Researching Russia with digital trace data[1]

Alexey Bessudnov (University of Exeter)

This is a pre-copyedited, author-produced PDF of an article accepted for publication in the Communist and Post-Communist Studies (https://online.ucpress.edu/cpcs).

Abstract. The Russo-Ukrainian war and the deterioration of the relationship between Russia and the West made the application of traditional social science research methods, such as surveys and fieldwork, difficult for researchers. It is likely that in the foreseeable future students of Russia will have to increasingly rely on the data that can be collected online. Using these data has defined the field of computational social science and has many analytical advantages and come limitations. This paper provides a review of the sources of digital trace data for Russia and of the studies that used them for the analysis of political communication and behaviour, education, labour markets, discrimination, and the Russo-Ukrainian war.

Keywords: Russia, Russo-Ukrainian war, digital trace data, digital methods, computational social science

In the last decade, the methodological landscape of the social sciences has changed rapidly. Traditionally, qualitative social scientists conducted fieldwork to collect interviews and engage in participant observation. Quantitative social scientists often worked with survey data or, more rarely, implemented field experiments. These methods, of course, remain valid and are still used in most of the social science research. However, much of our everyday life activity (communication, shopping, cultural consumption, interaction with the state and service providers, etc.) has now moved online. These activities leave behind digital trace data that are sometimes available for social research. The proliferation of digital trace data and the new methodological opportunities offered by online communication and activities have led to the emergence of the new interdisciplinary field of computational social science (CSS), at the intersection of social sciences and computer science.

In a review of the CSS approaches in sociology, Edelmann et al. (2020) define the CSS as follows: “Computational social science is an interdisciplinary field that advances theories of human behavior by applying computational techniques to large datasets from social media sites, the Internet, or other digitized archives such as administrative records” (p.62). Therefore, the focus is on methodology (“computational techniques”, as opposed to more traditional statistical analysis) and data sources (large data sets collected online and perhaps administrative data). In practice, the boundary between 'traditional' quantitative social science and CSS has become somewhat blurred,

[1] I am grateful to Maxim Alyukov, Olessia Koltsova, Lanabi La Lova and two anonymous reviewers for their helpful comments on this paper.

especially as digital methods find their way into the toolbox of social scientists who were not specifically trained in CSS.

Salganik (2018) in his pioneering textbook on the CSS provides a classification of the most popular research designs / strategies. First, these are new ways of observing human behaviour, using data that were not specifically created for the purpose of research (unlike traditional survey data). These can be data collected online (from the social media, internet search engines, digitised archives, etc.), but also data from mobile phone networks, supermarkets and online marketplaces, other service providers, as well as large administrative data collected by the governments. These data can be used then for providing more precise estimates of unknown quantities ('counting things', as Salganik described it), as well as for forecasting and approximating natural experiments.

Another CSS approach is using the internet for survey research. Obviously online surveys and panels are now a common place in the polling and market research industry, as well as in academic research. The innovation brought by the CSS is linking survey data with large online observational data sets (such as for example the social media data) enhancing them with new information and then potentially training machine learning models with the enhanced data to use them for prediction with the original observational data. Generally, machine learning (including more recently the large language models) has become one of the main methodological tools in the CSS-orientated social sciences (see Grimmer, Roberts, and Stewart 2021; Molina and Garip 2019).

Finally, according to Salganik's classification, there are experimental designs where experiments are conducted online (either using existing environments or tailor-made apps) and mass collaborations for distributed data collection and coding. To this list of tools, we could add social network analysis (a long-established field, however, recently enhanced by access to large network data and new computational approaches) and agent-based modelling.

In this paper, I will provide a review of the empirical studies that used CSS approaches and digital trace data to study Russia (see Koltsova, Porshnev, and Sinyavskaya (2021) for an alternative review of the studies that used data from the social media and La Lova (2023) and Libman (2023) for reviews of the methodological approaches in the Russian studies). I will start with describing the available data sources, and then continue with reviewing the studies of education, labour markets, political communication and behaviour. In a separate section, I review the studies related to the Russo-Ukrainian war.

This paper aims to make several contributions. First, it provides an overview of the sources of digital trace data available for Russia (at the time of writing in late 2024 – early 2025). Second, it reviews empirical studies that used these data across various disciplines, such as political communication, educational studies, labour economics, sociology, and others. Researchers working in these areas are not always aware of the studies conducted in the neighbouring disciplines, and hopefully a cross-disciplinary review will be useful for them. The review can also serve as an introduction to available data sources and methodological approaches for researchers of Russia who are not yet familiar with CSS. I mainly focus on data and methods that were used in empirical studies rather than theories and substantive conclusions. Third, the review discusses the future of digital methods in the Russian studies in the context of the Russo-Ukrainian war and increasingly repressive nature of the Russian state.

Data sources

Using digital trace data in social research provides some unique advantages: these data are bigger, produced in real time, and often directly reflect human behaviour. There are disadvantages, too: the data are incomplete (both in the sense of population coverage and missing information), biased, and not always available (see Salganik (2018) and Lazer and Radford (2017) for a more detailed discussion). Normally, we can complement these data sources with more traditional quantitative and qualitative data collected specifically for the purpose of research.

However, in authoritarian regimes in general, and in Russia after 2022 in particular, digital data may become indispensable (Kalsaas 2023). It is much more difficult now for Western-based researchers to conduct fieldwork in Russia. Access to the country is limited, risks are high, research trips are unlikely to be approved by ethics committees and IRBs, not to mention potential interference of the Russian authorities. It is also much more difficult to commission surveys in Russia. Currently many pollsters in Russia would not work with projects that are even vaguely political and have Western funding, especially if there are questions related to the war. Paying to Russian partners for data collection from outside Russia is complicated by international sanctions. There are still some on-going surveys with the data accessible for Western researchers. For example, the Russian Longitudinal Monitoring Survey conducted by the HSE University (https://www.hse.ru/en/rlms/) recently released the data for 2022-23. The questionnaire even includes some questions on how people's lives changed after the beginning of the 'special military operation' in Ukraine, although the question on the attitudes towards it is missing. In this context, using digital trace data that can be collected online from outside of Russia is often the only viable research strategy for Western-based academics and analysts. At some point the war will end and collecting data in Russia will likely become easier, but unless there are dramatic political changes, it is hard to envisage going back to the pre-2022 level of access.

What kind of data are available? First, these are data from the Russian social media, such as VK. VK is the largest and most popular social media website in Russia, having about 89 million people as its monthly audience (as of 2024). It is generally similar to Facebook: users can add other people as 'friends', write, share and comment on posts, create groups, as well as listen to the music, watch videos, play games, etc. Unlike Facebook that closed unrestricted data access after the Cambridge Analytica scandal, VK still provides ample opportunities for data collection (at the time of writing in 2025). It is possible, for example, to collect information about users from a particular group or a particular city, search users by name, download posts for users and groups, etc.

Another data source is Yandex, a search engine competing in Russia with Google. The Wordstat Yandex service, mostly developed for advertisers, allows users to collect statistics on search terms over time (both the number and the proportion of search requests), disaggregated at the regional and city level. The Yandex.Audience service allows for estimating the number of Yandex users in a particular location (that can be specified in a very precise way) potentially letting researchers trace population movements. Similar data from Facebook and other digital services were used by demographers in other countries (see Leasure et al. (2023) for a study of population displacement in Ukraine with Facebook data).

Telegram, a messaging app popular worldwide and in particular in Russia, has an application programming interface (API) that allows automatically downloading messages from public channels that have become especially popular since the start of the Russo-Ukrainian war in 2022.

YouTube is still one of the most popular web services in Russia, despite the attempts of the Russian authorities to limit its use. It is possible to collect statistics on the most popular videos in different categories, the number of views for videos, and automatically download comments, using packages developed both for R and Python.

Table 1 provides an overview of the most popular sources of digital trace data for Russia. In the following sections, I review examples of using these and other data for empirical social research.

Table 1. Main sources of digital trace data for Russia

Name	URL	API	R/Python packages / API wrappers	Notes	Examples of research
VK	https://vk.com	https://dev.vk.com/en/method	R: vkR, R:rvkstat, Python: vk-api	The largest social media website in Russia, with a well-developed API.	Bessudnov et al. (2023); Rykov, Koltsova, and Sinyavskaya (2020); Sivak and Smirnov (2019); Smirnov (2020); Sokolova (2023)
Odnoklassniki	https://ok.ru/	https://apiok.ru/en/		The second largest social media website.	

				The API is less developed and less useful for social research.	
Telegram	https://telegram.org	https://core.telegram.org	R:telegram, R:telegram.bot, Python: python-telegram-bot, Python: Telethon	An instant messaging tool popular in Russia. The API is of limited use for social researchers, but it allows automatically downloading the content of groups / channels.	Kuznetsova (2024); Urman and Katz (2022)
TikTok	https://www.tiktok.com	https://developers.tiktok.com/products/research-api/	R: traktok, Python: Pyktok, Python: TikTok-Api, Python:TikTokApi	A popular social media website. The Research API is available for researchers in the USA, EEA, UK and Switzerland, after project authorisation.	Bösch and Divon (2024); Molotov and Khlevniuk (2024); Primig, Szabó, and Lacasa (2023)
YandexWordstat	https://wordstat.yandex.com	https://yandex.com/dev/direct/doc/dg-v4/en/	R: ryandexdirect, Python:YandexWordstatAPI	The largest Russian search engine word search statistics and trends.	Anastasiadou, Volgin, and Leasure (2024)
Google Trends	https://trends.google.com/trends/		R:gtrendsR, R: gtrendsAPI, Python:pytrends	Word search statistics from Google.	
Yandex.Audience	https://audience.yandex.com	https://yandex.com/dev/audience		Allows estimating the size of the Yandex audience, including by fine grained location.	
Youtube	https://www.youtube.com	https://developers.google.c	R:tuber, Python: python-	Search and download information	Bodrunova et al. (2021)

		om/youtube/v3	youtube, Python: youtube-data-api	about Youtube videos, download comments.	
Twitter / X	https://x.com	https://developer.x.com/en/products/x-api	R:rtweet, Python: Tweepy	The API has been widely used in social research, but currently requires paying a large fee for data collection.	Chen and Ferrara (2023); Stukal et al. (2017, 2022)

Notes: All the information in this table is as of February 2025.

Education, social capital, labour market, discrimination

The VK's API allows users to search user accounts and collect information by school and university, making it a good data source for educational research. To show the reliability of VK data, Smirnov, Sivak, and Kozmina (2016) matched the data obtained directly from a Moscow school and a Moscow university with the data found on VK. Using a variety of methods, they were able to find VK accounts of 88% of the school students and 93% of the university students (although direct matching by full name was only successful in 18% of the cases). Using the same data, Smirnov and Thurner (2017) analysed the evolution of student networks over time (where links between students were operationalised as 'likes' they left on each other's VK pages) and showed that they changed their friends following changes in the academic performance, and not the other way round. Another study involved using the VK data to look at the school segregation in the digital space (Smirnov 2019).

Often the most interesting research findings result from a combination of digital and survey data. The Educational and Career Trajectories survey (TrEC, https://trec.hse.ru/en/) is a longitudinal study of young people that started in 2011 when students were in 8th grade (Malik 2019). By 2021, ten waves of data were collected. The study included the results of the international mathematics and reading assessments: TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment). Additionally, participants were asked for consent to using data from their social media profiles for research purposes. Smirnov (2018, 2020) used data on VK public page subscriptions and VK posts to reasonably accurately predict student performance in PISA tests. Note that once the model has been built with the TrEC sample it can be extended to all VK users of school age, although the predictive power of such models is generally low. Using VK public posts, Sivak and Smirnov (2019) demonstrated that parents in Russia mention their sons more often than daughters on the social media (also see Leontyeva, Koltsova, and Verhoeven (2024) on gender stereotyping in the films released in Russia in 2008-19 with the data collected online on Kinopoisk, a Russian online film data base).

In another application of VK data, Alexandrov et al. (2018) looked at educational migration from Russia showing links between the Russian Far East and Siberia and China, the Russian North West and Nordic countries, and the North Caucasus and Muslim ethnic republics in Volga region and the Middle East (also see Alexandrov, Karepin, and Musabirov (2016)) and ethnic communities on VK (Alexandrov, Gorgadze, and Musabirov 2016).

Rykov, Koltsova, and Sinyavskaya (2020) collected the complete VK data for Vologda (about 200,000 users and 10 million links between them) to analyse social capital (defined via social network characteristics) in a medium sized city. Note that this type of research would be impossible with virtually any other social media where the opportunities for data collection are much more restricted. Rykov, Meylakhs, and Sinyavskaya (2017) downloaded and analysed data from a VK AIDS denialist community (about 15,000 members) identifying the community structure and conducting content analysis of main narratives. Panicheva et al. (2022) used a novel approach combining the VK data and data from a mental health app to predict subjective well-being (also see Semenov et al. (2015) who only used VK data). Koltsova et al. (2021) explored the communication pattern of VK users looking at the volume and stability of communication over time.

In another methodologically innovative application of VK data, Milkova, Rudnev, and Okolskaya (2023) developed a machine learning classified of value-expressive social media posts, using both human coders from Toloka.ai (more on this below) and GPT models to annotate the training data set, A similar approach was previously used by Bessudnov et al. (2023) to create a tool to detect perceived ethnicity from personal names, with the data on names collected on VK and processed on Toloka.

Studies of the Russian labour market using online data are rarer. Job search and recruitment in Russia are now often conducted online (Roshchin, Solntsev, and Vasilyev 2017) using such websites as Headhunter, Superjob, and others. Both Headhunter and Superjob have APIs, but they were largely developed for employers and applicants, and the amount of information publicly available for researchers is limited. As an interesting exception, Shevchuk, Strebkov, and Tyulyupo (2021) scraped data from FL.ru, the largest website for freelance workers, to describe work patterns of freelancers in Russia. Bessudnov and Shcherbak (2020) used Superjob and Headhunter to conduct a field experiment to study ethnic discrimination in the labour market. While the experiment was partially conducted offline the websites were used to create the profiles of fictitious job applicants, respond to job ads and interact with the employers.

Cian.ru is the largest Russian website for real estate sales and rentals. Avetian (2022) and Veterinarov and Ivanov (2018) used a similar research design to study ethnic discrimination in the rental market (also see Avetian and Veterinarov (2024)). They scraped rental ads from Cian.ru and looked at the proportions and characteristics of ads where the discriminatory language was used (such as “for Slavs only” etc.). (In 2022, Cian.ru banned the use of discriminatory language in rental ads.) Bodrunova et al. (2017) applied topic modelling to over 360,000 posts automatically collected in 2013 from Livejournal.com (a blogging platform popular in Russia in the 2000s-2010s) and analysed the characteristics of the xenophobic and anti-migrant discourse on the social media (also see Pronoza et al. (2021)).

Applications of scraped data for criminological research include the study of Hydra, a darknet marketplace for drugs shut down in 2022 (Goonetilleke, Knorre, and Kuriksha 2023), and the study of sentencing disparities in Russian criminal courts based on a data set that includes about 3 million criminal cases from 2009 to 2013 (Volkov 2016).

Political communication and behaviour

A large body of research with the social media data is devoted to political communication and preferences. Most of it used data from Twitter / X that until 2023 had an API allowing for ample opportunities for data collection (in 2023, under Elon Musk, those opportunities were severely restricted) (see Murthy (2024) for a review of data collection on Twitter / X). Given public attention to the Russian interference in the 2016 US election, it is unsurprising that several studies looked at Russian political bots / trolls on Twitter. Stukal et al. (2017) suggested a methodology for detecting bots on Twitter with the data from 2014-15. Stukal et al. (2019) proposed a machine learning classifier to detect political orientation of the bots (identifying 35% as pro-Kremlin, 18% as pro-opposition, 18% as pro-Ukrainian, and remaining 29% as neutral). In a paper published in the American Political Science Review, Stukal et al. (2022) used about 32 million tweets about Russian politics in Russian from 1.4 million Twitter users in 2015-18 to show that pro-Kremlin bots were primarily employed to control the Russian online domestic political agenda rather than to react to offline events and protests of the opposition.

Other examples of research on Russia with Twitter / X data include Badawy, Ferrara, and Lerman (2018) and Bail et al. (2020) analysing the effect of the posts by Russian trolls on the online behaviour of the US Twitter users, Filer and Fredheim (2016) comparing Twitter narratives about the murders of the special prosecutor Alberto Nisman in Argentina and the Russian oppositional politician Boris Nemtsov, Zherebtsov and Goussev (2021) mapping the Russian political Twitter applying social network analysis methods, Crilley et al. (2022) analysing Twitter accounts that followed the RT, and Zakhlebin et al. (2016) looking at political polarisation on Twitter during the 2011 Moscow protests.

While research with Twitter data dominates the field of online political communication it is perhaps less relevant for the study of Russian domestic politics. According to the Levada Centre data, only about 3% of the Russian adult population used Twitter in 2021, and in 2024 the Twitter / X penetration decreased to 1%. This compares with 50% for VK, 37% for Youtube, 25% for Odnoklassniki and 20% for Tik Tok (as of March 2024) (Levada 2024).

Examples of using data other than from Twitter for the study of political communication include Koltsova and Koltcov (2013) who used topic modelling with Livejournal data to study public agenda and Koltsova and Shcherbak (2015) exploring political preferences of Livejournal users. Urman (2019) downloaded data from a random sample of 55,000 VK users to find out that only about 15% of them followed any of the major Russian news media sources on VK and applied social network analysis to study political polarisation. Urman and Katz (2022) mapped the far-right landscape on Telegram. This study was not focussed on Russia, but it shows the methodological potential of using data from Telegram that is the second most popular messenger in Russia (after Whatsapp) used in March 2024 by 51% of the adult population (Levada 2024). Orttung and Nelson (2019) used a data set of Youtube video titles to analyse RT's strategy on Youtube, and Bodrunova et al. (2021) collected and analysed Youtube comments for the videos on the Moscow 2019 protests. A number of studies used Yandex.News and Yandex search data to explore potential biases in Yandex output, often comparing it to Google (Erbsen and Põldre 2023; Kravets and Toepfl 2022; Makhortykh, Urman, and Wijermars 2022). Other studies relied on the data from traditional media (such as TV channels and newspapers), either scraped online or purchased from archives such as Integrum (Lankina and Watanabe 2017; Otlan et al. 2023; Rozenas and Stukal 2019).

Another strand of research looked at the effect of social media on political behaviour. In an influential paper published in Econometrica, Enikolopov, Makarin, and Petrova (2020) showed that higher penetration of VK increased the probability of a protest in a city during the protest wave in December 2011 and also increased the number of protesters, most likely by boosting the opportunities for online political communication rather than by providing critical information about the government. They applied a sophisticated identification strategy to disentangle the causal effect of VK penetration using the information on the city of origin of students who studied together with the VK founder (and thus were VK's early adopters) as an instrument. The study was not exclusively based on online data (although the authors collected the information on the city of origin of all VK users who joined the social media before 2011, as well as the data from Odnoklassniki and Facebook, and analysed the content of all posts made on VK before the 2011 election), but also used a manually collected data set on political protests in 2011-12, a large survey and official statistics from the census. The combination of digital trace and traditional survey data is often a feature of high-quality studies in computational social science. Bursztyn et al. (2024) used a similar research design and identification strategy to study the effects of VK penetration on xenophobia showing that it increased the share of people holding xenophobic attitudes and led to more hate crimes in cities with high levels of nationalism. Using offline survey data from 2011 only, Reuter and Szakonyi (2015) demonstrated that being on Facebook or Twitter was correlated with higher perception of electoral fraud during the 2011 parliamentary election, while being on VK or Odnoklassniki did not have such an effect.

The Russo-Ukrainian war

The Russian invasions in Ukraine in 2014-15 and in 2022 and the following war increased the use of online data sources in research on Russia. This could be explained by restrictions on field work in the country, government censorship of the information related to the war and the lack of trust to traditional surveys in the context of the war and increasingly repressive state. I review four strands of research in this section: a) on the public opinion and attitudes to the war in Russia, b) on military casualties, c) on political communication during the war, d) on the consequences of the war for Ukraine. This research was conducted not only by academics, but also by data journalists in a number of Russian independent media outlets where the research and publication cycles are shorter compared to the academia and the public impact is arguably higher.

There have been several traditional surveys (including by independent and antiwar organisations) showing high levels of public support for the war in Russia. The question is to what extent these surveys can be trusted given that they were conducted in the context of military censorship in the authoritarian repressive state. Chapkovski and Schaub (2022) implemented a list experiment trying to identify the real level of support for the war in Russia. List experiments are a technique for indirect measurement of sensitive indicators (Rosenfeld, Imai, and Shapiro 2016). They found that when asked directly, 71% of respondents supported the war (in a non-nationally representative sample), while in the list experiment this share dropped to 61%. However, note the methodological critique of using list experiments in Russia that argues that they are subject to the artificial downward bias in the estimates (Frye et al. 2023).

Interestingly, Chapkovski and Schaub (2022) conducted the experiment on Toloka (https://toloka.ai), an online crowdsourcing platform similar to Amazon Mechanical Turk. Mechanical Turk has been actively used by researchers (mostly in the US); Toloka provides similar capabilities and access to users in Russia (Chapkovski 2023). Before 2023, Toloka used to be part of Yandex, a Russian digital giant, but it has since become independent and re-focussed on the international audience. Yandex has been developing a similar tool, Yandex.Tasks (https://tasks.yandex.ru). Some other online experiments conducted in Russia employed Yougov (Krishnarajan and Tolstrup 2023) and Cint (Alyukov and Zavadskaya 2024) panels. Other studies recruited participants via ads on VK and Facebook (Bryanov et al. 2023; Shirikov 2024).

Sokolova (2023b) collected and analysed VK posts from May 2022 to January 2023 to identify and compare the characteristics of pro-war and anti-war users in Russia (among those who publicly expressed their views, with the data from about 10,000 user profiles). Then she created a machine learning classifier that predicted the attitudes to war of a user from their subscriptions to VK groups. This classifier was applied in Sokolova (2024) where she conducted cluster analysis of VK user profiles to identify 12 clusters and estimate the pro-war sentiment in each of them. This study also looked at the network structure of pro-Kremlin pages that disseminated pro-war propaganda. Some of these studies were published in Novaya Gazeta Europe allowing them to reach a much wider audience than purely academic research. For example, Sokolova (2023a) looked at the dynamics of pro-war and anti-war posts on VK, Sokolova (2023c) matched the names of about 700 Russian military who died in Ukraine with their VK profiles and analysed them, and Sokolova et al. (2024) conducted network analysis of about 500 pro-war Telegram channels. Urman and Makhortykh (2022) looked at over 6 million public posts on VK posted in February-June 2022 to explore how the war was perceived by VK users.

Unsurprisingly, neither Russian nor Ukrainian governments provide reliable data on war casualties. For Russia, a group of volunteers created a crowdsourcing project that crawled the social media (along with using other data sources) to collect the names of the Russian military who died in Ukraine. The project was supported by two media organisations, the BBC News Russia and Mediazona. At the time of writing (March 2025), about 96,000 names were collected, which provides the lowest bound for the estimates of the Russian losses (Mediazona 2022). In 2025, the data were made public (https://200.zona.media, accessed in March 2025). For Ukrainian fatalities, similar projects include UALosses (https://ualosses.org/, about 132,000 deaths recorded by March 2025) and War Tears (https://wartears.org, about 66,000 deaths recorded by March 2025; War Tears likely overestimates Ukrainian losses in the statistical model provided on the website). Bessudnov (2023) used the BBC/Mediazona data set to estimate ethnic and regional inequalities in the Russian fatalities showing that some regions and ethnic categories were heavily overrepresented.

The consequences of the war for Russia included mass outmigration to other countries (both of the individuals with anti-war views and people escaping mobilisation) and increased state repression against anti-war activists. Anastasiadou, Volgin, and Leasure (2024) analysed Yandex search data to look at the trends and patterns of immigration intentions in Russia post-2022. OVD Info, an independent human rights project, collected and provided public access to the data on individuals persecuted in Russia for political reasons, containing about 5,000 cases since 2012, as in March 2025 (https://ovd.info/politpressing).

Telegram channels became prominent during the war and often offered a more detailed narrative than the traditional media, especially in the pro-war segment (Russian *voenkory*). Oleinik (2024) collected and analysed a corpora of Telegram channels covering the war both in Russia and Ukraine and compared them to the narrative in the traditional media in both countries, as well as in the USA, UK, and France. Kuznetsova (2024) explored the pro-Russian government activity on Telegram in framing the anti-war protests in 2022, and Dollbaum and Kim (2024) used VK and Telegram data to look at the position in relation to the war of the major Russian political parties. Also see Bareikytė and Makhortykh (2024) on the representation of the war in Telegram channels and Bareikytė et al. (2024) on archiving Telegram data. Tschirky and Makhortykh (2024) looked at how the siege of Mariupol was represented on Twitter / X, using both quantitative and qualitative approaches to data analysis. Haq et al. (2022) and Chen and Ferrara (2023) released public data sets of tweets related to the Russo-Ukrainian war and posted in 2022. In an earlier study, Koltsova and Pashakhin (2020) compared how the Russo-Ukrainian conflict in 2013-14 was framed on the Russian and Ukrainian TV.

Khokhlov (2024) collected the VK posts made by Russian regional governors in 2022-23 and applied structural topic modelling to analyse them. He found that the governors who were weaker politically, from less economically developed regions with high mobilisation rates were more likely use war-related messages in their public communication.

Researchers also used data from TikTok, a growing social media platform focussing on short video clips. Recently, TikTok introduced an API for researchers (see Table 1); in most published studies data were collected manually using hashtags. Bösch and Divon (2024) focused on analysing sound as the medium of war-related messaging. Primig, Szabó, and Lacasa (2023) performed a qualitative analysis of TikTok clips on the topic of Russo-Ukrainian war. Molotov and Khlevniuk (2024) conducted content analysis of representations of Stalin on Tiktok.

Although research on the consequences of the war in Ukraine is not in the realm of Russian studies, it is worth mentioning in this sections the studies that applied the digital trace approaches to study Ukraine. Leasure et al. (2023) used Facebook advertising data, combined with the administrative data on the population size and migration, to estimate that about 5 million Ukrainians were internally displaced in the first three weeks of the war. Abramenko, Korovkin, and Makarin (2024) utilised Twitter and VK data to show the shift from using the Russian language to Ukrainian in Ukraine, both in 2014-22 and especially after February 2022 (also see Kulyk (2024) documenting the same pattern with more traditional survey data). Golovchenko (2022) used the VK ban in Ukraine in 2017 as a natural experiment and showed that the ban did indeed lead to reduced VK activity in mainland Ukraine compared to Crimea.

Conclusion

With the advance of computational social science, the use of data that can be collected online and analysed with modern computational and statistical methods has been rapidly spreading in the social sciences. It is even more relevant for Russia and other authoritarian countries where access to traditional social science research methods may now be difficult. As usual, new digital data sources and methods bring both advantages and disadvantages that I reviewed earlier in this paper. There are several limitations of these data that are worth discussing in more detail.

First, it is the representativeness of the data. Traditionally, social scientists relied on research methods based on probability sampling (i.e. sampling where all units in the population have a known non-zero probability to be selected). This was considered the best method to ensure that the data were representative for the entire population. In quantitative research on Russia, we normally used nationally representative samples (unless focusing on a specific region or subgroup) from surveys conducted either face-to-face or on the telephone. The use of digital trace data is a deviation from this principle. Not everyone in Russia is a regular internet user, and not every internet user is active on the websites where the data are collected. However, about 80% of Russians now have regular internet access, with 60% using the social media (VK, Odnoklassniki, Youtube, etc.) on the daily basis (Levada 2024). When it comes to younger and middle-aged people, the internet and social media access is almost universal. Arguably, the coverage is not fundamentally different then compared to the usual face-to-face and telephone surveys with their high non-response rates. However, more methodological work needs to be done to show whether and how findings based on the digital trace data can be extended to the Russian population, given the self-selected nature of the data. This is similar to working with online survey data where methodological advances have been made in the last two decades (Couper 2017).

Another issue with online and social media data is ethical concerns. The use of data collected from the Russian social media and other websites remains in the ethically grey zone. For example, VK's API provides access to a lot more data compared to Facebook and other Western social media websites. However, getting informed consent from VK users to utilise their data for research is usually not possible (due to the volume of the data), and VK would not directly endorse using their data for research purposes (although wouldn't prohibit this either). Using leaked data sets containing personal information is even more controversial as this may directly violate the Russian (and international) data protection legislation. Practically, this should not prevent researchers from conducting online data collection and using publicly available data sets, but all such studies must be subject to review by ethics committees in the UK and IRBs in the USA (see Salganik (2018), ch.6, for a review of ethics in digital research).

The availability of digital trace data can change rapidly. Twitter / X data used to be widely utilised for social research; this changed after the purchase of the social media platform by Elon Musk when fees for data collection were introduced. Facebook data availability was also restricted for data protection reasons. Currently, much of digital trace research on Russia uses the VK API. It still offers wide opportunities for data collection at the time of writing (March 2025), but this can change in the future. After February 2022, the Russian government has been consistently trying to limit the use of Western social media platforms and services in the country. The internet service providers were required to block access to Twitter, LinkedIn, Facebook and Instagram (Kadlecová and Paggio 2025). In 2024, access to Youtube was restricted, with the government promoting the local alternative, RuTube. While restrictions and bans can be avoided by tech savvy Russians by using the virtual provide networks (VPN), this further reduces the representativeness of the data that can be collected on these platforms. Further restrictions can be introduced – or dropped – at any time. This may limit opportunities for digital data collection in Russia; however, it is hard to imagine that the bans and regulations would eliminate them completely (even in the absence or restrictions of APIs the data can often be scraped).

To conclude, I'll outline some promising directions for future research with digital data in Russia. First, while applying research designs with data collected from the social media only may certainly lead to interesting research findings, often the best studies use designs where the social media data are matched with external data sets, for example, from surveys (Koltsova (2022) notes the importance of linking digital trace and survey data in the case of VK). The reason is that social media data do not necessarily contain all the information necessary to answer research questions at hand, and surveys may significantly enhance these data. Of course, this requires having capabilities to commission and conduct surveys in Russia, which is at the time of writing became difficult. Before the war, Western researchers would normally use Russian pollsters for data collection, but currently this is complicated both for practical and political reasons.

Second, conducting online experiments (including survey experiments) is a promising area of research, especially when the focus is on identifying causal relationships. Currently participants from Russia can be found on some platforms, but creating a new online panel available for academics (similar to Prolific[2]) would increase opportunities for experimental research.

Finally, digital research in Russia would benefit from further development of data collection and storage capacities. Currently, individual researchers collect and store data for their research. Some of these data sets are publicly available, while others are not. A more centralised and systematic approach to digital data collection and archiving (perhaps in the form of a digital data archive for Russia) would be beneficial.

Bibliography

Abramenko, Serhii, Vasily Korovkin, and Alexey Makarin. 2024. 'Social Media as an Identity Barometer: Evidence from the Russia-Ukraine War'. *AEA Papers and Proceedings* 114:70–74.

Alexandrov, Daniel, Alexey Gorgadze, and Ilya Musabirov. 2016. 'Virtual Caucasus on VK Social Networking Site'. Pp. 215–17 in *Proceedings of the 8th ACM Conference on Web Science*. Hannover Germany: ACM.

Alexandrov, Daniel, Viktor Karepin, and Ilya Musabirov. 2016. 'Educational Migration from Russia to China: Social Network Data'. Pp. 309–11 in *Proceedings of the 8th ACM Conference on Web Science*. Hannover Germany: ACM.

Alexandrov, Daniel, Viktor Karepin, Ilya Musabirov, and Daria Chuprina. 2018. 'Educational Migration from Russia to the Nordic Countries, China and the Middle East. Social Media Data'. Pp. 49–50 in *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*. Lyon, France: ACM Press.

Alyukov, Maxim, and Margarita Zavadskaya. 2024. '"It's Not That Simple, We Don't Know the Whole Truth": The Effects of Disinformation Discourse in Wartime Russia'. Available at SSRN: https://ssrn.com/abstract=4681826 or http://dx.doi.org/10.2139/ssrn.4681826

Anastasiadou, Athina, Artem Volgin, and Douglas R. Leasure. 2024. 'War and Mobility: Using Yandex Web Searches to Characterize Intentions to Leave Russia after Its Invasion of Ukraine'. *Demographic Research* 50:205–20.

Avetian, Vladimir. 2022. *Consider the Slavs: Overt Discrimination and Racial Disparities in Rental Housing*. Working Paper. https://www.tse-fr.eu/sites/default/files/TSE/documents/conf/2022/echoppe/avetian.pdf

[2] Prolific (https://www.prolific.com) is a platform providing access to research participants, mostly in OECD countries.

Avetian, Vladimir, and Viktor Veterinarov. 2024. 'Slavs Only: Open Xenophobia and Racial Disparities in Rental Housing'. Available at SSRN: https://ssrn.com/abstract=4983808 or http://dx.doi.org/10.2139/ssrn.4983808

Badawy, Adam, Emilio Ferrara, and Kristina Lerman. 2018. 'Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign'. Pp. 258–65 in *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*. IEEE.

Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. 2020. 'Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017'. *Proceedings of the National Academy of Sciences* 117(1):243–50.

Bareikytė, Miglė, and Mykola Makhortykh. 2024. 'Digitally Witnessable War from *Pereklychka* to Propaganda: Unfolding Telegram Communication during Russia's War in Ukraine'. *Media, War & Conflict* 17506352241255890.

Bareikytė, Miglė, Mykola Makhortykh, Alexander Martin, Taras Nazaruk, and Yarden Skop. 2024. 'How Should Platforms Be Archived? On Sustainable Use Practices of a Telegram Archive to Study Russia's War against Ukraine'. *Media, Culture & Society* 01634437241245915.

Bessudnov, Alexey. 2023. 'Ethnic and Regional Inequalities in Russian Military Fatalities in Ukraine: Preliminary Findings from Crowdsourced Data'. *Demographic Research, 48 (31): 883-898*.

Bessudnov, Alexey, and Andrey Shcherbak. 2020. 'Ethnic Discrimination in Multi-Ethnic Societies: Evidence from Russia'. *European Sociological Review* 36(1):104–20.

Bessudnov, Alexey, Denis Tarasov, Viacheslav Panasovets, Veronica Kostenko, Ivan Smirnov, and Vladimir Uspenskiy. 2023. 'Predicting Perceived Ethnicity with Data on Personal Names in Russia'. *Journal of Computational Social Science* 6(2):589–608.

Bodrunova, Svetlana S., Olessia Koltsova, Sergey Koltcov, and Sergey Nikolenko. 2017. 'Who's Bad? Attitudes toward Resettlers from the Post-Soviet South versus Other Nations in the Russian Blogosphere'. *International Journal of Communication* 11:23.

Bodrunova, Svetlana S., Anna Litvinenko, Ivan Blekanov, and Dmitry Nepiyushchikh. 2021. 'Constructive Aggression? Multiple Roles of Aggressive Content in Political Discourse on Russian YouTube'. *Media and Communication* 9(1):181–94.

Bösch, Marcus, and Tom Divon. 2024. 'The Sound of Disinformation: TikTok, Computational Propaganda, and the Invasion of Ukraine'. *New Media & Society* 26(9):5081–5106.

Bryanov, Kirill, Reinhold Kliegl, Olessia Koltsova, Alex Miltsov, Sergei Pashakhin, Alexander Porshnev, Yadviga Sinyavskaya, Maksim Terpilovskii, and Victoria Vziatysheva. 2023. 'What Drives Perceptions of Foreign News Coverage Credibility? A Cross-National Experiment Including Kazakhstan, Russia, and Ukraine'. *Political Communication* 40(2):115–46.

Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova. 2024. 'Social Media and Xenophobia: Theory and Evidence from Russia'. Working paper. https://home.uchicago.edu/bursztyn/Social_Media_and_Xenophobia_2024.pdf

Chapkovski, Philipp. 2023. 'Conducting Interactive Experiments on Toloka'. *Journal of Behavioral and Experimental Finance* 37:100790.

Chapkovski, Philipp, and Max Schaub. 2022. 'Solid Support or Secret Dissent? A List Experiment on Preference Falsification during the Russian War against Ukraine'. *Research & Politics* 9(2):20531680221108328.

Chen, Emily, and Emilio Ferrara. 2023. 'Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War between Ukraine and Russia'. Pp. 1006–13 in *Proceedings of the International AAAI Conference on Web and social media*. Vol. 17.

Couper, Mick P. 2017. 'New Developments in Survey Data Collection'. *Annual Review of Sociology* 43(Volume 43, 2017):121–45.

Crilley, Rhys, Marie Gillespie, Bertie Vidgen, and Alistair Willis. 2022. 'Understanding RT's Audiences: Exposure Not Endorsement for Twitter Followers of Russian State-Sponsored Media'. *The International Journal of Press/Politics* 27(1):220–42.

Dollbaum, Jan Matti, and Seongcheol Kim. 2024. 'Going Jingo: A Classification of the Wartime Positions of Russia's "Systemic Opposition" Parties'. *Post-Soviet Affairs* 40(3):222–41.

Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. 'Computational Social Science and Sociology'. *Annual Review of Sociology* 46(1):61–81.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. 'Social Media and Protest Participation: Evidence From Russia'. *Econometrica* 88(4):1479–1514.

Erbsen, Heidi, and Siim Põldre. 2023. 'Is All Russian News the Same? Framing in Russian News Media Generated by the Yandex News Algorithm for the United States, Estonia, and Russia'. *Journalism* 24(8):1789–1816.

Filer, Tanya, and Rolf Fredheim. 2016. 'Sparking Debate? Political Deaths and Twitter Discourses in Argentina and Russia'. *Information, Communication & Society* 19(11):1539–55.

Frye, Timothy, Scott Gehlbach, Kyle L. Marquardt, and Ora John Reuter. 2023. 'Is Putin's Popularity (Still) Real? A Cautionary Note on Using List Experiments to Measure Popularity in Authoritarian Regimes'. *Post-Soviet Affairs* 39(3):213–22.

Golovchenko, Yevgeniy. 2022. 'Fighting Propaganda with Censorship: A Study of the Ukrainian Ban on Russian Social Media'. *The Journal of Politics* 84(2):639–54.

Goonetilleke, Priyanka, Alex Knorre, and Artem Kuriksha. 2023. 'Hydra: Lessons from the World's Largest Darknet Market'. *Criminology & Public Policy* 22(4):735–77.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2021. 'Machine Learning for Social Science: An Agnostic Approach'. *Annual Review of Political Science* 24: 395–419.

Haq, Ehsan-Ul, Gareth Tyson, Lik-Hang Lee, Tristan Braud, and Pan Hui. 2022. 'Twitter Dataset for 2022 Russo-Ukrainian Crisis'. http://arxiv.org/abs/2203.02955

Kadlecová, Lucie, and Viktor Paggio. 2025. 'Russia's Weak Spots in Cyber Sovereignty: How the West Can Keep Russian Citizens' Access to Online Information Free from the Kremlin Interference'. *Democracy and Security* 0(0):1–26.

Kalsaas, Johanne. 2023. 'Area Studies Online? Opportunities and Challenges When Researching" Digital Russia" during the War on Ukraine'. *Russian Analytical Digest* (293):6–10.

Khokhlov, Nikita. 2024. '"We Don't Abandon Our Own People": Public Rhetoric of Russia's Governors during the Full-Scale Invasion of Ukraine'. *Post-Soviet Affairs* 40(4):278–95.

Koltsova, Olessia. 2022. 'Social Media for Joint Experimental, Survey, and Observational Data Collection: The Case of VKontakte'. *Social Media Research Methods* 633.

Koltsova, Olessia, and Sergei Koltcov. 2013. 'Mapping the Public Agenda with Topic Modeling: The Case of the Russian Livejournal'. *Policy & Internet* 5(2):207–27.

Koltsova, Olessia, and Sergei Pashakhin. 2020. 'Agenda Divergence in a Developing Conflict: Quantitative Evidence from Ukrainian and Russian TV Newsfeeds'. *Media, War & Conflict* 13(3):237–57.

Koltsova, Olessia, Alexander Porshnev, and Yadviga Sinyavskaya. 2021. 'Social Media-Based Research of Interpersonal and Group Communication in Russia'. *The Palgrave Handbook of Digital Russia Studies* 335–52.

Koltsova, Olessia, and Andrey Shcherbak. 2015. '"LiveJournal Libra!": The Political Blogosphere and Voting Preferences in Russia in 2011–2012'. *New Media & Society* 17(10):1715–32.

Koltsova, Olessia Y., Larisa V. Mararitsa, Maxim A. Terpilovskii, and Yadviga E. Sinyavskaya. 2021. ‘Social Signature in an Online Environment: Stability and Cognitive Limits’. *Computers in Human Behavior* 122:106856.

Kravets, Daria, and F. Toepfl. 2022. ‘Gauging Reference and Source Bias over Time: How Russia’s Partially State-Controlled Search Engine Yandex Mediated an Anti-Regime Protest Event’. *Information, Communication & Society* 25(15):2207–23.

Krishnarajan, Suthan, and Jakob Tolstrup. 2023. ‘Pre-War Experimental Evidence That Putin’s Propaganda Elicited Strong Support for Military Invasion among Russians’. *Science Advances* 9(45):eadg1199.

Kulyk, Volodymyr. 2024. ‘Language Shift in Time of War: The Abandonment of Russian in Ukraine’. *Post-Soviet Affairs* 40(3):159–74.

Kuznetsova, Daria. 2024. ‘Broadcasting Messages via Telegram: Pro-Government Social Media Control During the 2020 Protests in Belarus and 2022 Anti-War Protests in Russia’. *Political Communication* 41(4):509–30.

La Lova, Lanabi. 2023. ‘Methods in Russian Studies: Overview of Top Political Science, Economics, and Area Studies Journals’. *Post-Soviet Affairs* 39(1–2):27–37.

Lankina, Tomila, and Kohei Watanabe. 2017. ‘“Russian Spring” or “Spring Betrayal”? The Media as a Mirror of Putin’s Evolving Strategy in Ukraine’. *Europe-Asia Studies* 69(10):1526–56.

Lazer, David, and Jason Radford. 2017. ‘Data Ex Machina: Introduction to Big Data’. *Annual Review of Sociology* 43:19–39.

Leasure, Douglas R., Ridhi Kashyap, Francesco Rampazzo, Claire A. Dooley, Benjamin Elbers, Maksym Bondarenko, Mark Verhagen, Arun Frey, Jiani Yan, Evelina T. Akimova, Masoomali Fatehkia, Robert Trigwell, Andrew J. Tatem, Ingmar Weber, and Melinda C. Mills. 2023. ‘Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data’. *Population and Development Review* 49(2):231–54.

Leontyeva, Xenia, Olessia Koltsova, and Deb Verhoeven. 2024. ‘Gender (Im) Balance in the Russian Cinema: On the Screen and behind the Camera’. *Journal of Cultural Analytics* 9(1).

Levada, Centre. 2024. ‘The Audience of the Users of the Internet, Social Networks, Messengers and VPN Services’. https://www.levada.ru/2024/04/23/auditoriya-polzovatelej-interneta-sotsialnyh-setej-messendzherov-i-vpn-servisov/

Libman, Alexander. 2023. ‘Credibility Revolution and the Future of Russian Studies’. *Post-Soviet Affairs* 39(1–2):60–69.

Makhortykh, Mykola, Aleksandra Urman, and Mariëlle Wijermars. 2022. ‘A Story of (Non) Compliance, Bias, and Conspiracies: How Google and Yandex Represented Smart Voting during the 2021 Parliamentary Elections in Russia’. *Harvard Kennedy School Misinformation Review* 3(2):1–16.

Malik, Valeriya. 2019. ‘The Russian Panel Study “Trajectories in Education and Careers”’. *Longitudinal and Life Course Studies* 10(1):125–44.

Mediazona. 2022. ‘Russian Casualties in Ukraine. Mediazona Count, Updated’. *Mediazona*. https://en.zona.media/article/2022/05/11/casualties_eng

Milkova, Maria, Maksim Rudnev, and Lidia Okolskaya. 2023. ‘Detecting Value-Expressive Text Posts in Russian Social Media’. http://arxiv.org/abs/2312.08968

Molina, Mario, and Filiz Garip. 2019. ‘Machine Learning for Sociology’. *Annual Review of Sociology* 45(1):27–45.

Molotov, Kirill, and Daria Khlevniuk. 2024. ‘“Five Unknown Facts About…” How Stalin Is Represented on Russian-Language TikTok’. *Communist and Post-Communist Studies* 57(3):81–103.

Murthy, Dhiraj. 2024. ‘Sociology of Twitter/X: Trends, Challenges, and Future Research Directions’.

Oleinik, Anton. 2024. ‘Telegram Channels Covering Russia’s Invasion of Ukraine: A Comparative Analysis of Large Multilingual Corpora’. *Journal of Computational Social Science* 7(1):361–84.

Orttung, Robert W., and Elizabeth Nelson. 2019. ‘Russia Today’s Strategy and Effectiveness on YouTube’. *Post-Soviet Affairs* 35(2):77–92.

Otlan, Yana, Yulia Kuzmina, Aleksandra Rumiantseva, and Katerina Tertytchnaya. 2023. ‘Authoritarian Media and Foreign Protests: Evidence from a Decade of Russian News’. *Post-Soviet Affairs* 39(6):391–405.

Panicheva, Polina, Larisa Mararitsa, Semen Sorokin, Olessia Koltsova, and Paolo Rosso. 2022. ‘Predicting Subjective Well-Being in a High-Risk Sample of Russian Mental Health App Users’. *EPJ Data Science* 11(1):21.

Primig, Florian, Hanna Dorottya Szabó, and Pilar Lacasa. 2023. ‘Remixing War: An Analysis of the Reimagination of the Russian–Ukraine War on TikTok’. *Frontiers in Political Science* 5:1085149.

Pronoza, Ekaterina, Polina Panicheva, Olessia Koltsova, and Paolo Rosso. 2021. ‘Detecting Ethnicity-Targeted Hate Speech in Russian Social Media Texts’. *Information Processing & Management* 58(6):102674.

Reuter, Ora John, and David Szakonyi. 2015. ‘Online Social Media and Political Awareness in Authoritarian Regimes’. *British Journal of Political Science* 45(1):29–51.

Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro. 2016. ‘An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions’. *American Journal of Political Science* 60(3):783–802.

Roshchin, Sergey, Sergey Solntsev, and Dmitry Vasilyev. 2017. ‘Recruiting and Job Search Technologies in the Age of Internet’. *Foresight and STI Governance* 11(4):33–43.

Rozenas, Arturas, and Denis Stukal. 2019. ‘How Autocrats Manipulate Economic News: Evidence from Russia’s State-Controlled Television’. *The Journal of Politics* 81(3):982–96.

Rykov, Yuri G., Peter A. Meylakhs, and Yadviga E. Sinyavskaya. 2017. ‘Network Structure of an AIDS-Denialist Online Community: Identifying Core Members and the Risk Group’. *American Behavioral Scientist* 61(7):688–706.

Rykov, Yuri, Olessia Koltsova, and Yadviga Sinyavskaya. 2020. ‘Effects of User Behaviors on Accumulation of Social Capital in an Online Social Network’. *PLOS ONE* 15(4):e0231837.

Salganik, Matthew J. 2018. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

Semenov, Aleksandr, Alexey Natekin, Sergey Nikolenko, Philipp Upravitelev, Mikhail Trofimov, and Maxim Kharchenko. 2015. ‘Discerning Depression Propensity Among Participants of Suicide and Depression-Related Groups of Vk.Com’. Pp. 24–35 in *Analysis of Images, Social Networks and Texts*. Vol. 542, *Communications in Computer and Information Science*, edited by M. Yu. Khachay, N. Konstantinova, A. Panchenko, D. Ignatov, and V. G. Labunets. Cham: Springer International Publishing.

Shevchuk, Andrey, Denis Strebkov, and Alexey Tyulyupo. 2021. ‘Always on across Time Zones: Invisible Schedules in the Online Gig Economy’. *New Technology, Work and Employment* 36(1):94–113.

Shirikov, Anton. 2024. ‘Fake News for All: How Citizens Discern Disinformation in Autocracies’. *Political Communication* 41(1):45–65.

Sivak, Elizaveta, and Ivan Smirnov. 2019. ‘Parents Mention Sons More Often than Daughters on Social Media’. *Proceedings of the National Academy of Sciences* 116(6):2039–41.

Smirnov, Ivan. 2018. ‘Predicting PISA Scores from Students’ Digital Traces’. in *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 12.

Smirnov, Ivan. 2019. ‘Schools Are Segregated by Educational Outcomes in the Digital Space’. *PloS One* 14(5):e0217142.

Smirnov, Ivan. 2020. ‘Estimating Educational Outcomes from Students’ Short Texts on Social Media’. *EPJ Data Science* 9(1):27.

Smirnov, Ivan, Elizaveta Sivak, and Yana Kozmina. 2016. ‘In Search of Lost Profiles: The Reliability of VKontakte Data and Its Importance in Educational Research’. *Educational Studies Moscow 4*:106–22.

Smirnov, Ivan, and Stefan Thurner. 2017. ‘Formation of Homophily in Academic Performance: Students Change Their Friends Rather than Performance’. *PloS One* 12(8):e0183473.

Sokolova, Alesya. 2023a. ‘Armchair Peacekeepers’. *Novaya Gazeta Europe*. https://novayagazeta.eu/articles/2023/07/21/armchair-peacekeepers-en.

Sokolova, Alesya. 2023b. ‘Opponents and Proponents of the War in Ukraine in Russian Social Media: Who Are They?’ http://arxiv.org/abs/2308.04473

Sokolova, Alesya. 2023c. ‘Those Who Leave and Those Who Stay’. *Novaya Gazeta Europe*. https://novayagazeta.eu/articles/2023/11/03/those-who-leave-and-those-who-stay-en.

Sokolova, Alesya. 2024. ‘Tailored Propaganda: How Russia Manipulates Public Opinion in VK’. *CEDAR*. https://cedarus.io/research/undefined/research/tailored-propaganda.

Sokolova, Alesya, Svyatoslav Kovalev, Georgy Vasiliev, and Katya Orlova. 2024. ‘The Z Universe’. *Novaya Gazeta Europe*. https://novayagazeta.eu/articles/2024/02/16/vselennaia-z.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. ‘Detecting Bots on Russian Political Twitter’. *Big Data* 5(4):310–24.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2022. ‘Why Botter: How pro-Government Bots Fight Opposition in Russia’. *American Political Science Review* 116(3):843–57.

Stukal, Denis, Sergey Sanovich, Joshua A. Tucker, and Richard Bonneau. 2019. ‘For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia’. *SAGE Open* 9(2):215824401982771.

Tschirky, Michael, and Mykola Makhortykh. 2024. ‘#Azovsteel: Comparing Qualitative and Quantitative Approaches for Studying Framing of the Siege of Mariupol on Twitter’. *Media, War & Conflict* 17(2):163–78.

Urman, Aleksandra. 2019. ‘News Consumption of Russian Vkontakte Users: Polarization and News Avoidance’. *International Journal of Communication* 13(0):25.

Urman, Aleksandra, and Stefan Katz. 2022. ‘What They Do in the Shadows: Examining the Far-Right Networks on Telegram’. *Information, Communication & Society* 25(7):904–23.

Urman, Aleksandra, and Mykola Makhortykh. 2022. ‘My War Is Your Special Operation: Engagement with pro- and Anti-Regime Framing of the War in Ukraine on Russian Social Media’. https://osf.io/67snk

Veterinarov, Viktor, and Vladimir Ivanov. 2018. ‘Slavs Only: Ethnic Discrimination and Rental Prices’. *Available at SSRN: https://ssrn.com/abstract=3249624 or http://dx.doi.org/10.2139/ssrn.3249624*

Volkov, Vadim. 2016. ‘Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia’s Criminal Courts’. *Journal of Empirical Legal Studies* 13(4):637–65.

Zakhlebin, Igor, Aleksandr Semenov, Alexander Tolmach, and Sergey Nikolenko. 2016. ‘Detecting Opinion Polarisation on Twitter by Constructing Pseudo-Bimodal Networks of Mentions and Retweets’. Pp. 169–78 in *Information Retrieval*. Vol. 573, *Communications in Computer and Information Science*, edited by P. Braslavski, I. Markov, P. Pardalos, Y. Volkovich, D. I. Ignatov, S. Koltsov, and O. Koltsova. Cham: Springer International Publishing.

Zherebtsov, Mikhail, and Sergei Goussev. 2021. ‘Tweeting Russian Politics: Studying Online Political Dynamics’. *The Palgrave Handbook of Digital Russia Studies* 537–67.