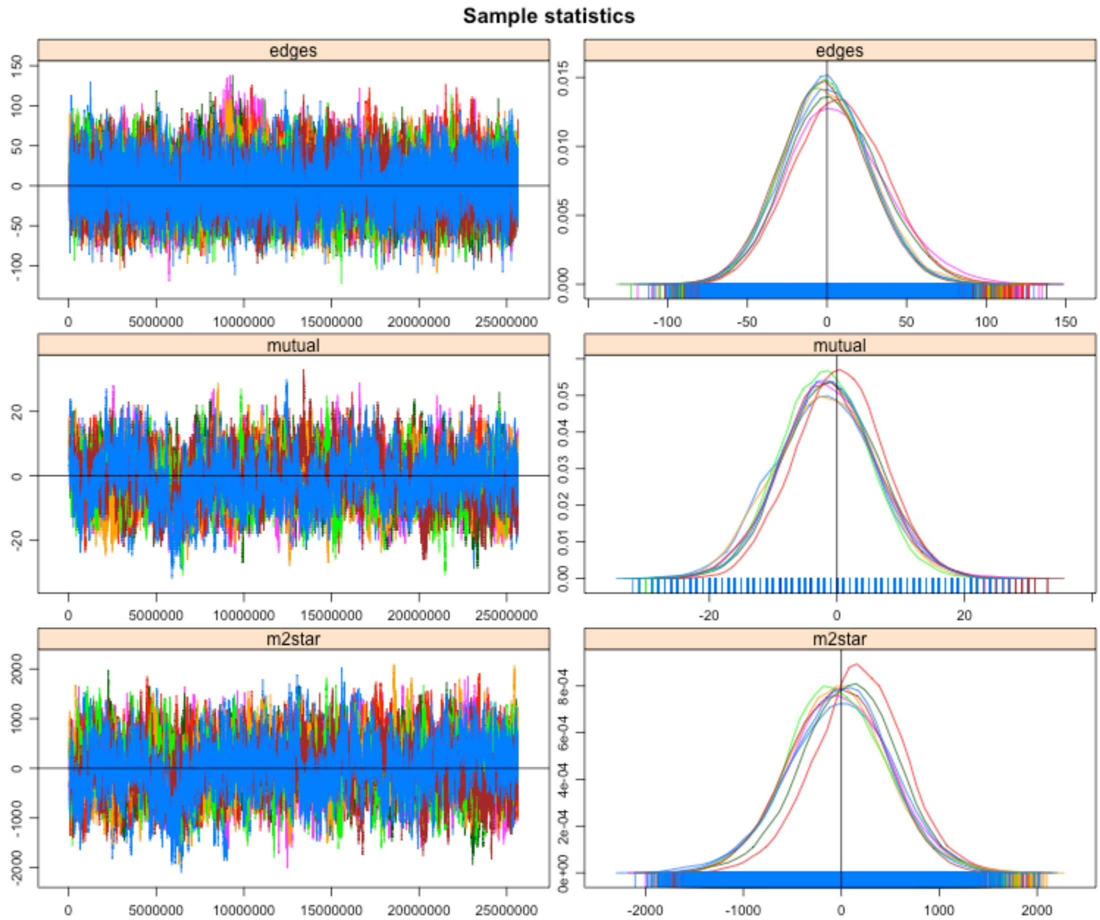


Figure 5.8: MCMC diagnostic plots for edges, mutual, and mixed two star parameters.

Summary results for remaining final models

In this section, the summary results for the remaining December, February, March, and May models are described and changes over time highlighted. Summary results for all final models are shown in Table 5.9.

December

In the final December model, the GWESP parameter has the largest odds ratio in the model, indicating that there is a very strong tendency towards triadic closure and clustering in the network. The strong and positive GWESP parameter also indicates the presence of multiple shared partners. Again, this type of structure is conducive to protest in risky contexts since multiple individuals can reinforce the importance of participating in collective action and confirm the validity of protest information (Centola & Macy, 2007). The mutual term is quite small and is not significant, indicating that links are not more likely than chance to be reciprocated in the network. GWIDEGREE is negative and significant, which provides evidence against a core-periphery structure. Taken together these terms indicate a greater tendency towards a polycentric network structure than a core-periphery one. There is also a tendency towards homophily by region, but not by country or media type. Nodes from Astrakhan and Novgorod are more likely than chance to receive inlinks, controlling for all other terms in the model. Nodes from Russia's largest cities, Moscow and St. Petersburg, are not more likely than chance to receive inlinks, providing evidence against the argument that the protest were of interest only to middle class workers in major cities (Dmitriev, 2015). Social networking sites and communities are more likely than other categories of Web sites to attract inlinks. This provides evidence of the importance of social networking sites such as Vkontakte and Facebook in this initial protest period, as well as online communities of users with similar interests, often around politics, sharing information about the protests and falsified elections (Reuter & Szakonyi, 2015; Smyth & Oates, 2015; S. White & McAllister, 2014). There is also a weak but positive and significant effect for nodes that publish many stories to receive links.

February

In the February time period, GWESP again has the largest odds ratio of any parameter in the model. In contrast to the December model, the reciprocity (mutual) term is positive and significant in February. The mutual term has the second highest odds ratio in the model, indicating that nodes that extend a link are 2.97 times more likely than chance to have that tie reciprocated, controlling for all other terms in the model. The negative parameter for geometrically weighted indegree provides evidence against a core-periphery network. The positive and significant GWESP term in combination with the negative and significant GWIDEGREE term indicates a polycentric network. Similar to the December model, there is again more evidence of tendencies towards a polycentric network than a core-periphery network. In terms of other controls, there is again a weak and positive effect for nodes that publish many stories to receive links compared to chance. There is also evidence of homophily in this model, with nodes that are from the same region as well as those that are of the same media type more likely than chance to link to one another other. Unlike the December model, there is not an indication of nodes from a particular region, large or small, to attract inlinks. Among the media type categorical indegree terms, only news sites have a positive and significant effect, being 1.24 times more likely than chance to attract links. This last finding is more in line with scholars who argue that traditional media and traditional gatekeepers have retained their power and influence in the networked public sphere (Hindman, 2008).

March

The structural effects in the March model are quite similar to the February model. The strongest effect of all terms in the model is for GWESP, followed by reciprocity. There is strong tendency therefore for transitive closure, clustering of nodes together in the network, and for pairs of friends to have many friends in common. Like the December and February models, there is evidence of a polycentric structure on the positive and significant GWESP estimate in combination with the negative and significant GWIDEGREE estimate. On its own, the negative GWIDEGREE estimate also provides evidence against a core-periphery structure. Also similar to both previous time periods, there is a weak but positive and significant effect for nodes that publish many stories to receive links compared to chance in this model. There is evidence of homophily by media type, but not by country or region. Among the categorical indegree terms, news sites again have a positive and significant odds ratio. Unlike previous models, user generated content sites also are more likely than chance to attract links in this time period. This model is more similar to the December model in terms of categorical indegree terms by region, with nodes from Astrakhan, Khakasiya, and Moscow more likely than chance to attract links.

Table 5.9: Final Models Summary Table

	December		February		March		April		May	
Parameter	OR	CI	OR	CI	OR	CI	OR	CI	OR	CI
Edges	0.003***	(0.002,0.003)	0.004***	(0.003,0.004)	0.003***	(0.003,0.004)	0.006***	(0.004,0.008)	0.004***	(0.002,0.008)
Mutual	0.84	(0.56,1.28)	2.97***	(1.78,4.96)	2.39***	(1.73,3.3)	39.82***	(26.23,60.44)	12.22***	(5.47,27.33)
GWESP	14.55***	(13.48,15.69)	10.09***	(9,11.3)	9.18***	(8.26,10.21)	4.31***	(3.52,5.27)	6.41***	(4.44,9.25)
GWDSP	0.94***	(0.94,0.95)	0.92***	(0.91,0.93)	0.94***	(0.94,0.95)	1.70***	(1.40,2.06)	0.80***	(0.72,0.89)
GWIDEGREE	0.13***	(0.11,0.15)	0.65***	(0.54,0.78)	0.67***	(0.55,0.81)	1.38	(0.93,2.06)	2.27**	(1.38,3.74)
GWODEGREE			0.09***	(0.08,0.11)	0.10***	(0.09,0.12)	0.23***	(0.16,0.31)	0.38***	(0.25,0.58)
Story Count	1.002***	(1.002,1.003)	1.004***	(1.003,1.004)	1.005***	(1.004,1.006)	1.007***	(1.006,1.008)	1.006***	(1.003,1.008)
Mixed Two Star							0.56***	(0.47,0.68)		
Cyclic Triple							0.25***	(0.14, 0.45)		
<u>Homophily</u>										
Media Type	0.8†	(0.62,1.03)	1.21***	(1.09,1.35)	1.09*	(1,1.19)	1.29**	(1.08,1.55)	1.75***	(1.33,2.32)
Country	0.99	(0.94,1.05)			0.91**	(0.85,0.97)				
Region	1.23***	(1.1,1.39)	1.29***	(1.16,1.42)	1.05	(0.95,1.17)	1.13	(0.94,1.38)	1.19	(0.89,1.61)
<u>Categorical Indegree (Media Type)</u>										
Blog	0.79***	(0.68,0.91)			0.94	(0.75,1.19)	0.40**	(0.20,0.79)	1.98	(0.74,5.27)
Commercial	1.11	(0.91,1.36)			1.15	(0.89,1.48)				
Community	1.13*	(1.02,1.24)			0.92	(0.72,1.18)	0.43**	(0.23,0.79)	0.66	(0.22,2.01)
LiveJournal			0.68***	(0.59,0.78)	0.75***	(0.63,0.88)	0.44***	(0.31,0.62)	0.61	(0.27,1.34)
News	1.07†	(0.99,1.15)	1.24**	(1.07,1.44)	1.19*	(1.01,1.4)	1.65**	(1.19,2.29)	1.80	(0.8,4.03)
NGO	1.09	(0.97,1.23)	1.09	(0.88,1.35)	1.17	(0.96,1.42)			2.13	(0.86,5.3)
SNS	1.25*	(1.03,1.5)	1.17	(0.81,1.7)			7.82***	(3.09,19.83)	4.02†	(0.83,19.35)
Tools	1.13	(0.91,1.41)								
User	1.16	(0.97,1.4)			1.47***	(1.18,1.83)				
<u>Categorical Indegree (Region)</u>										
Astrakhan	1.34***	(1.21,1.47)			1.70***	(1.49,1.94)	2.28***	(1.68,3.08)		
Chelyabinsk									2.47	(0.7,8.76)
Khakasiya					1.76***	(1.36,2.27)				
Krasnoyarsk	0.74†	(0.53,1.05)								

Table 5.9 (Continued)

Moscow	0.84***	(0.79,0.89)			1.11*	(1.02,1.21)	1.27*	(1.03,1.57)	1.19	(0.86,1.66)
Novgorod	1.36***	(1.14,1.62)								
Petersburg	0.83**	(0.73,0.93)	0.87	(0.73,1.04)	0.97	(0.83,1.14)	0.68	(0.36,1.27)	1.16	(0.53,2.57)
Samara	1.13	(0.96,1.33)							2.09	(0.67,6.47)
Smolensk							9.04***	(4.47,18.29)		
Tartarstan			1.52	(0.92,2.5)					2.54*	(1.13,5.68)

*Note. All models use fixed alpha of 0.25 for geometric terms. Blank cells indicate parameter was not significant and was removed from model to improve fit, or for categorical indegree that term did not have met a minimum threshold to be included. OR = Odds Ratio. CI = 95% Confidence Interval. Significance codes: *** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1*

April

The April model was discussed in detail above, but here I compare the model's parameter estimates to other time periods. There is a much stronger reciprocity parameter in April than in previous models. There is also a weaker, although still positive and significant, effect for the GWESP parameter. There is still a tendency towards triadic closure and clustering in the networks, but it is not as strong as in previous models. The very strong reciprocity (mutual) parameter, and the lower GWESP statistic may be evidence of a move towards less cohesion in the network compared to earlier time periods. This is the only model that has a positive GWDSP estimate. Although it is not significant, there is a positive GWIDEGREE term in this model for the first time, indicating that there might be a core-periphery structure compared to previous models. In summary, for the first time there is more evidence of a core-periphery network structure than a polycentric one in this time period, but the evidence is not clear cut given the insignificance of the GWIDEGREE term. Like all previous models there is a weak but positive tendency towards nodes that publish many stories to receive links compared to chance. Similar to the February and March models there is a positive tendency towards homophily by media type. Similar to the December model, there is a positive effect for social networking sites to receive inlinks. However, the effect is much stronger in the April model, as the odds ratio for social networking sites is fairly large. Protest pages on social networks gaining links in this period may explain this result.

May

The May network is the smallest of any of the protest periods. There are 254 nodes and 259 edges in the network. Similar to the April model, but in contrast to the earlier December, February, and March models, the strongest effect of any parameter in the model is for reciprocity. Also similar to the April model, the GWESP term is the second strongest term in the model after reciprocity. Still, this GWESP term has a smaller odds ratio relative to other terms in the model compared to earlier time periods. Unlike other models, there is a positive and significant GWIDEGREE term in May. Interpreted together, these parameter estimates suggest a tendency towards a core-periphery network structure. Similar to all other models, nodes that publish many stories are more likely to attract links than chance, although it is not a very strong effect. There is again a tendency towards homophily in the May network, but only by media type, which is similar to the April model. There are no significant categorical indegree terms at the standard acceptable alpha level of 0.05, although there is a strong positive effect for social networking sites at the 0.1 level. While several regions have positive categorical indegree terms, only Tartarstan is significant in this model.

Model fit

Model fit is assessed through goodness-of-fit diagnostic plots for each time period, as described above for the April model. The MCMC diagnostic plots are not reproduced as they generally demonstrate a normal distribution for all parameters in all models, and therefore indicate stable models. However, the December, February, and March models have some parameters that do not vary around a mean of zero, although this appears less important than a normallyshaped distribution in the plots, based on published political science research that fits ERGMs to large networks (Ringe, Victor, & Carman, 2014). Based on the MCMC diagnostic plots, all models are stable and the MCMC chains seem to mix well.

In general, model fit for the December period is not ideal, but the model does not appear degenerate. For the December model, the GOF plots, which are found in Appendix A for all models, show that the simulated networks are capturing the observed triad census values reasonably well. GOF statistics indicate that nine of the 16 triad types have statistically different p-values. This indicates that simulated values for several triad types are not statistically similar to observed values. However, as the GOF plot indicates, even though simulated values are not statistically similar for some triad types, they are not wildly different from the observed network. The minimum geodesic distance plot shows at short distances (low values) that the simulated network is capturing the observed network statistic well, but generally is underestimating the minimum geodesic distance of the observed network. This result for minimum geodesic distance is true of all the networks. As previously noted, observed geodesic distances can be difficult to replicate in simulated networks.

The size of this network may explain the difficulty in fitting this model, since networks with a large number of nodes and many edges can be more difficult to fit than small networks with few edges, such as the April and May models (Snijders et al., 2006). Future research could attempt to improve fit by creating smaller networks by either breaking the December data into weeks and modeling each period. It may also be helpful to increase the MCMC sample size. Analysis of the MCMC p-values for a test of how different sample statistics are from observed values indicates that some parameters have low p-values, meaning the starting values are statistically different from observed values and may indicate degeneracy. The typical solution for this issue is to further increase the MCMC sample size. However a second degeneracy test (the Geweke test) indicates that the model is not degenerate, with p-values for all parameters indicating that the simulated values are not different than the observed network.

The February GOF plots indicate that the simulated networks are representing the observed network parameters for triad census well. GOF statistics for triad census show that all but four of the 16 structures are statistically similar to the triad census in the observed network. Minimum geodesic distance is again off except at very low or very high values of the statistic. Since some of the MCMC diagnostic plots do not vary around a mean of zero, as in the December model, I investigated the p-values from the two degeneracy tests provided in the statnet MCMC summary function. These provided similarly mixed results as in December, with the starting values having low p-values for several parameters, a possible sign of degeneracy. Again though, the Geweke test p-values indicate that the model is not degenerate.

The March model was quite difficult to fit and probably has the worst fit of any model based on the GOF plot and statistics for triad census. The GOF statistics indicate that nine of the 16 triad types are not statistically similar to the observed data. However as the plots and statistics indicate, none of the simulated values are wildly off from the observed network. This model does a bit better at simulating geodesic distance than other time periods, but still underestimates the middle range of values for this parameter.

The April model fit is discussed in detail above and is generally well fit.

The May model GOF plots shows excellent fit for triad census. The GOF statistics indicate that p-values are above 0.05 for every triad type, indicating that the count for each triad type is statistically similar to the observed network. Like most other models, the minimum geodesic distance is not represented well by the simulated networks, except at very low and very high values of that statistic.

In general, the earlier, larger models with many edges are less well fit than the smaller models in April and May. In future research, decreasing the size of the networks or adjusting the controls for the MCMC algorithm could improve fit of the December, February, and March networks, and provide additional assurance that the models are not degenerate.

Discussion and conclusion

Analysis of the five ERGM models shows that initially the network has tendencies towards a polycentric network structure. However, over time the networks begin to show a tendency towards hierarchical network structures. Polycentric networks have both strong and weak tie structures, and are seen a conducive to collective action in risky political environments. Hierarchical networks can theoretically quickly spread information and more easily turn out members for protests, but are less robust and more susceptible to defections or removal of key nodes in high-risk political environments. Polycentric structure is indicated by large GWESP statistics early in the protests, negative GWDSP estimates, negative GWIDEGREE terms, and insignificant or small reciprocity parameter estimates. Later in the protests, in April and May, the networks are smaller, have relatively lower GWESP statistics, significantly higher reciprocity scores, and significant and positive GWIDEGREE statistics, indicative of hierarchical tendencies.

These results provide support for the main hypothesis tested in this chapter. The ERGM models of earlier time periods show tendencies toward polycentric structure, while the later time periods show a move towards a less effective hierarchical structure as the protests lose popular support. The time periods when the network has a hierarchical structure are correlated with smaller protests, a die-hard core of protest participants, and fewer regional protests. It is important to note that I am not making any causal claims about network structure and protest success, simply that the network structure aligns with what would be expected based on what we know about the composition of protests at different time periods.

In terms of controls, GWOUTDEGREE is negative and significant in every period in which the term is included (all but December), indicating that nodes generally have similar outdegree values and that the network is not centered around high outdegree nodes. In all time periods there is a weak, but positive and significant, tendency for nodes that publish many stories to receive more links than random chance. This is as expected for an online environment, but the effect is not as strong as some might predict.

There is a tendency towards homophily by media type, country or region in every model. There is a greater tendency towards homophily by region in the early models, and a greater tendency towards homophily by media type in later models. So, nodes from the same region are more likely to link to each other early in the protests, coinciding with the time periods when regional protests were more prevalent. The homophily by country term is necessary to include for model fit in two of the models, but does not address any research question of interest.

Among the categorical indegree (nodeifactor) terms by media type we see that early in the protests, in particular in the December model, that social networking sites and online communities have positive and significant parameter estimates. Over all periods, the social networking site parameter estimates are probably the most interesting. The SNS parameter has positive odds ratios in four of the five models and is significant in three. It has quite a large effect in the April model and also has a large effect in the May model. Although the SNS term is not significant at the standard 0.05 alpha level in May, it is significant at the less robust 0.1 level. These results support the findings of several Russia area-studies scholars that social networking sites such as Facebook, Vkontakte, Odnoklassniki, and Twitter played a role in raising awareness about election falsification and protests while also aiding coordination of protest activities (Reuter & Szakonyi, 2015; S. White & McAllister, 2014). However, it is not possible to tease out the influence of those different social networking platforms without additional modeling. User generated content sites have positive and significant odds ratios in two of the models. LiveJournal dominates the network, but has negative odds ratios and is significant in just two time periods.

Finally, categorical indegree terms by region show that Astrakhan has a positive and significant odds ratio in three models, including in the April model when the protests in that region are the center of attention. Astrakhan is the only region that is positive and significant in more than two models. Moscow is positive and significant in two time periods, April and March. St. Petersburg is significant in just the first model but has a negative odds ratio in that time period. Nodes from the two largest cities in Russia do not appear to attract links more than chance in the majority of models, undercutting somewhat the assertion that the protest movement concerns only the middle class residents of Moscow and St. Petersburg, when taking into account other terms in the models (Dmitriev, 2015). Otherwise, no clear patterns emerge from the categorical indegree results for regions. However, these categorical terms are helpful to include in models that have related homophily terms, since they can control for especially 'social' groups that may appear to be homophilous but are simply more social, or more active in their linking behavior, than other groups.

Future research

There are several possible avenues for future research that could improve these models and perhaps reveal the influence of hidden variables. First, no edge covariates are included in these models. Since the network has to be simplified to accommodate the statnet software, it is possible that adding a term for edge weight to each model could yield better, or at least different, results. It is also possible that additional homophily or nodeifactor terms could improve the models. Some likely candidates include support or opposition for the protests, but that would probably require the use of some sort of machine learning algorithm that, for example, assigns each node to a category of support, opposition, or neutral sentiment based on the text of the stories for each node.

Adding a political affiliation node attribute, such as whether nodes are affiliated with nationalists, environmentalists, or supporters of the democratic opposition could also improve or alter the results of the models. However, a quick test building off the final December model that added political affiliation as a node attribute to LiveJournal nodes did not improve the model fit, although democratic opposition and nationalist blogs were significant categorical variables in that test. The data used in that test was drawn from a previous large-scale Russian blog study that identified groups of nationalist, environmentalist, and democratic opposition blogs, among other political and cultural affiliations, based on their long-term linking behavior (Etling et al., 2010). However, only about half the nodes in the December network could be matched with data from that previous study, so a more concerted effort to code each node in the network for political affiliation could lead to better results in terms of model fit. The coding results generally would be more robust if multiple coders manually coded node attribute data, after conducting inter-coder reliability tests. Differential homophily models could also reveal the influence of specific media types or regions, although that would lead to additional terms in what are already complex models.

Finally, it may be that more recent models in statnet for temporal networks are able to handle significant change in networks between time periods. If so, that would be a useful area for future research using this data. It is also possible that selecting ego networks and simulating network change over time in statnet would be useful for understanding change over time in the protest network.

The next chapter continues the focus on the importance of networks and network structure for collective action in hybrid and authoritarian regimes. However, instead of focusing on tie formation I investigate the role of brokers, those actors in the network who sit at the end of weak ties. Brokers have been found to play important roles in the emergence and success of revolutionary and social movements. Their importance stems from their ability to access diverse information from different communities, and to introduce novel information into clusters of homophilous, strongly tied actors. However, this key assumption is rarely actually tested. In the next chapter, I use Latent Semantic Analysis to test whether brokered information is actually novel, and if there is an association between network structure and information novelty.

Chapter 6 Is Brokered Information Actually Novel? Using Latent Semantic Analysis to Test Topic Heterogeneity among Actors Connected by Brokers

> *It is hardly possible to overstate the value in the present state of human improvement of placing human beings in contact with other persons dissimilar to themselves, and with modes of thought and action unlike those with which they are familiar.*
>
> *-- John Stuart Mill*

Brokers play essential bridging roles in social movements, firms, and all sorts of social networks. Their importance stems from their ability to connect people with ideas, information, and individuals that they are not exposed to through their close friends or family (Burt, 1992, 2004; Granovetter, 1973, 1974, 1983; Putnam, 2000). Empirical evidence from studies of social and revolutionary movements show that brokers are able to tie together diverse political groups and social classes, aid in the diffusion of revolutionary ideas, and contribute to the emergence of broad-based, geographically dispersed movements instead of isolated, local ones (D. Faris, 2008; Han, 2009; Hedstrom et al., 2000; Kim & Pfaff, 2012). For example, leaders such as Paul Revere and Joseph Warren used overlapping memberships in diverse revolutionary organizations and worked in professions that allowed them to bridge the many social, political, and economic fault lines that existed during the American revolution (Han, 2009). It has also been shown that students from critical university communities were key to the diffusion of ideas during the protestant reformation (Kim & Pfaff, 2012), and that evangelical labor organizers in Sweden were able to create a coordinated, national labor movement by tying together otherwise disconnected working communities (Hedstrom et al., 2000).

Brokers in high-risk political settings can help solve the strong tie conundrum for protest movements. Strong-tie bonds are necessary for risky collective action, but have negative externalities associated with them. On the positive side, strong ties build internal group solidarity, are more likely to influence behavior, reinforce the normative importance of participation in collective action, and allow for repeated affirmation of protest information (Centola, 2010; Centola & Macy, 2007; McAdam, 1986, 1988; McAdam & Paulsen, 1993). But strong ties can lead to networks that have clusters of isolated, homophilous actors that are disconnected from each other, and whose actions are therefore likely to remain uncoordinated. This can lead to localized protests that have less impact than coordinated national ones, or a movement made up of a narrow group of participants that are more easily targeted and neutralized by the state. Brokers are able to help overcome these negative externalities by tying together disparate protest constituencies, ensuring that information travels between those groups, and also help translate information and ideas between diverse groups (Borgatti & Lopez-Kidwell, 2011; Burt, 1992, 2004; Granovetter, 1973, 1974; Han, 2009). For protests in high-risk political settings, it is essential that movements are broad-based and the efforts of diverse constituencies well coordinated if they are to overcome the advantages of powerful entrenched regimes with more resources at their disposal. Finally, for protest movements that emerge due to election fraud or other public, national scandals, brokers are necessary to tie together and coordinate the efforts of the diverse oppositional groups that are frequently mobilized around a common grievance under such scenarios (Tucker, 2007).

The main benefit of brokers is that they introduce or have access to novel information from actors in the communities they connect. This is because information and opinions are more homogenous within groups than between them (Burt, 1992, 2004; Granovetter, 1973, 1974; Putnam, 2000). It is generally believed that there is great value in being exposed to novel information. Those with access to novel information are better positioned to broker opportunities, make better decisions, apply that knowledge to solve problems, have a wider knowledge base, and are more persuasive (Aral, 2016; Aral & Van Alstyne, 2011). Information within groups tends to be homogenous due to homophily, the tendency for individuals with similar characteristic to form ties and cluster together in groups in networks. This leads to homogenous personal networks and can limit access to information, lead to polarization, and negatively impact the attitudes people have towards other political, social, and ethnic groups (McPherson et al., 2001; Putnam, 2000). The internal focus on the information within one's own group creates gaps, or structural holes, in the information flow between groups. Brokers bridge these structural holes, and therefore have unique access to novel information found in the diverse groups they connect.

Brokerage studies in the social movement and organizational studies fields rarely test whether the information that brokers access from the different groups they connect is actually novel, and how network structure might affect their access to novel information. This chapter attempts to help fill this gap in the literature. This chapter addresses two research questions. The first asks if brokered information in a protest communication network discussing high-risk collective action is actually novel. I focus in particular on gatekeeper and representative-type brokerage relationships as defined by Gould and Fernandez (1989), where a brokerage relationship only exists when a broker connects nodes from different, non-overlapping groups. The first hypothesis is that the text of nodes from different communities that are connected by brokers will be novel, since information and ideas shared between groups tends to be novel compared to that shared within groups. The second hypothesis states that the text of brokers should be novel compared to other nodes in their own communities, since brokers are linked to and have access to novel information from outside groups.

The second research question investigates the role of macro network structure on the novelty of information that brokers have access to locally. Specifically, is brokered information more novel when a network has a polycentric structure than when it has a hierarchical structure? The third hypothesis states the text of brokers will be the most novel when the network has a polycentric structure, since such structures will have many communities of different types discussing a diverse range of topics, more structural holes, and more brokers and weak ties connecting disparate communities than a hierarchical network.

In the next section of this chapter, I explain in more detail the different types of brokerage relationships that are possible and justify the focus on gatekeepers and representative-type brokers as defined by Gould and Fernandez (1989). I then describe the empirical network and textual data. Next, I explain the methods used to automatically partition the network into different communities, identify the brokers that link those communities, extract the brokers' ego networks, and the use of Latent Semantic Analysis (LSA) to measure novelty (topic similarity) between the text of brokers and the nodes they connect. In the method section, I also describe the major communities in the Russian protest network as well as its major brokers. The findings are then provided, followed by a conclusion and discussion section that also explores possible future directions for this work.

Brokerage types and measurement

There are several different measures and definitions of local brokerage. Burt's (1992, 2004) constraint is an index that measures the opportunity for a node to act as a broker by identifying nodes that are in a position to span structural holes. Burt's constraint measures the extent to which a node's ties are concentrated in a single group of interconnected nodes. Constraint is high if a person has few contacts and those contacts are strongly connected to one another, either directly or indirectly. Actors that are brokers and span structural holes therefore have low constraint. Another popular measure of brokerage is betweenness centrality, which measures the degree to which a node falls on the shortest paths between other nodes (Freeman, 1977). Yet another set of measures developed by Gould and Fernandez (1989) takes into account the direction of links as well as the groups that a broker is connecting, requiring that nodes in the network be divided into a set of non-overlapping groups. This is important since a key assumption of novelty within networks is that the information between groups is less similar than that within groups. I use the Gould and Fernandez method in this chapter. This method is also preferable to Burt's constraint since it actually identifies nodes that broker flows of information or resources between groups in a directed network, instead of just the potential for brokerage. This method is also preferable to betweenness methods because betweenness takes geodesics of any length into account and treats them equally. In reality, nodes connected by very long paths are unlikely to have a strong impact on an actor compared to one they are directly connected to (Bruggeman, 2016; Burt, 2010; Gould & Fernandez, 1989).

Gould and Fernandez develop definitions of five structurally distinct types of brokers. It is important to note that a given node can occupy more than one type of brokerage position, and they frequently do. Also, the nodes that occupy a given brokerage role can vary across time, since each network is treated as independent at each time period in my analysis. I use four of the five brokerage roles described by Gould and Fernandez in this chapter, and describe each next. Images and a short description of each are provided in Figure 6.1.

In the first type of brokerage role, a broker acts as a gatekeeper for his group, controlling the flow of information or resources coming in from nodes that belong to other groups. The top image in Figure 6.1 shows a gatekeeper type structure, where node A belong to a different community or group than the gatekeeper. I use the shorthand 'indiff' to describe node A. From the broker's perspective, node A links in to the broker or gatekeeper (node B). In these images, I also refer to the broker as ego, since the ego networks of every node in a brokerage position are extracted and used to determine the novelty of text. This is explained in more detail in the method section. The gatekeeper node links out to node C, and belongs to the same community as gatekeeper. I use the shorthand 'outsame' to describe node C. For this study I have modified the original Gould and Fernandez definition to include the rest of the nodes in the

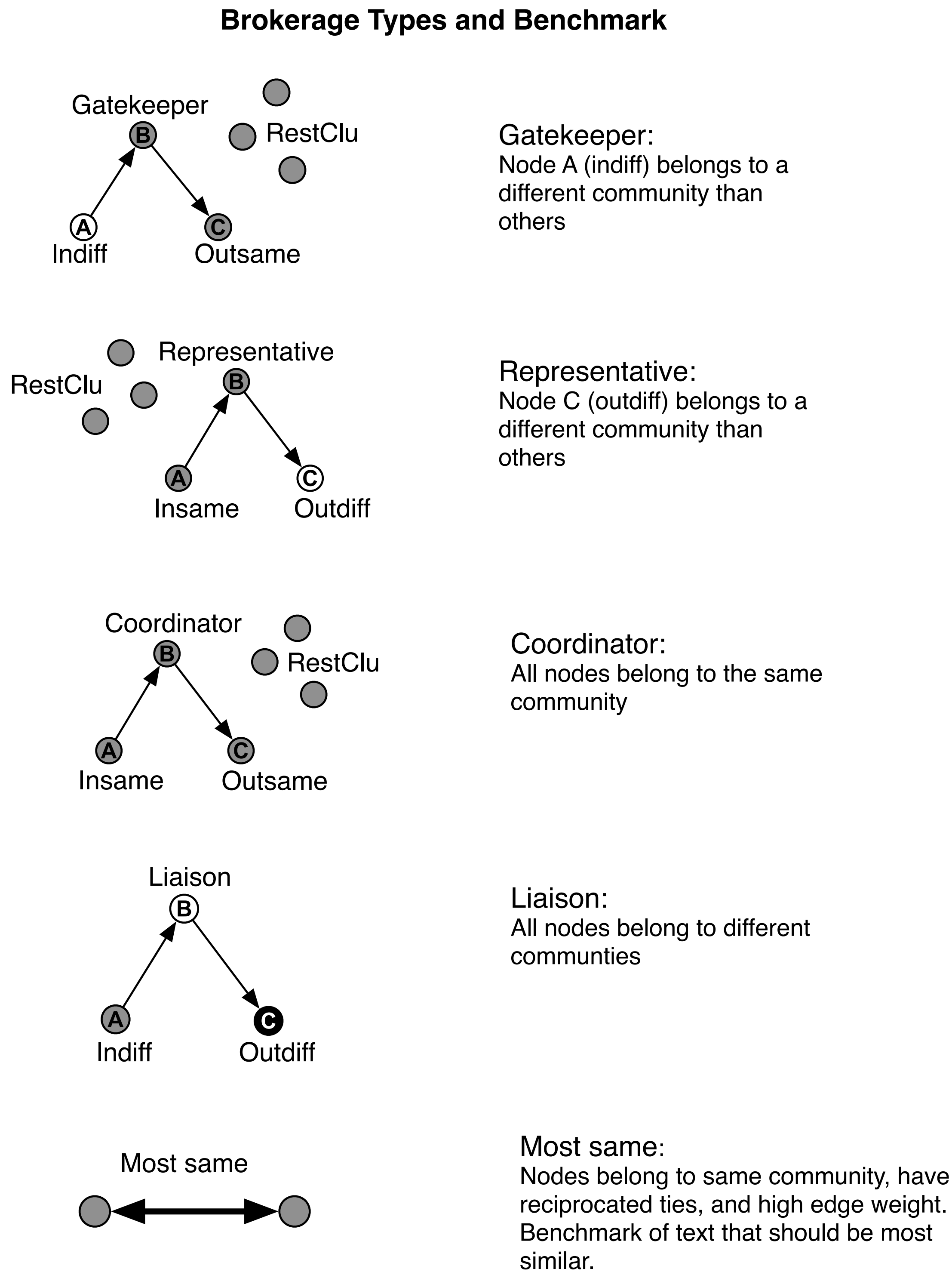


Figure 6.1: Brokerage types and benchmark of similarity. The top node (node B) is the broker in the first four images. The bottom image (mostsame) is not a brokerage type, but a benchmark of a relationship where the text should be the most similar.

gatekeeper's community that are not directly connected to the gatekeeper. This structure allows us to investigate the second hypothesis. If a broker has access to novel information outside his community, then the text of the broker should be novel compared to the text of the rest of the nodes in its cluster.

The second brokerage type is similar to the gatekeeper role just explained, in that the broker is creating a connection between nodes from two different communities. In representative type brokerage, node A (insame) and B (representative) are in the same community. The broker or representative node links out to node C (outdiff), which belongs to a different community. The general idea here is that the broker is a representative of his community to those outside it, since the broker controls the flow of information or resources out to the nodes belonging to different communities. As before, the restclu nodes are nodes that are in the same community as the broker, but are not connected directly to the broker.

Since novelty is relative, I use three benchmarks of similarity to compare brokered text against. The first two benchmarks are brokerage types developed by Gould and Fernandez (1989), the coordinator and liaison brokerage types. In coordinator type brokerage, all of the nodes belong to the same community. A key assumption of the brokerage literature as it relates to novelty is that the information shared among nodes in the same community will be more homogenous than the information shared among nodes that belong to different groups. In the coordinator brokerage relationship, Node A (insame) links in to the broker, and the broker then links out to node B (outsame). The restclu nodes are also part of the same community as those in the directly brokered relationship (nodes A, B, and C). Again, the point of including this type of brokerage in the analysis is to determine if the similarity of the text of the nodes in the coordinator type brokerage is higher than in the gatekeeper and representative types.

For the liaison brokerage type, each node belongs to a different community. In general, this type of relationship is analyzed to see if the overall levels of similarity between the nodes in this type of relationship are lower than other brokerage types (gatekeepers, representatives, and coordinators). Again, theory predicts that the text and information between different groups should be more heterogeneous than that shared within groups.

I use one final benchmark to compare novelty scores to. This is a benchmark of node pairs whose information should be the most similar of any relationship tested in this chapter. Network theory tells us that homophily will be strongest among people that are similar to each other, who reciprocate their communication, and who communicate frequently. This type of strong tie structure between similar nodes usually indicates high levels of cohesion, social capital, and trust (Blau, 1964; Burt, 1992, 2004; Granovetter, 1973, 1974; Putnam, 2000). We should therefore expect that the text of nodes that are in the same community, that have reciprocated ties, and that have a high edge weight should be very similar. For each time period, I therefore identify ten node pairs that are in the same community, that have reciprocated ties, and that have the highest edge weight of all node pairs in the network at that time period. I then take the average similarity scores of these node pairs as a benchmark of text that should be the most similar. I use the shorthand 'mostsame' to describe these node pairs (see bottom image of Figure 6.1). In the next two sections, I describe the data and methods used to test the chapter's hypotheses.

Data

One of the main reasons that the novelty of text has rarely been tested in brokerage studies is because, until recently, it was it rare to have to access to such granular data. This has begun to change with the rise of the Internet and electronic, networked communication. In this chapter, I use both network and textual data from the Russian protest data set described in Chapter 4. This unique dataset includes nearly 30,000 Russian documents and the linking structure of over 3,000 Web sites that discussed the Russian election protests between December 1, 2011 and May 31, 2012.

For the network data in this chapter, I create directed, temporal networks focused on the same time periods as the ERGM models that were built in Chapter 5. The results of those models indicate whether the networks at each time period are polycentric or hierarchical. In Chapter 5, five periods of peak protest activity were modeled and compared. These peaks were identified based on the distribution of documents (stories, blog posts, and other online content included in the data) between December 1, 2011 and May 31, 2012. These peaks coincide with major protest events and the legislative and presidential elections. These time periods coincide roughly with the months of December, February, March, April and May. However, since so few brokers were found in the May network, that month is not used in the analysis in this chapter. For example, there are only four nodes that play gatekeeper or representative brokerage roles in the May network, and just one who plays a liaison role. It is hard to draw any conclusions from such a small number of observations, so that month is not included in the analysis in this chapter. The network for each time period includes all those nodes that sent or received at least one link during the given period of interest. The text used in the analysis includes all the documents (stories, blog posts, social media posts, etc.) created by the nodes during each time period. Table 6.1 provides some descriptive statistics for these networks.

Some of the descriptive statistics in Table 6.1 can be used to help us understand if networks are polycentric or hierarchical. Centralization is a good descriptive measure of how hierarchical a network is, and refers to the extent that a network is centered on a single node. A centralization score of one indicates that all nodes are connected to one node in the center of the network. Table 6.1 shows that the April network is the most centralized and December the least centralized. The February and March networks have centralization score closer to December than April.

Table 6.1: Descriptive Statistics for Temporal Protest Networks

Network	Nodes	Edges	Degree	Density	Recip.	Transitivity	Centrality
December	1493	4584	6.14	0.0021	0.055	0.146	0.043
February	1118	2042	3.65	0.0016	0.063	0.120	0.063
March	1043	2167	4.16	0.0020	0.086	0.125	0.059
April	339	616	3.63	0.0054	0.214	0.055	0.204

Note. Recip. = Reciprocity.

Transitivity and density scores can also be interpreted together to help us understand whether the temporal networks are polycentric or hierarchical. Polycentric networks tend to have higher transitivity and lower density scores (Brooks, Hogan, Ellison, Lampe, & Vitak, 2014). This is because a network with higher transitivity will have many closed two paths, or triangles, which indicates a network with distinct clusters of well-connected groups. The lower density scores indicate that there are many possible connections that are not being made between these clusters. Lower transitivity scores, on the other hand, are likely to indicate more open two paths between these clusters, while higher density scores indicate more connections are being made between clusters. Table 6.1 indicates that, similar to the ERGM models, the December network is the most polycentric and the April model the most hierarchical, with the February and March models somewhere in between, but closer to a polycentric structure than a hierarchal one. With the description of the data complete, I move in the next section to a discussion of the method used to determine the novelty of brokered text, and also provide descriptions of the major communities and brokers in the Russian protest network.

Method

I use both social network analysis and automated text analysis methods in this chapter. In broad strokes, my method is to partition the network into communities, identify brokers, extract their ego nets, and use latent semantic analysis (LSA) to determine the similarity, or novelty, of the text of the nodes in each brokerage relationship. My specific methodological approach is as follows. First, since the Gould and Fernandez method of identifying brokers requires that nodes belong to non-overlapping groups, I first partition each network into communities using the Louvain community detection algorithm. This algorithm returned the best partitioning of the network, based on modularity scores, compared to several commonly used community detection algorithms that I tested. A comparison of the community detection algorithms and description of the communities identified are discussed in greater detail below, and the results are shown in Table 6.2. Next, I used the brokerage function in the sna package in R to identify nodes that play each of the brokerage roles of interest. Any node with a brokerage score above zero was included in the analysis for each brokerage type. I then extracted the ego network with both in and out links at a distance of one for each broker. This means that a subgraph is created for each broker, with the broker at its center and every node it is directly connected to. For each brokerage type of interest, I then deleted those nodes that did not meet the definition of the brokerage type being analyzed. So for a gatekeeper ego network, only those nodes from communities that are not in the same community as the broker and that link into the broker, and only those nodes that ego links out to that are in the same community. For the restclu nodes, I identified all the nodes that belonged to each broker's community, as determined by the Louvain community detection algorithm, and then deleted all those nodes in the broker's ego network from that set. I then matched the document text to each node in each ego network for every time period of interest based on the publication date of each document.

I use well-established methods from the information retrieval field to measure the novelty of text in brokered relationships. In broad strokes, I create a term-document matrix, weigh the terms in the term-document matrix by term frequency-inverse document frequency, identify latent topics with Latent Semantic Analysis (LSA), and calculate the similarity of topics among documents using cosine similarity. Specifically, I first preprocess the text corpus by converting all text to lowercase, removing stopwords, removing numbers, and removing white space. The text is then normalized by stemming, reducing words to their common root or stem. The TM package in R was used to preprocess and normalize text. The text is then put into a term-document matrix, where rows are terms and documents are columns. Documents are grouped by node type within each brokerage relationship of interest (e.g., indiff, broker, outsame and restclu for gatekeeper type brokerage). To improve the signal to noise ratio, I next weight terms in the term-document matrix by term frequency-inverse document frequency (TF-IDF) instead of using raw term counts. TF-IDF identifies important words by comparing the frequency of a word within a document to its frequency within the entire corpus. Specifically, TF-IDF assigns the most weight to terms that occur many times in a small number of documents, less weight to terms that occur less frequently in a document or that occur in many documents, and the lowest weight to terms that occur in virtually all documents (Manning et al., 2008). Latent Semantic Analysis (LSA) is a method used to identify latent topics in documents through the co-occurrence of words (Landauer & Dumais, 1996, 1997; Landauer et al., 1998). The basic idea behind LSA is that topics or concepts can be found in patterns of words that occur together in documents. So the co-occurrence of the words Obamacare, health, and insurance could represent a document discussing the Affordable Care Act and troop, war, Taliban and Afghanistan could indicate discussion of the war in Afghanistan (Grimmer & Stewart, 2013). LSA uses a mathematical technique called singular value decomposition (SVD), which is similar to factor analysis, to improve the signal to noise ratio in the term-document matrix. SVD reduces the dimensional representation of the matrix and emphasizes the concepts that are most important while eliminating the words that are noise. Finally, cosine similarity is used to rank the similarity of the latent topics. Cosine similarity is a method used to measure the similarity of any two vectors or documents by measuring the cosine of the angle between those two vectors. This results in a score between 0 and 1, where documents closer to 1 are more similar and those closer to 0 are the more dissimilar, or in this analysis, novel.

One can reasonably ask why I do not use Aral and Van Alstyne's methodological approach, since our research questions are similar. I chose not to replicate Aral and Van Alstyne's method because they use network autocorrelation models. These models have been shown to have some problems under certain conditions. Specifically, network autocorrelation models that use maximum likelihood estimation have been found to negatively bias the network effect under several conditions, including in dense networks. These problems have been found to occur across all common network structures (Dow, Burton, & White, 1982; Mizruchi & Neuman, 2008; Neuman & Mizruchi, 2010). More recent work has shown that this bias is less of an issue when networks are large, leading to the recommendation that networks should consist of at least 40 nodes (Wang, Neuman, & Newman, 2014). However, since I am dealing with ego networks, which tend to be small, I choose not to use network autocorrelation models. Instead, I use the method as described in this section, but acknowledge that this makes direct comparison with previous studies more difficult.

Community detection

The partitioning of the network into non-overlapping groups is an important part of the research method. I use an automated community detection algorithm to do so. I chose to use the popular Louvain algorithm (Blondel et al., 2008) after comparing it to several alternatives . Others algorithms tested included Girvan-Newman's Edge Betweenness (Girvan & Newman, 2002), Fast Greedy (Clauset, Newman, & Moore, 2004), Walktrap (Pascal & Latapy, 2006), Spin Glass (Reichardt & Bornholdt, 2006), Leading Eigenvector (Newman, 2006), Label Propagation (Raghavan, Albert, & Kumara, 2007) and Infomap (Rosval & Bergstrom, 2007). I experimented with each of these for the overall network and for each temporal network. Table 6.2 shows the results for each time period (the results for the overall time period are broadly similar to the December network). As Table 6.2 indicates, the Louvain method had the highest average modularity score over all time periods, and had the highest modularity score each month (including ties). Table 6.2 also shows that the fast greedy modularity scores are only slightly lower on average to Louvain, and are identical to Louvain in February and April. Modularity is a measure of the density of links that occur within communities compared to the density of links between communities, and is a common measure of how well a given algorithm is able to partition a network into communities (i.e., the quality of a partition) (Blondel et al., 2008; Girvan & Newman, 2002; Newman & Girvan, 2004).

Table 6.2: Comparison of Community Detection Algorithms

	December		February		March		April	
Algorithm	mod	# com	mod	# com	mod	# com	mod	# com
Louvain	0.737	80	0.748	66	0.716	43	0.567	28
Fast Greedy	0.735	81	0.748	69	0.713	40	0.567	28
Label Propagation	0.705	157	0.687	141	0.680	58	0.506	51
Walktrap	0.644	91	0.663	147	0.667	55	0.505	30
Edge Betweenness	0.638	83	0.620	79	0.649	35	0.486	32
Leading Eigenvector	0.622	97	NA	NA	0.660	43	0.479	22
Infomap	0.551	267	0.575	245	0.523	72	0.499	55
Spin Glass	0.370	23	0.257	25	0.314	19	0.229	17

Note. mod = modularity; # com = number of communities. Algorithms ranked by average modularity across all time periods. The leading eigenvector algorithm did not converge in February, so that algorithm has NA values for that time period.

The Louvain algorithm is divided into two phases that are repeated iteratively. In the first iteration of phase one, each node is assigned to its own community. Then each node is placed into a different community to see if this improves modularity. If an improvement is not made it is left in its original community. This is done iteratively until modularity has been maximized; that is, when no individual move improves the modularity. The second phase builds a new network where the nodes are the communities that were found in the first phase, and the edge weight between the nodes is the sum of the weight of the links between nodes in the original two communities being connected. Modularity is then maximized for this 'network of communities' as it was in the first phase of the algorithm. Blondel et al. call a combination of these two phases a pass of the algorithm. Passes are iterated until changes no longer improve modularity. This method might therefore be thought of as a combination of modularity maximization and hierarchical clustering methods. The main benefits of this method are that it is unsupervised, can be used on very large networks with millions of nodes, and is extremely fast (Blondel et al., 2008). One possible limitation of using the Louvain method is that the algorithm treats the network as undirected, so information on link direction in community detection is not used. The Infomap and Walktrap methods both take link direction into account, but neither had modularity scores as high as Louvain. As described in the previous chapter, Chapter 5, this network is not very stable over time. Nodes drop in and out of the network at each time period, and the structure changes over time. To obtain the best partitioning of the network at each time period, I therefore rerun the Louvain method and identify communities in networks for each time period.

Russian protest communities

Here I briefly describe the major communities that are found in this network, using the December time period as an example. How similar these communities are to each other can also be expected to have an impact on the novelty of information. Given that the nodes in these networks all discuss the same general topic of the election protests, the similarity scores may be higher than a network that discuss politics more generally, or one that discusses a range of topics.

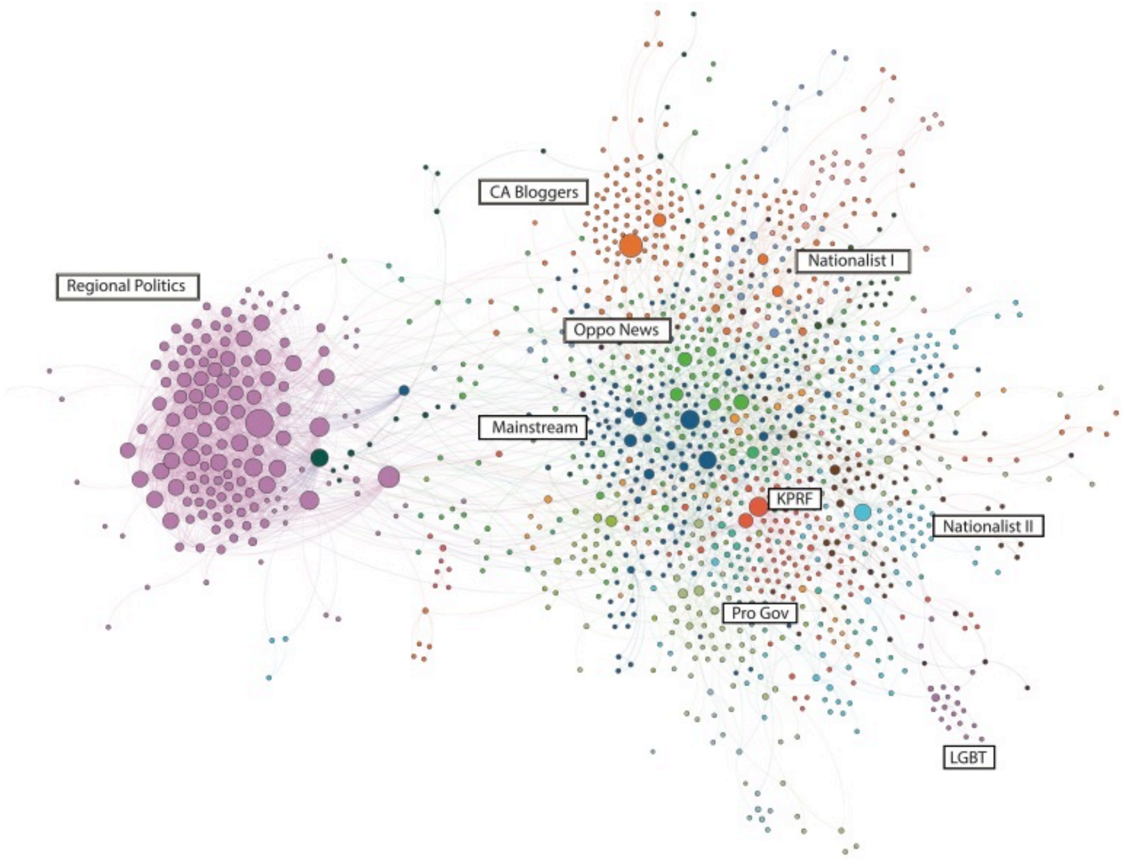


Figure 6.2: December network communities. Colors represent different communities and node size is determined by indegree.

Figure 6.2 shows the giant connected component for the December time period, which shows the core of the network. Here I describe some of the major communities in this network. The largest community in this network sits at the center of this network map. I call this the ***Mainstream*** community as it consists primarily of the mainstream core of the Russian protest movement. This includes blogs of protest leaders such as Alexei Navalny and Evgenia Chirikova. It also includes forums and Web sites affiliated with the umbrella Solidarity movement that helped to organize, publicize, and handle the logistics of the protests, and namarsh.ru, a site which helps to publicize and record civic protest event in Russia (Lankina, 2015; Lankina & Voznaya, 2015). This community also includes major social media sites that hosted protest pages and videos of events, including Vkontakte and YouTube. The Mainstream community also includes a couple independent news sites, such as lenta.ru. However, the community just to the north of the mainstream group on the map, which I call ***Oppositional News***, consists almost entirely of independent, oppositional news sites. Many independent journalists were key players in the protest movement (Greene, 2013; D. White, 2015). This community includes sites such as grani.ru, Novaya Gazeta, RBC daily, and the community news site ridus.ru. Independent but less oppositional news sites include the business paper Vedomosti and inopressa.ru.

The next largest community on the map is clearly visible off on its own on the left side of the network. I label this cluster ***Regional Politics***. This community consists almost entirely of political LiveJournal bloggers, many focused on regional issues in places such as Pskov, Astrakhan, Voronezh, and Kirov. While the majority of sites in the network are located in Moscow or St. Petersburg, the majority of nodes in this community are located in Russian regions. There are some nationalist and extreme leftist blogs in this group, but many more belong to mainstream regional politicians, activists, and close observers of regional politics. Oleg Shein, a leftist politician from Astrakhan that took part in a hunger strike in April, is a major node in this community.

The next major cluster found in the December network is located at the top of the map. I label this ***California Bloggers***, because this group is centered on a couple expatriate Russian bloggers who live in California. This community is made up almost exclusively of LiveJournal bloggers based in Russia, however. These bloggers discuss the protests extensively this month, including suggested tactics. This group also includes a couple national political actors associated with the protest, including Ilya Yashin and Ilya Ponomarev. Yashin is a leader of the umbrella Solidarnost democratic group, and Ponomarev was one of the only members of the federal legislature that came out in support of the protests.

On the right side of the network map are two closely related communities of nationalists. ***Nationalist I*** includes the major nationalist figures that participated in the protests and led the nationalist constituency within the broader protest movement. This includes individuals such as Vladlen Kralin (aka Vladimir Tor), leader of the since banned Movement Against Illegal Immigration (DPNI), Eduard Limonov, founder of the also banned National Bolshevik Party, and Konstantin Krylov of the Nationalist Democratic Party, who was also a member of the Opposition Coordinating Council (Popescu, 2012; Umland, 2012). While the first group is focused around individual leaders, ***Nationalist II*** is focused around the nationalist DPNI Web site, which serves as a sort of information clearinghouse for nationalists in Russia. A different algorithm, or tweaks to the Louvain algorithm, could collapse these communities together.

At the bottom of the graph are two smaller ***Pro Government*** clusters. These includes 'patriotic' blogs, forums focused on the Kremlin and Kremlin politics, and conservative, pro-government media such as Vzglad, Russia Today, NTV, Pravda, and Argumenti i Fakti (Greene, 2013). Writers and commentators affiliated with these conservative media sites are also found here.

Two smaller clusters include a group centered around Web sites and forums affiliated with the Communist Party of the Russian Federation ***(KPRF)***, as well as a more general Russian politics forum. Finally, on the far bottom right of the graph there is a small, but clear and tightly connected community of ***LGBT*** activists and gay rights Web sites. While not a large group, they form a clear community in the network. The influence of this community will be explored more in Chapter 7.

With the exception of the regional politics community, these automatically identified clusters very closely match the groups used in Chapter 7 that are believed to have played major roles in the protest movement. They can be thought of as representing different constituencies within in the broader movement, and were already described in Chapter 3. The communities identified in the December network are also very similar to those in the overall network, which includes all nodes active at any time in the six months of the protests. Since there is significant variation between networks I ran the community detection algorithm separately for each temporal network. Future research could use the networks created from the overall time period and hold those communities constant across all temporal networks. The Louvain algorithm could also be tweaked to see if it is possible to further improve the partitioning of the network.

Russian brokers

What individuals or Web sites served as the top brokers in the Russian protest network? Since this is a link network, a broker can be any type of Web site in the data, be it an individual's blog, a newspaper, a political forum, or a social networking site. Substantively, since this network included many different constituencies, we would want to know who, or what type of site, stitches this network together and bridges across these different groups. Table 6.3 lists the top brokers over all months under study, based on a total brokerage score. This score is the sum of the gatekeeper, representative, coordinator, and liaison scores for each node. As can be seen from Table 6.3, Oleg Shein is easily the top broker in the network. Again, Oleg Shein is a leftist politician from the region of Astrakhan who led protests against election falsification in the mayoral race in Astrakhan in the month of April. One might expect that protest leader Alexei Navalny might be an important broker, given his contacts with different constituencies, including democratic liberals, environmentalists, and nationalists. Navalny comes in at number seven based on total brokerage, which is not as high as one might expect. For example, Navalny is not nearly as important a broker as Oleg Shein, a protest leader with a much lower profile than Navalny. In terms of media types that serve brokerage roles, there is a mix of blogs of protest leaders, writers, and nationalists, as well as newspapers and forums. While most of these sites are affiliated at least loosely with the different parts of the opposition protest movement, the conservative, pro-government Vzglad newspaper is also on the list of top brokers.

Table 6.3: Top Brokers

Rank	Name/Title	Type	Brokerage	Notes
1	Oleg Shein	Blog	12153	Leftist, protest leader
2	Lenta	News	3961	Independent news site
3	Newsru	News	2312	News site
4	Gazeta.ru	News	1990	Independent news site
5	Irek Murtazin	Blog	1870	Echo of Moscow affiliate
6	Dissenters March	Forum	1616	Protest news and info.
7	Alexei Navalny	Blog	1348	Protest leader, activist
8	Ru Politics	Forum	1114	Russian politics forum
9	Vzglad	News	782	Conservative news site
10	Coordination & Revolution	Forum	762	Nationalist blog and news
11	Solidarity	Forum	655	Umbrella democratic org.
12	Ilya Ponomarev	Blog	645	Member of parliament
13	Andrei Mal'gin	Blog	597	Writer, journalist, critic
14	Basmanov	Blog	551	Nationalist, DPNI affiliate
15	Drug ZhZh	Blog	540	Nationalist

Note. Brokerage score is combined total of all brokerage types over all time periods.

In this section, I described my method of partitioning the network and the use of LSA to identify the similarity of topics among brokers and those they connect. I justified the use of the Louvain community detection algorithm and also provided a description of the communities identified by that algorithm in the December Russian protest network. I also provided a description of the top brokers in the Russian network. With the description and justification of my method complete, I next turn to the findings.

Results

This chapter seeks to understand if brokered text is novel, and if macro structure has an impact on the novelty of information available locally. I hypothesize that the text of nodes from different communities that are connected by brokers will be novel, since information and ideas shared between groups tends to be novel compared to that shared within groups. I also hypothesize that the text of brokers will be novel compared to the rest of the nodes in their communities, because brokers have connections to outside groups and, therefore, should also have access to novel information from those outside groups. Finally, the third hypothesis states that brokered text will be most novel when the network has a polycentric structure, since such networks have many communities of different types discussing a diverse range of topics, more structural holes, and more brokers connecting disparate communities via weak ties than hierarchical networks.

Gatekeeper results

The gatekeeper results are shown in Figure 6.3. That figure shows the mean gatekeeper similarity scores for each month. The points indicate the bootstrapped mean of each relationship tested for this brokerage type for each month. The error bars indicate bootstrapped bias-corrected accelerated (BCa) confidence intervals at the 95% confidence level, since many of the distributions are skewed. The BCa method of bootstrapping confidence intervals adjusts the percentiles to correct for bias and skewness (Efron, 1987).

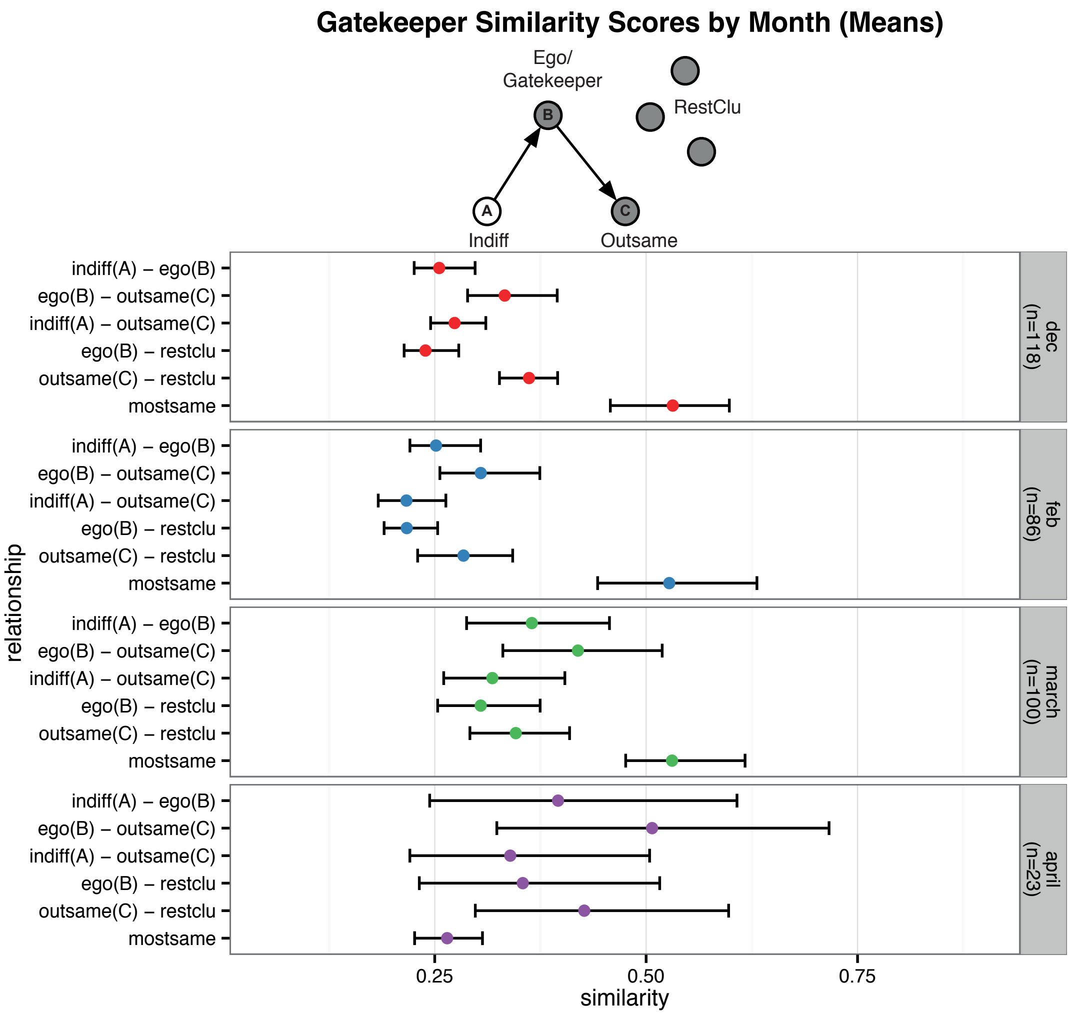


Figure 6.3: Mean gatekeeper similarity scores by month. The number of gatekeeper ego networks for each month is found on each facet label. Error bars indicate the bootstrap bias-corrected accelerated (BCa) confidence intervals at the 95% confidence level.

For the gatekeeper brokerage type, on average, the ego-restclu and the indiff-outsame relationships have the lowest similarity scores. So, the text of brokers compared to the other nodes in their communities is the most novel of any relationship tested. The next most novel relationship is for nodes that belong to different communities that are connected by the gatekeeper. These results provide support for the first and second hypotheses. On the other end of the spectrum, the most similar text on average belongs to nodes that are from the same community, have reciprocated ties, and that the highest edge weight. The text among nodes in this type of relationship, which I refer to as ‘mostsame’, is a benchmark of similarity, so this result is as expected. Looking at each time period, the December, February, and March results are consistent with these average results. The ego-restclu and indiff-outsame relationships are the most novel and the mostsame nodes have the highest similarity scores. However, it is difficult to say anything conclusive about the results from the April time period. Given that there are fewer gatekeeper relationships to test in this smaller, more hierarchical network, the confidence intervals are very wide and overlap for most of the relationships tested. This makes it difficult to say anything about this time period’s results.

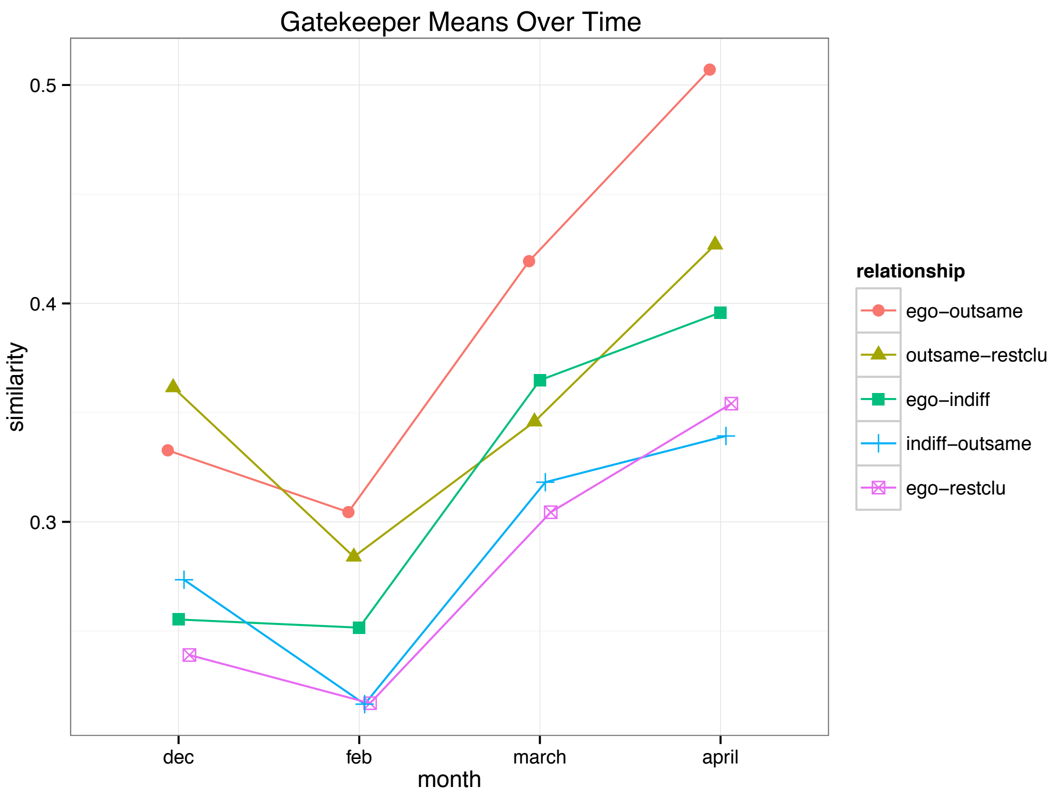


Figure 6.4: Gatekeeper similarity means over time.

Next, I investigate whether macro network structure affects the novelty of information available locally. The hypothesis (hypothesis three), is that that brokered text will be most novel when the network has a polycentric structure, since such networks have many communities of different types discussing a diverse range of topics, more structural holes, and more brokers connecting disparate communities via weak ties than hierarchical networks. Figure 6.4 shows the average similarity scores over time for node pairs that are part of the gatekeeper brokerage relationship. That figure demonstrates that similarity means for each relationship increase over time, although there is a dip in February for most relationships. For each type of relationship compared, the similarity is lower in December when the network has a polycentric structure and higher in April when the network has more of a hierarchical network structure. These findings provide general support for the third hypothesis; text similarity seems to increase in brokered relationships when the network is more hierarchical. However, as Figure 6.3 shows, the confidence intervals for each of the relationship means have overlapping confidence intervals, so it is unlikely that these changes over time are significantly different for nodes in the gatekeeper brokerage relationship.

Representative results

Figure 6.5 shows the similarity means for the representative brokerage type. The results in Figure 6.5 show that, on average, the ego-restclu relationship is the most novel of those tested. So the text of brokers compared to the nodes in their own communities is the most novel of any relationship tested, just as it was in the gatekeeper brokerage relationship. However, unlike the gatekeeper results, the ego-outdiff relationship is on average more novel than the insame-outdiff relationship, although the differences are not large, and in most months have overlapping confidence intervals.

On average, the benchmark mostsame node pairs have the most similar text of any relationship tested. As in the gatekeeper results, the ego-restclu and indiff-outdiff confidence intervals do not overlap with the mostsame confidence intervals in the December, February, and March time periods. Brokered text is indeed novel compared to the text of node pairs that are expected to have high similarity in these time periods, that is, those nodes that are in the same community, that have reciprocated ties, and that have a high edge weight relative to others. These findings for the representative brokerage type provide additional support for the first two hypotheses. As with the gatekeeper relationship, however, it is difficult to say anything conclusively about the differences among the relationships tested in April given the wide confidence intervals in that month. This is due at least in part to the smaller sample size in April and the limited number of brokers.

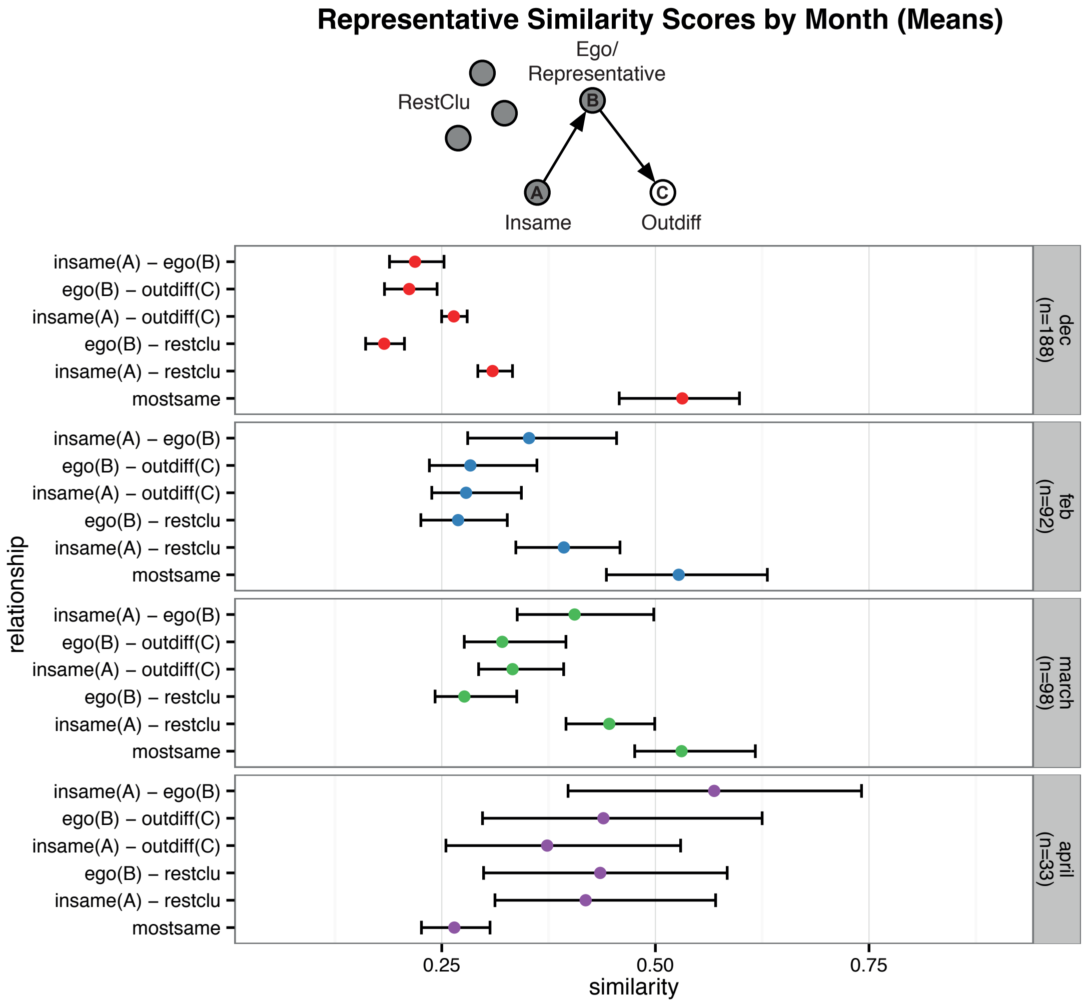


Figure 6.5: Representative similarity means by month. The number of representative ego networks for each month is found on each facet label. Error bars indicate the bootstrap bias-corrected accelerated (BCa) confidence intervals at the 95% confidence level.

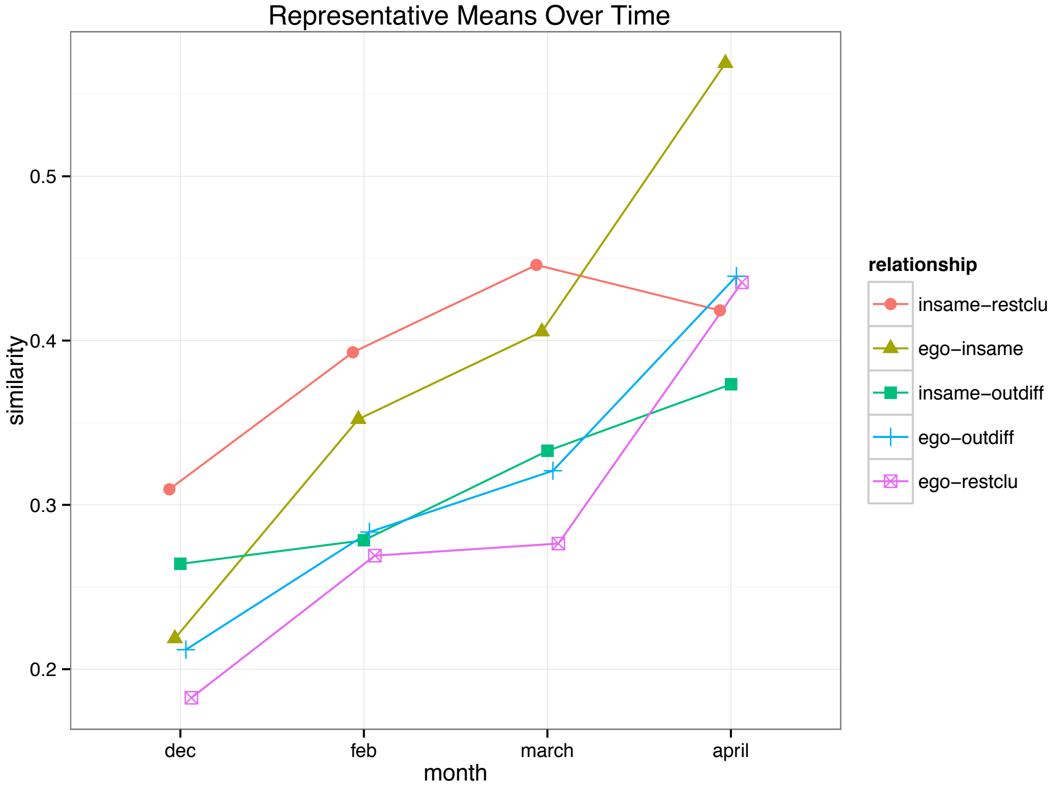


Figure 6.6: Representative similarity means over time.

The third hypothesis looks at that impact of macro network structure on the novelty of information available locally, and anticipates that text will become less novel as network structure becomes more hierarchical. In general, for representative type brokers, the similarity scores for each of the relationships tested tend to increase over time as the network becomes more hierarchical (Figure 6.6). Except for the insame-restclu relationship, which decreases slightly in April, each relationship tested increases monotonically over time. The similarity of each relationship is lowest in December when the network is most polycentric and highest in April when the network has the strongest tendencies towards a hierarchical network. The confidence intervals between the December and April time period do not overlap for the ego-restclu, ego-outdiff, and ego-insame relationships. The results from the representative brokerage type therefore provide better evidence than the gatekeeper type relationship that brokered information becomes less novel as the network becomes more hierarchical. These findings provide additional support for hypothesis three.

Comparison to coordinator and liaison brokerage types

The focus of this chapter is on the novelty of text in the gatekeeper and representative brokerage relationships. However, since novelty of information is relative, I also analyzed the novelty of text among nodes in the coordinator and liaison brokerage types. Since the nodes in the coordinator brokerage type all belong to the same community, the scores for this type of relationship should, on average, be higher than any other brokerage type tested in this chapter. And the similarity scores for the liaison type relationship should be less similar (or more novel) than the gatekeeper or representative types, since each node in the liaison relationship belongs to a different community. Figure 6.7 shows the combined average of the two-path relationship for each brokerage type. That is, the similarity scores for the broker and the two nodes it is directly connected to. The similarity of text in brokered relationships for the coordinator role is, on average, higher than others tested. However, as Figure 6.7 shows, it is not significantly higher than the representative and gatekeeper types, as was expected. Indeed, the results for the coordinator type of brokerage are in general quite similar to the results for the gatekeeper and representative types.

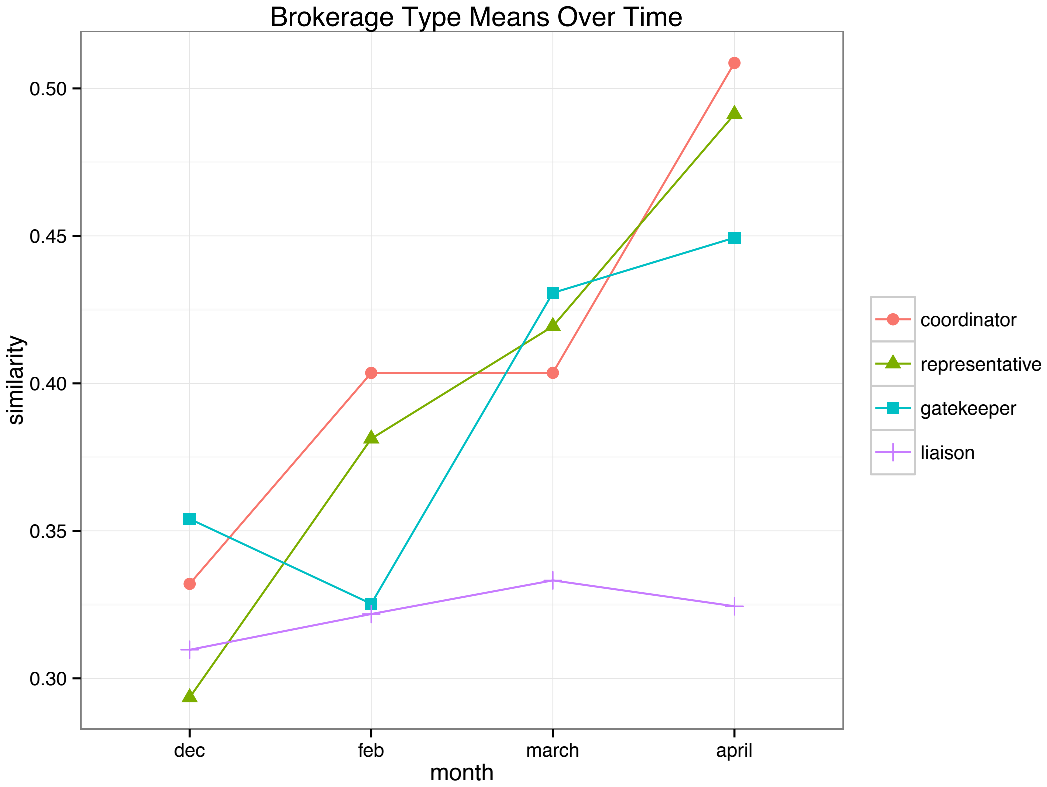


Figure 6.7: Average similarity of the two-path relationship for each brokerage type. (Excludes similarity to RestClu nodes).

The liaison brokerage type was expected to have the least similar text compared to all other brokerage types, since each node in that brokerage type belongs to a different community. There is evidence that this is indeed the case. As Figure 6.7 shows, the text among nodes in the liaison type of brokerage is, on average, the least similar of any type tested in this chapter. It is substantially lower than other types, especially in the later time periods. It is possible that this is a result of there simply being fewer liaison brokers in the later time periods, however.

Additional information on the similarity of text among nodes in the coordinator and liaison brokerage types is provided in the appendices. Specifically, Appendix C includes a chart with mean coordinator similarity scores and Appendix D includes a chart for liaison similarity scores by month. Appendix

C shows that for the coordinator brokerage type, the ego-restclu relationship has the most novel text of those tested. This result provides additional evidence for the second hypothesis, which predicts that the text of brokers will be novel compared to the rest of the nodes in their community. In general, compared to other brokerage types, the similarity of the text of nodes in the liaison brokerage relationship show less variation over time and tend to stay in a similar, low range. Having now provided evidence for or against each hypothesis, I move in the next section to concluding thoughts and ideas for future research.

Conclusion

This chapter tested if brokered text is novel, and if macro network structure has an impact on the novelty of information available at the local level. The brokerage literature rests on the assumption that brokers have access to information and ideas that are novel. Yet this assumption has almost never been tested, and, as far as I am aware, has never been tested in a protest setting in a country with an electoral authoritarian regime.

The findings provide support for the three hypotheses tested in this chapter. First, the text of the nodes that brokers connect from different communities is generally novel compared to a number of benchmarks of text similarity, especially when compared to nodes that are in the same community, that have reciprocated ties, and that have a high edge weight. Second, the findings also provide evidence that the text of brokers is novel compared to nodes in the rest of their communities with whom they are not connected. Indeed, the text of brokers compared to the rest of their communities tends to be the most novel of any relationship tested in this chapter. Third, the findings also show that macro network structure has an impact on the novelty of brokered text available at the local level. As the protest network moves from a polycentric to a hierarchical structure, text similarity increases; in other words, the novelty of information decreases. Text becomes less novel over time among all the relationships tested in the gatekeeper and representative brokerage types, where the broker is connecting nodes that belong to different communities. While there is some variation, for most relationships this increase in text similarity is generally monotonic, increasing every month as the network becomes more hierarchical. The December time period, when the network is the most polycentric, has the lowest similarity scores, while the April time period, when the network is most hierarchical, has the highest similarity scores. There are many variables that could impact the similarity of text, but the results indicate that there is an association between networks structure and brokered information. Further testing with confounding variables should be carried out in the future.

A note of caution is necessary with these findings. While the average findings hold for the December, February, and March time periods, it is difficult to say anything conclusive about the results from the April time period given the relatively small number of brokers compared to other time periods, and the relatively wide confidence intervals around the mean similarity scores for the relationships tested in that time period. Further, the nodes that are strongly connected and in the same community, and should have the most similar text actually have relatively low similarity in April. It is possible that this finding is explained by the work of Aral and Van Alstyne (2011), who argue that networks with redundant, high bandwidth ties actually provide more novel information than sparse networks with low bandwidth ties. But generally it is difficult to interpret the results for April time period given the small number of brokers in this time period.

Future research could also identify methods that can provide more reassurance that random chance has not led to the results shown here, and to help ensure their robustness. The main problem in finding an appropriate statistical method is that the independence of observations cannot be assumed since network data is used. This is a key assumption for most traditional statistical methods. I would further argue, based on the findings, that most of the observed similarity scores are likely influenced in some way by the presence, absence, or strength of a tie. Among the statistical models for networks, the problems with network autocorrelation models were already discussed in the method section. Exponential Random Graph Models or Stochastic Actor Oriented Models have as the dependent variable the existence or absence of a tie instead of a text based similarity score. It would also be impractical to run hundreds of these very complex, time-consuming models given the large number of ego networks tested in this chapter. A Quadratic Assignment Procedure (QAP) model might be a possibility, although again that would require running hundreds of models given the many ego networks in the data. Instead of a network model, the best option might be a more robust bootstrapping method, or another sort of permutation method. So, for example, take thousands of samples with replacement from each relationship of interest and calculate the similarity statistic and confidence intervals from that larger sample. Future research should also explore data sets from a variety of protest contexts and different countries, so that the results are more generalizable. That work is beyond the scope of this thesis.

This study has some additional limitations that I discuss next. The first is that I chose to use an automated method to identify communities, which based on their linking practices. Different results are possible if I had more data available on each node, for example their political affiliation or support or opposition to the protests, and used that information to assign nodes to communities. Also, individual nodes would not necessarily agree with the community assignment given to them by the community detection algorithm, so self-selection of nodes into groups would be preferable to one imposed from the outside. That is not possible with the data set or approach I use, although it could be possible for subject matter experts to assign nodes to groups based on more than just linking patterns. Another limitation of this study is that I did not attempt to identify if brokers had access to a novel piece of data first, and exclusively, from outside groups that they then shared with their own communities. Instead, I showed whether the aggregate text of nodes in brokered relationships at given time periods was distinct. This brings up a general limitation of the data and method. Traditional brokerage theory assumes a much more fine-grained data transfer process than what I was able to track in this study. This chapter is therefore just a first step in understanding the novelty of brokered information. Future research could address any of these limitations with more granular, temporal data.

This chapter empirically tested whether the information of brokers and those they connect from different communities is actually novel, which is a surprising rare test in brokerage studies. In the next chapter, I further explore the text produced and shared by different protest constituencies and the government. Specifically, I investigate the role of framing during the Russian election protests. I attempt to determine whether the government or opposition is more effective at promoting its frames online, and within the movement, whether moderate or extremist framing is more prevalent.

Chapter 7 Framing Contention: Using Supervised Learning Methods to Investigate Framing Battles During the Russian Election Protests

Framing matters. Frames serve to define social problems, diagnose their causes, and suggest remedies to correct them. Framing also helps social movements to identify targets of action, mobilize adherents, and attract new supporters (Entman, 1993; Goffman, 1974; Halfmann & Young, 2010; Snow et al., 1986). Theoretically, framing also helps to explain how emotion, grievances, and culture impact individual decisions to participate in collective action. Sociologists define frames as mental models that allow individuals to interpret events in a certain meaningful way, attribute blame, and assign responsibility (Crossley, 2002; Goffman, 1974; Opp, 2009; Snow et al., 1986). This definition implies that social movements have three core framing tasks, which are diagnostic, prognostic, and motivational (Benford & Snow, 2000). Diagnostic framing refers to the identification of the cause of social problems and the assignment of blame to those who created them. Prognostic framing entails the articulation of a plan to attack those social ills and the development of tactics to carry out that plan. Finally, motivational framing includes calls to action that mobilize both supporters and casual bystanders to participate in collective action. In the social movement literature, the result of these framing processes are known as collective action frames (Benford & Snow, 2000).

Framing is frequently contentious. Different constituencies within movements argue over which social problems to focus on, who should be held accountable for them, and which tactical responses are necessary and appropriate (Benford & Snow, 2000). Framing battles often also emerge between movements and those they are making claims against. For example, authoritarian government frequently engage in counter-framing to undermine oppositional protest movements.

This chapter investigates the framing battles that emerged during the Russian election protests, a case of high-risk collective action in a hybrid regime. Two research questions are addressed in this chapter. The first asks if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames. The second research question asks if moderate or extremist elements within a protest movement in a hybrid regime are more effective at spreading their preferred frames. The first hypothesis tested in this chapter states that the Russian protest election movement will be more effective than the government at spreading its preferred frames online, except when the government organizes large pro-government rallies. The second hypothesis states that moderates will be more successful than extremists in the Russian election protest movement at spreading their frames, unless violence occurs at protest events.

This chapter begins with a discussion of framing during the Russian election protests, followed by a short description of the various oppositional constituencies that participated in the Russian election protests. I then discuss the supervised learning method used to determine what proportion of text each week is similar to the government or each constituent part of the protest movement, and whether moderate or extremist frames are more popular. The process used to create the training set and the selection of the categories is described next. Qualitative descriptions of the documents in each category are also provided. The validation method and results are then discussed. Results are then provided and tied back to the hypotheses. The conclusion summarizes the findings and suggests possible paths for future research.

Framing during the Russian protests

The Russian election protests provide a good example of framing battles in a high-risk political setting. In the Russia case, negative framing by the government of the protest movement and its leaders was an overt strategy aimed at undermining the protests (Koesel & Bunce, 2012). The government used a mix of negative frames to describe election protesters and the movement. This included highlighting extremist, nationalist elements and slogans in the protest movement instead of moderates. The government also framed the protesters as a minority, and as part of a spoiled and poorly behaved, cosmopolitan 'creative class' in major cities. The government also employed anti-Western rhetoric, claiming that the protests were orchestrated and paid for by the United States, and that student protesters were like cattle herded around by their paid masters. Protest leaders and participants were also called traitors like Judas Iscariot who were part of a fifth column bent on U.S.-supported regime change. Anti-Putin protesters were also described as unpatriotic and anti-Russian, preferring Western European values and rejecting Russian culture and tradition (Koesel & Bunce, 2012; Smyth et al., 2013; Volkov, 2015).

Tactically, the government and its proxies also orchestrated a series of pro-government, 'Anti-Orange' rallies that also coopted the opposition's demand for 'honest elections.' The rallies were labeled 'Anti-Orange' because organizers believed the opposition protesters were trying to replicate the 2004 Ukrainian Orange Revolution, which Putin and his supporters argue was orchestrated by the United States. The implication was that such a revolution in Russia would lead to economic decline, political instability, and civil war (Koesel & Bunce, 2012). These rallies were also intended to demonstrate overwhelming support for Putin, especially among 'average Russians.' Pro-government events also took place at symbolically important sites, such as Manezh Square in the shadow of the Kremlin walls, and at Reverence Hill (Poklonnaya Gori). Opposition protesters, on the other hand, were denied permits for events at sites such as Red Square and Revolution Square, and instead forced to protest at places like Swampy Square (Bolotnaya Ploshad), on an island in the Moscow River. Some government rallies also took place on symbolically important dates that emphasized Russian war victories and sacrifice, such as the Day of the Defenders of the Motherland (Koesel & Bunce, 2012; Smyth et al., 2013).

Framing of the government and Putin was also a strategic imperative for the protest movement, both before and during the protests. One of the best examples of successful, bottom-up, oppositional framing included blogger and protest leader Alexei Navalny's framing of Putin's United Russia party as a 'party of crooks and thieves,' and his call for Russians to vote for 'anyone but United Russia.' In a mark of the success of this campaign, an April 2013 poll found that a majority of Russians (51%) agreed with the statement that United Russia is a party of crooks and thieves (Levada Center, 2013). Some Russia area scholars argue that framing the government in this way, as self-serving and corrupt, resulted in a unifying theme for the disparate groups that made up the protest movement (Robertson, 2013). Corruption and mismanagement was also an undercurrent in many of the civic protests that occurred before the elections (Greene, 2013; Koesel & Bunce, 2012). The large, peaceful, and orderly protests early in the protest cycle also helped to undercut the government's framing that the protests would lead to chaos and violence. Further, the protest movement had an additional advantage early on in the protests since the government was largely silent until after the first major protests took place.

Protest constituencies

The Russian protest movement consisted of a wide range of oppositional groups and individuals, some of which made for strange bedfellows. More overtly political groups spanned the ideological spectrum, and included communists, liberal democrats, far-right nationalists, and even anarchists. Actors not aligned with formal political parties included independent journalists, writers, musicians, and activists associated with issue-specific causes. Leaders of issue-specific civic groups played an important role in the organization of the election protests, including anti-corruption activist Alexei Navalny, environmentalist Evgenia Chirikova, and drivers' rights leader Sergei Kanaev (Greene, 2013). The movement drew heavily from the non-systemic opposition in Russia, which is formally or informally prevented from competing in elections (Smyth, 2013; Taylor, 2011).

Among the more traditional democratic parties, the umbrella political group Solidarity played a key role in the logistics and organization of protests. Solidarity represents individuals and parties affiliated with the democratic opposition, such as RPR-PARNAS and Yabloko. According to Levada Center polling conducted at several protest events, 60 to 70% of protest participants identified themselves as democrats or liberals (Volkov, 2015). This mix of independent civic groups, parties, journalists, writers, and artists form what I call the mainstream core of the protest movement.

More extreme elements in the movement included ultra-nationalists and far-leftists. On the far left, Sergei Udaltsov played a significant organizational role in protests. Udaltsov leads the radical socialist group Left Front and is a self-proclaimed revolutionary and calls for a social-democratic revolution in Russia. On the far right, Russian ultra-nationalists that participated in the protests include Eduard Limonov, a leader in the Strategy 31 movement and founder of the banned National Bolshevik Party, Vladlen Kralin (a.k.a. Vladimir Tor) of the Movement Against Illegal Immigration (DPNI), Ilya Lazarenko, former head of the fascist party National Front, Dmitry Dyomushkin of Slavic Union, and Konstantin Krylov of the Nationalist Democratic Party (Umland, 2012). These individuals and groups make up the core of the extremist constituency in the protests movement.

The systemic opposition played a more limited role in the protests. The systemic opposition includes the three 'opposition' parties besides United Russia that are allowed to compete in elections and hold seats in the Duma. They are the Communist Party of the Russian Federation (KPRF), Just Cause (Spravedlivaya Rossiya), and the Liberal Democratic Party of Russia (LDPR). The

Liberal Democratic Party of Russia (LDPR) remained a steadfast supporter of Putin throughout the protests, while The Communist Party of the Russian Federation (KPRF) and Just Cause initially backed the protest movement (Sakwa, 2015). Additional details on the different protest constituencies are available in Chapter 3.

In this chapter, I focus the empirical analysis on five different elements of the protest movement: the Mainstream, Environmental, Extremist, LGBT, and KPRF constituencies. Texts from Web sites and blogs associated with each constituency, as well as pro-government supporters, are used to identify the framing employed by each group. This chapter relies on textual data generated by 3,583 Web sites that discussed the Russian election protests or election falsification between December 1, 2011 and May 31, 2012. The data collection process and possible limitations are described in detail in the data and methods chapter (Chapter 4). With the constituencies described, I move next to a discussion of the method used to track the frames used by each of these groups over the course of the protests.

Method

The increased adoption of the Internet and social media has led to an opportunity and challenge for collective action scholars. This shift has led to a tremendous amount of text, images, and links from protest participants, leaders, and traditional media. New online data sources also allow academics to track government reaction through official channels such as government Web sites and state-controlled media, as well as through blogs and social media accounts of government supporters and elected officials. In the past, scholars could manually code small sets of documents drawn from speeches, official literature, and press clippings to measure the reach of social movement frames. Today, however, computer-assisted methods are required to sift through the massive amounts of protest data now available online.

Automated text analysis approaches that can assist in this task fall into two general categories, supervised and unsupervised methods. Unsupervised methods, such as Latent Dirichlet Allocation (LDA), aim to automatically identify topics that exist in a set of texts based on the co-occurrence of words in documents that are then analyzed by a researcher to determine what broader topics those clusters of words represent (Blei, Ng, & Jordan, 2003). Such methods can be useful when researchers do not have a sense of the topics in their data set, and for exploratory research. Supervised methods, on the other hand, rely on a set of hand coded documents that are used to train a 'machine', or algorithm, which then determines what proportion of the text in the larger corpus matches that in the training set. Most methods use a 'bag of words' approach, meaning that syntax plays no role in the analysis and that the text in documents goes through a natural language pipeline. This process cleans and preprocesses text, reduces words to common stems, and creates a document term matrix with documents as rows of the matrix and word stems as the columns (Grimmer & Stewart, 2013).

In this chapter, I use a nonparametric supervised learning method created by Hopkins and King (2010) and implemented through the 'ReadMe' package in R (Hopkins, King, Knowles, & Melendez, 2013). Like other supervised learning methods, such as naïve Bayes, the goal of this method is classification of text in the larger corpus based on theoretically relevant categories established by researchers beforehand. A training set is built with documents that represent text used by each of these predefined categories. The Hopkins and King method takes a relatively small set of hand coded documents and provides approximately unbiased and consistent estimates of the proportion of all documents in each category chosen by the researcher. Hopkins and King refer this to as 'document category proportions.' Most supervised methods classify individual documents into categories, assess the level of accuracy of that classification process and further tune the machine for acceptable levels of true and false positives (precision and recall), and then aggregate individual classifications across the entire corpus. The Hopkins and King method skips this intermediate classification and aggregation step. Instead, it directly estimates document category proportions in the corpus based on the word profiles for each category in the training set (Hopkins & King, 2010).

I use this method to test the two main hypotheses in this chapter. The first hypothesis states that the Russian protest election movement will be more effective than the government at spreading its preferred frames online, except when the government organizes large pro-government rallies. The second hypothesis states that moderates will be more successful than extremists in the Russian election protest movement at spreading their frames, unless violence occurs at protest events. To test these hypotheses, I create a training set with documents that represent the framing of the protests by different opposition protest constituencies, as well as a category representing pro-government frames, and estimate the proportion of text accounted for by each category in the larger text corpus.

This method can be broken down into three primary tasks. The first step is text preparation, cleaning, and normalization. Second, a small set of documents is hand coded to create a training set of between 100 and 500 documents that are used to train the machine. Finally, the ReadMe software estimates the document category proportions in the entire text corpus. In this chapter, I break the text corpus into weeks in order to determine weekly document category proportions for each category over the six months of protests (Hopkins & King, 2010).

Text cleaning and data preparation

The text cleaning and data preparation phase described here is similar across nearly all supervised and unsupervised methods. First, a set of individual text documents of interest are collected and put into a directory on a computer for retrieval and manipulation. All words are then converted to lowercase, punctuation is removed, and the text is then normalized through stemming, which reduces words to a common root or stem. So, consist, consistency, consisted, consistent, consistently and consisting are reduced to their stem, consist. Numbers are also frequently removed, but in this analysis, because numbers played an important part in the framing of protest success or failure (did a protest have 10,000 or 200,000 participants, did United Russia received 34 or 54 percent of the vote) I did not remove numbers from the text corpus. The processed and normalized text is then converted into a document term matrix where rows are documents and columns are word stems. Matrix cell values are dichotomous, either one or zero, indicating the presence or absence of the stem in the document, although it is possible to use a weighted matrix where cell values are counts of the frequency of a term in documents. Document term matrices tend to be very sparse, with most cells in the matrix having a value of zero. To reduce the sparseness of the matrix, stemmed unigrams appearing in fewer than 1% or greater than 99% of all documents are deleted (Grimmer & Stewart, 2013; Hopkins & King, 2010).

My textual data required additional cleaning and preparation. During the initial data exploration phase, which included manually reading a random subset of stories and analysis of term frequencies during the most active phases of the protests, it became clear that additional text cleaning was required. This is primarily because the Media Cloud extractor, which extracts substantive document text from other html and navigation text, advertisements, comments and other cruft, had failed to remove some of this unwanted text in the documents. I used a series of regular expressions to identify sentences that were clearly not related to the substance of documents and deleted those problematic sentences from the text corpus.

The use of Russian language data also required extra text cleaning and preparation. The ReadMe text preparation function, in particular stemming, is designed only for English text. Instead of relying on the ReadMe package for data preprocessing, I did almost all text preparation in the TM (short for Term Matrix) package in R, which is the standard R package used to prepare text for automated text analysis (Feinerer & Hornik, 2008, 2015). Similar to the standard ReadMe text processing approach, I used the TM package to convert all text to lower case, remove extra white space, remove punctuation, and remove Russian stop words (common words which occur frequently and provide little to no value in automated text analysis). I then used the SnowballC package to stem the Russian text, which relies on the popular Porter stemmer but adapts the algorithm for Russian text. It is possible that lemmatization would work better than stemming for Russian text, and this could be attempted in future research. In lemmatization, each part of speech is identified and the actual dictionary root or lemma for each term is used for normalization. The Porter stemmer uses a more algorithmic approach, iteratively removing word endings and reducing terms to their stems. These stems are recognized by a computer but are not always recognizable to a human. Lemmatization can help text normalization in languages that have many different cases and endings that change for different parts of speech. However, stemming is usually sufficient in most languages, especially English. Finally, to ensure that the data was lined up properly for the remaining ReadMe functions I modified the undergrad function, which does the preprocessing of text, to call a modified Python3 script directly from R to create the document term matrix and remove words used in less than 1% or more than 99% of documents. With some additional work, the text cleaning and data preparation could be done entirely in R.

Creation of the training set

Next, as in almost all supervised methods, a training set is created in which the researcher places a number of documents into categories of theoretical interest. Hopkins and King have found that with their method as few as 100 documents can be used as the training set, and that diminishing marginal returns begin to set in for training sets larger than 500 documents (Hopkins & King, 2010).

Unlike other supervised learning methods, the hand coded set does not need to be randomly drawn from the larger corpus.

The main assumption for the Hopkins and King method is that the documents in the hand coded training set contain sufficiently good examples of the language used in each category in the larger corpus. In other words, the word profiles for a given category in the training set should be similar to those in the corpus (Hopkins & King, 2010). This can be an issue if one is tracking an issue over a long period of time and the way it is discussed changes or 'drifts.' To guard against conversational drift impacting estimates of document category proportions, I drew training data from across the entire six months of protests. Additional assumptions for the Hopkins and King method are that researchers must be able to place documents into well-defined categories. These categories should be mutually exclusive, exhaustive, and relatively homogenous. Finally, a sufficient number of documents need to be included in each category for a reliable word profile to be created for that document, although Hopkins and King do not state what a sufficient number of documents might be (Hopkins & King, 2010). Once a hand-coded training set is compiled, cross-validation is used to determine how well the machine is able to estimates actual, know values. The cross-validation method and results are discussed at the end of this section.

I now describe the process used to create the training set used in this analysis. For analyzing political texts it is common to use metadata or other knowledge of the source of the text to indicate political slant or framing. For example, conservative or liberal text could be identified based on whether the source of a given text is from a conservative or liberal newspaper's op-ed page, party affiliation of U.S. congressional speeches, or the text of political blogs where social network analysis or other methods can be used to divine the political position of the speaker (Diermeier, Godbout, Yu, & Kaufmann, 2012). I use such an approach to identify texts that represent the framing of different constituencies in the Russian protest movement, as well as texts that are good examples of pro-government counter framing. This is justified because I argue that instead of asking coders to subjectively guess whether a text reflects the framing of a certain constituency, it is better to draw text directly from the leaders of those constituencies, Web sites of organizations that represent different protest groups, and those that frequently link to those sites.

I first identified what I call media sources in the protest data set (e.g., blogs, organizational Web sites, social media pages, or traditional media sources) that are clearly affiliated with different constituencies in the protest movement. This varied somewhat with each constituency based on the type of online content available in the protest data, but I attempted to include documents from different types of media sources to prevent text from one type of media source (for example, blogs compared to TV news transcripts) from potentially skewing the results. I extracted documents from across the entire six months of protests for each of these varied media sources. I also read each document to ensure that the language used in a particular document was indeed a good example of the framing used by a given constituency, and to also ensure that the assumptions of the Hopkins and King supervised learning method were met.

Documents were coded into eight categories, which were introduced and justified earlier: mainstream, extremist, environmental, LGBT, KPRF (Communist Party of the Russian Federation), pro-government, neutral, and off topic. These categories are mutually exclusive. The neutral and off topic categories are included as catchall categories to ensure that the category list is also exhaustive. This process resulted in a training set of 504 documents. I also created a Multi-Dimensional Scaling plot with the text from documents in the training set grouped by category to ensure that the categories were indeed different (Figure 7.1). A classical MDS algorithm was used to calculate the distances between the categories based on eigenvalues. These eigenvalues are then used as coordinates in a two-dimensional scatter plot. Categories with similar text will cluster closer together on the plot, while those with dissimilar text will be found further from each other. Figure 7.1 shows that the environmental, LGBT, and neutral categories are the most similar, while the KPRF and off topic category are the most different compared to other categories. The similarity of the environmental, LGBT, and neutral categories may be due to the smaller number of documents in the training set for each of those categories.

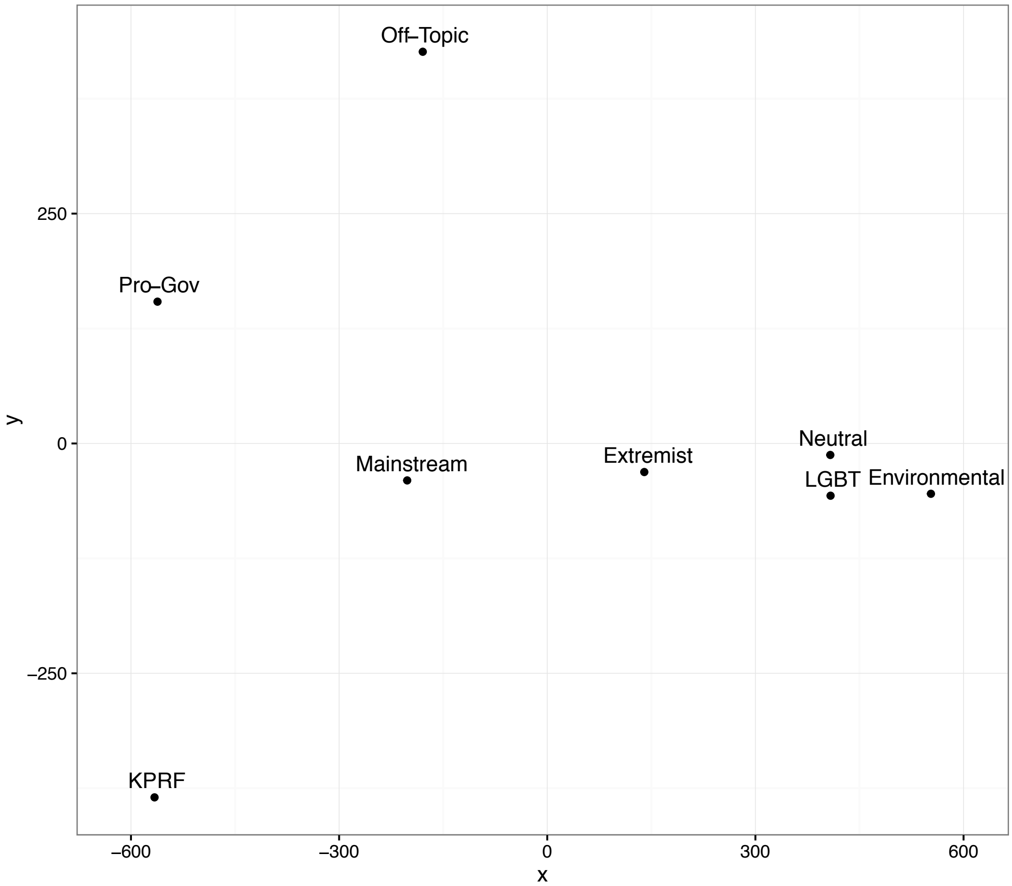


Figure 7.1: Multidimensional scaling plot of category text in the training set.

Below I provide more details on the types of sites that were used to identify texts for each category and qualitative descriptions of the content that emerged in each category. Appendix E also includes a translated word cloud for each substantive category weighted by term frequency-inverse document frequency (TF-IDF). TF-IDF identifies important words by comparing the frequency of a word within a document to its frequency within the entire corpus. Specifically, TF-IDF assigns the most weight to terms that occur many times in a small number of documents, less weight to terms that occur less frequently in a document or that occur in many documents, and the lowest weight to terms that occur in virtually all documents (Manning et al., 2008).

1. Mainstream

This category represents the framing of the major independent civic groups, democratic parties, independent journalists, writers, and artists that played leading roles in the organization of the protests. As noted in the protest constituencies' section of this chapter, I argue that this mix of actors and organizations formed the mainstream core of the protest movement. Some of the sites that documents were drawn from include blogs and Web sites affiliated with mainstream protest leaders such as Alexei Navalny, Boris Nemtsov, and Ilya Yashin, democratic political groups and parties such as Solidarity and Yabloko, and election monitoring sites such as Golos. Documents were also drawn from media sites with sympathetic coverage of the protests including TV Rain, Grani.ru, Echo of Moscow, Kasparov.ru, and independent journalists such as Oleg Kashin.

Documents in this category include discussion of preparations for election monitoring and get out the vote efforts. During the run up to Duma election, Russians are encouraged to vote for anyone but United Russia. Documents in this category include evidence of election fraud including videos, testimonials, and statistical analysis showing it was impossible for United Russia to win 54% of the vote. Texts also include the main demands of the protest movement: New elections must be held; Central Election Commission chairman Churov must step down and be replaced by a truly independent election commissioner (one of the heads of Yandex is mentioned as a worthy candidate), all parties must be allowed to participate, political reforms must be made, and peaceful protests must be allowed. Documents in this category include clear evidence of protest mobilization and organization. Around the time of the Presidential election, discussion pivots to efforts aimed at preventing Putin's reelection and demands for new presidential election laws. Protest leaders in this category emphasize that a broad coalition supports the movement, from leftists to nationalists. Documents also emphasize the unity of the movement and counter government 'propaganda' that says there are splits in the movement. Texts in the mainstream category also emphasize the large size and success of the protests. General movement strategy is also discussed as well as plans for what to do next after each major protest event. There is also criticism of the government strategy of paying participants to attend pro-government rallies. Texts in this category include frequent references to the campaign against the party of crooks and thieves, which is how Navalny and other protest leaders framed United Russia.

2. Environmental

Data for the environmental category was drawn from the blog of protest leader and environmental activist Evgenia Chirikova, sites affiliated with environmental organizations such as the Ecological Defense of the Moscow Oblast, and blogs and Web sites that link to them. Given that it was difficult to find documents for this category, I also used a number of regular expressions to identify documents that included text patterns indicative of discussion of ecological issues, and also reviewed blogs identified in previous research as frequently linking to sites focused on environmental and social activism (Etling et al., 2010).

Documents in this category include discussion of how election falsification is linked to the broader problem of corruption in Russia. Participation in the protests is also seen as part of the longer history of protest and civic actions in defense of the environment in Russia, especially efforts to save the Khimki forest (a campaign to stop construction of a highway between Moscow and St. Petersburg through an old growth forest in Khimki, a suburb of Moscow). Texts also include evidence of protest mobilization and organization specifically among environmentalists. For example, a Zelonaya Kolona (green column) protest was organized as part of larger protest at the first major oppositional protest at Bolotnaya Square. There is an emphasis on support for Chirikova in her role as a leader of both the environmental movement and the election protests. Documents also show support for Alexei Navalny and others leaders who were arrested as part of the protests. They also use Navalny's party of crooks and thieves framing to describe United Russia, writing, for example, "Eco Defenders against PZHiV" (PZHiV is a transliterated acronym for party of crooks and thieves – Partiya Zhulikov i Vorov).

3. LGBT

Data for this category was drawn from gay rights leaders and gay rights organizations in Russia, such as Nikolai Alexayev, gayrussia.org, and blogs and Web sites that frequently link to such sites.

Authors of documents in this category discuss how LGBT votes were stolen in the falsified elections, just like many other Russians, and that an appropriate response was to participate in the protest movement. There is evidence of an internal debate over participation in the protests alongside nationalists, who are typically referred to as fascists, and concerns over personal safety because of possible attacks by nationalists. Authors of documents in this category note that they do not agree with nationalists, but are coming together as part of a common cause for free and fair elections. Writers also argue that liberals who lead the election protest movement have typically supported the LGBT movement, while the government that stole the election has consistently attacked and made scapegoats of the LGBT community. Documents in this category note that the LGBT community in Russia sees participation in the election protests as a continuation of a long history of gay rights marches and activism. Participation in the protests is also seen as a way to publicize the movement, and to show that members of the LGBT community are Russian citizens, just like everyone else upset about election fraud, who have a right to have their votes counted and their voices heard. There is also internal debate about whether the LGBT community and formal organizations advocating for gay rights in Russia should be involved in broader political debates at all, and if they should instead focus on issues of specific concern to the LGBT community. Like the environmentalists, they also adopt Navalny's framing of United Russia, writing LGBT against PZHiV. Later in the protests, there is also discussion about how the movement's representatives are not given prominent speaking roles in some cities, and concerns that mainstream protest leaders are sidelining the LGBT community.

4. Extremist (Nationalists and Far Left)

Data for this category comes from prominent Russian nationalists and far left activists, including Vladimir Tor, Konstantin Krylov, Eduard Limonov, and Sergei Udaltsov. Texts were also drawn from Web sites of organizations these activists were affiliated with at the time of the protests, such as the Movement Against

Illegal Immigration (DPNI) and Left Front. Generally, this category has more texts connected to ultranationalists than far leftists.

Documents in this category reveal agreement among nationalists and nationalist organizations that they should participate in the larger protest movement for free and fair elections, even if they disagree with the political positions of many of the other groups involved. They argue that elections cannot be fair if nationalists are not allowed to participate. There are calls to participate in the elections and for nationalists to throw their support behind KPRF or the Liberal Democratic Party of Russia (LDPR). There is discussion about the need to show their presence at the protests with nationalist flags and symbols. Throughout the protest cycle, but especially at later protest events, there are complaints that nationalist speakers are not allocated enough time on the podium. There are also concerns that nationalist leaders do not have enough influence in the protest movement's decision-making process. There is evidence that some leaders want to break from the larger protest movement and also calls to organize separate rallies at more symbolic sites like Revolution Square instead of Bolotnaya. There are frequent demands to free those arrested during election protests as well as nationalist and far left protests held earlier in the year. Documents in this category include calls for changes to laws against extremism that are used to arrest nationalist leaders and prevent their participation as candidates in elections. It is noted that many nationalists helping to organize the election protests have also been key organizers of the annual 'Russian March.' There is significant criticism of Putin and United Russia. There are also demands that pro-American democrats and liberals should be prevented from dominating the protests and the organizing committee. There are calls to free Sergei

Udaltsov from jail, ways to support him, and writers in this group closely follow his court cases and confrontations with the government.

5. Communist Party of the Russian Federation (KPRF)

I initially did not include a category for the KPRF, since scholars argued that the party played a marginal role in the protests and expressed concerns about the potential for an Orange Revolution in Russia (March, 2012; Sakwa, 2015). However, after an initial review of a random set of documents in the protest data, it became clear that many documents that used framing sympathetic to the communists existed in the data, that the KPRF saw the elections as fraudulent, and that they supported the protests. Data for this category comes from KPRF-affiliated Web sites and blogs, including regional party affiliates, and online communities that support the communists.

Documents in this category include discussion of how the elections were not free, fair, or democratic. Writers argue that there is no way that United Russia won more than 30% of the vote in the Duma elections. Authors in this category also note that the KPRF had election observers in every region of the country, and that many found serious election violations. There is also generally more discussion of regional protests among the communists than other oppositional categories. Texts in this category also attack Putin and others in the government for verbal attacks and insults slung at the protesters. In reaction to the government framing of the protests as supporting an 'Orange Revolution' in Russia, documents in this category highlight that communists do not want an Orange Revolution and that they do not agree with liberals politically, but they have decided to join the protest because they want to be able to compete in free and fair elections. Authors of documents in this category also argue that communists would win if elections were free and fair. There is discussion defending communist agitation outside the pro-government, Anti-Orange protests in February at Poklonnaya Gori. This is justified because the party believes that many participants at the Anti-Orange rallies were likely past supporters of KPRF, and could be supporters again in the future. Texts in this category indicate that it is obvious that the pro-government protesters at Poklonnaya were bused in and paid to participate, and that the Poklonnaya protests show that falsification and fraud remain weapons of the government. They write that just because pro-government youth groups such as Nashi pay protesters to participate in pro-government demonstrations, this does not mean that the opposition does so. Compared to other categories, there is more reaction to TV programs that put KPRF in negative light, complaints about government control of TV, and complaints that not enough time is given to Gennady Zhuganov (the long time leader of the party) during the presidential campaign. Texts in this category argue that no other leader in the opposition or in the protest movement is strong enough to actually defeat Putin and United Russia, except the KPRF and Zhuganov. There are also calls to support Zhuganov in presidential election. The texts in this category are unique in that they do not appear to use party of crooks and thieves language to describe United Russia, while nearly every other oppositional category does.

6. Pro-government

Pro-government data was drawn from conservative, pro-government media sites including Vzglad, Izvestia, Komsomolskaya Pravda, Channel 1, and government

Web sites. Texts from blogs, including blogs of leaders of the pro-government youth group Nashi, were also included in the training set for this category. However, this category does include more documents from traditional media sites than others categories.

Documents in this category frequently criticize the election protests and protest leaders. The size of anti-government protest events is frequently questioned. Generally, documents in this category emphasize the lower limit of participants attending anti-government protests and the upper limit for pro-government protests. This category includes texts generally supportive of Putin and his leadership, especially in the run-up to the Presidential elections. Documents in this category frequently highlight disagreements and splits among protest leaders and different protest groups. Texts also frequently warn of violence that is likely to occur at election protests, then credit the police for efficient organization and professionalism when violence does not occur. One government official pleads that protesters should not allow themselves to be manipulated by 'provocateurs' or to become 'cannon fodder.' Pro-government framing is also frequently anti-American. Texts in this category argue that protesters and protest leaders are supported and even organized by the United States. Writers in this category argue that the United States is interfering in the internal affairs of Russia and that U.S. Secretary of State Hillary Clinton 'sent a signal' that was picked up by protest leaders. They note that Clinton claimed the Duma (legislative) elections were falsified before the final results were even tabulated. Texts in this category include discussion of funding documents and other evidence showing that the National Endowment for Democracy (NED) and the U.S. Agency for International Development (USAID) fund protest groups and individual protest leaders, including the election monitoring group GOLOS. Pro-government framing also includes support for the pro-government, Anti-Orange protests and emphasizes that, according to official estimates, well over 200,000 protesters participated in the Anti-Orange protests. Pro-government framing also frequently focuses on extremists as leaders of the protest movement instead of other opposition leaders. Alexei Navalny is rarely mentioned by name or quoted as a protest leader, while the 'extremist' leader Sergei Udaltsov nearly always is. Texts in this category frequently use language about fascists, extreme nationalists, and revolution, and argue that the protest will lead to the destruction of Russia. The view is also expressed that the protests will not lead to change, and that protesters are just expressing the 'trendy' view that rich Muscovites and 'hipsters' are unhappy. Finally, it is argued that no real politicians or actual political leaders are taking part in protests, and that protesters are not serious about governing.

7. Neutral

This category includes content that did not express an opinion or reveal any obvious framing or slant that would easily align it with either a protest constituency or the government. This was often factual reporting of election results or protest events without opinion, but this was fairly rare. This category was required to ensure that the list of categories was exhaustive. Given that neutral content is relatively rare in the data, future research could collapse neutral content into the off topic category.

8. Off topic

This category includes documents that are on topics totally unrelated to the protests. Documents in this category are also about election falsification and protests in other parts of the world, 'mixed content' where a sentence or two may be about the protests but the majority of the document is on a totally different topic, and text related to advertisements and navigation that did not get removed during the data cleaning phase. This category also includes a set of documents related to election fraud that took place in Russia before the period of interest – especially in Chechnya – which I chose to place in this category as it was not the motivation for the large election protests following the Duma election.

Estimation of document category proportions

Estimation is done through the readme function after invariant columns in the document term matrix are removed. In the data preparation step discussed above, a document term matrix was created with documents as rows and word stems as columns. For each document in this matrix there is now a row with a one in the cell if a word exists in the document and a zero if not. Hopkins and King call this a 'word stem profile' because it provides a summary of all the word stems in a document. For documents in the training set there is now also a word stem profile for each category based on the aggregated word stem profiles in the training set. An assumption of the method is that the word stem profile for each category in the training set is similar to that in the corpus. This is a critical assumption because the word stem profile for each category is used to directly tabulate the document category proportions. As Hopkins and King (2010) write "Among all documents in a given category, the prevalence of particular word profiles in the labeled set should be the same in expectation as in the population set." (p. 237) The documents do not need to have the same text, but the documents in the test and training sets for a category do need to have the same probability of using similar word profiles. So, for example, the language that Russian bloggers use to criticize Putin in the training set must be at least a subset of the way he is talked about in the test set (Hopkins & King, 2010).

The ReadMe software estimates the document category proportions by tabulating what proportion of documents in the test set have word profiles similar to those in the hand-coded training set for each category. As it applies to this chapter, the argument is that a document category represents a blogger's framing, which they have beforehand and set out in words when writing their blog post. This blogger's framing is summarized in a word profile for that given blog post and can be used as a variable in statistical analysis. Hopkins and King justify this approach because a blogger does not start writing a post and then discover that they have an opinion about a candidate for office. Instead, they have an opinion, which is represented in their word profile for a given document. The words chosen by a blogger are therefore by definition a function of their framing, meaning their word profile can be used to represent a category (Hopkins & King, 2010).

Statistically, document category proportions are tabulated through ordinary least squares regression using the standard regression formula $Y = X\beta$ (with no error term) where the population document category proportions are the unknown regression coefficients β, the probability of word profiles occurring within a given document category are the explanatory variables matrix X, and the probability of each possible word profile occurring is the dependent variable Y. The software then solves for our quantity of interest, the document category proportions β, via the usual regression calculation: $\beta = (X^T X)^{-1} X^T y$. However, because the number of possible word profile combinations is so large, subsets of between 5 and 25 word stems, or features, are selected (the default is 16). Three hundred random draws of 16 features are taken and the results from each draw are averaged. Cross validation can be used to test whether larger numbers of features or a larger number of random draws improve how well the software estimates true versus estimated document category proportions. I experimented with sets of 16 to 20 word stems and larger random draws, but this did not noticeably improve results. In practice, setting the number of subsets too high has been found to be inefficient (Hopkins & King, 2010).

One benefit of this method is that its assumptions are much less restrictive than other supervised learning approaches. As long as the method's key assumptions (discussed above) are met, the training set does not need to be randomly drawn from the larger corpus of documents. Also, the distribution of word profiles across documents and the distribution of documents across categories can be completely different in the training and test sets. So, no bias is introduced if the training set has many more documents in a given category than in the corpus, or vice versa (Hopkins & King, 2010). The major limitation of the method, for some research questions, is that it does not classify individual documents. This is not a concern here since my research questions and hypotheses relate to document category proportions and not individual documents.

Validation

Cross-validation is necessary with machine learning methods to ensure that the supervised methods can reliably replicate human coding (Grimmer & Stewart, 2013). Cross-validation is a process in which a small subset of data where the true or actual outcome is known is removed and the model is fit to understand how well the algorithm predicts the known values (Hanna, 2013). As recommended by Hopkins and King (2010), I randomly split the training set in two, treating half of the documents as the training set and the rest as the test set. The estimated proportions generated by the software are then plotted vertically against the known true category proportions on the horizontal axis (Figure 7.2). Estimates are colored circles in Figure 7.2. Circles that are closer to the 45-degree line are more accurate. Vertical lines represent 95% confidence intervals and are generated by standard bootstrapping techniques. The results of the cross-validation show that all estimates are near the 45-degree line, and that all categories except for one (LGBT) have confidence intervals that cross the 45-degree line (representing no difference between estimates and the truth). These results indicate approximately accurate estimates with no obvious systematic estimation errors, although it would be preferred if confidence intervals for all categories crossed the 45-degree line. This cross-validation test used a training set of 504 documents. This training set had the best results of three that were built and tested, including tests with smaller training sets of 414 and 485 documents. Therefore, the training set with 504 documents is used to estimate and report category proportions in the rest of this chapter (Grimmer & Stewart, 2013; Hopkins & King, 2010; King & Lu, 2008).

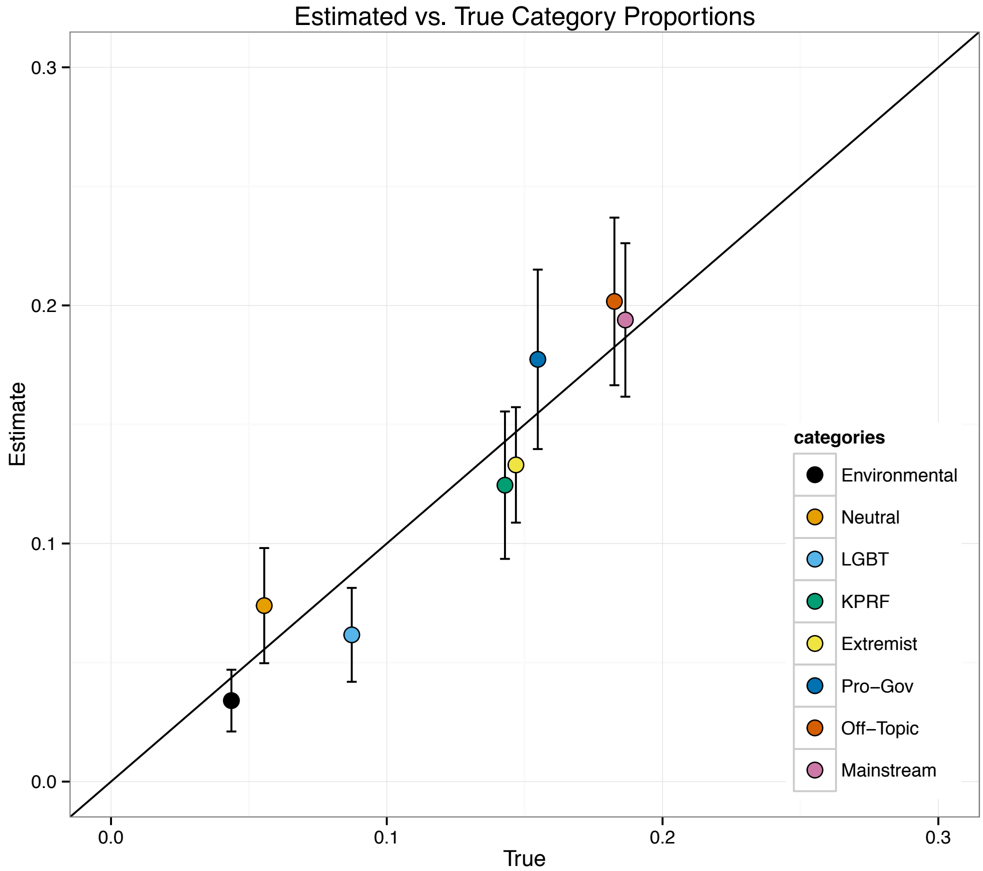


Figure 7.2: Estimated vs. True Category Proportions. This cross-validation plot shows software estimates of document category proportions against actual category proportions. Colored circles are estimates. Estimates closer to the 45-degree line are more accurate. Vertical lines represent 95% confidence intervals generated via standard bootstrapping procedures.

As mentioned above, cross-validation was also used to test whether using larger subsets of words during estimation improved the results beyond the default 16. However, increasing the number of features and number of random draws did not noticeably improve the results. Additional cross-validation tests could be conducted with different training sets, or changes to the underlying data (for example, removing numbers during the text preparation step or using lemmatization instead of word stemming). However, cross-validation is time intensive due to bootstrapping, and can take many hours per bootstrapping run. Tweaks made to the underlying data so far have not substantively improved cross-validation results or resulted in document category proportion estimates outside the range of standard errors.

Results

This chapter tested two hypotheses. The first hypothesis states that the Russian protest election movement will be more effective than the government at spreading its preferred frames online, except when the government organizes large pro-government rallies. The second hypothesis states that moderates will be more successful than extremists in the Russian election protest movement at spreading their frames, unless violence occurs at protest events. To test these hypotheses, I estimate the proportion of text that falls into each of the eight categories every week across the entire six months of protests. The results show that the first hypothesis can be rejected, since the government never accounts for more text than the opposition. However, the results do provide evidence in support of the second hypothesis, since the text of moderates account for a larger proportion of text throughout the protest, until May, when violence occurs between the protesters and the police. I discuss the results in more detail next.

The results are summarized in Figure 7.3. In Figure 7.3, the category proportion is plotted on the vertical axis of each panel and time is plotted on the horizontal axis. I use facets, or panels, to isolate each category's results since they are difficult to read if included in just one plot given the large number of categories and the fact that these are proportions, not raw counts of documents. The facets are ordered in Figure 7.3 based on the average document category proportion across all weeks, with those categories having the highest average proportion located at the top of the chart and the lowest on the bottom.

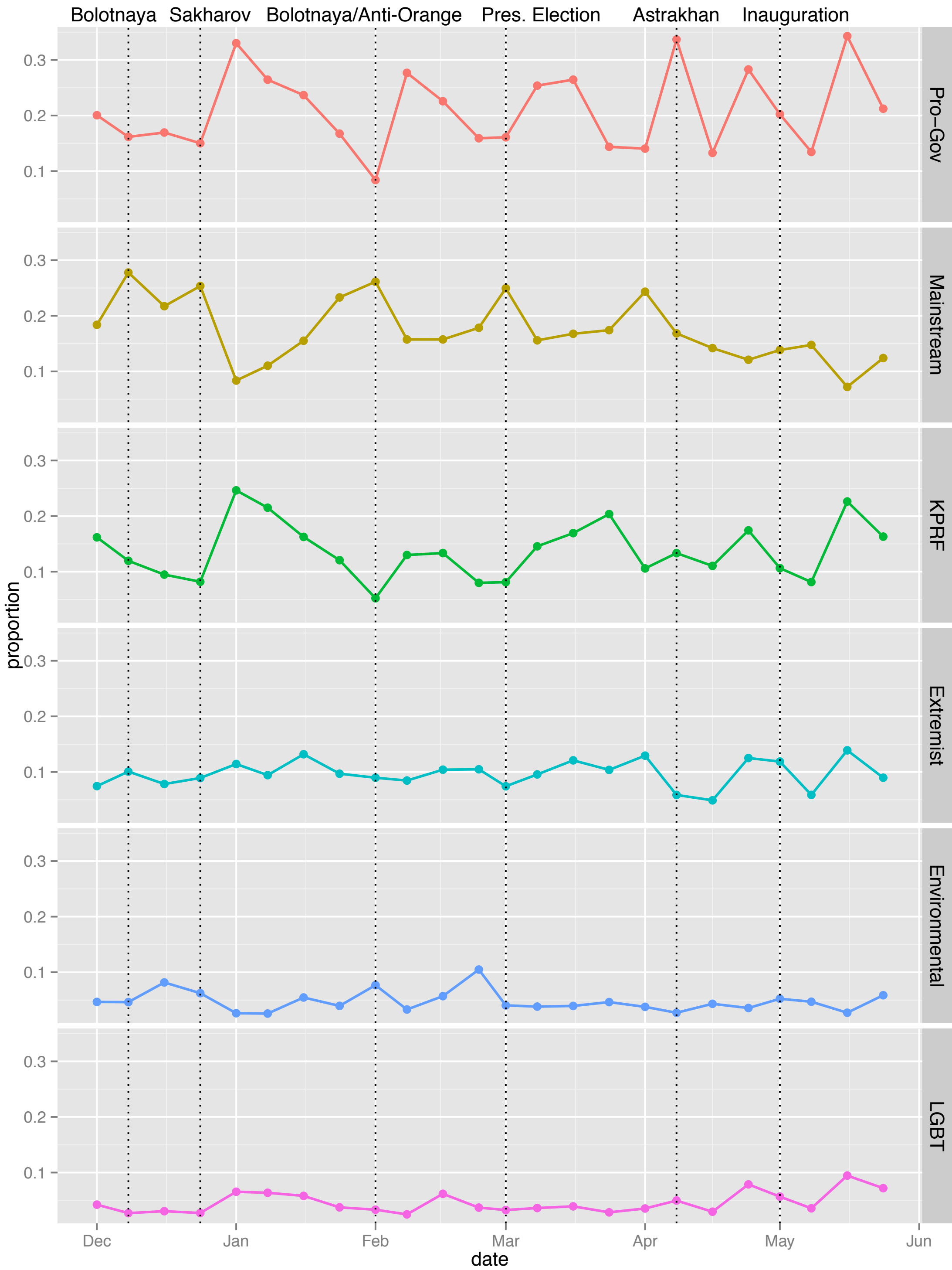


Figure 7.3: Monthly document category proportions for each substantive category. Dotted vertical lines represent major protest events or elections. Neutral and off-topic category proportions are not shown to prevent over plotting.

Dotted vertical lines mark major protest events. To aid in interpretation and comparison of substantive categories, the catchall neutral and off topic categories are not included in Figure 7.3. On average, the neutral category accounted for 8% of content and the off topic category accounted for 20%. The off topic category tends to account for a larger proportion of text during weeks of low protest activity and few documents in the data, such as the New Year's holiday during the first week of January. Results for each week for all categories, including the neutral and off topic categories, are provided in Appendix F.

Taken together, the five oppositional categories that make up the constituent parts of the protest movement (the mainstream, KPRF, extremist, environmental, and LGBT categories) easily account for a higher proportion of documents than the government throughout the entire protest cycle. On average, these five oppositional categories cumulatively account for 50% of the content over the six months of protests, while the government accounts for only 21%. Figure 7.4 shows the cumulative oppositional document category proportions compared to the government. The mainstream category accounts for the largest average proportion of documents of all the oppositional categories, at 17%.

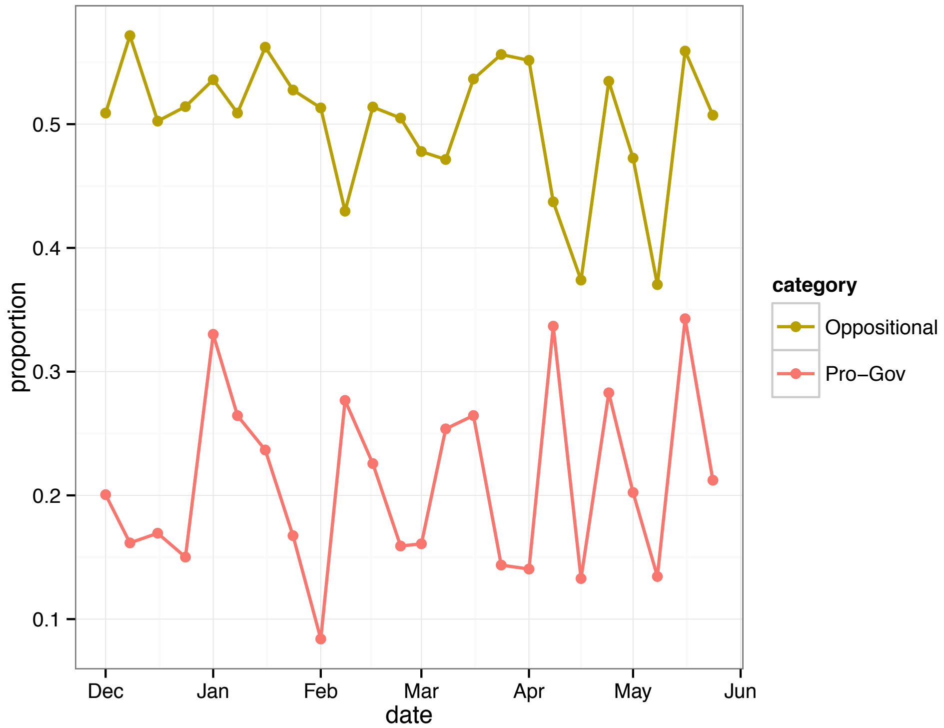


Figure 7.4: Collapsed oppositional categories compared to pro-government category proportions.

On its own, the government category accounts for the largest average proportion of text. However, it is important to note that Figures 7.3 and 7.4 are somewhat misleading since they do not show the number of documents in each week. To address this shortcoming, Figure 7.5 plots document frequency against the mainstream and pro- government category proportions. The top panel of Figure 7.5 shows the frequency of documents in the data for each week during the protests. The off-line events that drive increases in document frequencies are also labeled in the top panel. The bottom two panels show the proportion of documents accounted for by the mainstream and pro-government categories over time. It is clear from this chart that the mainstream category proportions very closely track changes in document frequency. Major oppositional protest events drive increases in the number of documents, and these spikes lead to increases in the proportion of text accounted for by the mainstream category.

The pro-government category moves in the opposite direction, with category proportions at their lowest when major protest events occur and peaking during weeks when there are no protests and very few documents in the data. Specifically, the three weeks in which the pro-government category accounts for the largest proportion of text (the first week in January and the last two weeks of May) are the three weeks with the fewest number of documents in the data. The off topic category (not shown) follows a similar pattern, rising when there is little text data and little protest activity.

Another interpretation of the results shown in Figure 7.5 is that the government is purely reactive; with peaks occurring typically one week after major oppositional protest events. Since the data represents the production of text, one could argue that the mainstream category and other oppositional groups are also serving an agenda setting function in the online space during the protests, while the government is only able to react to that oppositional agenda and its related framing. While I am not able to measure the preponderance of protest topics compared to all other topics discussed during this time period with my data, Koltsova and Koltcov (2013) have shown, using topic modeling techniques, that the protests were indeed a major agenda item on Russian LiveJournal during this time period.

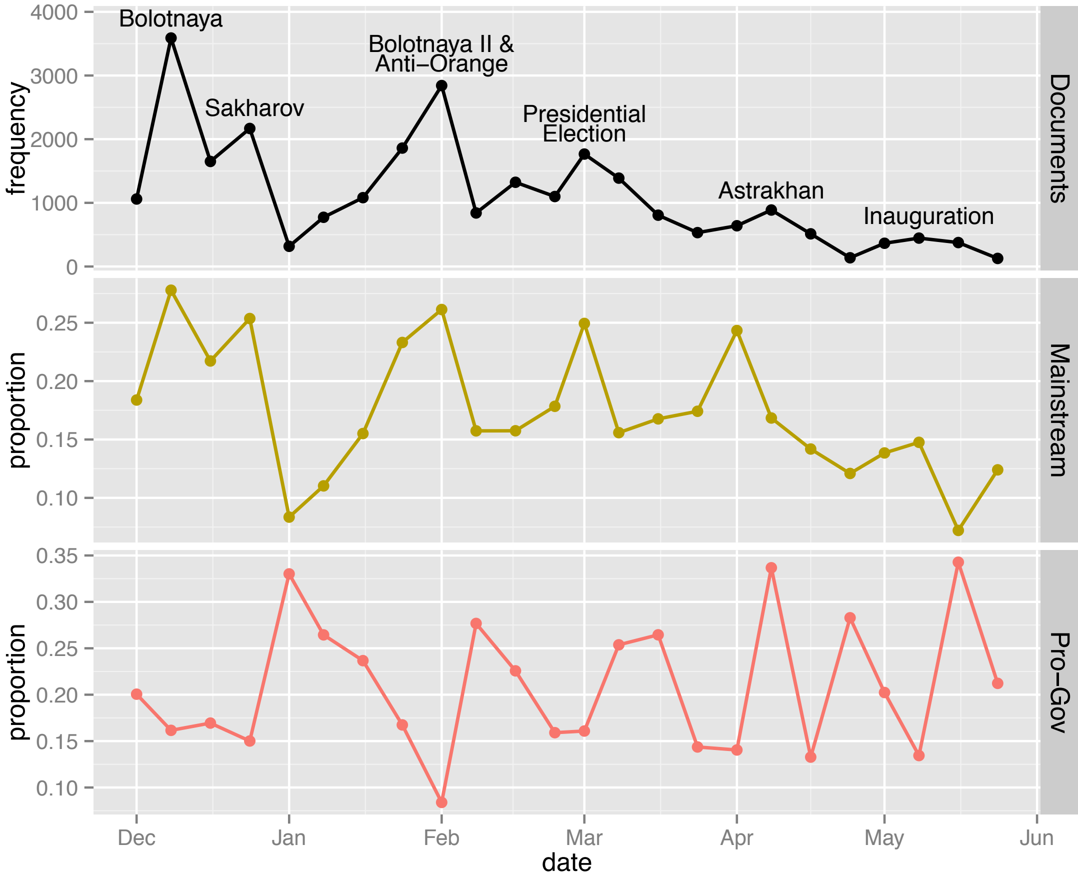


Figure 7.5: Document frequency compared to mainstream and pro-government category proportions. The top panel shows the frequency of online documents over time and affiliated offline events. The bottom two panels plot category proportions over time for the mainstream and pro-government categories.

There is one exception to the general rule that government category proportions peak the week after major protest events. Two protests took place in the second week of April in Astrakhan (on April 10 and 14), when many protest leaders from Moscow and other large cities traveled to Astrakhan to support Oleg Shein in his hunger strike and fight against falsified mayoral elections in that city (Balmforth, 2012b; Herszenhorn, 2012a). In this week the pro-government category accounts for a larger proportion of documents than the mainstream category. This anomaly could be explained by the Orthodox Easter holiday, which took place the day after the second protest in Astrakhan. Oppositional category proportions tend to be much lower during holidays. The mainstream category also accounts for a larger proportion of text the week before the Astrakhan protests, indicating that it was still an important topic of conversation that week. And in any case, when all oppositional categories are collapsed, the opposition still accounts for more text than the government during the second week of April.

Given that major oppositional protest events coincide with the mainstream and other oppositional categories accounting for a larger proportion of text documents, we can also ask if major pro-government protests, which were in part intended to demonstrate overwhelming support for Putin among average Russians, lead to a larger proportion of text falling into the pro-government category. The answer is clearly no. The best example of this finding is the large, pro-government Anti-Orange protest that took place on February 4, 2012 at Poklonnaya Hills. These rallies were labeled Anti-Orange because organizers believed the opposition protesters were trying to replicate the 2004 Ukrainian Orange Revolution, which Putin and his supporters argue was orchestrated by the United States. The Anti-Orange protest was also held on the same date as another major oppositional rally at Bolotnaya Square in Moscow (Barry & Kramer, 2012; Parfit, 2012). It is clear from Figure 7.5 that the mainstream category accounts for the largest proportion of text during this first week in February, when both pro- and anti-government demonstrations took place. This week represents one of the highest peaks for the mainstream category (26%), while the proportion of text accounted for by the government category plunges to 8%, the lowest of any week for the pro-government category. On December 24, a smaller Anti-Orange protest also took place at Sparrow Hills (Vorobiyevych Gori) on the same date as the opposition protest at Sakharov Prospect (Smyth et al., 2013). Figure 7.5 tells a similar story for this week: an increase in document frequency, a rise in the proportion of text accounted for by the mainstream category, and a drop in that for the pro-government category. Pro-Kremlin youth groups also organized pro-government protests in central Moscow on December 6, United Russia organized a rally in support of the president on December 12, and President Putin's presidential campaign organized additional pro-government events at Moscow's Luzhniki stadium on February 23 and at Manezhnaya Square on March 4 (Elder, 2011a; Herszenhorn, 2012b; "Pro-Putin supporters rally in Moscow," 2011; Smyth et al., 2013; Weir, 2012). The pro-government category never accounts for a larger proportion of text than the cumulative oppositional categories on any of these dates, and the mainstream category on its own accounts for a larger proportion of text than the government on all but two of those dates. Again, this evidence allows us to reject the first hypothesis.

The second research question addressed in this chapter asks whether framing by moderates is more prevalent than that of extremists within a movement. I hypothesize that moderates will be more effective at spreading their frames than extremist elements in the Russian protest movement, unless violence occurs at protests. The results provide evidence in support of this hypothesis.

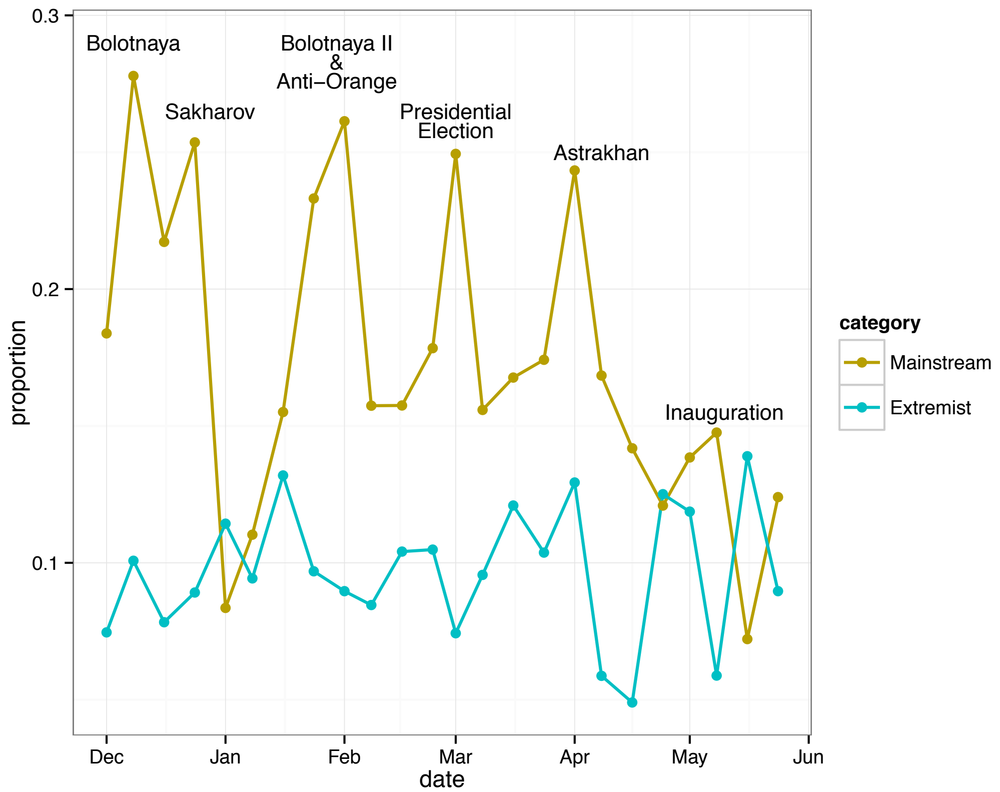


Figure 7.6: Mainstream versus extremist category proportions.

Overall, as Figure 7.6 shows, the framing used by democrats, civic groups, and independent journalists, who represent the moderate, mainstream core of the movement, account for more text than extremists. Again, the extremist category represents framing by ultranationalists and those on the far left. On average, the mainstream category accounts for 17% of text and the extremists 10%. On the rare weeks when the extremists account for more text than moderates, this is during low traffic weeks when there are few documents in the data, such as the first week of January and the last week in April. Of more substantive interest, there is also more extremist language the third week of May. This is one week after the May Inauguration protests, when violence occurred between the police and protesters, and is likely a reflection of more radical language this week. This week also accounts for the highest proportion of text for the extremists of any week in the data, at 14% of text. The increase in extremist language and the decrease in moderate language in the last month of the protests coincide with a general decrease in support for the protest movement and a decrease in protest size, as theory predicts. Still, the proportion of text that falls into the extremist category overall is less than expected given the large number of nationalist blogs that previous studies have found in the Russian blogosphere, the prominent role played by extreme leftist Sergei Udaltsov in the protests (Etling et al., 2010), and the significant support for ethno-nationalism and authoritarianism among those backing the protest movement (Chaisty & Whitefield, 2013).

A couple final observations based on the results. The KPRF category, on average, accounts for a larger proportion of text than expected given that some Russia area scholars argue that the party played a marginal role in the protests (March, 2012; Sakwa, 2015). It is possible that KPRF's status in the systemic opposition, which allows the party to both participate in elections and appear in traditional media, helps to explain why to this category accounts for a larger proportion of text than expected. However, similar to the pro-government category, some of the weeks when the KPRF category accounts for a higher proportion of text than other categories are those weeks when there is little protest activity and few text documents in the data. Specifically, the first week of January is one of the lowest traffic weeks in the data with just 317 documents, but this is the week when KPRF accounts for the largest proportion of text of any week in the data (25%). So, the low traffic weeks make the KPRF appear more influential than they actually are in this analysis. To be fair, though, the KPRF does retain influence and has significant support in the country. There may also be more KPRF supporters online than one might assume, given that communist party supporters tend to skew older.

Finally, the environmental and LGBT categories clearly account for the smallest proportion of text. Both, on average, account for about 5% of the documents. The environmental category never exceeds 11% and the LGBT category never accounts for more than 9% of documents. The environmental category could be considered to be part of the mainstream protest community, and so splitting them out might be an artificial divide. Future research could try to collapse this category into the mainstream group. Given the extremely difficult situation for gay rights activists in Russia, and attacks by pro-government and nationalist elements against those that speak out on LGBT issues, it is not surprising that the LGBT category accounts for such a small proportion of text.

Conclusion

This chapter showed that despite its control over mainstream media and oft-cited concerns that pro-Kremlin bloggers and paid online supporters will drown out oppositional voices online (Deibert et al., 2010; Gunitsky, 2015), the government framing of the protests online is never more prevalent than that put forward by the protest movement. Government counter-framing of protests was not as effective as some would have us believe (Koesel & Bunce, 2012; Smyth et al., 2013). Even during the large, pro-government, Anti-Orange protest in February 2012, the government does not account for a larger proportion of text than the oppositional categories. This is even true at the end of the protests, when violence broke out between the police and the protesters, and one might expect government framing to overtake that of the protest movement. With the exception of this final month, the mainstream category on its own accounts for more text in the larger corpus than the government. The government is also largely reactive, with spikes in pro-government category proportions occurring typically in the weeks following major oppositional protest events, and never beforehand. However, pro-government framing is more prevalent when protest discourse is minimal and in the absence of major oppositional protest events. The results provide evidence in support of the second hypothesis tested in this chapter. Moderate voices account for a larger proportion of text on average than extremist ones, except when there is violence at the final protest event. The predominance of mainstream over extremist framing also serves to undermine the government's framing of the protesters as led by extremists and radicals who did not represent the grievances and demands of the population. The increase in extremist framing at the end of the protests also coincides with the least effective phase of the protests.

In this chapter, I showed that the framing of moderate, mainstream protest leaders was more likely to be adopted online than that of extremists within the movement, as well as counter-framing put forward by the government and its supporters. The previous substantive chapter (Chapter 6) empirically tested whether the information of brokers and those they connect from different communities is actually novel, which is a surprising rare test in brokerage studies. Chapter 6 also explored how macro network structure appears to impact the novelty of information at the local level. The brokerage chapter built off the first empirical chapter (Chapter 5), which used Exponential Random

Graph Modeling to show that the network had a polycentric structure early in the protests, but moved to a theoretically less effective hierarchical structure as the energy behind the movement waned. In the next chapter, I conclude this thesis by reflecting further on the findings from these three substantive chapters and their implications for the study of collective action in hybrid regimes.

Chapter 8 Conclusion

In this thesis, I set out to understand how networked communication technologies facilitate high-risk collective action. To unpack this broad topic, I focused on three areas that have a strong foundation in the social sciences and that can be empirically tested: the role of network structure, brokerage, and framing in high-risk collective action. By doing so, I was able engage with several claims made by those who question the power of the Internet and social media to influence high-risk collective action: that the Internet creates flat networks that are not conducive to high-risk collective action, that authoritarian governments are better than opposition movements at using social media, that the Internet empowers extremists at the expense of moderates; and that the Internet creates echo chambers of the like-minded that limit exposure to diverse information and ideas.

I addressed three core research questions in this thesis. The first asked if a protest network challenging a hybrid regime will have a polycentric or hierarchical structure, and if that structure changes over time. Those skeptical of the Internet's ability to facilitate risky collective action argue that the Internet and social media facilitate the creation of flat or polycentric networks. Critics further contend that hierarchical networks are necessary to take on entrenched powers. The second research question sought to understand if brokered information is actually novel, and if brokered text is more novel when a network has a polycentric structure than when it has a hierarchical one. Brokers are the individuals or nodes in a network that, via weak ties, connect disparate groups and close structural holes. The ability to access, share, and translate novel information between diverse individuals and groups is the primary explanation for the many benefits that accrue to brokers and the groups they connect, but this is rarely actually tested. The last research question asked if a protest movement in a high-risk political setting can be more successful than the government at spreading its preferred frames, and within such a movement, whether moderate or extremist framing is more prevalent. Skeptics claim that authoritarian governments are able to shape online discourse in their favor, while movements that are led by extremists instead of moderates are less likely to gain widespread public support, and thus, less likely to be successful.

This concluding chapter begins with a discussion of the major themes that emerge from this thesis's substantive findings. I then discuss the implications of the general findings for Russia and highlight some Russia-specific findings. Next, I discuss the methodological and theoretical contribution of the thesis and how computational social science methods and 'big data' aid the study of collective action. I then highlight the benefits of looking at high-risk collective action in hybrid regimes. This is followed by suggestions for future research. Before concluding, I also discuss the study's main limitations.

Major themes emerging from the findings

Polycentric network structures are associated with positive outcomes for movements in high-risk political settings

Skeptics claim that networked communication technologies facilitate the creation of flat, or polycentric, networks, which they argue are not conducive to risky collective action. These skeptics further argue that hierarchical networks are necessary to take on dictatorships or entrenched elites, because such networks have clear leadership and decision-making structures, and are better at strategy development and implementation. However, the collective action and network science literatures find that there are benefits to both polycentric and hierarchical networks. Polycentric networks are believed to be conducive to high-risk collective action because they are robust to defections or the removal of key nodes compared to hierarchical networks. Movements with such structures have also been found to be able to recruit supporters from different segments of society and to be more innovative than those with a hierarchical structure. On the other hand, hierarchical networks are theoretically more efficient at spreading information and can more easily mobilize the entire network to participate in protests. However, they are less beneficial in risky political contexts since they are not robust to the targeted removal of key nodes.

The results from Chapter 5 provide strong evidence against the claim that networked communication technologies facilitate the creation of just one type of network, and that those networks are flat. If one looks at the protest movement's communication network as a static snap shot, it would indeed have a flat, polycentric network structure. However, by using data that allows us to look at network structure over time, I showed that there is significant variation in network structure as time passes. The early time periods, when the network has the greatest tendencies towards a polycentric structure, coincide with the time periods when the protests had a larger, more diverse base of participants from across the country. This was shown in Chapter 5, based on a series of Exponential Random Graph models built over the course of the protests.

However, at the end of the protests, when the network consisted of die-hard activists and supporters, and when the movement had the lowest levels of enthusiasm and support from the broader public, it has tendencies towards a hierarchical or core-periphery structure. This implies that protest movements in high-risk political environments should create flatter, more diffuse networks instead of rigid, hierarchical ones. As Gerlach (2001) and Baldassari and Diani (2007) have shown, drawing on previous examples of civic action, decentralized networks can be powerful. It is important to note, however, that I cannot say anything about causation. My methods and data do not allow me to say that a given network structure led directly to positive outcomes for the protest movement, merely that there is an association between tactical success, such as large, well-attended protest events with a broad base of participants, and polycentric network structure.

One of the reasons polycentric networks are believed to be advantageous for social movements is because they are able to tap into and recruit from diverse segments of society. The lack of centralized decision-making means that disparate groups are able to think more freely about how to enact change. This leads to more innovative movements with a diverse range of actors, ideas, and tactics, which also allows for a certain level of tactical experimentation. Local groups are also able to identify and carry out tactics that are more appropriate for their specific regional conditions than those imposed from a central headquarters, which is often located in a far off capital. The results from Chapter 6 show that polycentric networks do not simply have different types of actors involved, but that the information flowing between different clusters of strongly connected actors is more novel when the network has a polycentric structure.

This provides support for the notion put forward by Gerlach (2001), that part of the reason why polycentric networks are powerful is because they have novel ideas flowing through the network that allow for innovation and creative problem-solving, instead of one top-down approach. In high-risk political environments, movements that consist of multiple groups and different centers of power are more difficult for the state to dismantle compared to those with just one leader sitting at the top of a hierarchical organization.

A possible lesson for protest organizers from the results of this study is the need to create a broad-based, loose coalition of supporters instead of a rigid hierarchical form of organization. This calls for a greater focus on bottom-up, informal organization and less on creating formal bureaucratic groups. These types of 'leaderless' protest movements also appear to be more prevalent in recent protest movements around the world whose work has been facilitated by networked communication technologies. These include movements in Egypt during the Arab Spring, the global Occupy movement, and the 2014 Euromaidan protest in Ukraine (Castells, 2012). This type of organization and associated polycentric network structure should be embraced for its innovation, lack of formal organization or leadership structure, and ability to reach into and mobilize many different segments of society. The findings in this thesis suggest that formal bureaucratic civic organizations are no longer required for protest movement success. Scholars should therefore also move towards understanding the strengths and weaknesses of these types of new organizational forms and their role in collective action facilitated by networked communication technologies.

Authoritarian governments are not able to shape online discourse more effectively than oppositional movements

A common claim by those skeptical of the ability of the Internet and social media to aid protest movements, especially in authoritarian settings, is that these tools are actually more beneficial to governments than they are to oppositional movements (Gunitsky, 2015; Kalathil & Boas, 2003). Specifically, critics argue that authoritarian governments can use paid armies of online supporters to shape political discourse in their favor, and by attacking critics, force them into silence. The results from Chapter 7 show that there is no point in the entire six months of the protests when the Russian government framing of the protests is more prevalent than that put forward by the protest movement itself. The government also organized a series of pro-government rallies that were meant to demonstrate overwhelming support for the government by average Russians. Even during the largest of these rallies, the pro-government, Anti-Orange protest in February 2012, government framing does not account for a larger proportion of text than that of oppositional categories. Further, the results from Chapter 7 show that the government is reactive instead of proactive. Government framing tends to peak only after major protest events, while that put forward by the opposition peaks leading up to and during the weeks when major anti-government protests occur. In this way, the protest movement is also playing an agenda-setting function in the online space, while the government has to follows their lead, filling in the discourse afterwards with their preferred description of the events and the protesters.

Certainly, there is evidence that authoritarian governments do attempt to shape discourse in the online space (Deibert et al., 2010; Fossato & Lloyd, 2008;

King, Pan, & Roberts, 2016). However, the mere existence of these efforts, or framing that is supportive of authoritarian governments, does not mean that such efforts have their intended impact. A contribution of this thesis was to take an online data set that included sources and text from both the opposition and those supportive of the government, and to compare frames put forward by both. Negative framing of the protests by the government and its supporters was an overt strategy meant to undermine the movement (Koesel & Bunce, 2012). This approach allows us to test if it was effective. The evidence provided in Chapter 7 provides strong evidence against the notion that the government is more successful in advancing its framing of the protests than the opposition. Again, instead of merely identifying the existence of pro-government frames or actors online, future studies should continue to determine if pro-government actions online have their intended effect, including how widely the ideas put forward by pro-government actors are adopted.

To win support among the broader public, movements should advance moderate framing

There is a general concern that the Internet is empowering extremist voices at the expense of moderate ones (Copsey, 2003; Perry & Olsson, 2009; Wojcieszak, 2010). For protest movements in hybrid regimes this is especially important because movements that are seen as being led by extremists are less likely to win broad support among the public or to draw large crowds to protest events, limiting their ability to win concessions from those in power (Lohmann, 1994). The results from Chapter 7 show that, overall, the framing of the protests put forward by moderates accounted for a larger proportion of text than extremists. The predominance of mainstream over extremist framing also provides additional evidence that the Russian government was not successful in its attempt to frame the protest movement as one led by extremists and radicals who did not represent the grievances and demands of the general population. However, the results from Chapter 7 also show that at the very end of the protests, when the energy behind the movement was at its nadir and the size of the protests were small, that the proportion of extremist speech began to rise and even overtook the framing of the mainstream core of the movement. These time periods also saw the gap between pro-government and mainstream framing narrow. The rise in extremist speech also coincided with violence between protesters and the police at the last major protest event. Except for this last time period, moderate framing was significantly higher than that of extremists around each major protest event.

Brokered text is novel, but information novelty decreases when networks have a hierarchical structure

There is a general concern that the networked communication technologies are leading people to isolate themselves into echo chambers of the likeminded, to filter out information they find disagreeable, and to surround themselves with those who hold similar opinions (Pariser, 2011; Sunstein, 2007). It is believed that this is contributing to polarization and exacerbating racial and ethnic tensions, as we increasingly sort ourselves into ever more homophilous, strongly tied groups. However, when networks have a polycentric structure, brokers can, in theory, bridge otherwise isolated clusters of strongly connected individuals and facilitate the spread of novel information. Empirical evidence from social movements show that brokers are able to tie together diverse political groups and social classes, aid in the diffusion of revolutionary ideas, and contribute to the emergence of broad-based, geographically dispersed movements instead of isolated, local ones (D. Faris, 2008; Han, 2009; Hedstrom et al., 2000; Kim & Pfaff, 2012). The main reason that brokers are advantageous is because they provide access to novel ideas and information that we would not be exposed to from our close family and friends. However, this foundational assumption in the brokerage literature is almost never tested using node level textual data.

The results from Chapter 6 show that brokered text is indeed novel. The text of the nodes that brokers connect from different communities is generally novel compared to a number of benchmarks of text similarity. Further, the results from Chapter 6 also show that the text of brokers is novel compared to nodes in the rest of their communities. Indeed, the text of brokers compared to the rest of their communities tends to be the most novel of any relationship tested. These findings suggest that brokers are able to alleviate some of the negative externalities associated with communities that consist of strongly connected actors; what critics often refer to as echo chambers. The findings imply that network structures that include brokers can help protest ideas and information to spread widely across diverse movements, allow different constituencies to be exposed to each other's grievances and tactics, and generally prevent individual communities from becoming isolated from the rest of the movement.

I also showed that macro network structure is associated with the novelty of brokered text at the local level. The results from Chapter 6 indicate that as the protest network moves from a polycentric to a hierarchical structure, that text similarity scores increase; in other words, the novelty of information decreases. The results show that text becomes more similar over time among all the relationships tested in the gatekeeper and representative brokerage types, where the broker is connecting nodes that belong to different communities. While there is some variation, for most relationships this increase in text similarity is monotonic, increasing every month as the network becomes more hierarchical. The December time period, when the network is the most polycentric, has the lowest similarity scores, while the April time period, when the network is most hierarchical, has the highest similarity scores. There are many variables that could impact the similarity of text, but the results from Chapter 6 indicate that there is an association between networks structure and brokered information. Further testing with confounding variables should be carried out in the future.

Finally, brokers are especially important for movements that emerge as a result of electoral fraud. Electoral fraud gives an entire nation a common grievance and brings together diverse, often competing political interests. Brokers are able to tie these diverse groups together and serve as conduits for novel information flowing between them. In short, brokers help movements to become broad-based and national, instead of isolated and local.

Protest movements are not static, and should not be studied as if they are

As many of the findings already discussed in this concluding chapter have shown, there is significant change over time in each of the aspects studied in this thesis: network structure, the novelty of brokered text, and protest frames. Conclusions drawn from any single time period do not necessarily hold in others, especially at the beginning and end of the protests. This is especially true if one were to draw conclusions based on data only from the December time period, when the movement was most successful, or only in May, when the movement was weakest. I would have also come to different conclusions if I had treated the data as static, and looked only at the overall results from each of the analyses undertaken. This is another contribution of this thesis, and calls for more studies in the future to seek out data and to study protests and social movements over much longer time periods than just a single month or protest event. Unfortunately, it is often difficult to go back in time and gather online data if one is not constantly collecting relevant data before protest events occur. Doing so is also often prohibitively expensive. In those cases where it is not possible to gather data over the full arc of a movement, it is important for scholars to underscore that this is a limitation in their research design, and that the findings may not hold at other time periods.

Implications of the Russia case

Although it failed to achieve its strategic goals, the Russian protest movement had a number of tactical successes

Although my objective in this concluding chapter is to identify themes that apply broadly to movements in a range of high-risk political settings, the use of the

Russia case allows for some conclusions specific to that case. First, many viewed the Russian protest movement as a failure because it was not successful in removing Putin from power or forcing the government to rerun the falsified Duma elections. While the movement did fail to achieve many of these strategic goals, it did have a number of important tactical successes. The best example from the empirical evidence provided in this thesis is probably from Chapter 7, where I anticipated that the government would likely win the framing battle against the opposition by holding major pro-government rallies and actively engaging in anti-protest framing, once the government realized the movement could not simply be ignored. Instead, the oppositional categories were more successful in having their framing adopted online throughout the entire protest cycle, even in the face of pro-government rallies, government control over national TV stations, and an overt strategy by the government to frame the protests in a negative light.

More generally, the protest movement was also able to create a national, broad-based movement that consisted of participants from across the country and across the oppositional political spectrum. Beforehand, civil society in Russia was seen as weak, and the opposition as too divided to launch a successful attack against the government. Despite numerous challenges and roadblocks put in the movement's way by the Russian government, as well as the systemic challenges in place long before the protests, the movement was able to hold a series of high profile, well-attended national protests for three months. Unfortunately, as the results from Chapter 5 show, the movement was only able to hold together this polycentric network temporarily. The breaking point seems to have been the election of Putin in March, and the absence of evidence of electoral fraud at the federal level. Afterwards, the movement moves to a core of die-hard activists, reflected in a more of a hierarchical structure as the protests wind down, instead of a broad-based, national movement. This also suggests that focusing on United Russia, with its lower levels of support and more easily targeted local functionaries, was possibly a more successful strategy than taking on Putin directly.

Leftists, including the Communist Party of the Russian Federation (KPRF), played a more prominent role in the Russian protest movement than many think

Another key finding emerging from the Russia case is that the Communist party was more influential than many scholars believe. The results from Chapter 7 show that, while the framing put forward by Communists is not the largest, it is also far from the smallest. Among the six substantive categories tested, Communist framing accounts for the third highest proportion of text overall, after the government and mainstream categories. Communist framing beats out that by extremists, environmentalists, and gay-rights activists. It is somewhat surprising that the Communists have such a large reach online, given that supporters of the Communists tend to be older and, as a general rule, this makes them much less likely than young people to use the Internet. It may be that the status of KPRF in the formal opposition explains their higher than expected levels of support online, since this means that they are allowed to take part in elections, hold seats in the parliament, and to appear on state-controlled television. However, I would argue it is also possible that Russia area studies scholars underestimated the level of support for Communists within the protest movement (March, 2012; Sakwa, 2015).

It is also worth highlighting that the one actor who tied together the entire movement is a leftist, Oleg Shein. This was shown in Chapter 6, based on his combined brokerage scores. Shein has the highest brokerage scores of any actor in the movement. While this is partly due to Shein's emergence as the center of the movement's attention during the April protests in Astrakhan, he has high brokerage scores throughout the protests. As a regional politician he is also able to attract links from a range of other regional actors outside of the major capitals. These findings suggest Shein, and actors like him, played a more important role in tying together the movement than is often appreciated. These findings also suggest that it might be possible for the opposition in Russia to put together a sizable coalition with liberals and moderate leftists, without having to align with far-right nationalists. This is not to say that leftists were the largest part of the movement or the most important constituency: polls and the results from Chapter 7 indicate that liberal democrats hold that title (Volkov, 2015). However, the role of the Communists and leftists does seem to be underappreciated.

The findings from this thesis also suggest that it may not be necessary for the future protest movements in similar political settings to create a formal, centralized social movement organization. The limited usefulness of the Russian Opposition Coordinating Committee bears this out. The meetings of this group tended to publicly highlight the schisms in the opposition movement instead of signaling its unity in purpose. While traditional social movement scholarship sees an important roll for formal social movement organizations, the results from this thesis show it may be better for contemporary networked social movements in high-risk political settings to create flatter, polycentric networks. This can be achieved, in part, by ensuring that brokers exist and have sufficient standing with different parts of the movement that they are able to tie these groups together into a larger movement. Further, protest leaders should ensure that no individual constituency becomes so internally focused that it becomes isolated from the broader movement

Assessing the success and failure of protest movements

How does one assess the success or failure of a social movement, especially a protest movement in a high-risk political setting? When taking on an entrenched hybrid regime, I would argue that determining success or failure based exclusively on whether a movement overthrows the government is a poor choice. A point I have tried to make in this thesis is that is as much to learn from 'failed' movements in this sense as there are in the cases where regimes are overthrown. I have already sketched out the lessons that should be taken based on the evidence from this thesis. Here I explore some of the possible longer-term implications of the movement, and its current struggles.

Perhaps as a result of the success of the protest movement in moving election fraud to the top of the political agenda and a sustained six months of protests, the Russian government has responded very harshly to the challenge put forward by the protest movement. There has been a severe curtailment of civil society in Russia since the end of active protests in May 2012, with a particular focus on the protest movement and its leaders. Protest leaders or their family members have been arrested and tried on trumped up charges. NGOs have been starved of overseas funding due to a new law requiring them to register as 'foreign agents'. Boris Nemstov, an important democratic opposition politician and protest leader, was assassinated within view of the Kremlin. Many believe his murder was a result of his continued activism and criticism of the Russian invasion of Crimea and eastern Ukraine. Common protest participants were also targeted, with a number of them swept up in the final inauguration day protest in May 2012 where fights broke out between the police and protesters. The movement has not been able to reinstate itself significantly and push back against the state as effectively as it did during the election protests. This seems to suggest that fear and repression, at least in the short term, are effective tools for authoritarian leaders. This also suggests that digitally enabled collective action may be more successful in less authoritarian hybrid regimes.

However, as we have seen before, increased repression often eventually leads to a backlash. When dissenting voices within and outside the government are not heard, this makes it difficult for the state to understand how much support it does or does not actually have (Lohmann, 1994). Authoritarian governments also often lack the agility required to respond to emergent economic or political challenges when they have to continue to appease political supporters. And increased repression is unlikely to reverse many of the chronic problems that led many to sympathize with the protest movement in the first place, such as corruption and abuse of power at the local level. Relaxation of repression over time is also not out of the question in Russia. This could be due to a situation in which the state becomes more comfortable in their level of control, and no longer sees the opposition or popular protests as a major threat.

It might also be necessary to loosen control over civil society and dissent if the economy does not improve, if sanctions are not lifted, or if overseas adventures such as those in Ukraine and Syria lead to broader public disapproval. These and other possible scenarios could allow the opposition to again gain some room to operate. If that happens, then the opposition in Russia may be in a better position to act than might be apparent at the moment. Of course, it is also possible that Russia becomes an entrenched authoritarian regime, along the lines of Iran or China, but this seems less likely given the levels of freedom that Russians have become accustomed to since the fall of the Soviet Union.

Whatever the reaction of the state, the protests have had a number of important longer-term implications. The Internet and the protests they enabled in Russia helped create an alternative public sphere, where counter-publics can emerge and challenge, if not necessarily overturn, state authority. The Internet has also allowed for new types of citizen engagement. This type of engagement was seen as an important building block to the Tunisian revolution. In that successful revolution, Zayani (2015) argues that the role of the Internet was not necessarily how far or fast the image of a fruit seller's self-immolation was able to travel, serving as the spark for the broader Arab Uprisings, so much as the ability of the online sphere to gradually become an alternative space for political communication and engagement. The online space also allowed youth to become involved in everyday politics, pushing back against the conventional wisdom that they were apathetic. Zayani's evidence shows many parallels to the Russia case, where before the protests Russian civil society was seen as weak and apathetic, but where there existed a strong, vibrant, contentions alternative public sphere engaged in everyday discourse about politics, often from a non-aligned perspective (Etling et al., 2010). Over time these groups began to engage in small-scale, successful protests and citizen actions, such as organizing volunteer efforts connected to natural disasters (Asmolov, 2014a, 2014b). These civic efforts formed the foundation for the larger protest movement. Similarly, a major implication of the protests in Russia is the continued building of civil society and democratic practice in Russia. As Bennett and Segerberg (2013) remind us, we should not understate the importance of protests as a form of democratic practice. This is even more crucial for protest movements in nondemocratic contexts. In Russia, these civic efforts will have to be small scale to remain under the radar of the government, but it is not because the opposition does not have aspirations to make larger changes in Russia.

The opposition in Russia also appears more willing and better prepared than others, especially compared to activists in countries connected to the Arab Spring or the Gezi protests in Turkey (Tufekci, 2014), to take over the reigns of power if political opportunities change. In the longer term, while civil society and the movement have been largely neutralized, we do see protest leaders that are willing to compete for political power through elections. While Alexei Navalny was unsuccessful in his bid for mayor of Moscow, the authorities were shocked that he did as well as he did even though he was largely denied access to TV and state support for the incumbent heavily tilting the scales against him. Approximately 15,000 volunteers supported Navalny's campaign (Lipman, 2013). Recent empirical evidence has also shown that major Vkontakte protest groups have become institutionalized into civic initiatives and NGOs focused on election observation, anti-corruption activity, and other local issues of concern (Sherstobitov, 2014). These protests were also unique in that they showed how a broad swath of oppositional groups, from nationalists to LGBT activists, with very different political agendas, were able to put aside their differences and work together toward a common goal.

The protests were also important because they served as the most significant challenge to Putin since he came to power. Part of the lasting impact of the protests will be the damage done to the Kremlin's carefully honed image of Putin. The protests helped to undermine the idea that there is no alternative to Putin or that his support is as deep as he has proclaimed. This included with constituencies that Putin argued wholeheartedly supported him, such as Russian nationalists. The pro-government rallies also made clear to many that much of Putin's support may be coerced or manufactured.

The relative success or failure of the movement must also be seen in the extraordinarily difficult space that Russian activists operated in, which is even more restrictive now. When compared to what is viewed as success for movements in advanced liberal democratic contexts, such as gaining attention in the media, putting issues on the public agenda, changing legislation or enacting laws, and showing their worthiness, unity, numbers, and commitment (Tilly's WUNC), the movement was a success (Bennett & Segerberg, 2013). The movement elevated election fraud to the top of the public agenda, forced even government controlled TV channels to cover the protests, used the Internet to organize the largest protests since the fall of the Soviet Union, countered government critics, and forced the government to change election laws to make them more inclusive. In the end, it may take several years before we are able to judge the ultimate success or failure of the movement.

Contribution of the thesis

Methodological contribution

This thesis contributes to a growing body of work that uses online data and computational social science methods to study collective action (Benkler et al., 2015; Gonzalez-Bailon et al., 2013; Gonzalez-Bailon & Wang, 2016; Greene, 2012; Hanna, 2013; King et al., 2013; Lewis, Gray, & Meierhenrich, 2014; Margetts, John, Hale, & Yasseri, 2015). I used a novel online data set and a unique combination of social science methods in this thesis. The data set included text and links from nearly 30,000 documents drawn from over 3,500 Russian-language Web sites. This data was also drawn from a number of different platforms, including online news sites, hybrid blog/social network platforms like LiveJournal, and protest pages from Russian social networks like Vkontakte. This data set is much more representative of the networked public sphere than studies that focus on just one platform. Scholars studying collective action should continue to expand their data collection efforts beyond single platforms, especially Twitter. This work is also a departure from more historical or journalistic accounts of the impact of the Internet on contentious politics (Soldatov & Borogan, 2015; Zayani, 2015). Instead of asking political bloggers what they think their influence is (or is not), we need to use digital methods to test their influence, whether through social network analysis or modeling, or by tracking how widely their preferred frames are adopted in the broader online space under study. This thesis is also an important addition to academic studies that seek to understand the impact of the Internet on the 2011-2012 Russian election protests, but that frequently rely on surveys or interviews instead of online data (Reuter & Szakonyi, 2015; Smyth & Oates, 2015; Smyth et al., 2013; S. White & McAllister, 2014).

To analyze this data set I used a range of computational social science methods. This work goes far beyond basic web scraping (Schneider & Foot, 2004) and descriptive social network analysis (Adamic & Glance, 2005; Benkler et al., 2015). While I agree on the need to continue to move away from early studies focused on Web culture and toward the use of online data to answer more traditional social science questions (R. Rogers, 2013) , I also showed how to apply specific advanced computational methods with a unique foreign-language data set. The use of Exponential Random Graph (ERG) modeling is a contribution to network studies as it allows the field to continue to move beyond descriptive network statistics and visual interpretation of network maps, and towards hypothesis testing. ERGMs also allow us to test network structure more robustly and consider many variables, instead of using fairly simplified simulations (Watts, 1999). This thesis also contributes to ERG modeling by offering a new theoretical interpretation of the Geometrically Weighted Edgewise Shared Partner (GWESP) statistic as it applies to collective action. This term is often just treated as a control or a nuisance parameter that has to be included for models to converge. I argue that this term is a good indicator of the type of structures in a social network that can allow for repeated affirmation of protest information, as well as structures that can allow for reinforcement of the normative importance of participation (Centola & Macy, 2007; McAdam, 1988). Chapter 5 also answered the call among network scientists to interpret multiple parameters together in ERGMs (Robins & Lusher, 2013; Robins et al., 2007). In Chapter 5, I identified and tested a combination of structural terms that indicate tendencies towards hierarchical or polycentric networks. The results of the ERGM chapter also show the importance at looking at the evolution (and dissolution) of networks over time, a major focus of the field. There is a clear need for better network statistical models that can handle data of the size and volatility of that collected online. Since the future of data collection for network science is moving towards larger and more volatile data, we need models that are better suited to handling this type of data. Finally, as suggested by Gonzalez-Bailon and Wang (2016), I drew on network theory and findings from empirical data instead of using networks as a metaphor, as is common among some theorists (Bennett & Segerberg, 2012; Castells, 2012).

I also showed that by using machine learning methods, it is possible to quantitatively test who wins framing battles during high-risk collective action. Past studies have tended to take a more qualitative approach to framing (Marshall, 2001; A. Morris & Braine, 2001), look only at oppositional frames (Tarrow, 2013), rely only on framing as carried out by traditional media outlets (Boykoff & Laschever, 2011; Entman, 2004), or rely on expensive hand coding of a relatively small number of documents (Hamdy & Gomaa, 2012). I showed in Chapter 7 that the use of supervised learning methods allows for analysis of much larger data sets than would be feasible exclusively through hand coding of documents. This approach makes it possible to leverage the massive amount of data now available online connected to social movements and collective action.

The combination of both social network analysis and automated text analysis is another methodological contribution. Too often, computational social scientists develop expertise in either networks or automated text analysis methods, and rarely combine the two in a robust manner. In Chapter 6, I showed how it was possible to test a bedrock assumption of the brokerage literature, that brokered information is actually novel, using a combination of social network analysis and automated text analysis techniques. Future research could identify ways to improve the robustness of the method I developed in that chapter.

Contribution to framing theory

This thesis contributes to framing theory by drawing on common strands in both the political communication and sociology literatures to evaluate framing battles during the Russian election protests. I contend that frames matter because they allow activists to identify social problems, attribute blame, gain new supporters, and create tactical plans to take on those in power that social movements are making claims against. I see this as a common strand in the sociology and political communication literatures on framing, and helps explain why framing matters for protest movements. I also showed why it not necessarily useful to focus exclusively on framing by the media, as Entman (1993) does. The results of the framing chapter (Chapter 7), in which movement framing is never overtaken by government framing, also suggests that framing is likely just one factor that contributes to a social movement's success. It is hard to imagine a contemporary, networked movement that is successful today that does not win its framing battles. However, other factors, especially political opportunity structures and regime response, will also factor into the movement's ultimate success or failure. Finally, I also contributed to framing theory by looking at how moderates and extremists within a movement frame their actions. This again calls for moving away from research that focuses on framing by the media of movements, and towards a greater understanding of framing by individual actors and constituencies within movements.

Contribution to connective action theory

Bennett and Segerberg's (2013) logic of connective action theory focuses on networks and personal action frames in ways that are similar to my focus on network structure and collective action frames. I offered some contributions and challenges to connective action theory in this thesis by more closely drawing on network theory, instead of treating networks as metaphors for social movements. I argue that looking more explicitly at network structure could strengthen Bennett and Segerberg's typology of collective action and connective action networks. Specifically, I would test whether the network types they describe exhibit tendencies towards hierarchy or polycentrism, instead of classifying each based on the presence or absence of organizations and personal action frames. I also believe that a better way to measure hierarchy in a network is to use the combination of terms I used in the ERGM models in Chapter 5, instead of visual interpretation of social network graphs or power law curves of static networks. An interesting future test for this theory would be to use ERGMs to see whether the empirical networks Bennett and Segerberg reference in their work show tendencies towards hierarchy or polycentrism as one would expect, i.e., with organizationally brokered networks having the greatest tendencies towards hierarchy and crowd-enabled networks showing the greatest tendencies towards polycentrism. I also showed how the structure of the Russian protest network changes over time, even though the Russian protest network would probably be classified as a hybrid, organizationally-enabled connective action network within their typology. This opens up some interesting theoretical questions about how protest movements can transition between ideal types over time within their typology.

I think my method of tracking frames over time using supervised learning techniques could also strengthen theoretical approaches to personal action frames. Future connective action research could create a more formal coding scheme for defining what is and is not a personal action frame, and use supervised learning to track the prevalence of personal action frames compared to more organizationally brokered frames. This would allow one to see whether inclusive personal action frames of the type Bennett and Segerberg describe (e.g., 'We are the 99%') are more prevalent or used among broader groups of participants than those put forward by organizations operating in the same protest space. In my view, this would be far more persuasive and informative than looking at the opportunities for personal engagement offered on organizational Web sites. Also, simply classifying a protest frame as personalized because it is inclusive is probably insufficient, since such a frame might also be classified as a successful bridging frame in the more traditional framing literature.

Finally, I also showed the value in creating networks from multiple platforms instead of just organizational Web sites. If connective action networks, especially crowd-enabled networks, consist of 'multiple layers of networks', then data collection and network mapping should use data from multiple platforms and create one large network from them based on a common linking structure. Again, I believe this would be especially beneficial for the study of the more crowd-enabled networks of the Occupy Wall Street variety, and should show the greatest tendencies towards polycentric structures of the three ideal network types.

Collective action in hybrid regimes

This thesis also contributes to studies of digitally mediated collective action by looking at hybrid regimes. Most previous studies of protest movements and networked communication technologies are drawn from cases in the United States or other advanced Western democracies. This makes sense, since Internet penetration is highest in those countries and the use of online tools is therefore more likely among movements in those countries. At the other end of the spectrum, studies of collective action in authoritarian countries have tended to focus on China. In this thesis, I analyzed a case of high-risk collective action in a hybrid regime. This is useful since, after democracies, hybrid regimes are the most common regime type in the world. Hybrid regimes include some elements of democratic governance, but are largely authoritarian in nature (Diamond, 2002). They are therefore important to understand as distinct from democracies, where the costs of collective action and protest are relatively small, as well as fully authoritarian regimes, where speech and collective action are restricted or tightly regulated. The use of a protest movement that began as a result of election fraud is also important, as this is a frequent spark for protest movements in hybrid regimes. Since hybrid regimes must hold routine elections to maintain domestic and international legitimacy, but must manipulate them to win, there is a high likelihood that these protests will continue to occur in the future.

Future research questions

In each of the substantive chapters, I included ideas for how each chapter and its methods could be improved, as well as suggestions for future research. Here I mention a few specific research questions that stem from the work completed in this thesis.

What is the influence of different social media platforms on network structure? Most studies of protest movements that rely on Twitter have a hierarchical structure (Barbera et al., 2015; Gonzalez-Bailon & Wang, 2016). Is this an artifact of the platform, or does it have to do more with the political, social, or media environment in which the protests occur? My expectation is that Twitter in particular will have more hierarchical networks than other social network platforms, such as Facebook. Twitter, while having friend and follower type relationships, is in some ways closer to a broadcast type platform than Facebook, with some very popular actors with high indegree using the platform as broadcast platform, instead of an interactive platform with back and forth among users (Cha, Haddadi, Benevenuto, & Gummadi, 2010).

Do protest movements in western democratic contexts tend to have hierarchical or polycentric network structures? One of the major advantages of polycentric networks is their robustness; their resilience even if key nodes are removed. However, this is less of an advantage in a democratic context, where governments are not typically trying to arrest or otherwise remove protest or social movement leaders. Hierarchical networks may therefore be more advantageous in democratic settings. Based on existing research, my expectation is that communication networks related to protests in democratic settings are more likely to be hierarchical (Benkler et al., 2015; R. Faris et al., 2016). The findings from this thesis also show that network structure changes over time, so it would also be useful to understand whether this hypothesized hierarchical structure in democratic contexts is consistent, or if it changes over time.

Do successful movement networks develop over time as predicted by traditional diffusion theory, with a small network of dedicated activists transforming over time into a mass movement as it gains broader support? Theory suggests that movements will begin as small hierarchical networks, with a core of hard-core supporters, and grow over time into movement with more of a polycentric network structure (Granovetter, 1978). This would be the opposite of what I found in this thesis, but might be more common among movements that largely achieve their goals, are not put down by authorities, or that operate in a democratic context.

As networks become more hierarchical, does the text among nodes in communities become more similar? Another way to think of this question is whether the existence of a broker in a community leads to more diverse language or topics compared to communities without them. One way to test this could be to manually remove brokers from a community in an empirical network, and see if this leads to less diverse text in that community. Based on the findings from Chapter 6, I expect that communities with fewer brokers will have less novel text, and since hierarchical networks tend to have fewer brokers, hierarchical networks should therefore also have less novel text.

Are extremists more likely than moderates to cluster together in isolated, strong-tie communities with a lack of brokers and connections to outside groups? Intuitively, one would expect that those who hold extremist views that are not discussed openly in public are more likely to have communication networks with strong tie, isolated community structures. Future research could compare the moderate and extremist communities identified in this thesis, to see if moderate, mainstream communities tend to have more brokers and more weak-tie connections to different outside groups than extremists.

Limitations

The main limitation of this study is its reliance on just one case. However, I would argue that by focusing on the sociological processes at play in a movement in a high-risk political setting, the findings should be applicable far beyond Russia's borders. This case was internally also very comparative, by looking at different constituencies within the movement, and by also looking at the evolution of the movement over time. Nonetheless, the findings from this thesis will be strengthened by comparison to additional cases of collective action in high-risk political settings outside of Russia. It is also important to note that the findings in this thesis are most directly relevant to high-risk political settings, and may not be as relevant to democratic contexts. Where movements are not fighting against entrenched state actors who can jail or otherwise forcibly remove protest activists opposed to them, the risks to participation are much lower, and the type of network structure might be much different. Indeed, other studies that use cases of collective action drawn from Western democratic contexts show that communication networks are more likely to have a hierarchical structure (Barbera et al., 2015; Benkler et al., 2015; R. Faris et al., 2016). A commercially driven, but otherwise free media system, could also present different challenges for movements as they try to set the media agenda and advance their preferred frames.

Another limitation relates to the data and was raised in Chapter 4. It is possible that more data from additional social media sources could change the results and conclusions I arrived at in this thesis. Although I drew the initial data set from a very large and diverse set of blogs, online media, and government Web sites, I did not include data from Twitter, Facebook, or Vkontakte in the initial crawl. While protests pages on Vkontakte were discovered in that crawl and are highly linked to, Twitter and Facebook data were minimal. Future research could attempt to include data from these sources. Future research could also look at the differences of the discussion on different platforms, and how information flows between these platforms.

This thesis focused on collective action. While I did directly take on one of the claims made by skeptics, that authoritarian regimes are able to shape online discourse in ways that benefit them, I did not explore other ways that governments use the Internet and social media to gain an advantage over opposition movements. Clearly, governments have other online and offline methods at their disposal, including DDOS attacks (Soldatov & Borogan, 2015), surveillance of users and blocking of online content (Deibert, Palfrey, Rohozinski,

& Zittrain, 2008), the use of bots on Twitter (J. Kelly et al., 2012; Sanovich, Stukal, Penfold-Brown, & Tucker, 2015), employing trolls to undermine regime critics (Chen, 2015), blackmail and anonymous harassment of the opposition (Pearce, 2015), and enacting laws that limit critical online speech (Deibert et al., 2010; Soldatov & Borogan, 2015). As this thesis was not a study of how authoritarian governments use the Internet to further entrench their rule, I did not have space to take on all these claims. Governments may therefore have the upper hand in other ways that are not explored in this thesis. As already noted, it is important to understand the impact of such tactics, and to go beyond merely identifying the existence of government surveillance or filtering of online content.

Conclusion

Examples of collective action facilitated by networked communication technologies continue to emerge and demand our attention. In this thesis, I showed how it is possible to use online data and computational social science methods to understand the role of network structure, brokerage, and framing in collective action. I focused on high-risk collective action in a hybrid regime, as a contribution to other studies that have tended to focus on more democratic or authoritarian contexts. The findings suggest that movements in high-risk political contexts should try to create flatter, more diffuse networks by ensuring that brokers tie together diverse protest constituencies. The findings also provide evidence against those who claim that authoritarian governments are more effective in shaping online discourse than oppositional movements, and also suggest that movements should advance moderate framing in order to attract a wider base of support among the general population. I showed the importance of looking at movements over the longer term, and investigating their entire cycle of contention. My hope is that I have offered a contribution to the growing body of work that uses computational methods and empirical evidence to understand the role of networked communication technologies in collective action. Much work remains for this growing and exciting field.

References

Adamic, L. A., & Glance, N. (2005). *The political blogosphere and the 2004 U.S. Election: Divided they blog.* Paper presented at the LinkKDD-2005: Proceedings of the the 3rd International Workshop on Link Discovery, Chicago, IL.

Alexanyan, K., Barash, V., Etling, B., Faris, R., Gasser, U., Kelly, J., . . . Roberts, H. (2012). *Exploring Russian cyberspace: Digitally-mediated collective action and the networked public sphere*. Berkman Center Research Publication. Berkman Center for Internet & Society. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2014998

Ansell, C. (2003). Community embeddedness and collaborative governance in the San Francisco Bay Area environmental movement. In M. Diani & D. Mcadam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 123-146). Oxford, UK: Oxford University Press.

Aral, S. (2016). The future of weak ties. *American Journal of Sociology, 121*(6), 1931-1939.

Aral, S., & Van Alstyne, M. (2011). The diversity-bandwidth trade-off. *American Journal of Sociology, 117*(1), 90-171.

Asmolov, G. (2014a). Natural disasters and alternative modes of governance: The role of social networks and crowdsourcing platforms in Russia. In S. Livingston & G. Wlater-Drop (Eds.), *Bits and atoms: Information and communication technology in areas of limited statehood*. New York, NY: Oxford University Press.

Asmolov, G. (2014b). Virtual rynda - the atlas of help: Mutual aid as a form of social activism. In E. Zuckerman & L. LeJeune (Eds.), *Global dimensions of digital activism*. Cambridge, MA: MIT Center for Civic Media.

Baldassari, D., & Diani, M. (2007). The integrative power of civic networks. *American Journal of Sociology, 113*(3), 735-780.

Balmforth, T. (2012a, February 19). The paradox of Russia's left. *Radio Free Europe/Radio Liberty*. Retrieved from http://www.rferl.org/content/the_paradox_of_russias_left/24488988.html

Balmforth, T. (2012b, April 24). The political metamorphisis of Oleg Shein. *Radio Free Europe/Radio Liberty*. Retrieved from http://www.rferl.org/a/political_metamorphosis_of_oleg_shein/24558401.html

Barakovskaia, K. (2011, December 6). They protested at Chistye Prudy with ipads in their hands. *Vedomosti*. Retrieved from http://www.vedomosti.ru/politics/news/1444277/policiya_presekla_pohod_oppozicii_na_cik

Barbera, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & Gonzalez-Bailon, S. (2015). The critical periphery in the growth of social protests. *Plos One, 10*(11).

Barry, E. (2011, December 10). Rally defying putin's party draws tens of thousands. *New York Times*. Retrieved from http://www.nytimes.com/2011/12/11/world/europe/thousands-protest-in-moscow-russia-in-defiance-of-putin.html?pagewanted=all

Barry, E., & Kramer, A. (2012, February 4). In biting cold, protesters pack the center of Moscow. *The New York Times*. Retrieved from http://www.nytimes.com/2012/02/05/world/europe/tens-of-thousands-protest-putin-in-moscow-russia.html?_r=0

Barry, E., & Schwirtz, M. (2012, May 6). Arrests and violence at overflowing rally in Moscow. *New York Times*. Retrieved from http://www.nytimes.com/2012/05/07/world/europe/at-moscow-rally-arrests-and-violence.html

Batty, D. (2011, December 10). Russian election protests -- follow live updates. *The Guardian*. Retrieved from http://www.theguardian.com/global/2011/dec/10/russia-elections-putin-protest

Benford, R., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology, 26*, 611-639.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2015). Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. *Political Communication, 32*(4), 594-624.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action. *Information Communication & Society, 15*(5), 739-768.

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action*. Cambridge, UK: Cambridge University Press.

Bennett, W. L., Segerberg, A., & Walker, S. (2014). Organization in the crowd: Peer production in large-scale networked protests. *Information, Communication & Society, 17*(2).

Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *American Sociological Review, 62*(3), 366-385.

Bimber, B., Flanagin, A., & Stohl, C. (2012). *Collective action in organizations: Interaction and engagement in an era of technological change*. New York, NY: Cambridge University Press.

Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Block, P. (2015). Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks, 40*, 163-173.

Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), 10008.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. Los Angeles: Sage.

Borgatti, S. P., & Lopez-Kidwell, V. (2011). Network theory. In J. Scott & P. J. Carrington (Eds.), *The sage handbook of social network analysis* (pp. 40-55). London: Sage.

Boyd, D., & Ellison, N. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230.

Boykoff, J. (2006). Framing dissent: Mass-media coverage of the global justice movement. *New Political Science, 28*(2), 201-228.

Boykoff, J., & Laschever, E. (2011). The Tea Party movement, framing, and the US media. *Social Movement Studies: Journal of Social, Cultural and Political Protests, 10*(4), 341-366.
Brooks, B., Hogan, B., Ellison, N., Lampe, C., & Vitak, J. (2014). Assessing structural correlates to social capital in Facebook ego networks. *Social Networks, 38*, 1-15.
Bruggeman, J. (2016). The strength of varying tie strength: Comment on Aral and Van Alstyne. *American Journal of Sociology, 121*(6), 1919-1930.
Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*(2), 349-399.
Burt, R. S. (2010). *Neighbor networks: Competitive advantage local and personal*. Oxford: Oxford University Press.
Cartwright, D., & Harary, F. (1966). Structural balance: A generalization of Heider's theory. *Psychological Review, 63*, 277-293.
Castells, M. (2012). *Networks of outrage and hope: Social movements in the Internet age*. Cambridge, UK: Polity Press.
Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*, 1194-1197.
Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology, 113*(3), 702-734.
Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). *Measuring user influence in Twitter: The million follower fallacy*. Paper presented at the Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Menlo Park, CA.
Chadwick, A. (2007). Digital network repertoires and organizational hybridity. *Political Communication, 24*(3), 283-301.
Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford, UK: Oxford University Press.
Chaisty, P., & Whitefield, S. (2013). Forward to democracy or back to authoritarianism? The attitudinal bases of mass support for the Russian election protests of 2011-2012. *Post-Soviet Affairs, 29*(5), 387-403.
Chen, A. (2015). The agency. *The New York Times Magazine*.
Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), 066111.
Coalson, R. (2013, July 28). Russia's Aleksei Navalny: Hope of the nation -- or the nationalists? *Radio Free Europe/Radio Liberty*. Retrieved from http://www.rferl.org/content/russia-navalny-nationalist-fears/25059277.html
Copsey, N. (2003). Extremism on the Net: The extreme right and the value of the Internet. In R. Gibson, P. Nixon, & S. Ward (Eds.), *Political parties and the Internet: Net gain?* (pp. 218-233). London, UK: Routledge.
Crossley, N. (2002). *Making sense of social movements*. Buckingham, UK: Open University Press.
Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (2008). *Access denied*. Cambridge, MA: MIT Press.

Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (2010). *Access controlled: The shaping of power, rights, and rule in cyberspace*. Cambridge, MA: MIT Press.

Denisova, A. (2016). Democracy, protest, and public sphere in Russia after the 2011-2012 anti-government protests: Digital media at stake. *Media, Culture & Society, Advance online publication*. doi:https://doi.org/10.1177/0163443716682075

Diamond, L. (2002). Thinking about hybrid regimes. *Journal of Democracy, 13*(2), 21-35.

Diamond, L. (2008). *The spirit of democracy: The struggle to build free societies throughout the world* (1st ed.). New York, NY: Times Books/Henry Holt and Co.

Diani, M. (2003). 'Leaders' or brokers? Positions and influence in social movement networks. In M. Diani & D. McAdam (Eds.), *Social movements and social networks: Relational approaches to collective action* (pp. 105-122). Oxford, UK: Oxford University Press.

Diani, M. (2010). Social movement networks virtual and real. *Information Communication & Society, 3*(3), 386-401.

Diermeier, D., Godbout, J.-F., Yu, B., & Kaufmann, S. (2012). Language and ideology in Congress. *British Journal of Political Science, 42*(1), 31-55.

Dmitriev, M. (2015). Lost in translation? The geography of protests and attitude change in Russia. *Europe-Asia Studies, 67*(2), 224-243.

Dow, M. M., Burton, M. L., & White, D. R. (1982). Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks, 4*(2), 169-200.

Earl, J. (2015). The future of social movement organizations. *American Behavioral Scientist, 59*(1), 35-52.

Earl, J., & Kimport, K. (2011). *Digitally enabled social change: Activism in the Internet age*. Cambridge, MA: MIT Press.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*(397), 171-185.

Elder, M. (2011a, December 6). Pro-Kremlin activists outnumber pro-democracy demonstrators in Moscow after disputed election. *The Guardian*. Retrieved from http://www.theguardian.com/world/2011/dec/06/russian-election-anti-putin-protests

Elder, M. (2011b, December 9). Putin faces wave of protests as opposition calls for new Russian elections. *The Guardian*. Retrieved from https://www.theguardian.com/world/2011/dec/09/vladimir-putin-protests-opposition-elections

Elder, M. (2013, May 5). Russia protests: One year on, anti-Putin activists are still awaiting trial. *The Guardian*. Retrieved from https://www.theguardian.com/world/2013/may/05/russia-anti-putin-protests-arrests

Elder, M., & Parfit, T. (2011, December 24). Russian anti-Putin protests draw thousands to Moscow again. *The Guardian*. Retrieved from https://www.theguardian.com/world/2011/dec/24/russia-europe-news

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51-58.

Entman, R. M. (2004). *Projections of power: Framing news, public opinion, and U.S. Foreign policy*. Chicago: University of Chicago Press.
Etling, B. (2014). *Russia, ukraine, and the west: Social media sentiment in the euromaidan protests*. Berkman Center Working Paper Series. Berkman Center for Internet & Society. Cambridge, MA.
Etling, B., Alexanyan, K., Kelly, J., Faris, R., Palfrey, J., & Gasser, U. (2010). *Public discourse in the Russian blogosphere: Mapping runet politics and mobilization*. Berkman Center Research Publication 2010-11. Berkman Center for Internet & Society, Harvard University. Cambridge, MA.
Etling, B., Roberts, H., & Faris, R. (2014). *Blogs as an alternative public sphere: The role of blogs, mainstream media, and TV in Russia's media ecology*. Berkman Center Research Publication 2014-8. Berkman Center for Internet & Society, Harvard University. Retrieved from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2427932
Faris, D. (2008). Revolutions without revolutionaries? Network theory, Facebook, and the Egyptian blogosphere. *Arab Media & Society*(6).
Faris, R., Roberts, H., Etling, B., Othman, D., & Benkler, Y. (2016). The role of the networked public sphere in the U.S. Net neutrality policy debate. *International Journal of Communication, 10*, 5839-5864.
Farrell, H. (2012). The consequences of the Internet for politics. *Annual Review of Political Science, 15*, 35-52.
Feinerer, I., & Hornik, K. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1-54.
Feinerer, I., & Hornik, K. (2015). Tm: Text mining package (Version 0.6-2) [R package]. Retrieved from http://CRAN.R-project.org/package=tm
Fossato, F., & Lloyd, J. (2008). *The Web that failed: How opposition politics and independent initiatives are failing on the Internet*. RISJ Challenges. Reuters Institute for the Study of Journalism, University of Oxford. Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/fileadmin/documents/Publications/The_Web_that_Failed.pdf
Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35-41.
Gans, H. (1961). *The urban villagers*. New York: Free Press.
Gel'man, V. (2007). The transformation of Russia's party system. *Russian Analytical Digest, 19*(7), 12-14.
Gerlach, L. (2001). The structure of social movements: Environmental activism and its opponents. In J. Arquilla & D. Ronfeldt (Eds.), *Networks and netwars: The future of terror, crime and militancy* (pp. 289-310). Santa Monica, CA: RAND.
Giles, J. (2012). Computational social science: Making the links. *Nature, 488*(7412), 448-450.
Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821-7826.
Gitlin, T. (1980). *The whole world is watching: Mass media in the making and unmaking of the new left.* . Berkeley, California: University of California Press.

Gladwell, M. (2010). Small change: Why the revolution will not be tweeted. *The New Yorker*.
Goffman, E. (1974). *Frame analysis*. Cambridge, MA: Harvard University Press.
Goldstone, J. A., & Useem, B. (1999). Prison riots as microrevolutions: An extension of state-centered theories of revolution. *American Journal of Sociology, 104*(4), 985-1029.
Gonzalez-Bailon, S., Borge-Holthoefer, J., & Moreno, Y. (2013). Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist, 57*(7), 943-965.
Gonzalez-Bailon, S., & Wang, N. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks, 44*, 95-104.
Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., & Morris, M. (2008). A statnet tutorial. *Journal of Statistical Software, 24*(9).
Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography, 46*(1), 103-125.
Gould, R. (1991). Multiple networks and mobilization in the Paris commune, 1871. *American Sociological Review, 56*, 716-729.
Gould, R., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *American Journal of Sociology, 19*, 89-126.
Graeff, E., Stempeck, M., & Zuckerman, E. (2014). The battle for "Trayvon Martin": Mapping a media controversy online and off-line. *First Monday, 19*(2).
Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380. doi:Doi 10.1086/225469
Granovetter, M. (1974). *Getting a job: A study of contact and careers*. Chicago: University of Chicago Press.
Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*(6), 1420-1443.
Granovetter, M. (1983). The strength of weak ties: A network thoery revisited. *Sociological Theory, 1*, 201-233.
Greene, S. A. (2012). *Twitter and the Russian street: Memes, networks and mobilization*. CNMS Working Paper 2012/1. Center for New Media and Society. Moscow. Retrieved from http://www.newmediacenter.ru/2012/05/22/twitter-and-the-russian-street-memes-networks-mobilization/
Greene, S. A. (2013). Beyond Bolotnaia: Bridging old and new in Russia's election protest movement. *Problems of Post-Communism, 60*(2), 40-52.
Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267-297.
Gunitsky, S. (2015). Corrupting the cyber-commons. *Perspectives on Politics, 13*(1), 42-54.
Halfmann, D., & Young, M. P. (2010). War pictures: The grotesque as a mobilizing tactic. *Mobilization, 15*(1), 1-24.
Hamdy, N., & Gomaa, E. (2012). Framing the Egyptian uprising in Arabic language newspapers and social media. *Journal of Communication, 62*(2), 195-211.

Han, S. K. (2009). The other ride of Paul Revere: The brokerage role in the making of the American revolution. *Mobilization, 14*(2), 143-162.
Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2003). Statnet: Software tools for the statistical modeling of network data. Retrieved from http://statnetproject.org
Hanna, A. (2013). Computer-aided content analysis of digitally enabled movements. *Mobilization, 18*(4), 367-388.
Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside.
Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge accross organizational subunits. *Administrative Science Quarterly, 44*(1), 82-111.
Harris, J. K. (2014). *An introduction to exponential random graph modeling*. Los Angeles, CA: Sage.
Hedstrom, P., Sandell, R., & Stern, C. (2000). Mesolevel networks and the diffusion of social movements: The case of the Swedish Social Democratic Party. *American Journal of Sociology, 106*(1), 145-172. doi:Doi 10.1086/303109
Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
Herszenhorn, D. (2012a, April 10). Moscow protesters try to expand movement. *The New York Times*. Retrieved from http://www.nytimes.com/2012/04/11/world/europe/astrakhan-hunger-striker-gets-support-from-moscow-protesters.html
Herszenhorn, D. (2012b, March 4). Putin wins, but opposition keeps pressing. *The New York Times*. Retrieved from http://www.nytimes.com/2012/03/05/world/europe/russia-votes-in-presidential-election.html
Hindman, M. (2008). *The myth of digital democracy*. Princeton, NJ: Princeton University Press.
Holland, P. W., & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology, 76*(3), 492-513.
Hopkins, D., & King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science, 54*(1), 229-247.
Hopkins, D., King, G., Knowles, M., & Melendez, S. (2013). ReadMe: Software for automated content analysis. Retrieved from http://gking.harvard.edu/readme
Huang, R., & Yip, N.-M. (2012). Internet activism in urban China: A case study of protests in Xiamen and Panyu. *Journal of Comparative Asian Development, 11*(2), 201-223.
Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks, 29*(2), 216-230.
Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics, 15*(3), 565-583.
Hussain, M. M., & Howard, P. N. (2013). What best explains successful protest cascades? ICTs and the fuzzy causes of the Arab Spring. *International Studies Review, 15*(1), 48-66. doi:Doi 10.1111/Misr.12020

I was a participant in the falsification of the 2011 elections. (2011, December 12, 2011). *Novaya Gazeta*. Retrieved from http://www.novayagazeta.ru/inquests/49943.html
Ioffe, J. (2011a). The Decembrists. *Foreign Policy*.
Ioffe, J. (2011b). Putin's big mistake? *The New Yorker*.
Ioffe, J. (2011c). Won't get fooled again. *Foreign Policy*.
Ioffe, J. (2012). Protest and pretend in Moscow. *The New Yorker*.
Kalathil, S., & Boas, T. (2003). *Open networks, closed regimes: The impact of the Internet on authoritarian rule*. Washington, DC: Carnegie Endowment for International Peace.
Karpf, D. (2012). *The moveon effect: The unexpected transformation of American political advocacy*. Oxford, UK: Oxford University Press.
Karpf, D. (2016). *Analytic activism*. Oxford, UK: Oxford University Press.
Kelly, J., Barash, V., Alexanyan, K., Etling, B., Faris, R., Gasser, U., & Palfrey, J. (2012). *Mapping Russian Twitter*. Berkman Center Research Publication No. 2012-03. Berkman Center for Internet & Society. Cambridge, MA. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2028158
Kelly, L., & De Carbonnel, A. (2012). Hundreds of anti-Putin protesters detained in Russia. *Reuters*.
Kim, H., & Pfaff, S. (2012). Structure and dynamics of religious insurgency: Students and the spread of the Reformation. *American Sociological Review, 77*(2), 188-215.
King, G., & Lu, Y. (2008). Verbal autopsy methods with multiple causes of death. *Statistical Science, 23*(1), 78-91.
King, G., Pan, J., & Roberts, M. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review, 107*(2), 326-343.
King, G., Pan, J., & Roberts, M. E. (2016). *How the Chinese government fabricates social media posts for strategic distraction, not engaged argument*. Working Paper. Harvard University. Cambridge, MA. Retrieved from http://j.mp/1Txxiz1
Koesel, K. J., & Bunce, V. J. (2012). Putin, popular protests, and political trajectories in Russia: A comparative perspective. *Post-Soviet Affairs, 28*(4), 403-423.
Koltsova, O., & Koltcov, S. (2013). Mapping the public agenda with topic modeling: The case of Russian LiveJournal. *Policy & Internet, 5*(2), 207-227.
Koskinen, J., & Daraganova, G. (2013). Exponential random graph model fundamentals. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods, and applications* (pp. 49-76). Cambridge, UK: Cambridge University Press.
Koskinen, J., & Snijders, T. A. B. (2013). Simulation, estimation, and goodness of fit. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods, and applications* (pp. 141-166). Cambridge, UK: Cambridge University Press.
Krebs, B. (2011). Twitter bots drown out anti-Kremlin tweets. Retrieved from http://krebsonsecurity.com/2011/12/twitter-bots-drown-out-anti-kremlin-tweets/

Krinsky, J., & Crossley, N. (2014). Social movements and social networks: Introduction. *Social Movement Studies: Journal of Social, Cultural and Political Protests, 13*(1), 1-21.
Landauer, T. K., & Dumais, S. T. (1996). *Latent semantic analysis approaches to categorization.* Paper presented at the Proceedings of the 19th Annual Conference of the Cognitive Science Society, Mahwah, NJ.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259-284.
Lankina, T. (2015). The dynamics of regional and national contentious politics in Russia: Evidence from a new dataset. *Problems of Post-Communism, 62*(1), 26-44.
Lankina, T., & Voznaya, A. (2015). New data on protest trends in Russia's regions. *Europe-Asia Studies, 67*(2), 327-342.
Lapse in airtime targets suggests effort to improve Putin's image. (2013). Retrieved from Washington, DC: http://archive.constantcontact.com/fs053/1102820649387/archive/1102911694293.html
Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford, UK: Oxford University Press.
Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as social process. *Freedom and Control in Modern Society, 18*, 18-66.
Lazer, D., Pentland, A. S., Adamic, L. A., Aral, S., Barabasi, A. L., Brewer, D., . . . Van Alstyne, M. (2009). Life in the network: The coming age of computational social science. *Science, 323*(5915), 721-723.
Levada Center. (2011). *Poll on Sakharov avenue, December 24*. Retrieved from http://www.levada.ru/26-12-2011-opros-na-prospekte-sakharova-24-dekabrya
Levada Center. (2013). *More than half of the country considers UR a "party of crooks and thieves"*. Retrieved from http://www.levada.ru/29-04-2013/svyshe-poloviny-strany-schitaet-er-partiei-zhulikov-i-vorov
Levitsky, S., & Way, L. (2002). The rise of competitive authoritarianism. *Journal of Democracy, 13*(2), 51-65.
Levitsky, S., & Way, L. (2010). *Competitive authoritarianism: Hybrid regimes after the cold war*. Cambridge: Cambridge University Press.
Lewis, K., Gray, K., & Meierhenrich, J. (2014). The structure of online activism. *Sociological Science, 1*, 1-9.
Lipman, M. (2013). Alexey navalny's miraculous, doomed campaign. *The New Yorker*. Retrieved from http://www.newyorker.com/news/news-desk/alexey-navalnys-miraculous-doomed-campaign
Lipmann, M., & McFaul, M. (Eds.). (2010). *The media and political developments*. Lanham, MA: Rowman and Littlefield.
Litvinenko, A. (2012). *Role of social media in political mobilization in Russia: On the example of the parliamentary elections 2011*. Paper presented at the Conference for E-Democracy and Open Government (CeDEM12), Krems, Austria. http://www.epractice.eu/files/CeDEM12 Proceedings of the

International Conference for E-Democracy and Open Government.pdf - page=183
Lohmann, S. (1994). The dynamics of information cascades: The Monday demonstrations in Leipzig, East Germany, 1989-1991. *World Politics, 47*(1), 42-101.
Lubell, M., Robins, G., & Wang, P. (2014). Network structure and institutional complexity in an ecology of water mangement games. *Ecology and Society, 19*(4).
Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications* (D. Lusher, J. Koskinen, & G. Robins Eds.). Cambridge, UK: Cambridge University Press.
Lusher, D., & Robins, G. (2013). Formation of social network structure. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks* (pp. 16-28). Cambridge, UK: Cambridge University Press.
MacFarquhar, N. (2014, May 6). Russia quietly tightens reins on Web with 'bloggers law'. *The New York Times*. Retrieved from http://www.nytimes.com/2014/05/07/world/europe/russia-quietly-tightens-reins-on-web-with-bloggers-law.html
Mackey, R., & Kates, G. (2012, January 13). Russian ad compares Putin foe to hitler. *The New York Times*. Retrieved from http://thelede.blogs.nytimes.com/2012/01/13/russian-ad-compares-putin-foe-with-hitler/
Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.
March, L. (2012). The Russian Duma 'opposition': No drama out of crisis? *East European Politics, 28*(3), 241-255.
Margetts, H., John, P., Hale, S., & Yasseri, T. (2015). *Political turbulence: How social media shape collective action*. Princeton, NJ: Princeton University Press.
Marin, A. (2012). Don't mention it: Why people don't share job information, when they do, and why it matters. *Social Networks, 34*(2), 181-192.
Marsden, P. V., & Campbell, K. E. (1984). Measuring tie strength. *Social Forces, 63*(2), 482-501.
Marshall, A.-M. (2001). A spectrum of oppositional consciousness: Sexual harassment plaintiffs and their lawyers. In J. Mansbridge & A. Morris (Eds.), *Oppositional consciousness: The subjective roots of social protest* (pp. 99-145). Chicago, IL: Chicago University Press.
McAdam, D. (1986). Recruitment to high-risk activism - the case of Freedom Summer. *American Journal of Sociology, 92*(1), 64-90.
McAdam, D. (1988). *Freedom Summer*. New York ; Oxford: Oxford University Press.
McAdam, D., & Paulsen, R. (1993). Specifying the relationship between social ties and activism. *American Journal of Sociology, 99*(3), 640-667.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444.
Miller, B. (2014). Ukraine protests: At least two dead as president viktor yanukovich meets opposition leaders. *ABC News*. Retrieved from http://www.abc.net.au/news/2014-01-22/three-dead-in-ukraine-as-president-agrees-to-talks/5213724

Mizruchi, M. S., & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks, 30*(3), 190-200.
Morris, A., & Braine, N. (2001). Social movements and oppositional consciousness. In J. Mansbridge & A. Morris (Eds.), *Opositional consciousness: The subjective roots of social protest* (pp. 20-37). Chicago: Chicago University Press.
Morris, M., Handcock, M. S., & Hunter, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software, 24*(4), 1-24.
Neuman, E. J., & Mizruchi, M. S. (2010). Structure and bias in the network autocorrelation model. *Social Networks, 32*(4), 290-300.
Newman, M. E. J. (2006). Finding community structure using the eigenvectors of matrices. *Physical Review E, 74*(3), 036104.
Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.
Oates, S. (2013). *Revolution stalled: The political limits of the Internet in the post-soviet sphere*. Oxford: Oxford University Press.
Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge: Harvard University Press.
Opp, K.-D. (2009). *Theories of political protest and social movements: A multidisciplinary introduction, critique, and synthesis*. London ; New York: Routledge.
Osa, M. (2003). Networks in opposition: Linking organizations through activists in the Polish People's Republic. In M. Diani & D. Mcadam (Eds.), *Social movements and networks: Relational approaches to collective action* (pp. 77-104). Oxford, UK: Oxford University Press.
OSCE. (2012). *Russian Federation: Elections to the state Duma 4 December 2011. OSCE/ODIHR election observation mission final report*. Retrieved from Warsaw:
Owen, T. (2015). *Disruptive power*. Oxford, UK: Oxford University Press.
Parfit, T. (2011, December 17). Vladimir Putin's persecution campaign targets protest couple. *The Guardian*. Retrieved from http://www.theguardian.com/world/2011/dec/17/putin-protest-organiser-hunger-strike
Parfit, T. (2012, February 4). Anti-Putin protesters march through Moscow. *Guardian*. Retrieved from https://www.theguardian.com/world/2012/feb/04/anti-putin-protests-moscow-russia
Pariser, E. (2011). *The filter bubble*. New York: Penguin Press.
Pascal, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications, 10*(2), 191-218.
Pearce, K. E. (2015). Democratizing kompromat: The affordances of social media for state-sponsored harassment. *Information, Communication & Society, 18*(10), 1158-1174.
Perry, B., & Olsson, P. (2009). Cyberhate: The globalization of hate. *Information & Communications Technology Law, 18*(2), 185-199.
Popescu, N. (2012). The strange alliance of democrats and nationalists. *Journal of Democracy, 23*(3), 46-54.

Pro-Putin supporters rally in Moscow. (2011, December 12). *Euornews*. Retrieved from http://www.euronews.com/2011/12/12/pro-putin-supporters-rally-in-moscow/
Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.
Putnam, R. D., & Goss, K. A. (2002). Introduction. In R. D. Putnam (Ed.), *Democracies in flux: The evolution of social capital in contemporary societies*. Oxford: Oxford University Press.
Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E, 76*(3), 036106.
Razumovskaya, O. (2011, December 8). Russia seeks to block protest groups on social network. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052970203413304577086153882152884.html
Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E, 74*(1), 016110.
Reuter, O. J., & Szakonyi, D. (2015). Online social media and political awareness in authoritarian regimes. *British Journal of Political Science, 45*(1), 29-51.
Ringe, N., Victor, J. N., & Carman, C. J. (2014). *Bridging the information gap: Legislative member organizations as social networks in the United States and the European Union*. Ann Arbor: University of Michigan Press.
Roberts, H., & Etling, B. (2011). Coordinated DDoS attack during Russian Duma elections. Retrieved from https://blogs.law.harvard.edu/idblog/2011/12/08/coordinated-ddos-attack-during-russian-duma-elections
Roberts, H., Seymour, B., & Fish, S. (2015). *Examining tools for online public health media analysis: An Ebola case study*. Paper presented at the American Public Health Association Annual Meeting, Chicago, IL.
Robertson, G. B. (2013). Protesting Putinism: The election protests of 2011-2012 in broader perspective. *Problems of Post-Communism, 60*(2), 11-23. doi:Doi 10.2753/Ppc1075-8216600202
Robins, G., & Lusher, D. (2013). Illustrations: Simulation, estimation, and goodness of fit. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods, and applications* (pp. 167-185). Cambridge, UK: Cambridge University Press.
Robins, G., Pattison, P., & Wang, P. (2009). Closure, connectivity, and degree distributions: Eponential random graph (p*) models for directed social networks. *Social Networks, 31*(2), 105-117.
Robins, G., Snijders, T. A. B., Wang, P., Handcock, M. S., & Pattison, P. E. (2007). Recent developments in exponential random graph (p*) models for social networks *Social Networks, 29*, 192-215.
Rogers, E. (2003). *Diffusions of innovations* (Fifth Edition ed.). New York: Free Press.
Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
Rosenberg, S. (2012, March 13). Moscow protest against Putin win sees drop in support. *BBC*. Retrieved from http://www.bbc.com/news/world-europe-17323565

Rosval, M., & Bergstrom, C. (2007). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(4), 1118-1123.
Sakwa, R. (2015). Questioning control and contestation in late Putinite Russia. *Europe-Asia Studies, 67*(2), 192-208.
Sanovich, S., Stukal, D., Penfold-Brown, D., & Tucker, J. (2015). *Turning the virtual tables: Government strategies for addressing online opposition with and application to Russia*. Paper presented at the Midwest Political Science Association Annual Conference, Chicago, IL.
Schneider, S. M., & Foot, K. (2004). The Web as an object of study. *New Media & Society, 6*(1), 114-122.
Schwirtz, M. (2012a, May 12). A Russian protest leader takes center stage. *The New York Times*. Retrieved from http://www.nytimes.com/2012/05/12/world/europe/sergei-udaltsov-russian-protest-leader-takes-center-stage.html?pagewanted=all
Schwirtz, M. (2012b, February 26). Thousands ring central Moscow in anti-Putin protest. *New York Times*. Retrieved from http://www.nytimes.com/2012/02/27/world/europe/thousands-join-anti-kremlin-protest-in-moscow.html
Sherstobitov, A. (2014). *The potential of social media in Russia: From political mobilization to civic engagement*. Paper presented at the Proceedings of the 2014 Conference on Electronic Government and Open Society: Challenges in Eurasia, St. Petersburg, Russia.
Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York: Penguin.
Shirky, C. (2011). The political power of social media: Technology, the public sphere, and political change. *Foreign Affairs, 90*(1), 28-41.
Shvetsova, L. (2012). Putinism under siege: Implosion, atrophy or revolution? *Journal of Democracy, 23*(4), 19-32.
Simmel, G. (1908). *Sociology: Studies of the societal forms*. Berlin: Duncker & Humblot.
Simmel, G. (1950). *The sociology of Georg Simmel*. New York: Free Press.
Smyth, R. (2013). *Two models of mobilization: The organization of pro- and anti-regime street actions in Russia*. Paper presented at the Talk at Russian and Eurasian Studies Centre Seminar, St. Antony's College, University of Oxford.
Smyth, R., & Oates, S. (2015). Mind the gaps: Media use and mass action in Russia. *Europe-Asia Studies, 67*(2), 285-305.
Smyth, R., Sobolev, A., & Soboleva, I. (2013). A well-organized play. *Problems of Post-Communism, 60*(2), 24-39.
Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*(1), 99-153.
Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44-60.
Snow, D. A., Rochford, B., Jr., Worden, S., & Benford, R. (1986). Frame alignment processes, micromobilization, and movement participation. *American Sociological Review, 51*(4), 464-481.

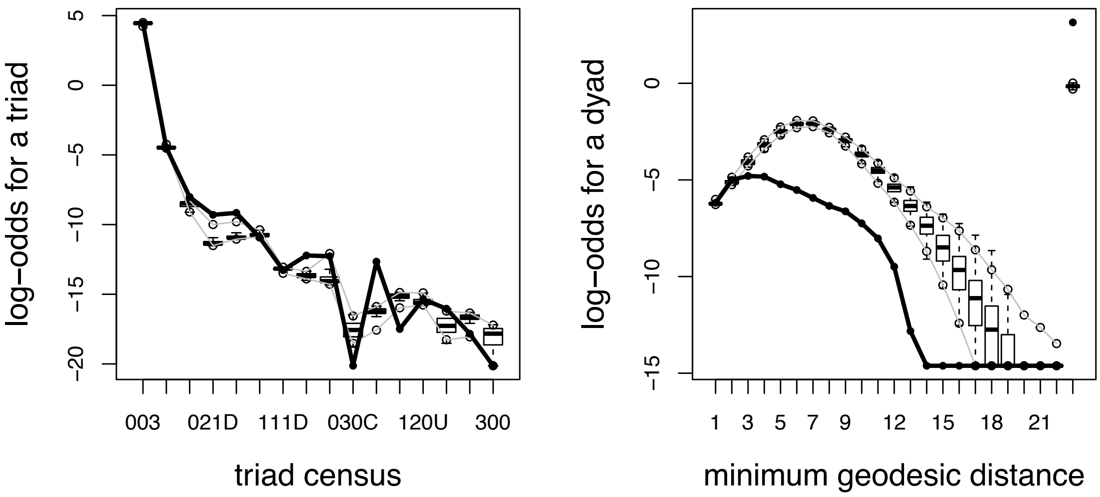
Soldatov, A., & Borogan, I. (2015). *The red web*. New York, NY: Public Affairs.
Suk-Young Chwe, M. (1999). Structure and strategy in collective action. *American Journal of Sociology, 105*(1), 128-156.
Sullivan, A. (2009). The revolution will be twittered. Retrieved from http://www.theatlantic.com/daily-dish/archive/2009/06/the-revolution-will-be-twittered/200478/
Sunstein, C. (2007). *Republic 2.0*. Princeton: Princeton University Press.
Tarrow, S. G. (1998). *Power in movement: Social movements and contentious politics* (2nd ed.). Cambridge: Cambridge University Press.
Tarrow, S. G. (2013). *The language of contention: Revolutions in words, 1688-2012*. Cambridge, UK: Cambridge University Press.
Taylor, A. (2011, December 12). Russian election protests. *The Atlantic*. Retrieved from http://www.theatlantic.com/photo/2011/12/russian-election-protests/100206/
Tucker, J. A. (2007). Enough! Electoral fraud, collective action problems, and post-communist colored revolutions. *Perspectives on Politics, 5*(3), 535-551.
Tufekci, Z. (2010). What Gladwell gets wrong: The real problem is scale mismatch (plus, weak and strong ties are complementary and supportive). Retrieved from http://technosociology.org/?p=178
Tufekci, Z. (2014). After the protests. *New York Times*. Retrieved from https://www.nytimes.com/2014/03/20/opinion/after-the-protests.html?_r=0
Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication, 62*, 363-379.
Umland, A. (2012). Could Russia's ultranationalists subvert pro-democracy protests? *World Affairs Journal*.
Volkov, D. (2012). The protesters and the public. *Journal of Democracy, 23*(3), 55-62.
Volkov, D. (2015). The protest movement in Russia 2011-2013. In C. Ross (Ed.), *Systemic and non-systemic opposition in the Russian Federation: Civil society awakens?* (pp. 35-50). Surrey, England: Ashgate.
Wang, W., Neuman, E. J., & Newman, D. A. (2014). Statistical power of the social network autocorrelation model. *Social Networks, 38*, 88-99.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
Watts, D. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology, 105*(2), 493-527.
Watts, D. (2003). *Six degrees: The science of a connected age*. New York: W.W. Norton & Company.
Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.
Wegren, S. K., & Herspring, D. R. (2010). *After Putin's Russia: Past imperfect, future uncertain*. Lanham, MA: Rowman and Littlefield Publishers.
Weir, F. (2012). Pro-Putin rally in Moscow bring outs lots of people, little enthusiasm. *Christian Science Monitor*.

White, D. (2015). Political opposition in Russia: The challenges of mobilisation and the political-civil society nexus. *East European Politics, 31*(3), 314-325.
White, S., & McAllister, I. (2014). Did Russia (nearly) have a Facebook revolution in 2011? Social media's challenge to authoritarianism. *Politics, 34*(1), 72-84.
Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology, 116*(2), 583-642.
Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society, 12*(4), 637-655.
Yang, G. (2009). *The power of the Internet in China: Citizen activism online*. New York, NY: Columbia University Press.
Zayani, M. (2015). *Networked publics and digital contention: The politics of everyday life in tunisia*. Oxford, UK: Oxford University Press.

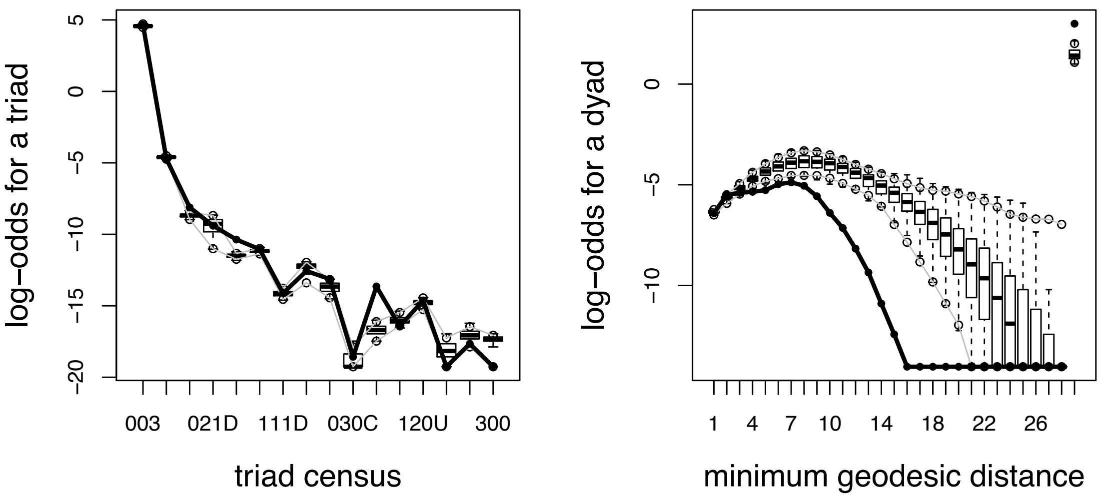
Appendices

Appendix A: Goodness-of-Fit Diagnostic Plots for Each Time Period

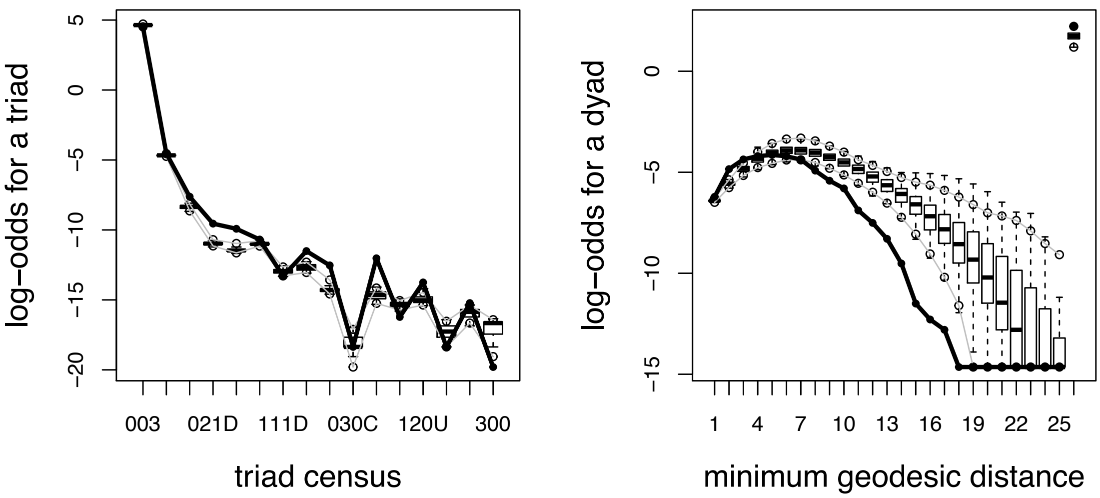
December



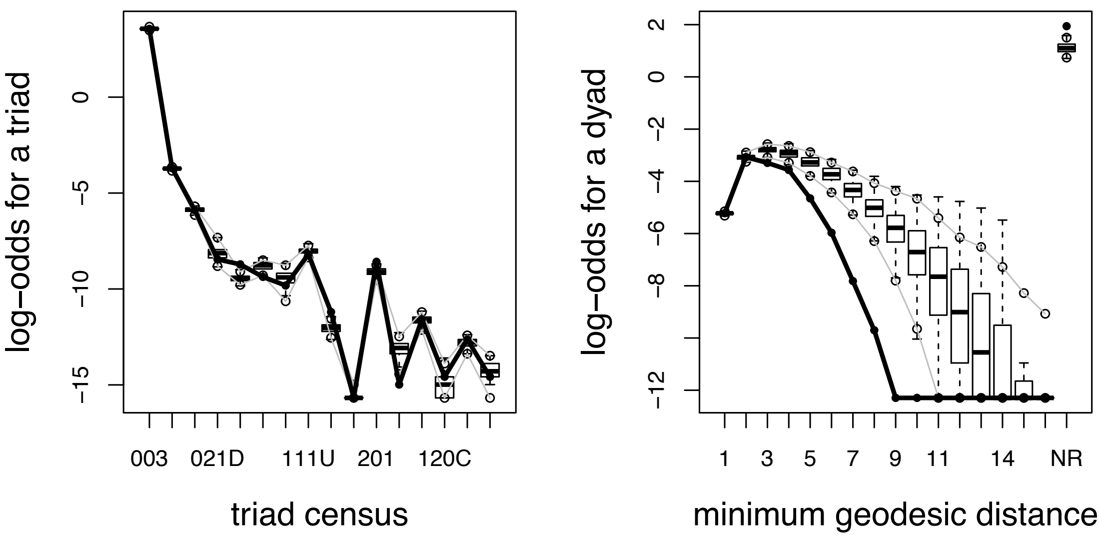
February



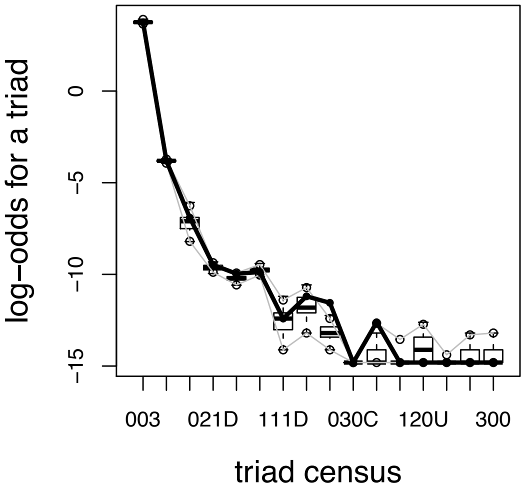
March

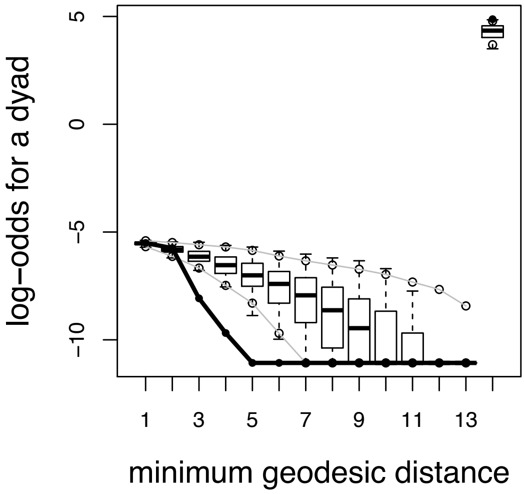

log−odds for a triad
triad census
003
021D
111D
030C
120U
300
log−odds for a dyad
minimum geodesic distance
1
4
7
10
13
16
19
22
25

April

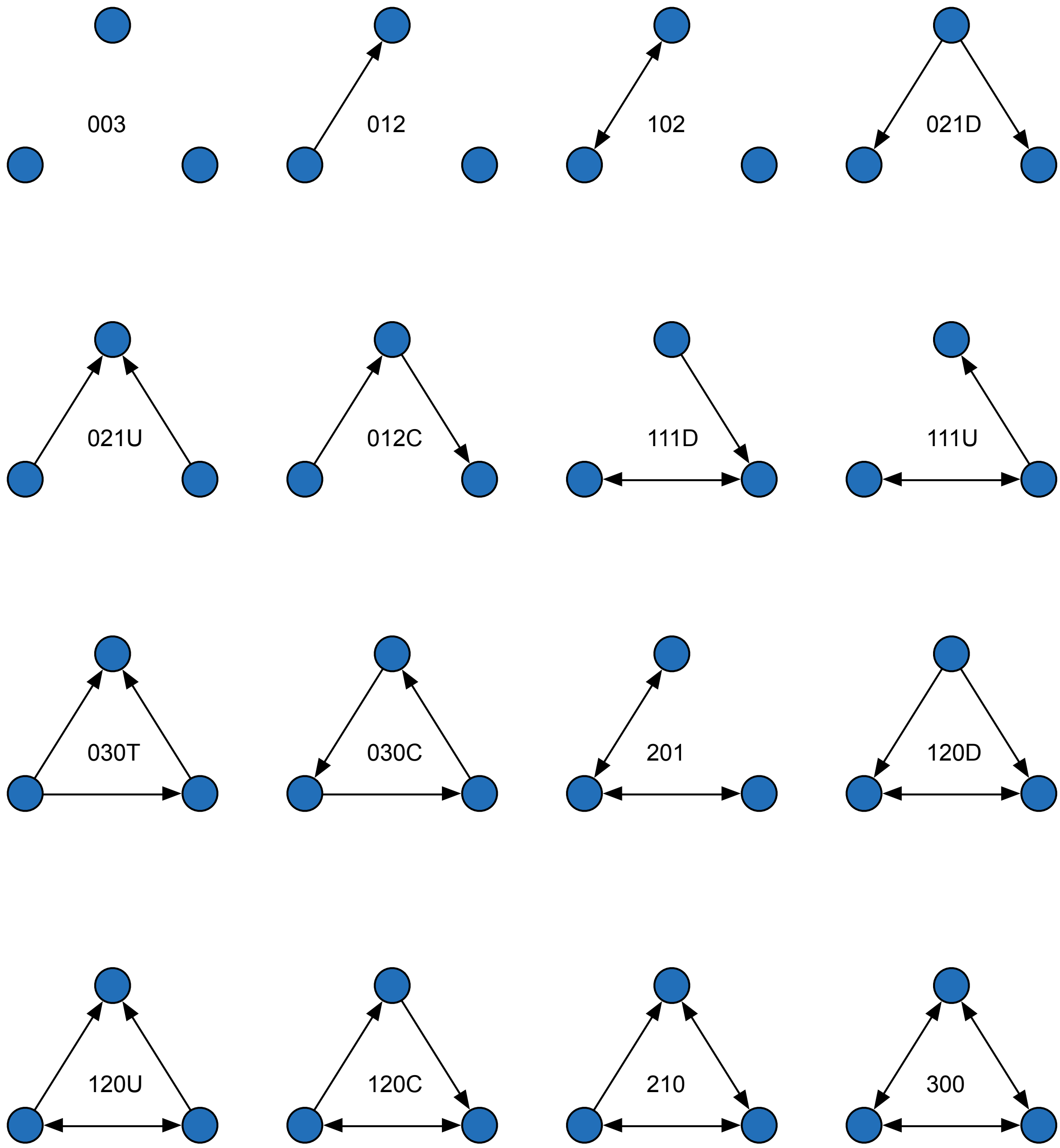

log−odds for a triad
triad census
003
021D
111U
201
120C
log−odds for a dyad
minimum geodesic distance
1
3
5
7
9
11
14
NR

May

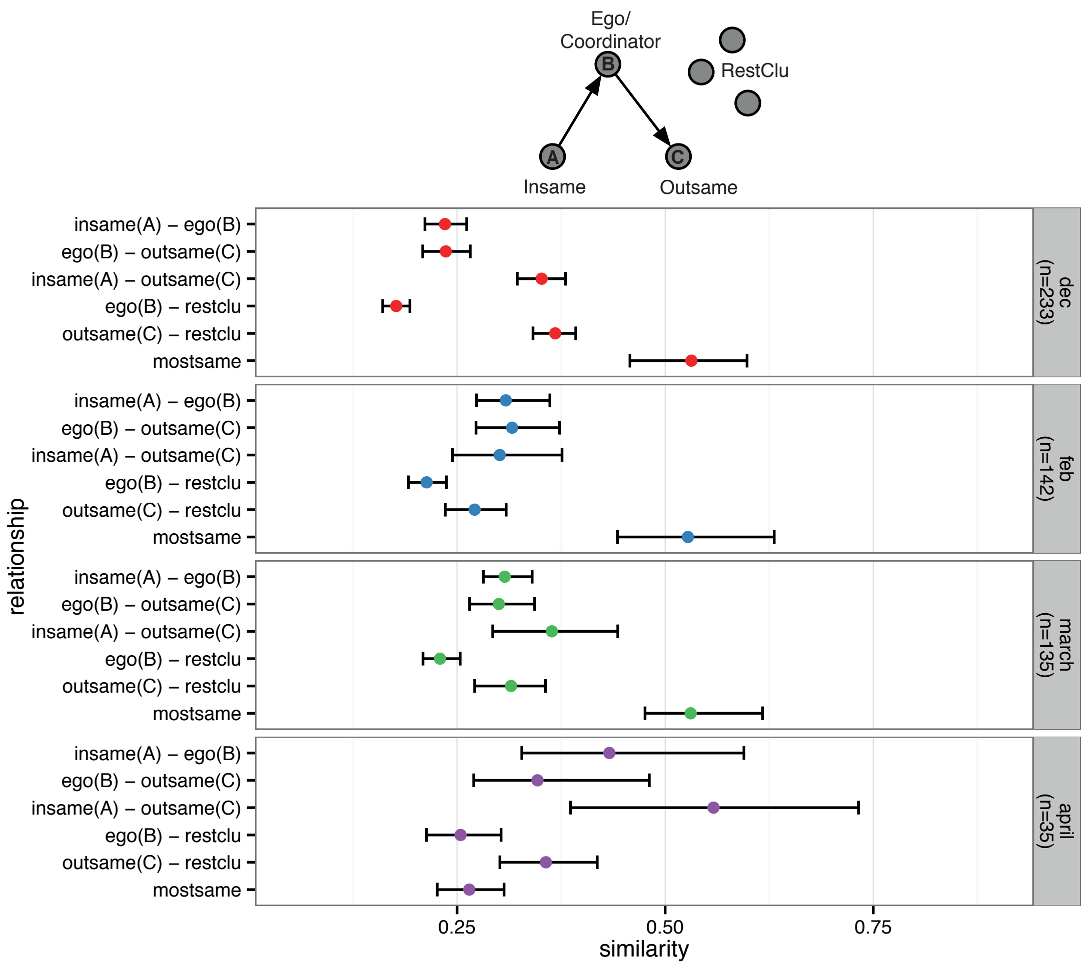

log−odds for a triad
0
−5
−10
−15
003
021D
111D
030C
120U
300
triad census


log−odds for a dyad
5
0
−5
−10
1
3
5
7
9
11
13
minimum geodesic distance

Appendix B: Triad Types for Directed Network

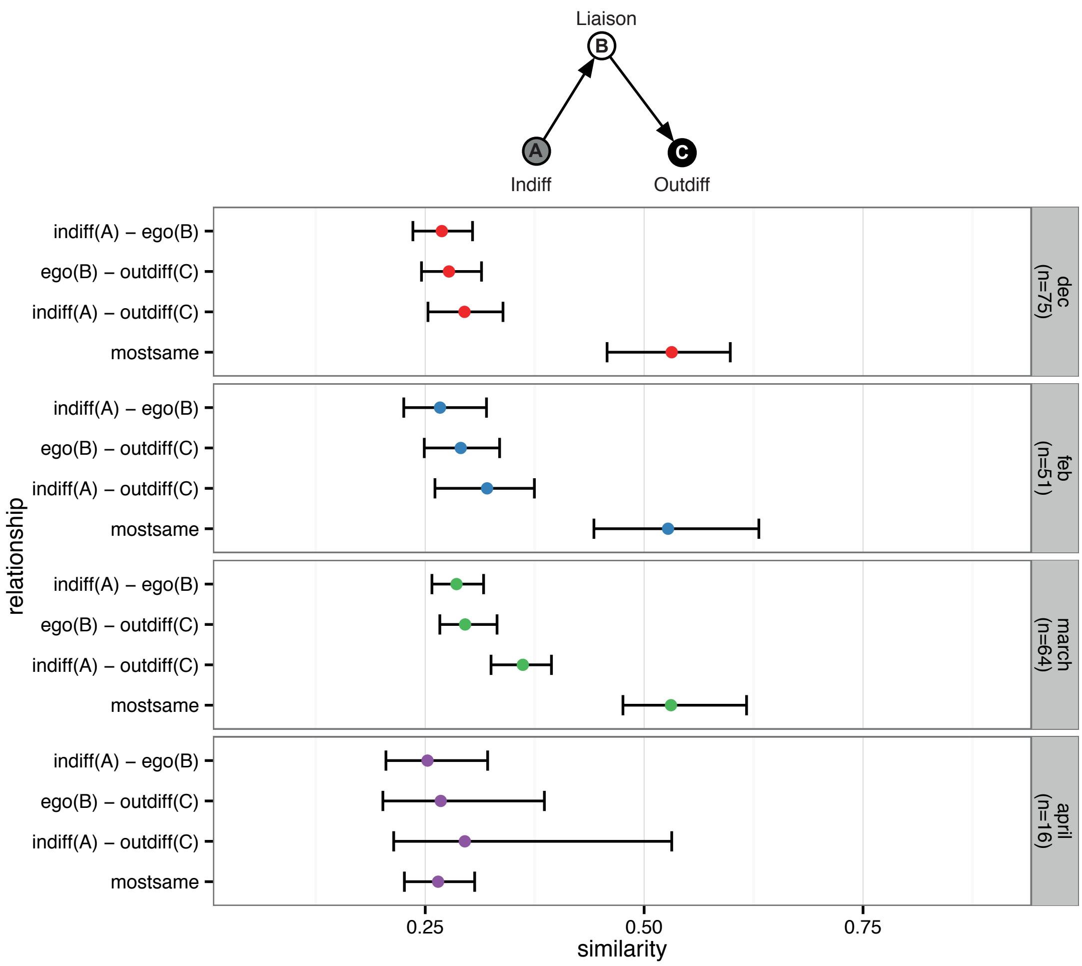


Appendix C: Coordinator similarity scores by month (means)



Mean coordinator similarity scores by month. The number of coordinator ego networks for each month is found on each facet label. Error bars indicate the bootstrap bias-corrected accelerated (BCa) confidence intervals at the 95% confidence level.

Appendix D: Liaison similarity scores by month (means)



Mean liaison similarity scores by month. The number of liaison ego networks for each month is found on each facet label. Error bars indicate the bootstrap bias-corrected accelerated (BCa) confidence intervals at the 95% confidence level.

Appendix E: Word Clouds with Translated Text from Training Set

Mainstream

Pro-Gov

Environmental

LGBT

Extremist

KPRF

Appendix F: Weekly Category Proportions

Table A1: Weekly Category Proportions

Week	Pro-Gov	Off Topic	Mainstream	KPRF	Nationalist	Neutral	Environment	LGBT
Dec 1-7	0.20	0.21	0.18	0.16	0.07	0.08	0.05	0.04
Dec 8-14	0.16	0.18	0.28	0.12	0.10	0.08	0.05	0.03
Dec 15-21	0.17	0.22	0.22	0.09	0.08	0.10	0.08	0.03
Dec 22-31	0.15	0.23	0.25	0.08	0.09	0.10	0.06	0.03
Jan 1-7	0.33	0.09	0.08	0.25	0.11	0.04	0.03	0.07
Jan 8-14	0.26	0.19	0.11	0.22	0.09	0.03	0.03	0.06
Jan 15-21	0.24	0.14	0.16	0.16	0.13	0.06	0.05	0.06
Jan 22-31	0.17	0.19	0.23	0.12	0.10	0.11	0.04	0.04
Feb 1-7	0.08	0.26	0.26	0.05	0.09	0.14	0.08	0.03
Feb 8-14	0.28	0.22	0.16	0.13	0.08	0.07	0.03	0.02
Feb 15-21	0.23	0.20	0.16	0.13	0.10	0.06	0.06	0.06
Feb 22-29	0.16	0.19	0.18	0.08	0.10	0.14	0.10	0.04
Mar 1-7	0.16	0.26	0.25	0.08	0.07	0.11	0.04	0.03
Mar 8-14	0.25	0.22	0.16	0.15	0.10	0.06	0.04	0.04
Mar 15-21	0.26	0.14	0.17	0.17	0.12	0.06	0.04	0.04
Mar 22-31	0.14	0.24	0.17	0.20	0.10	0.06	0.05	0.03
April 1-7	0.14	0.24	0.24	0.11	0.13	0.07	0.04	0.04
April 8-14	0.34	0.18	0.17	0.13	0.06	0.04	0.03	0.05
April 15-21	0.13	0.30	0.14	0.11	0.05	0.19	0.04	0.03
April 22-30	0.28	0.11	0.12	0.17	0.13	0.07	0.04	0.08
May 1-7	0.20	0.18	0.14	0.11	0.12	0.15	0.05	0.06
May 8-14	0.13	0.29	0.15	0.08	0.06	0.21	0.05	0.04
May 15-21	0.34	0.07	0.07	0.23	0.14	0.03	0.03	0.09
May 22-31	0.21	0.22	0.12	0.16	0.09	0.06	0.06	0.07
Average	**0.21**	**0.20**	**0.17**	**0.14**	**0.10**	**0.09**	**0.05**	**0.05**