

Mapping Russian Twitter

By John Kelly, Vladimir Barash, Karina Alexanyan, Bruce Etling,
Robert Faris, Urs Gasser, and John Palfrey



Berkman

The Berkman Center for Internet & Society
at Harvard University

Berkman Center Research Publication No. 2012-3

March 20, 2012

About this paper

The Berkman Center for Internet & Society at Harvard University, with funding from the MacArthur Foundation, is undertaking a three-year research project to investigate the role of the Internet in Russian society. The study will include a number of interrelated areas of inquiry that contribute to and draw upon the Russian Internet, including the Russian blogosphere, Twitter, and the online media ecology. In addition to investigating a number of core Internet and communications questions, a key goal for the project is to test, refine, and integrate various methodological approaches to the study of the Internet more broadly. More information about the project is available on the Berkman Center Web site: <http://cyber.law.harvard.edu>.

The authors owe special thanks to Devin Gaffney who collected the Russian Twitter dataset used in this research. We would also like to thank Sam Gilbert and the Web Ecology Project for ideas and inspiration for the study of Russian Twitter.

Key Findings

- Drawing from a corpus of over 50 million Russian-language tweets collected between March 2010 and March 2011, we created a network map of 10,285 users comprising the ‘discussion core,’ and clustered them based on a combination of network features. The resulting segmentation revealed key online constituencies active in Russian Twitter.
- The major topical groupings in Russian Twitter include: Political, Instrumental, CIS Regional, Technology, and Music. There are also several clusters centered on Russian regions, which is significant given the limited reach of the Internet in the regions outside of Moscow and St. Petersburg.
- Russian Twitter features a great deal of activity generated by marketing campaigns and search engine optimization (SEO) initiatives, including both automated and coordinated human actors. After our initial mapping resulted in a network dominated by these ‘instrumental’ actors, we constructed a filter to limit their presence in the network and discover relationships among a wider variety of ‘organic’ actors.
- Similar to the Russian blogosphere, the Twitter network includes a democratic opposition cluster associated with Gary Kasparov and the opposition Solidarity movement.¹ In other respects the political clusters identified in Weblog and Twitter networks display interesting variation. Nationalists, who are very active in Russian blogs, do not appear to be organized in Russian Twitter (at least as of March 2011). Conversely, pro-Putin youth groups like the Young Guards and Nashi, and elected officials allied with them, have a distinct Twitter footprint.
- While other clusters within Twitter often mirrored those in Weblogs, such as one cluster focused on major bloggers and online personalities, there were some Twitter clusters that had no clear Weblog analog. Most notably, there are two clusters of Twitter users affiliated with local government administrations in Tver and Ivanovo, representing active outreach to citizens by local government actors.
- While the filtered version of the map successfully reduced the presence of SEO actors, it curiously eliminated a pro-government cluster as well. In the filtered map, whereas the number of political actors was greatly increased overall, a cluster in the original map that focused on President Medvedev’s economic modernization policy disappeared, along with related hashtags. One possibility is that, as we observed in the Russian blogosphere, some political initiatives have adopted the tactics and/or services of online marketers.

Основные выводы:

- Исследование базировалось на более чем 50-ти миллионов русскоязычных “твитов”, которые были собраны в период с марта 2010-ого по март 2011-ого годов. На их базе было идентифицировано “дискуссионное ядро” - сетевая карта включающая 10.285 пользователей, состоящая из кластеров отражающих характеристики сетевой активности. Подобная сегментация позволила выявить ключевые структуры динамики интеракции в русскоязычном Твиттере.
- Основные тематические группы русскоязычного Твиттера включают: политический, инструментальный, региональный (СНГ), технологический и музыкальный кластеры. Также выделено несколько кластеров которые фокусируются на российских регионах, что является особенно важным, учитывая ограниченный доступ к Интернету за пределами Москвы и Санкт Петербурга.
- Анализ показывает, что значительная часть активности в Российском Твиттере сгенерирована в рамках маркетинговых компания и поисковой оптимизации (SEO), включая как использование автоматических программных решений, так и действия специалистов в этой области. После карты российского Твиттера созданного в рамках первичного анализа показавшей доминирование инструментальной активности, мы применили фильтры которые ограничили данную форму сетевой деятельности. Это позволило сфокусировать новую карту на разнообразие связей естественных участников сетевой интеракции.
- Подобно русскоязычной блогосфере, сфера Твиттера включает кластер демократической оппозиции, который связан с Гари Каспаровым и движением «Солидарность»¹. Другие политические кластеры проидентифицированные в блогосфере и Твиттере имеют ряд интересных различий. Националисты, которые проявляют высокую активность в российской блогосфере, не формируют явной структуры в Твиттере (по крайней мере на март 2011). В то же время, про-путинские молодёжные движения как например Молодая Гвардия и Наши, а также связанные с ними депутаты, имеют отчетливое присутствие в Твиттере.
- В то время как многие кластеры в Твиттере соответствуют схожим структурам в блогосфере, к примеру кластер который включает популярных блоггеров и сетевых персонажей, в Твиттере была обнаружены также ряд кластеров не имеющих аналогов в блогосфере. Стоит особенно отметить два кластера пользователей Твиттера, связанных с местными администрациями в Твери и Иваново, функцией которых являются общественные связи местных чиновников с гражданами.
- Отфильтрованная версия карты, на которой было успешно удалено большинство активностей связанных с искусственным продвижением контента (SEO), привело также к исчезновению про-правительственного кластера. После фильтрации, количество участников сетей связанных с политикой увеличилось, однако вместе с тем, кластер который фокусировался на политике экономической модернизации президента Медведева исчез вместе с относившимися к нему хештагами. Одно из возможных объяснений этому явлению может сводиться к тому что также как и в русскоязычной блогосфере, ряд политических инициатив в Твиттере использовали тактики и/ или сервисы Интернет-маркетинга

Introduction

Over the past several years, Twitter has become a global platform for online expression. In the United States, Twitter has been a phenomenally successful social media platform, adopted as a key mode of communication by a wide range of people and organizations, from GOP lawmakers and major corporations, to Justin Bieber fans and street demonstrators. It has been prominently implicated in social unrest across the world, particularly in the Middle East, where it featured in stories about Iran and Egypt. With little of the same fanfare, Twitter has come to play a role in Russia also, where the diversity of its users echoes the United States, and its employment in oppositional politics calls to mind the Arab Spring and the ongoing Occupy movement. But beyond these easy comparisons, the discourse network of Russian Twitter reflects important facts about the state of the online public sphere in this key global power.

Most of the popular discourse about the impact of the Internet on Russian social and political affairs has focused on the influence of the Russian blogosphere and social networking sites like Vkontakte and Facebook. This is not surprising given that the adoption of Twitter as a major online platform is relatively recent in Russia. The adoption rate, however, has been exponential. According to Yandex, in 2007 there were barely 1000 Russian Twitter users while today (March 2012) there are over 2.3 million.² The period covered by our study (March 2010 to March 2011) witnessed this pattern of growth, with many more users and tweets in the latter part of 2011. It is fair to say our study covers the Russian Twitter network during its early explosion of growth. The mature network may come to look somewhat different; its role in the Russian public sphere is actively evolving.

One of the first widely recognized instances of Twitter playing a key role in political affairs was during the December 2011 Duma elections. Russians used Twitter to share election results from their own polling stations and links to videos and stories about election violations, to help organize and promote protests following the election, and to share pictures snapped on smart phones by activists such as Alexey Navalny and Ilya Yashin after their arrest at a public demonstration on December 5, 2011.³ At one point shortly after his arrest, Navalny's Twitter account gained thousands of followers hourly.

During the 2011 Duma election, many prominent Internet sites were taken down by distributed denial of service (DDoS) attacks.⁴ Unlike popular blogging sites such as LiveJournal, the crowd-sourced election violation map kartanarusheniy.ru, and independent media sites, Twitter was not subject to DDoS attacks on Election Day. Many sites that were unavailable because of the DDoS attacks relied on Twitter accounts to share news and information on Election Day. Also, those sympathetic to the regime flooded protest hashtags with automated and human spam to limit their usefulness for coordinating protests.⁵ However, we are not aware of any successful DDoS attacks or hashtag spamming during the March 2012 presidential election.

While this recent volatile period marks a coming of age for Twitter in Russia, these activities reflect the configuration of online forces and tactics already evident between March 2010 and March 2011, the period of our study. From the use of Twitter by opposition activists, to the adoption of spam and automated online marketing techniques by pro-government actors, what

happened around the elections fit well with our earlier data. What we discovered was a network organized around a wide range of topical interests, but including specific (and clearly contending) political forces active within Russia itself, as well as the footprint of the wider former Soviet world, from Ukraine to Kazakhstan. This component of the Russian networked public sphere, while centered on Moscow, encompasses the wider scope of historical Russian empire. The sharp conflict between key political movements is only one part of this wider picture.

Background

Journalistic and popular discourse about the role of the Internet in social and political change has come to accept the premise that new media technologies are having a strong, nearly global effect on the practice of politics. But thus far, academic research has mainly illuminated just how complicated it is to discern the mechanisms of this change. New tools for measurement and analysis of Internet data must be leveraged against theoretical frameworks, and it is not clear that older disciplinary approaches are up to the challenge without some retooling.⁶ One compelling theoretical model is Benkler's concept of a *networked public sphere*, in which Internet-based technologies weave a fabric of many-to-many socio-political communication that allows a more diverse and less institutionally bound set of actors to contribute to society's agenda setting and policymaking functions.⁷ Whereas social power once emerged from a handful of mass media megaphones (and the barrels of guns, of course), now the formerly atomized public can coalesce into connected clusters of concern about salient issues, and gain power and influence in ways that are facilitated by online communication.

In addition to bridging theories of political communication and sociology, the *networked public sphere* model provides a path forward for empirical analysis of Internet data and other data on social groups and political institutions and processes. Networked communications can be examined to find and correlate structural and semantic features, discovering when and how emergent clusters of sociopolitical activity arise online, what kinds of issues they arise around, and how they relate to offline networks, institutions, and social forces. This mode of analysis promises insights into the political role of networked communications in particular societies, as well as traction on major theoretical questions about socio-technical change, such as the importance of technological affordances of online genres on the one hand, and culture and evolved practices on the other.

In our recent line of research on online social media networks, we have focused on the analysis of Weblog networks across several key societies: Iran, the Arab world, and Russia.⁸ This work has shown how online clusters emerge in each of these societies, and that the particular nature of these clusters – their topical constitution and structural topology – vary significantly from society to society. We also found evidence of major differences in the online behavior of political bloggers in different societies, such as the strength of political homophily and the consequent polarizing effects at the macro level. But if this research has contributed to understanding the diversity of networked public sphere processes in Weblogs across several societies, it has not yet addressed differences across online genres. Indeed, the very capacity for such diversity in Weblog network topology might relate to features of Weblogs, and not hold true for other forms of social media.

Do networks emerging within other social media genres replicate the topical and topological dynamics of Weblogs? Or do they have characteristics that diverge from Weblogs? How do other genres of online discourse engage the same or different sets of social actors?

We chose to look at Russian Twitter networks in order to address an alternative social media genre active in a society where we had done extensive work on Weblog networks. As discussed above, Twitter has become a very active platform within Russia, including for political discourse. Like our previous work on Weblogs, our method combined quantitative and qualitative analysis. The first step involved mapping the social network topology of a ‘discussion core’ of the most active and connected authors, and partitioning them into groups using statistical clustering of the authors’ connections and link histories. The next step involved qualitative analysis by subject matter experts of authors and discourse within the various clusters. The result is an interpreted map of the Russian Twitter discussion core, which reveals key constituencies active within the medium.

Methods

This study is based on a corpus of over 50 million Russian tweets collected by Devin Gaffney of the Web Ecology Project, and covers the period from March 2010 through March 2011. The Twitter data set was based on an index of Russian Twitter users listed by the popular Russian language search engine Yandex.⁹ From these tweets, we extracted the complete network of ‘mentions,’ which occur when Twitter user *A* ‘mentions’ Twitter user *B* by including *B*’s Twitter user name, preceded by an ‘@’ character, in *A*’s tweet. Popular forms of mentioning include ‘retweets’ and ‘replies.’ This allowed us to see the complete network of direct *dynamic* relationships (as opposed to *static* relationships like ‘following’) among Russian Twitter users during the course of one year. To identify and map the ‘discussion core’ of the most active, connected, and influential users, we performed a *k*-core breakdown (Seidman) of the mentions network.¹⁰ We picked a value of *k* such that the *k*-core consisted of 10,285 users, which we find to be a sufficiently large group to represent the major sets of highly active users, but not so large as to make clustering and visualization impractical. Next we used the Twitter API to pull follows and follower relationships for users in the core. We also extracted all hashtags from the tweet corpus, as well as all embedded URLs. Shortened URLs were expanded using a program that tracked URL redirects.

Using all of this data, we clustered and visualized the discussion core using a variation of the approach we have used on blogs in previous work. This involved clustering Twitter users according to a combination of their static (follows) and dynamic (mentions) relationships as well as their patterns of hashtag use and URL citations. Visualizations used the Fruchterman-Reingold physics model algorithm, creating a map of the discussion core that was used as a basis for viewing where within the network certain events (like the use of particular hashtags, or links to Web content) occurred.¹¹

The First Mapping

To our initial surprise, the first mapping revealed a discussion core structurally dominated by ‘instrumental nodes,’ accounts being run in the service of marketing and search engine optimiza-

tion (SEO) campaigns. Some of these appeared to be controlled by human beings, while others appeared to be controlled programmatically, by ‘bots.’ The nature of these clusters was evident in an analysis of the words most frequently used and the sites most frequently linked to in tweets by users from each cluster. Over half of the clusters identified tended to use extremely generic terms (‘hello,’ ‘good night’), as opposed to topically focused terms like ‘Putin,’ ‘democracy,’ ‘economic forum,’ and ‘hip hop.’ More tellingly, these clusters had a disproportionate number of links to pornographic, e-commerce, and advertising sites. Qualitative examination of Twitter users assigned to these clusters revealed that they were, indeed, spammers or SEO marketers. Upon reflection, the widespread use of Twitter by U.S. marketers, including use of robotically controlled accounts, should have led us to expect similar practices in Russia. Indeed, our mapping of the Russian blogosphere revealed a number of instrumental clusters operating there.¹² Russia clearly has an expert community of SEO and online marketing practitioners.

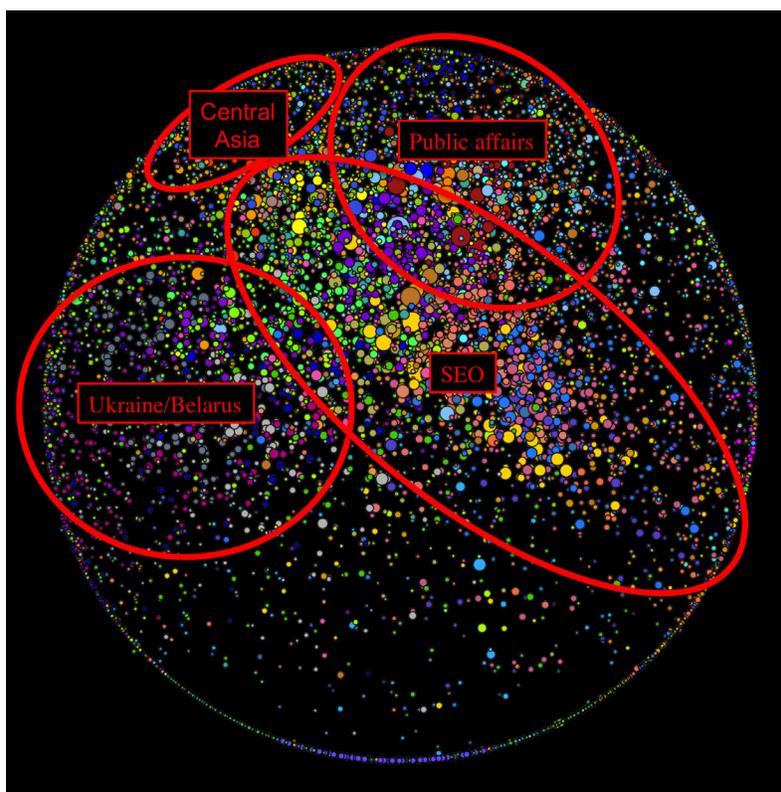


Figure 1: First Mapping

Despite the dominating presence of instrumental actors, the first Twitter network mapping did reveal a number of interesting ‘organic’ clusters. While SEO dominated the center of the network, there were clusters associated with former Soviet countries and regions, including Ukraine, Belarus, and Kazakhstan. There was also a set of Russia-centered clusters focused broadly on public affairs. Variations of most of these public affairs clusters (such as the ‘democratic opposition’) also appeared in the second, filtered map (described below). But one very interesting cluster did not. It was a generally pro-government cluster focused primarily on President Medvedev’s policy of ‘modernization’ of the economy, with a particular focus on Skolkovo, an attempt by the

government to create a Russian Silicon Valley. The cluster's top link targets included the Web site of the president's commission on modernization, the Kremlin Web site, sites connected to United Russia, and Rusnano, the government's venture capital fund. This cluster also linked heavily to the blogs of Marina Yudenich (a journalist and writer who formerly worked in PR roles for the Kremlin), Anatoly Chubais (head of Rusnano), and Sergei Mironov (former Chairman of the Federation Council and head of the Just Russia party). Top news links included Echo of Moscow, Kommersant, gazeta.ru, rian.ru and Vedomosti. Top users included Alexander Voloshin (former head of presidential administration under Yeltsin and Putin), Marina Yudenich, Nikita Belyk (governor of Kirov), and Margarita Simonova (an editor and host on Russia Today, the Kremlin's English-language TV channel). There were also some liberal and oppositional link targets (including Marina Litvinovich, Alexey Navalny, Echo of Moscow and the New Times, a liberal online newspaper favored by the opposition), which were perhaps linked to in order to criticize them. Top hashtags included #iRu, #GoRu, #Gosuslugi and #iRuz.

Because our ultimate goal was not to understand how Twitter serves tech-savvy marketers, but rather to focus on political and social discussion, we employed a network filtering approach to reduce the presence of instrumental nodes. While the disappearance of the modernization cluster left us with a fascinating mystery, the resulting second mapping contained a richer variety of non-instrumental clusters, which tell a more interesting story about the role of Twitter in the Russian networked public sphere.

The Filter

In constructing our filter, we drew on current academic research into automated discovery of instrumental accounts in Twitter, which has showed that it is possible to identify spammer accounts on Twitter using a set of features derived from user metadata, without having to rely on the content of the tweet.¹³ For instance, the ratio of followers to users followed tends to be very different for spam accounts (many more followed than following) than for organic ('real people') accounts. We followed this approach, constructing a simple classifier based on the following three metrics for each user u :

1. $\text{Log}(\text{number_of_friends})$ = the logarithm of the number of users followed by u
2. Friend_follow_rat = the ratio of the number of users followed by u and the number of users who follow u .
3. Log_replies = the logarithm of the number of distinct tweets by other users that reply to u . A reply tweet to one of u 's tweets always begins with u 's username preceded by the '@' character. Note that unlike in a retweet or non-reply mention tweet, where the username can appear anywhere in the text, a reply tweet has to begin with the username of the reply target.

We found that $\text{Spam_Score} = \text{Log}(\text{number_of_friends}) + \text{Friend_follow_rat} - \text{Log_replies}$ served as a simple linear classifier that can differentiate between spammers and non-spammers. Users with $\text{Spam_Score} > 0.25$ tended to be spammers, while users with $\text{Spam_Score} \leq 0.25$ tended to not be spammers. Furthermore, the higher the threshold is set, the greater the false positive rate

becomes than the false negative rate. Since the users with the highest spam score also tended to produce the most spam content, and thus were also precisely the users we wanted to get rid of, we applied the filter conservatively in order to minimize false positives while prioritizing elimination of the most active spam producers.

We applied our classifier to the full set of Twitter users in the dataset. Eliminating users with `Spam_Score > 0.25`, we created a new version of the map by repeating the process of generating a mentions network, performing a k-core breakdown to reduce the network size to around 10,000, and then building a map from the combination of follows and mentions relationships between the core users. Users were again clustered according to a combination of follows, mentions, hashtag, and URL citation profiles.

The Second Map

The second mapping greatly reduced the presence of instrumental Twitter accounts. And those that remained appeared to mainly be experts on social media marketing themselves, actual humans discussing their craft, as opposed to the bot networks deployed to execute their campaigns. There was a much greater diversity of ‘organic’ clusters, which can be divided into six major groupings: Political, Instrumental, CIS Regional, Technology, Music, and Russia Regional. Since our focus is on political speech within Russia, in this paper we focus primarily on description of the political zone and the clusters within it. However, an overview of the other major topical zones is also provided to give a sense of the diversity within the Twitter discussion core.

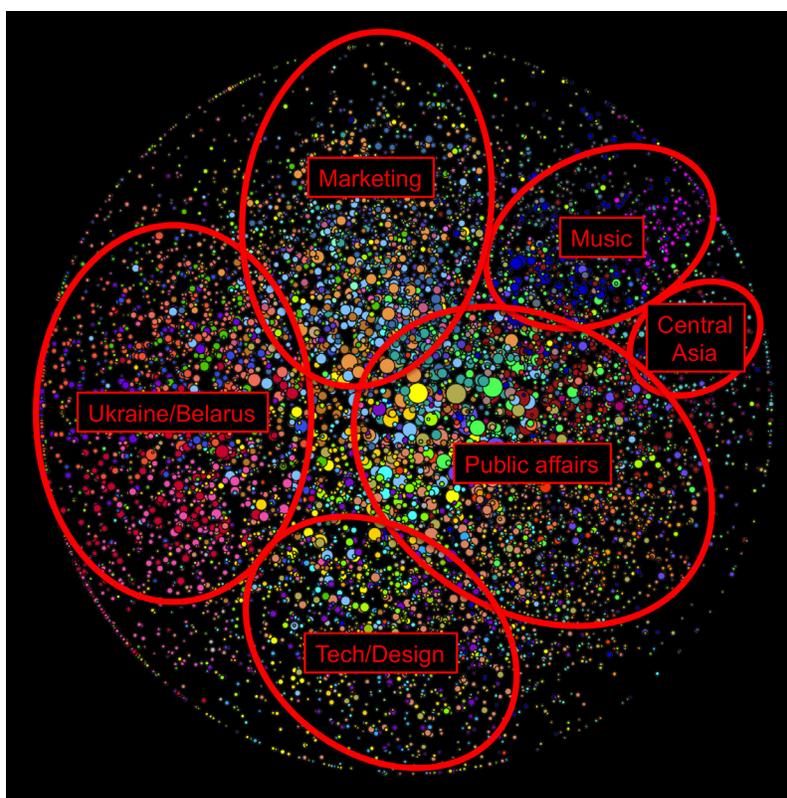


Figure 2: Second Mapping

Findings

Political Zone

The political zone includes six clusters of Twitter users that we labeled *Democratic Opposition*, *ProGov Youth*, *Non-aligned Politics*, *TverAdmin*, *IvanovoAdmin*, and *Echo of Moscow and Current Events*.

DEMOCRATIC OPPOSITION

This cluster of Twitter users is very similar to the cluster of democratic opposition bloggers we found in our Russian blogosphere research.¹⁴ Web sites preferred by this cluster include those of Russian political opposition figure Gary Kasparov and the Solidarity movement, a coalition of democratic opposition activists and parties. The oppositional but more nationalist National Bolshevik Party is also cited, though less frequently. Twitter users mentioned preferentially by this cluster predominately include activists, media strategists, and members of the Solidarity movement. The most preferentially used terms for this group include ‘Khimki forest’ (the source of a series of popular protests to preserve the old growth forest outside of Moscow that is being cut down to make way for a new highway), ‘Echo of Moscow,’ ‘United Russia,’ ‘Triumph Square,’ ‘protest,’ ‘Putin,’ ‘opposition,’ and ‘police.’ According to user profiles, most users in this cluster are located in Moscow. The top media outlinks for this group can be considered liberal or oppositional, including the oppositional news site grani.ru, Novaya Gazeta (a newspaper frequently critical of the government), and svobodanews.ru (the Web site of U.S. government-funded Radio Free Europe/Radio Liberty). The most popular hashtags for this cluster are #KSHN (referring to journalist and blogger Oleg Kashin, who was attacked in December 2010), #putin-pidoras (a slur against Putin), a number of hashtags related to the Strategy 31 protests (#31Aug, #31May, #31svoboda, etc.), and #ddt22, which refers to a protest concert headlined by the group DDT against the destruction of Khimki forest.

PROGOV YOUTH

One of the key findings of this paper is that, unlike the Russian blogosphere, Twitter features a distinct cluster of users who support the government and are connected to pro-government political structures and youth groups. This cluster is one of the largest in the political zone. Top links are to the dominant (at the time of our study) pro-government party United Russia, a number of links related to the Young Guards (United Russia’s youth wing), the pro-Putin youth movement Nashi, and kreml.org, a Web site run by political strategist and former Kremlin PR advisor Gleb Pavlosky and his Foundation for Effective Politics. The top media sites for this group are generally owned, supported by, or related to the government and its PR strategists. These include the government-funded news agency RIA Novosti (rian.ru), the Web site of the federal TV channel Vesti, another news and political discussion site created by Pavlosky (liberty.ru), and the news agency Interfax. Top terms for this group include ‘Young Guard,’ ‘Youth Policy,’ and ‘United Russia.’ Heavily used hashtags include a number of tags related to the Young Guards (#mger, #mger_revolution, #mgeRevolution), #Putin, #Seliger (the annual pro government youth forum), and #Nashi. The most mentioned Twitter users by this group are leaders of Russian youth move-

ments, young elected officials, and United Russia functionaries, such as Maxim Mischenko, Russlan Gattarov, Timur Prokopenko, and Robert Shlegel.

NON-ALIGNED POLITICS

This cluster includes links to major Russian Internet personalities and bloggers, such as Alexey Navalny, Anton Nossik, Oleg Kashin, Artemy Lebedev, and Maria Litvinovich. This group appears to focus heavily on Russian politics and corruption, but does not appear closely tied to any formal political party, which differentiates this cluster from the *Democratic Opposition* or *ProGov Youth* clusters. This ‘non-aligned’ political discussion is characteristic of the way most Russian political bloggers discuss politics, avoiding identification with any movement or faction. Top media outlinks include major traditional and Web-native news outlets, such as Echo of Moscow, gazeta.ru, Kommersant, Vedomosti, newsru.ru, inosmi, and newtimes.ru. This cluster also links heavily to Alexey Navalny’s anti-corruption website rospil.info. While this cluster occasionally links to the nationalist DPNI Web site, we did not find a clear nationalist cluster in Russian Twitter, as we did in the Russian blogosphere. Preferred terms indicate an interest in news sites (‘echo,’ ‘forbes,’ ‘newsru’) and politics (‘Putin,’ ‘government,’ ‘power,’ ‘Kashin,’ and ‘Domodedovo’). Top hashtags are also related to political topics and major news events during the period of our study: Khodorkovsky (#yukos, #mbkh, #khodorkovsky), #KSHN, #Luzkov, and #domodedovo.

TVERADMIN AND IVANOVOADMIN

These two smaller but similar clusters of Twitter users reflect the Russian regional reach of Twitter. These two clusters consist of users, links, and terms directly connected to the local administrations of Tver and Ivanovo oblasts. Unlike the other Russian regional clusters, these two are located next to each other on the map, and are also located close to the *ProGov Youth* cluster, and share many of the same outlinks, terms, and hashtags (i.e., United Russia, Young Guards). Most links are to local news outlets in the respective regions. These clusters also link preferentially to Nikita Belyk, a well-known governor of the Kirov oblast who has a significant online presence, but is not affiliated with either Ivanovo or Tver. Top terms include administrative issues such as ‘hotline,’ ‘police,’ ‘roundtable,’ and ‘hot water’ (the latter is likely a reflection of the lack of hot water during the summer months across most of Russia.)

ECHO OF MOSCOW AND CURRENT EVENTS

This small cluster includes many journalists and links connected to the Echo of Moscow radio station. Its topical focus is clearly news, current events, and major news stories, including many events and topics which are important to the *Democratic Opposition* cluster. Frequently used hashtags include news and political topics (#KSHN, #31Aug, #Khodorkovsky) and tags related to international economic forums (#G8, #SPIEF, #WEF). Most outlinks are to LiveJournal blogs, but also openspace.ru, blogs.reuters.com, Kashin’s LiveJournal blog, and his Web site olegkashin.ru. Top terms include ‘Echo of Moscow,’ ‘Navalny,’ ‘Kashin,’ ‘Putin,’ ‘Medvedev,’ and ‘Khimki.’

Non-Political Zones

The **CIS REGIONAL ZONE** contains two regions – Eastern Europe, comprised of eight *Ukrainian* and one *Belorussian* cluster, and Central Asia, comprised of one *Kazakhstan* and one *Kyrgyzstan* cluster. The primary distinguishing features of this zone are the location of its members, and the URLs they are paying attention to and citing, which, in each case, remain localized and internally oriented. In terms of location within the map, the Eastern European clusters are all in close proximity on the left side of the map, while the Central Asian clusters are near each other on the right side of the map. This arrangement makes sense in that it indicates discussion within the regions, but not between them.

The **RUSSIAN REGIONAL ZONE** contains six clusters. These clusters are spread out across the map, indicating, again, conversation within the cluster, but not between clusters. The primary cluster in this map is a relatively large *Major Cities* cluster. The primary distinguishing feature of this cluster is the predominance of users from various major cities in the Russian regions, and the lack of Moscow or St. Petersburg Twitter users. In addition to the *Major Cities* cluster, the areas around Samara, Chelyabinsk, Novosibirsk, Krasnodar, and Vladivostok are knit together tightly enough to segment into small clusters of their own.

Seven clusters group together into a relatively localized **TECHNOLOGY ZONE**, located toward the bottom of the map. The members of these clusters include technologically savvy individuals across Russia, and, in some cases, Russian speakers around the world. This zone includes a relatively large cluster of *Software Developers*, who discuss programming languages, Open Source and social coding topics; a midsize cluster of *Social Media Professionals*, mostly in Moscow; and a midsize cluster of *Mac and Apple Aficionados*. The remaining four clusters are oriented around *Mobile Technology*, *Design*, *Microsoft*, and *Yandex*.

Three midsize clusters group together in the small **MUSIC ZONE**, located in the upper right of the map. These three clusters break out by musical genre, discussing *Rap & Hip Hop*, *Club Music* and *Russian Pop*. Their members are located throughout Russia.

Finally, at least seven clusters are located within the **INSTRUMENTAL ZONE** (several additional clusters, which are difficult to characterize, may also belong here). This zone is comprised of clusters whose members are commercially oriented, attempting to generate income from social media use. This includes methods of search engine optimization, as well as social media marketing. Many Twitter users in this zone appear to be paid to tweet commercial or other content. These clusters are very similar to the Instrumental Zone we discovered in the Russian blogosphere.¹⁵

Two other clusters stand out, but do not fit into any particular zone. The *Media Celebrities* cluster is relatively large and spread across the map. It is oriented around major top media celebrity tweeters such as Anton Korobkov, Editor at Channel 1, or Tina Kandelaki, a TV producer, media company owner, and public figure. Key links and citations are to leading news and entertainment publications. We also found a relatively small English-speaking *Diaspora* cluster whose members are located in the United States and the United Kingdom (primarily California, the East Coast of the United States, and London). Their topics of discussion and key links are centered on American news and culture.

Discussion

Through large-scale social network analysis and data mining we are gaining an increasingly useful and more nuanced picture of public opinion and political discourse in democratic and semi-authoritarian countries. In Russia, state control over federal television makes that medium a good indicator of the political perspectives and priorities of the Kremlin. In contrast, the freedom of online discourse, including the use of social media platforms such as Twitter, blogs, and social networks, offers researchers access to the views and interests of an increasingly diverse slice of the Russian population, as well as the ability to assess the structure of naturally emerging online communities.

Twitter is becoming an increasingly important platform for the spread of ideas and for online organizing among the major political forces in Russia, complementing activities on other online platforms and genres. Each platform plays an important role in the larger online media ecology: social network services (like Facebook and Vkontakte) tie together friends and family in ways that typically reflect offline personal relationships, blogs support long form public political discussion, ‘hybrid’ platforms like LiveJournal combine social network and blog functions, and Twitter allows rapid information sharing among communities of followers. These platforms interact to create a networked public sphere that, although certainly used for entertainment and commercial purposes, also allows for important political functions: fact checking government claims, providing alternative news and information at odds with official and mainstream media, promotion and debate of salient political topics, and coordination of protests and other forms of collective action.

Similar to the Russian blogosphere, the Russian Twitter space extends beyond Russia proper, to include important former Soviet states and Russian expatriate communities. But the core of Russian language Twitter is dominated by users based within Russia. Interestingly though, many of these are well outside the centers of Moscow and St. Petersburg, an indication that a more diverse group of users are coming online over time. Participation in online public discourse is beginning to expand beyond principal urban elites. While the communities we find active in Twitter discuss a range of topics, there are several groups highly attuned to politics that argue from oppositional, non-aligned, and pro-government viewpoints, representing a wide spectrum of political opinions in Russia. Curiously, Russia’s several strains of nationalists appear to be under-represented in Twitter, as compared to their strong blogosphere footprint. It will be interesting to see if this changes in the aftermath of Russia’s recent elections, in which Nationalists joined a range of generally more liberal groups and individuals in the ‘White Ribbon’ protest movement in opposition to Vladimir Putin and election falsification. Indeed, recent political upheavals in Russia may lead to significant changes in the online political network.

Additional Research

Methods for studying the role of Twitter and other social media in political and social change are being actively developed and refined. Several recent Twitter studies have focused on the Arab Spring, addressing the prevalence and spread of hashtags representing political activity.¹⁶ These studies have made important contributions to our understanding of how Twitter is being used for social and political purposes in other global contexts. An objective of the current paper is to make an important methodological advance by identifying key structures and communities within Twitter that can be used to provide context for further analysis of discourse. Several forthcoming companion papers will build on the mapping and segmentation presented here. These papers include the following analyses:

- Analysis of patterns in the use of key hashtags, using metrics developed to identify concentration within both time and particular network segments. This will enhance our understanding of the structural flow of ideas in this medium.
- Alignment of offline issues and events associated with different political and social groups in Russia with social media activity within Twitter clusters. The events and topics analyzed include those associated with: 1) the Russian government, 2) grass roots social movements, and 3) major news events occurring between March 2010 and March 2011. Semantic markers related to these events and topics (including hashtags, links, names, and terms) are tracked across the network over time.
- Comparison of the activity above (in Twitter) with activity around the same issues in Russian blogs. In addition to shedding light on the increasingly important role of the Internet and social media in the political changes sweeping Russia today, this analysis contributes a 'multi-layer' framework for understanding the interplay between different platforms and online genres within the larger networked public sphere.

Endnotes

- ¹ Bruce Etling, Karina Alexanyan, John Kelly, Robert Faris, Urs Gasser and John Palfrey, “Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization,” *Berkman Center Publication 2010-11*, October 19, 2010, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Public_Discourse_in_the_Russian_Blogosphere_2010.pdf.
- ² Yandex, “Trends in the Russian Blogosphere,” Spring 2009, http://download.yandex.ru/company/ya_blogosphere_report_eng.pdf; For current rankings and the number of Russian Twitter accounts see Yandex Twitter rankings: <http://blogs.yandex.ru/top/twitter/>.
- ³ Alexey Navalny tweeted this picture and text on December 5, 2011: “Sitting with the boys in the OMON (police) bus. They say hello to everyone <http://instagr.am/p/XhTGt/>” (“Сижу с пацанами в омоновском автобусе. Они всем передают привет <http://instagr.am/p/XhTGt/>”).
- ⁴ Hal Roberts and Bruce Etling, “Coordinated attacks during Russian Duma elections,” Internet & Democracy blog, <http://blogs.law.harvard.edu/idblog/2011/12/08/coordinated-ddos-attack-during-russian-duma-elections/>.
- ⁵ Brian Krebs, “Twitter bots drown out anti-Kremlin tweets,” Krebs on Security blog, <http://krebsonsecurity.com/2011/12/twitter-bots-drown-out-anti-kremlin-tweets/>.
- ⁶ Karina Alexanyan, Vladimir Barash, Bruce Etling, Robert Faris, Urs Gasser, John Kelly, John Palfrey, and Hal Roberts, “Exploring Russian Cyberspace: Digitally-Mediated Collective Action and the Networked Public Sphere,” *Berkman Center Research Publication No. 2012-02*, March 2, 2012, <http://ssrn.com/abstract=2014998>; Sean Aday, Henry Farrell, Marc Lynch, John Sides, John Kelly, and Ethan Zuckerman, “Blogs and Bullets: New media in contentious politics,” *Peaceworks No. 65, U.S. Institute of Peace*, 2010, <http://www.usip.org/files/resources/pw65.pdf>; R. Kelly Garrett, Bruce Bimber, Homero Gil De Zuniga, Francois Heinderyckx, John Kelly, and Marc Smith, “New ICTs and the Study of Political Communication,” *International Journal of Communication* 6 (2012), <http://ijoc.org/ojs/index.php/ijoc/article/view/1112/694>.
- ⁷ Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006).
- ⁸ Etling et al, “Public Discourse in the Russian Blogosphere;” John Kelly and Bruce Etling, “Mapping Iran’s Online Public: Politics and Culture in the Iranian Blogosphere” *Berkman Center Research Publication 2008-01*, April 2007, http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Kelly&Etling_Mapping_Irans_Online_Public_2008.pdf; and Bruce Etling, John Kelly, Robert Faris, and John Palfrey, “Mapping the Arabic Blogosphere: Politics and Dissent Online,” *New Media and Society*, December 2010: 12(8) (An earlier version of this paper is available at: http://cyber.law.harvard.edu/publications/2009/Mapping_the_Arabic_Blogosphere).
- ⁹ Yandex Twitter Rankings: <http://blogs.yandex.ru/top/twitter/>.
- ¹⁰ Stephen B. Seidman, “Network structure and minimum degree,” *Social Networks* 5 (3) (1983): 269–287.
- ¹¹ Kelly and Etling, “Mapping Iran’s Online Public;” Etling et al, “Mapping the Arabic Blogosphere;” Etling et al, “Public Discourse in the Russian Blogosphere.”
- ¹² Etling et al, “Public Discourse in the Russian Blogosphere.”
- ¹³ Kyumin Lee, Brian David Eoff, and James Caverlee, “Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter” (Paper presented at the 5th International AAAI Conference on Weblogs and Social Media (ICWSM), Barcelona, July, 2011, <http://students.cse.tamu.edu/kyumin/papers/lee11icwsm.pdf>).
- ¹⁴ Etling et al, “Public Discourse in the Russian Blogosphere.”
- ¹⁵ Ibid.
- ¹⁶ Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, and danah boyd, “The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions,” *International Journal of Communication* 5 (2011), <http://ijoc.org/ojs/index.php/ijoc/article/view/1246>; Philip N. Howard, Aiden Duffy, Deen Freelon, Muzammil Hussain, Will Mari, and Marwa Mazaid, “Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?,” Project on Information Technology & Political Islam (pITI-PI), <http://pitpi.org/index.php/2011/09/11/opening-closed-regimes-what-was-the-role-of-social-media-during-the-arab-spring/>.