DISCUSSION PAPER SERIES

DP14877

SOCIAL MEDIA AND XENOPHOBIA: EVIDENCE FROM RUSSIA

Maria Petrova, Leonardo Bursztyn, Georgy Egorov and Ruben Enikolopov

PUBLIC ECONOMICS

ISSN 0265-8003

SOCIAL MEDIA AND XENOPHOBIA: EVIDENCE FROM RUSSIA

Maria Petrova, Leonardo Bursztyn, Georgy Egorov and Ruben Enikolopov

Discussion Paper DP14877
Published 12 June 2020
Submitted 12 June 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SOCIAL MEDIA AND XENOPHOBIA: EVIDENCE FROM RUSSIA

Abstract

We study the causal effect of social media on ethnic hate crimes and xenophobic attitudes in Russia and the mechanisms underlying this effect, using quasi-exogenous variation in social media penetration across cities. Higher penetration of social media led to more hate crimes in cities with a high pre-existing level of nationalist sentiment. Consistent with a mechanism of coordination of crimes, the effects are stronger for crimes with multiple perpetrators. Using a national survey experiment, we also find evidence of a mechanism of persuasion: social media led individuals (especially young, male, and less-educated ones) to hold more xenophobic attitudes.

JEL Classification: D7, H0, J15

Keywords: social media, Xenophobia, Hate crime, Russia

Maria Petrova - maria.petrova@upf.edu
Universitat Pompeu Fabra and CEPR

Leonardo Bursztyn - bursztyn@uchicago.edu
University of Chicago

Georgy Egorov - g-egorov@kellogg.northwestern.edu
Northwestern University

Ruben Enikolopov - ruben.enikolopov@upf.edu
ICREA-UPF and Barcelona GSE and CEPR

Acknowledgements
We are grateful to Garance Genicot, Matthew Gentzkow, Mike Golosov, Brian Knight, Andrew Little, Andreas Madestam, Natalya Naumenko, Nancy Qian, Aakaash Rao, Alvaro Sandroni, Daniel Spiro, David Strömberg, Alireza Tahbaz-Salehi, and Katia Zhuravskaya for very helpful discussions, to Danil Fedchenko for excellent research assistance, and to numerous seminar participants at AEA, BEROC, BI Norwegian, Chicago Harris, DCPEC Political Economy Seminar, EEA, HSE, Northwestern, Pompeu Fabra, Stockholm, Uppsala, and Social Media and Political Participation NYU conference for helpful comments and suggestions. Ruben Enikolopov and Maria Petrova acknowledge financial support from the BBVA Foundation grant.

Social Media and Xenophobia: Evidence from Russia*

Leonardo Bursztyn† **Georgy Egorov**‡
Ruben Enikolopov§ **Maria Petrova**¶

June 2020

Abstract

We study the causal effect of social media on ethnic hate crimes and xenophobic attitudes in Russia and the mechanisms underlying this effect, using quasi-exogenous variation in social media penetration across cities. Higher penetration of social media led to more hate crimes in cities with a high pre-existing level of nationalist sentiment. Consistent with a mechanism of *coordination* of crimes, the effects are stronger for crimes with multiple perpetrators. Using a national survey experiment, we also find evidence of a mechanism of *persuasion*: social media led individuals (especially young, male, and less-educated ones) to hold more xenophobic attitudes.

Keywords: social media, xenophobia, hate crime, Russia
JEL Classification: D7, H0, J15

*We are grateful to Garance Genicot, Matthew Gentzkow, Mike Golosov, Brian Knight, Andrew Little, Andreas Madestam, Natalya Naumenko, Nancy Qian, Aakaash Rao, Alvaro Sandroni, Daniel Spiro, David Strömberg, Alireza Tahbaz-Salehi, and Katia Zhuravskaya for very helpful discussions, to Danil Fedchenko for excellent research assistance, and to numerous seminar participants at AEA, BEROC, BI Norwegian, Chicago Harris, DCPEC Political Economy Seminar, EEA, HSE, Northwestern, Pompeu Fabra, Stockholm, Uppsala, and Social Media and Political Participation NYU conference for comments and suggestions. Ruben Enikolopov and Maria Petrova acknowledge financial support from the BBVA Foundation grant.

†University of Chicago and NBER, bursztyn@uchicago.edu

‡Northwestern University–Kellogg School of Management and NBER, g-egorov@kellogg.northwestern.edu

§New Economic School, ICREA-Universitat Pompeu Fabra, Barcelona IPEG, and Barcelona GSE, renikolopov@nes.ru

¶ICREA-Universitat Pompeu Fabra, Barcelona IPEG, New Economic School, and Barcelona GSE, petrova.ma@gmail.com

1 Introduction

In recent years, the world has witnessed a large increase in expression of hate and xenophobia.[1] Candidates and platforms endorsing nationalism and views associated with intolerance toward specific groups have also gathered increased popular support both in the U.S. and across Europe. Social media has been widely named a major factor in the increase in expression of hate, and hate crimes in particular.[2] In this paper, we document the causal effects of social media exposure on violent hate crimes and xenophobic attitudes in Russia, a country with more than 180 ethnic groups. We also combine observational data with survey experiments to provide evidence of the particular mechanisms behind these effects.

Conceptually, social media may affect expression of hate, and hate crimes in particular, through two main channels. First, social media can facilitate coordination and collective action. For example, it has been shown that social media facilitates the coordination of political protests.[3] Coordination through social media may be particularly relevant for illegal and stigmatized activities, such as hate crimes, as social media makes it easier to find like-minded people through online communities and groups, and possibly to out oneself as someone having such views. We refer to this as the *coordination* channel. Second, social media may influence individual opinions: previously tolerant individuals might become exposed to intolerant views, while intolerant individuals might end up in "echo chambers" (Sunstein, 2001, 2017; Settle, 2018) that might make their views even more extreme. We refer to this as the *persuasion* channel.

The main challenge in identifying a causal effect of social media is that access and consumption of social media are not randomly assigned. To overcome this challenge, we follow the approach

[1] For example, according to the Center for the Study of Hate and Extremism, across thirty major cities in the U.S., the total number of hate crimes in 2018 was 42% higher than in 2010, following five consecutive years of increases. See the Center's 2019 report.

[2] See, for example, "How Everyday Social Media Users Become Real-World Extremists," *New York Times*, April 25, 2018.

[3] See e.g. Enikolopov et al. (forthcoming), Manacorda and Tesei (2020), Fergusson and Molina (2020).

from Enikolopov et al. (forthcoming). This approach exploits the history of the main Russian social media platform, *VKontakte* (VK). This online social network, which is analogous to *Facebook* in functionality and design, was the first mover in the Russian market and secured its dominant position with a user share of over 90 percent by 2011. VK was launched in October 2006 by Pavel Durov, who was at the time an undergraduate student at Saint Petersburg State University (SPbSU). Initially, users could only join the platform by invitation through a student forum of the University, which had also been created by Durov. The vast majority of early users of VK were, therefore, Durov's fellow students of SPbSU. This, in turn, made friends and relatives of these students more likely to open an account early on. Since SPbSU attracted students from across the country, this sped up the propagation of VK in the cities these students had come from. As a result, the idiosyncratic variation in the distribution of the home cities of Durov's classmates had a long-lasting effect on VK penetration. This allows us to use fluctuations in the distribution of SPbSU students across cities as an instrument for the city-level penetration of VK.[4] We then evaluate the effect of higher VK penetration on hate crimes and attitudes towards other ethnicities using data on hate crimes, collected between 2007 and 2015 by an independent Russian NGO, *SOVA*, as well as newly collected data on hate attitudes from a survey experiment we designed and implemented.

Using this instrumental variables approach, we first show that higher penetration of social media led to more ethnic hate crimes, but only in cities with a higher baseline level of nationalist sentiment prior to the introduction of social media. To proxy for baseline local nationalist sentiment, we use the city-level vote share of *Rodina* ("Motherland"), an explicitly nationalist and xenophobic party, in the 2003 parliamentary election, the last one before the creation of VK. We show that the city-level impact of social media on hate crime victims positively and significantly depends on the strength of pre-existing support of nationalists in the city: for example, a 10% increase in VK penetration increased hate crimes by 25.8% in cities where Rodina received most votes, but had no

[4]To deal with possibility that cities with taste for social media were also more likely to send students to SPbSU, we control for the distribution of SPbSU students in cohorts several years older and several years younger than VK founder.

effect in cities where Rodina got minimal support. The effect being driven by areas with higher pre-existing nationalism is potentially consistent with a mechanism of social media facilitating the co-ordination of hate crimes among people already predisposed to disliking ethnic minorities. Further consistent with this mechanism is our additional finding that the effects of social media penetration in areas with higher support for Rodina was especially strong for crimes with *multiple* perpetrators.

This heterogeneity with respect to pre-existing levels of nationalism is also consistent with the literature on traditional media, which suggests that media effects often depend on a predisposition of the population. For example, Chiang and Knight (2011) show that the effect of newspaper endorsements was the most influential for moderate voters, while Adena et al. (2015) find that radio propaganda by the Nazis in the 1930s was effective only in areas with historically high level of antisemitism.[5]

To understand if social media simply serves as a coordination device or if it can also change opinions, we examine the *persuasion* mechanism by studying the effects of social media on xenophobic *attitudes*. To that end, we designed and conducted an online survey experiment in the summer of 2018, with over 4,000 respondents from 125 cities.[6] The survey was framed as a study of patterns of usage of social media and the Internet, to which we added our question of interest, related to ethnic hostility.

Given the possibility that a stigma, associated with directly reporting xenophobic views in a survey, can prevent our respondents from truthfully reporting their opinions, we use the list experiment technique. This is one of the main methods to elicit truthful answers to sensitive survey questions (Blair and Imai, 2012, Glynn, 2013), and it has been shown to perform particularly well in online surveys (Coutts and Jann, 2011).[7] Specifically, respondents were asked about their

[5]Relatedly, Yanagizawa-Drott (2014) documents that the effect of traditional media on conflict (radio in the context of 1994 Rwandan genocide) propagated through offline social interactions.

[6]This survey was pre-registered on the AEA RCT Registry website under entry AEARCTR-0003066.

[7]The intuition behind this technique is that the respondents are asked only to indicate the number of statements with which they agree from a list. By adding the statement of interest to a random subgroup of respondents, one can estimate the share of respondents agreeing with this statement without being able to identify who exactly agrees with it. We discuss the procedure in more detail in subsection 3.2.

agreement with a statement (that we borrowed from existing surveys): "' *I feel annoyance or dislike toward some ethnicities*."

We find a positive effect of social media penetration on elicited ethnic hostility, i.e., the share of respondents that hold xenophobic attitudes. The magnitude of the effect is particularly large in certain subsamples, specifically younger respondents and those with lower levels of education, as well as males. Those are precisely the groups more likely to use social media and to be engaged in hate crime.[8] Numerically, a 10% increase in VK penetration makes respondents 2.0% more likely to agree with the hateful statement in the list experiment, with this magnitude going up to 2.8% for younger respondents, to 4.4% for those with low education, and to 2.9% for males.

Furthermore, for the subset of respondents randomly assigned to the list *not* containing the statement of interest (that is, the "control" list) we added the question about agreement with the same statement as a direct question after the list experiment question. This allows us to examine the effect of social media on ethnic hostility, as reported without the cover provided by the list experiment. We do not find evidence a positive effect of social media on these direct, self-reported xenophobic preferences. We also obtain similar (null) results if we use the answers to the same direct question from a much larger, nationally representative survey of more than 30,000 respondents conducted in 2011 by one of the biggest Russian survey companies, FOM (*Fond Obschestvennogo Mneniya*, Public Opinion Foundation). Our findings, therefore, highlight the importance of properly eliciting sensitive attitudes – direct survey questions would have led to the incorrect conclusion that there was a null effect.[9]

Our paper contributes to a growing literature on the impact of social media on polarization, xenophobia, and hate crime. Allcott et al. (2020), Mosquera et al. (2020), and Yanagizawa-Drott

[8]This goes in line with the argument in Boxell et al. (2017) and Allcott and Gentzkow (2017) that the presumed impact of social media should be higher for groups more likely to be affected.

[9]One can also use this finding to infer whether social media changed the perceived acceptability of reporting ethnic hostility in a survey (by comparing the effects on the measures with and without the cover provided by the list experiment). Our results suggest no evidence that social media reduced the perceived stigma associated with reporting ethnic hostility in a survey; if anything, the effects point in the opposite direction.

et al. (2019) provide evidence that social media makes people's political opinions more diverging. Gentzkow and Shapiro (2011) finds that interactions online are less segregated as compared to interactions offline with friends, colleagues, family members, or neighbors. At the same time, Halberstam and Knight (2016) show that the segregation of communications on social media (Twitter) is more pronounced and closer to the segregation in offline interactions. In contrast to all these papers, we study more extreme outcomes, such as hate crime and hate attitudes.

Two concurrent papers also examine the relationship between social media and hate crimes. Müller and Schwarz (2018) show that anti-refugee sentiment on Facebook predicts day-to-day changes in crimes against refugees in Germany. These results, however, do not speak to longer-lasting changes in the patterns of hate crime with the arrival of social media, and could instead reflect displacements of hate crime towards days with more xenophobic content. Müller and Schwarz (2019) find that anti-Muslim hate crimes in the United States have increased in counties with high Twitter penetration users, but only since the start of Donald Trump's presidential campaign. These findings are consistent with the previous literature on traditional media, which has found that top politicians' speeches can have an effect on people's hate-driven behavior.[10] However, their results also imply that social media had no effect on hate crime in the absence of a xenophobic President, or at least a major presidential candidate. Thus, despite important recent works on social media and hate crimes, the question whether the introduction of social media *per se* can lead to a persistent increase in hate crimes and xenophobia has remained open. Our paper fills this gap by providing evidence of a persistent effect of social media penetration on both hate crimes and hate attitudes – and an effect that is not driven by top politicians' messages. Moreover, in contrast to the previous literature, we examine the mechanisms behind the results.

Our paper also contributes to a larger literature on the effects of media and, in particular, social media on individual behavior. Enikolopov et al. (forthcoming), using an identification approach

[10] E.g. Adena et al. (2015) show that Hitler's speeches was a major component of radio content in Germany since January 1933, while Yanagizawa-Drott (2014) reports that the speeches by key government officials , including Prime Minister Jean Kambanda, were an intergral part of RTLM radio propaganda in Rwanda.

similar to ours, show that higher social media penetration increased the probability of political protests in Russia in 2011. Focusing on the effects of the internet, Manacorda and Tesei (2020) show that 2G penetration in Africa led to stronger cell-level protest participation, and Campante et al. (2018) examine the impact of broadband internet on different forms of political participation in Italy.[11] In contrast to these studies, our paper looks at the impact on social media on hate crime, and adds a survey experiment to further investigate underlying mechanisms. We also contribute to the broader literature on the impact of social media on political economy outcomes.[12] Bond et al. (2012) show that political mobilization messages in Facebook increased turnout in the U.S. elections. Qin et al. (2017) find that publications in the Chinese microblogging platform Sina Weibo predict future protests, strikes, conflicts, while Qin et al. (2019) show that the spread of information over online social networks leads to the spread of offline protests and strikes in China. Enikolopov et al. (2018) show that anti-corruption blog posts by a popular Russian civic activist had a negative impact on market returns of targeted companies and led to a subsequent improvement in corporate governance. Sophie Hatte and Zhuravskaya (2020) and Sen and Yildirim (2016) show that social media affected reporting strategies of traditional media. Note that our contribution to this literature is not about developing a new identification strategy for the effects of social media; it is instead about understanding the longer-term effects of social media on an important set of outcomes, and getting at the mechanisms behind these effects.

We also add to a growing literature studying the recent rise in populism and nationalist attitudes. Bursztyn et al. (2019) and Enke (2019) study the 2016 U.S. election. Algan et al. (2017) show that the Great Recession triggered a trust crisis and led to higher voting shares of non-mainstream, particularly populist parties. Guriev et al. (2019) show that 3G penetration around the globe promoted populist voting and reduced government support.

[11] Relatedly, Acemoglu et al. (2018) find that protest-related activity on Twitter preceded actual protest activity on Tahrir Square in Egypt. Steinert-Threlkeld et al. (2015) show that the content of Twitter messages was associated with subsequent protests in the Middle East and North Africa countries during the Arab Spring.

[12] See Zhuravskaya et al. (2020) for a more detailed overview of this literature.

The remainder of this paper proceeds as follows. We discuss our identification strategy, data, and results on hate crimes in Section 2. In Section 3, we discuss our survey design and the results on xenophobic attitudes. Section 4 concludes.

2 Social Media and Hate Crimes

2.1 Identification Strategy

Our empirical strategy for identification of the causal effect of social media penetration follows the approach in Enikolopov et al. (forthcoming). In particular, we look at the penetration of the most popular social network in Russia, *VKontakte* (VK), which had substantially more users than Facebook throughout the whole period we analyze. For example, in 2011, te midpoint of our hate crime data, VK had 55 million users in Russia, while Facebook had 6 million users. VK was created in the fall of 2006 by Pavel Durov who at the time was a student at the Saint Petersburg State University (SPbSU). The first users of the network were largely students who studied with Durov at SPbSU. This made their friends and relatives at home more likely to open an account, which let to a faster spread of VK in these cities. Network externalities magnified these effects and, as a result, the distribution of the home cities of Durov's classmates had a long-lasting effect on VK penetration. In particular, the distribution of home cities of the students who studied at SPbSU at the same time as Durov predicts the penetration of VK across cities. This prediction is robust to controlling for the distribution of the home cities of the students who studied at SPbSU several years earlier or later. This effect persists throughout the period between 2007 and 2016 which we analyze, although the magnitude of the effect decreases over time. Thus, the effect of social media penetration is identified using cross-sectional variation in the number of VK users across Russian cities, and is driven by the number of students from different cities who happened to study at SPbSU at the time the network was created. The results of the first stage regression are reported in Table 1, and are also illustrated in Figure B1.[13] The number of students in Durov's cohort is positively and significantly (at 1% level) related to subsequent VK penetration, while the number of students in older or younger cohorts do not significantly predict VK spread. We also show that even though VK penetration is correlated with nationalistic party support, future VK penetration does not predict past nationalist party support, neither in the reduced form nor in the IV specifications in columns (2)-(3) of Table 1.

However, for outcomes observed in the late 2010s, the first stage becomes weaker over time. In most of our empirical tests, the strength of the instruments is not always high enough to make inference using conventional thresholds for the first stage F-statistics. Throughout the paper, we thus follow the recommendation in Andrews et al. (2019) and use the appropriate methods applicable in our particular case. In particular, in all tables we report weak instrument robust confidence sets developed by Chaudhuri and Zivot (2011) and Andrews (2017) and implemented in Stata by Sun (2018). Likewise, in all tables we denote the significance level of the endogenous coefficients based on these weak instrument robust sets and tests.

2.2 Data

Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis.[14] It is a Moscow-based Russian independent nonprofit organization providing information related to hate crimes, which is generally considered to be the most reliable source of information on that issue. The dataset covers incidents of violent hate crime, which include murder, battery, and death threat. These data have been collected consistently since 2007, with some incomplete data for 2004-2006. In the analysis we use data for 2007-2015. We classify all hate

[13]Note that we use a more succinct set of controls than Enikolopov et al. (forthcoming), because we have a much smaller number of cities and we face statistical power issues in the survey part of our analysis. The results of the analysis of the effect on hate crime (full sample of cities) are quantitatively and statistically similar if we use exactly the same list of controls as in Enikolopov et al. (forthcoming).

[14]The database can be found at https://www.sova-center.ru/en/database/violence/

crimes as "ethnic" or "non-ethnic" based on the type of victim reported in the database. Figures 1 and 2 show the number of hate crimes in general and ethnic hate crimes in particular in the 2007-2015 period across Russian cities on the map (for both the universe of cities and for the cities included in our survey). Table 2 presents more detailed information on the number of victims for each type. Based on the textual description of each incident in the database, we have also manually coded the number of perpetrators for every incident. Non-ethnic crimes are more likely to be committed by single perpetrators (Figure 3), whereas ethnic hate crimes (Figure 4) are more likely to be committed in groups, with the modal number of perpetrators being two. The average number of recorded hate crimes and hate crime victims has been declining over time (see Figures B2 and B3).[15]

A potential concern with this data is that there could be a differential likelihood of recording crimes in cities with different social media penetration in a way that is consistent with our results. For example, hate crimes might be more likely to be covered by mass media and get recorded in the database if they caught attention on social media. Although we do not have evidence to directly rule out this possibility, we believe that it is highly unlikely that ethnic hate crimes were disproportionately reported in areas with both higher penetration of VK *and* a higher baseline level of nationalist sentiment, *and* especially so for crimes with multiple perpetrators. We also do a couple of additional tests to ensure that this possibility is not biasing our results. First, we show that the effects are stronger in larger cities, where the role of social media on recording crimes is likely to be lower, since traditional media are more likely to cover crimes in these cities regardless of reports in social media. Second, the effect of social media becomes, if anything, smaller over time, while potential bias in reporting should become stronger over time, as social media penetration has been quickly increasing during the same time period (Figure B4).[16] Finally,

[15]This decline seems to be partly related to the large-scale government efforts to combat hate crimes and nationalistic organizations during this time period.

[16]The fact that the number of reported crimes and hate crime victims has been declining over time (Figures B2 and B3) is also inconsistent with reporting bias related to social media penetration.

our survey results on attitude changes are also consistent with social media having an effect beyond mere reporting of hate crimes. We discuss the evidence against differential reporting in more detail later in subsection 2.4, after summarizing the results.

The data on social media penetration and socioeconomic controls comes from Enikolopov et al. (forthcoming). The sample consists of 625 Russian cities with a population over 20,000 according to the 2010 Census.[17] To measure social media penetration we use information on the number of users of the most popular social media service in Russia, VK. In particular, we calculate the number of VK users who report a particular city as their city of residence in 2011, the midpoint for our hate crime data. We summarize the evolution of VK penetration over time in Figure B4.

We then use information on the city of origin of the students who studied at SPbSU. Specifically, we calculate the number of students coming from each city in five-year cohorts. We mostly focus on three cohorts in our analysis: i) those who were born the same year as the VK founder (1984) or within two years of it; ii) those who were born from three to seven years earlier than the VK founder; iii) those who were born from three to seven years later than the VK founder. Unfortunately, we do not have administrative data on students. Instead, following Enikolopov et al. (forthcoming), we collect this data based on the information provided in public accounts of the users of another social network, *Odnoklassniki* (Classmates), which was founded to help connect former classmates with each other. Note that more than 80% of Russian Internet users had an account in Odnoklassniki back in 2014, at the time these data were collected, and this share is likely to be even larger for recent students, who studied around the time Pavel Durov studied at the SPbSU. To deal with potential bias from our data collection, we control for Odnoklassniki penetration throughout the paper. We further investigate potential biases associated with our method of

[17]The exceptions are cities with similar names that caused problems with disambiguation in the data, as well as Moscow and Saint Petersburg, which are excluded from the sample as outliers. Both Moscow and St Petersburg have high levels of hate crime and high penetration of social media, thus their exclusion only leads to more conservative estimates. Besides, given that the majority of students in Saint Petersburg State University were from St Petersburg and Moscow, their year-to-year fluctuations exhibit less variation as compared to other cities, thus we cannot use our identification strategy for the sample with these cities included.

data collection in Appendix A.

As a measure of nationalist sentiment in a city *before* the creation of the VK social network we use the vote share of the *Rodina* ("Motherland") party in the parliamentary election of December 2003, the only election this party participated in and the last parliamentary election before the creation of VK. This party ran on an openly nationalist platform. It received 9.2 percent of the vote and got 37 of the 450 seats in the *State Duma*, the lower house of the Russian parliament. The data on electoral outcomes come from the Central Election Commission of the Russian Federation. We validate that the vote share for the party can serve as a proxy for nationalist sentiment in a city by showing that it is positively and significantly correlated with ethnic hate crime in the subsequent years, as well as with xenophobic attitudes revealed in the opinion polls (see Table B1).

City-level data on population, age, education, and ethnic composition come from the Russian Censuses of 2002 and 2010. Data on average wages come from the municipal statistics of RosStat, the Russian Statistical Agency. Additional city characteristics, such as latitude, longitude, year of city foundation, and the location of the administrative center, come from the Great Russian Encyclopedia.[18]

The data on attitudes towards other ethnicities come from a survey of over 4,000 individuals that we conducted in the summer of 2018 in 125 Russian cities. The survey was administered by a professional marketing firm, *Tiburon Research*, with a representative panel of urban Internet users in Russia. The survey was not designed to create a representative sample of the cities, and, to be able to conduct list experiment within each city, we tried to maximize the number of respondents per city. The resulting median number of respondents per city is 39.[19] The sample consists of 4,447 respondents, of which 2,221 were allocated to the control group and 2,226 to the treatment group.[20]

[18] The electronic version of the Encyclopedia can be found at https://bigenc.ru/

[19] On average, the cities in our survey sample were larger than the average city in the hate crime sample. It was done so as we aimed to have a sufficient number of respondents in treatment and control group, to be able to do within-city comparisons. Note that later on, to alleviate some concerns about the construction of our key variable, we repeat our hate crime analysis for the sample of larger cities with population above 50,000, and our baseline results only become stronger. Thus we do not expect any downward bias in the estimates because of our focus on the larger cities.

[20] More specifically, we collected the data in two batches, the pilot and the main experiment. As part of the pilot,

We also use data from the MegaFOM opinion poll conducted by FOM (*Fond Obschestvennogo Mneniya*, Public Opinion Foundation) in February 2011. This is a regionally representative survey of 54,388 respondents in 79 regions of Russia, of which 29,780 respondents come from 519 cities in our sample. In particular, we use information on answers to exactly the same direct question about hostility to different ethnicities that we asked in our survey in 2018.

2.3 Social Media and Hate Crime: Empirical Specification

Our main hypothesis is that social media penetration (specifically, VK penetration) has an impact on hate crime. Thus, we estimate the following model:

$$\text{HateCrime}_i = \beta_0 + \beta_1 \text{VKpenetration}_i + \beta_2 \mathbf{X_i} + \varepsilon_\mathbf{i}, \tag{1}$$

where $HateCrime_i$ is a measure of hate crime, which reflects either the total number of victims of hate crimes in city i during the period 2007-2015, or the number of victims of particular types of hate crime (ethnic or non-ethnic crimes, conducted by single or multiple perpetrators). VKpenetration_i is the logarithm of the number of VK users in city i in summer 2011.[21] This endogenous variable is instrumented using the number of students from each city in a five-year student cohort who have studied at the same year as the founder of VK, Durov, as well as one or two years earlier or later. $\mathbf{X_i}$ is a vector of control variables that include the number of students from the city in the other two five-year student cohorts, those that studied three to seven years

we surveyed 1,007 individuals from 20 cities. Individuals from this batch were randomized into three groups, with one containing a statement about ethnic minorities as part of the list experiment, another containing a statement about LGBTQ individuals, as well as a control group. As we found no reliable data on hate crimes against LGBTQ individuals, we dropped the second group of 336, leaving us with 671 individuals from the pilot. As part of the main experiment, we surveyed 4,034 individuals from 111 cities. In this batch, the cities were randomly chosen by the firm we were working with, and since we had the data on VK penetration for only 105 of these cities, we had to drop 246 observations from six cities. Additional 12 surveys were incomplete, which left us with 3,776 observations from the main part. In most analyses, we pool the two batches together, but our results are robust to looking at the second batch only. The survey was approved by the University of Chicago Institutional Review Board (IRB18-0858) and was pre-registered in the AEA RCT Registry (AEARCTR-0003066).

[21] We add one to the variable in our logarithm measures to deal with zeros.

earlier than Durov, and those that studied three to seven years later than Durov. It also includes the following socioeconomic controls: the logarithm of the population, the indicator for being a regional or a subregional (*rayon*) administrative center, the average wage in the city, the number of city residents of different five-year age cohorts, the share of the population with higher education in 2010 in each five-year age cohort, the indicator for the presence of a university in the city, ethnic fractionalization, and the logarithm of the number of Odnoklassniki users in 2014. For all specifications we report weak-instrument robust confidence sets.[22] Similarly, for our heterogeneity analysis we estimate the equation:

$$\text{HateCrime}_i = \beta_0 + \beta_1 \text{VKpenetration}_i \times \text{Nationalist Support}_i + \beta_2 \mathbf{X_i} + \varepsilon_\mathbf{i}, \tag{2}$$

where $\text{NationalistSupport}_i$ denotes the votes for the nationalist Rodina party in 2003 and $\mathbf{X_i}$ is a set of controls that includes direct effects of VKpenetration_i, $\text{NationalistSupport}_i$, and all the controls just described above.

2.4 Social Media and Hate Crime: Results

Table 3 summarizes the results of estimating Equation (1) for the average impact of exposure to VK on hate crime. While there is some significant relationship between social media penetration and hate crime in OLS specifications (Panel A), there is no consistent evidence of a significant effect of VK penetration on hate crime, for either ethnic- or non-ethnic- hate crime or for crimes conducted by single or multiple perpetrators in the IV specification (Panel B). At the same time, the confidence intervals do not allow us to rule out large effects (e.g., at maximum 66.0% increase,

[22] As discussed above, we report weak instrument robust confidence sets developed by Chaudhuri and Zivot (2011) and Andrews (2017) and implemented in Stata by Sun (2018) throughout the paper. We made a choice not to cluster standard errors, given that hate crime, in contrast to political protests, is primarily a local phenomenon, and hate crime patterns in cities 100-200 km apart are not related. Figures 1 and 2 illustrate these patterns. Moreover, if we regress hate crime on region fixed effects, we find that the fixed effects are not significantly related to the patterns of hate crime, neither individually nor collectively.

i.e., 0.69 of a standard deviation of the dependent variable in column 1, Panel B), and only one out of nine IV coefficients in the table is marginally significant.

However, these results mask an important heterogeneity of the effect with respect to the underlying level of nationalism. People in cities with very few nationalists to begin with and people from very nationalist cities can respond differently to the arrival of social media. To capture this dimension of heterogeneity into account, we interact VK penetration with a measure of pre-existing nationalist support, as captured by the Rodina party vote share in 2003.

Table 4 summarizes the results of estimating Equation (2). Panel A reports OLS results, and Panel B reports IV results. The nationalist party support variable is demeaned to simplify interpretation of the direct coefficients. The OLS relationship is positive and significant for all total hate crime and ethnic hate crime variables. It is presented visually in Figure 5, which illustrates that the link between the number of victims of hate crimes and VK penetration is significantly positive for cities with higher pre-existing levels of support for the nationalist party, while there is no significant relationship in cities with low levels for support of that party.[23]

Panel B presents the results for the IV specification. In all the IV specifications except one (column 8, Panel B), the effect of social media penetration on hate crime is significantly stronger in cities with higher pre-existing level of nationalism. Numerically, the results imply that the effect of a one standard deviation increase in social media penetration ranges from being close to zero (non-significant with different signs) at the minimum level of nationalist party support to a 25.8% increase in the total number of hate crime victims at the maximum level of nationalist support (column 1 of Table 4, Panel B).

The full version of this table with complete interactions between Rodina support and the number of students in the cohorts above and below Durov's is shown in Table B2. As one can see, there are no significant interaction coefficients for the older cohort and nationalistic party support. The

[23] In the picture, higher level of support is defined as being above the 85th percentile in the number of votes for the party, for illustrative purposes. The variable that captures pre-existing level of nationalism, used in Table 4 and others, is the actual vote share of the nationalist party (Rodina) in 2003, demeaned.

interaction terms for the younger cohort are occasionally significant (columns 2 and 5 in Table B2), but are never positive, consistent with idea that Durov's cohort was special in terms of generating hate crime in cities with high levels of pre-existing nationalism, and being among the first ones to get VK is the most likely channel for that. These results further strengthen our identification argument.[24] OLS coefficients (Panel A of Table 4) are approximately three times smaller than IV coefficients (Panel B of Table 4), which implies that some negative selection is going on; for example, social media users are more likely to have higher levels of education, while most hate crime perpetrators have lower levels of education.

The results indicate that in cities with high pre-existing level of nationalism, higher penetration of social media led to a higher total number of victims of hate crimes. Note that our results hold for the victims of ethnic and non-ethnic crimes, as well as of crimes conducted by either single or multiple perpetrators, i.e., for different types of hate crime. Another important takeaway from Table 4 is that the coefficients of interest are noticeably larger for the incidents that involved multiple perpetrators, i.e., acts of violence that require coordination.[25] At the same time, the results are also positive and significant for crimes committed by single perpetrators (with the exception of non-ethnic crime in column 8), which suggests that while social media may have facilitated coordination and thus contributed to hate crime, coordination alone may not fully explain the impact of social media (though coordination broadly speaking may take other forms, such as providing information on opportunities for committing crime to single perpetrators).

To interpret the evidence on the link between social media and hate crime victims presented in Tables 3 and 4, it is also important to distinguish between the intensive and extensive margins. In

[24] We also report reduced-form estimation results in Table B3. The interaction of pre-existing nationalism with the size of Durov's cohort is positive and significant in 8 out of 9 specifications. In contrast, the coefficients for older and younger cohorts are much smaller and are mostly insignificant; if anything, 2 out of 18 coefficients in this table are significant, and those are negative, not positive. These results are further consistent with VK's impact on hate crime, and, importantly, these results do not suffer from the weak instrument problem.

[25] In the seemingly unrelated regressions specification, the difference between the interaction coefficients in columns 2 (single perpetrator) and column 3 (multiple perpetrators) is statistically significant at the 10% level; the differences for ethnic and non-ethnic crimes are similarly large in magnitude.

Table B4, we estimate Equation (2) with the number of crimes rather than the number of victims as the dependent variable. The results suggest that the number of crimes responds to the introduction of social media and to the number of victims very similarly, both in terms of magnitude and statistical significance. For example, the impact of a 10% increase in social media penetration on the number of crimes is bounded by 24.8% for total crimes, a figure very similar to the maximal effect on the number of victims. In other words, the increase in the number of hate crime victims is well explained by the increase in the number of crimes, so it is the extensive margin that seems to play the role.

We also attempt to understand the evolution of the impact of social media over time. The beginning of our time period, 2007-2009, was the time of a rapid introduction of social media into people's lives, with the total number of VK users growing from about a hundred thousand to more than thirty million, while by 2013-2015 the exponential growth had already stopped and other platforms, such as Twitter, started to gain some popularity. At roughly the same time, following the Arab Spring and the protests in Russia in 2011-2012, the Russian government began to regulate online content, which prevented openly xenophobic communities from being created and sustained. We find that the effect of social media has been declining over time.[26]

Table 5 reports the results of placebo regressions for hate crime in the period 2004-2006, i.e., *before* the creation of the VK social network. The results indicate no significant effect of social media on hate crime even in cities with maximum level of support of the nationalist party, and the difference between these results and the results in Table 4 is statistically significant in seemingly unrelated regressions framework. These findings are consistent with the premise that social media

[26]If we examine the effect for the three 3-year sub-periods separately (see Table B5), one can see that the effects are similar in size in 2007-2009 and 2010-2012, but become noticeably smaller and statistically insignificant in 2013-2015. A reduced form version of this table is shown in Table B6. Note that in the latter table, the coefficients for the later period (2013-2015) are significantly larger than the coefficients for the middle period (2010-2012) in seemingly unrelated regressions framework, for total hate crimes, for total hate crimes, committed by multiple perpetrators, and for ethnic hate crimes, committed by single perpetrators. At the same time, for the results in IV version of this table, i.e. Table B5, we seem to lack the statistical power to make a proper comparison. We should also note that the predictive power of the instrument in the first stage regression is going down with time (see Figure B5).

has a causal effect on hate crime after the creation of VK in the end of 2006, in places with a higher level of nationalistic party support.[27]

As we mentioned above, a potential concern is that the results are driven by differential likelihoods of recording crimes that is correlated with explanatory variables. Although we do not have direct evidence directly ruling out differential likelihoods of recording crimes, we can do a couple of tests to alleviate this concern. First, we check if the effects that we identify negatively depends on the size of the cities. Arguably, in smaller cities with fewer traditional media news sources if any, reporting of hate crimes may be more dependent on whether they were discussed in social media or not. This should make the measurement error in hate crime data stronger in smaller cities. However, we find that, on the contrary, if we restrict the sample to cities with population above the median, the results become only stronger (see Table B8). Second, if indeed social media makes hate crimes more visible, this effect is supposed to be more pronounced as social media penetration goes up with time. However, if anything, we observe the opposite: the magnitude of the effect of social media is becoming smaller over time (see Tables B5 and B6). Finally, it is unlikely that ethnic hate crimes were disproportionately more reported in areas with both higher penetration of VK *and* a higher baseline level of nationalist sentiment, *and* especially so for crimes with multiple perpetrators. We should also note that our results on attitude changes are also consistent with social media having an effect beyond just the reporting of hate crimes (see the next section).

Overall, the results in Tables 3-5 indicate that social media had a positive effect on hate crime, but only in places where the level of nationalism was already sufficiently high before the creation of social media.

[27] Note that the null results in Table 5 may also be driven by the fact that the data for this time period are incomplete, in contrast to the later years. We also test whether the coefficients are statistically different for the period 2004-2006 and 2007-2015 in a pooled regression (see Table B7). In all specifications except one either the coefficient for the direct effect of VK penetration or its interaction with the support of the nationalistic party or both are statistically different from each other. Unfortunately, we cannot provide weak-instrument robust confidence intervals for this specification, as the triple-difference specification turns out to be too demanding and the confidence intervals often become degenerate and consisting of a single point.

3 Social Media and Hate Attitudes

While the results so far suggest a role of social media as a means of coordination of hate crime, it is important to evaluate whether social media has also played a persuasive role. Indeed, it is possible that social media made previously tolerant people more intolerant toward minorities, and previously intolerant people even more intolerant, and this could also contribute to an increase in hate crimes. As noted above, the results on crimes with a single perpetrator suggest that mechanisms other than coordination may be at play as well. To explore the persuasion channel, we designed and conducted a survey aimed at measuring the true level of underlying nationalism.

3.1 Survey Setup

To elicit xenophobic attitudes we conducted an online survey in 2018. The survey was administered in 125 cities chosen by the survey firm to create a representative sample of Russian urban Internet users and included a list experiment as part of it. This design (also called the "unmatched count" and the "item count technique") was originally formalized by Raghavarao and Federer, 1979 and further developed in recent works by Blair and Imai, 2012 and Glynn, 2013, among others. It is a standard technique for eliciting truthful answers to sensitive survey questions. The list experiment works as follows. First, respondents are randomly assigned to either a control group or a treatment group. Subjects in both groups are then asked to indicate the number of statements they agree with. In this way, the subject never reveals their agreement with any particular statement (unless the subject agrees with all or none, which is something the experimental design should try to avoid), only the total number of statements. In the control condition, the list contains a set of statements or positions that are not stigmatized. In the treatment condition, the list includes all the statements from control list, but also adds the statement of interest, which is potentially stigmatized (and in both cases, the positions of statements are randomly rotated). The support for the stigmatized opinion can then be inferred by comparing the average number of statements the subjects agree with in the treatment and control conditions. For recent applications of list experiments in economics, see Cantoni et al. (2019) and Enikolopov et al. (2017).

In our case, the survey participants were asked the following question: *"Consider, please, whether you agree with the following statements. Without specifying exactly which ones you agree with, indicate just the number of statements that you can agree with."* The respondents in the control group were given four statements unrelated to the issues of ethnicity.[28] The respondents in the treatment group were given the additional fifth statement: *"I feel annoyance or dislike toward some ethnicities."* Here, we took the exact wording used by one of the leading opinion polling firms in Russia in their regular large-scale surveys, which has the additional advantage of making our results comparable with the results of the opinion polls by this firm (see subsection 3.4 for more detail). Respondents in the control group, after answering the question on the number of statements they agreed with (which did not include the statement on ethnicities), were then asked a direct question about annoyance or dislike toward some ethnicities. Overall, the share of respondents who agreed with the xenophobic statement in the list experiment (i.e., the difference between the average number of statements with which respondents in treatment and control group agreed) was approximately 38 percent, while the percentage of respondents who admitted being xenophobic in the direct question was 33 percent.

Before proceeding with our analysis, it is important to note one important limitation we face. Due to the smaller number of cities in the survey sample (as compared to the full sample used in the analysis of hate crimes), we are not powered to study the impact of social media penetration interacted with pre-existing nationalism. Thus all the results in this section refer only to the average effects of social media penetration.

[28]The exact statements were the following: i) Over the week I usually read at least one newspaper or magazine; ii) I want to see Russia as a country with a high standard of living; iii) I know the name of the Chairman of the Constitutional Court of the Russian Federation; iv) Our country has a fairly high level of retirement benefits.

3.2 Elicited hostility, individual-level results

Given the randomization, comparing the mean number of positive answers between treatment and control groups provides a valid estimate of the percentage of respondents who agree with the sensitive statement about having xenophobic attitudes (Imai, 2011). However, our goal is to estimate the impact of an independent variable (social media penetration) on the answer to this sensitive question. Following Imai (2011) and Blair and Imai (2012), we use the regression model with interactions to estimate how answers to the list experiment question depend on other parameters, in our case, characteristics of the respondent's city. Formally, we estimate the following model:

$$\text{NumberOfStatements}_{ij} = \beta_0 + \beta_1 \text{T}_{ij} + \beta_2 \left(\text{T}_{ij} \times \text{VK}_j\right) + \beta_3 \text{VK}_j + \beta_4 \text{X}_{ij} + \varepsilon_{ij}, \tag{3}$$

where $\text{NumberOfStatements}_{ij}$ is the number of statements respondent i from city j agreed with, T_{ij} is the dummy variable for whether respondent i from city j was assigned to the treatment group, and VK_j is the measure of VK (social media) penetration in city j instrumented by the number of students from the city who studied at SPbSU together with the founder of VK. Other controls include older and younger cohorts of SPbSU students from a city, city-level controls, and the interaction of pre-existing nationalism with the treatment dummy to account for the differential response. Standard errors are clustered at the city level.

In this specification, the effect of social media penetration on the share of respondents in city j who have xenophobic attitudes is captured by the coefficient β_2. In what follows, we also look at subsamples, paying special attention to the groups more likely to be involved in hate crime: males, young respondents (below the median age in the sample, which is 32), and respondents with lower levels of education (below the median in our sample).[29]

[29]Note that we pre-registered heterogeneity by gender in our pre-analysis plan, but later we decided that these other simple characteristics (being young and low-educated) are also likely to predict being a hate crime perpetrator, and so

The results, presented in Table 6, indicate that social media increased elicited hostility to other ethnicities, both on average and for subgroups that are more likely to engage in hate crime (male, young, or low-educated).[30] The results in column 1 imply that, on average, the elasticity of elicited hostility with respect to social media penetration is 0.075. In other words, a 10% with this magnitude going up to 2.8% for younger respondents, to 4.4% for those with low education, and to 2.9% for males. increase in VK penetration increases the share of those agreeing with the statement in the list experiment by 2.0%. This magnitude goes up to 2.9% for males (column 2), 4.4% for those with low education (column 4), and 2.8% for younger respondents (column 6).[31] We do not find any significant effect of VK for females, those with higher education, or older respondents, and the magnitudes of the coefficients are considerably smaller for these groups. To put these numbers in a perspective, the average city in our sample has about 355,000 people, of which 27.1% are VK users. Our baseline coefficient (column 1 of Table 6) implies that 10% increase in VK penetration lead to 2654 more people with xenophobic views in the average city.[32]

3.3 Elicited hostility, city-level results

In this subsection, we repeat the analysis above at the city level. In what follows we refer to the variable $\text{NumberOfStatements}_{ij}$ as y_{ij}. Then, assuming that Equation (3) is a true data generating process, we derive the city-level specification we would like to estimate. More specifically, we first

we added them to the analysis.

[30] The number of cities varies slightly; it is 123 for older subsample, 124 for females and those with higher education, and full sample of 125 for the rest of the categories. That happens because for some cities we do not happen to have respondents in all categories.

[31] We got the first number by dividing one tenth of the effect, 0.0075, by 0.376, the average share of those agreeing with the xenophobic statement, as given by the difference between the number of options chosen by treatment and control groups. For the other columns, similar calculations apply.

[32] Note that we also have 58 people who do not use any social networks in our sample. If we repeat the estimation of equation (3) for this subsample of people, we get a negative but not significant effect of VK penetration interacted with a dummy for the list experiment option (magnitude -0.342, standard error 0.877). However, we are not able to compute weak instrument robust confidence sets for this (very small) subsample of non-users.

sum individual responses by city and treatment status:

$$\sum_{T_{ij}=0} y_{ij} = \beta_0 \sum_{T_{ij}=0} 1 + \beta_3 \mathrm{VK}_j \sum_{T_{ij}=0} 1 + \beta_4 \sum_{T_{ij}=0} \mathrm{X}_{ij} + \sum_{T_{ij}=0} \varepsilon_{ij};$$

$$\sum_{T_{ij}=1} y_{ij} = (\beta_0 + \beta_1) \sum_{T_{ij}=1} 1 + (\beta_2 + \beta_3)\, \mathrm{VK}_j \sum_{T_{ij}=1} 1 + \beta_4 \sum_{T_{ij}=1} \mathrm{X}_{ij} + \sum_{T_{ij}=1} \varepsilon_{ij}.$$

We then divide both sides of the last two equations by the number of respondents in each treatment group in a city ($\sum_{T_{ij}=a} 1$) and take the difference. We get

$$\frac{\sum_{T_{ij}=1} y_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} y_{ij}}{\sum_{T_{ij}=0} 1} = \beta_1 + \beta_2 \mathrm{VK}_j + \beta_4 \left[\frac{\sum_{T_{ij}=1} X_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} X_{ij}}{\sum_{T_{ij}=0} 1} \right] + \eta_j, \tag{4}$$

where we denoted $\left[\frac{\sum_{T_{ij}=1} \varepsilon_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} \varepsilon_{ij}}{\sum_{T_{ij}=0} 1} \right]$ as η_j to simplify notation.

All city-level controls that are not interacted with an extra treatment option T_{ij} cancel each other in (4). For a conservative estimation without simple demographic controls, the only term that was interacted and that differs between treatment and control group is $\mathrm{NationalistSupport}_j \times T_{ij}$. Thus, the city level specification reduces to

$$\frac{\sum_{T_{ij}=1} y_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} y_{ij}}{\sum_{T_{ij}=0} 1} = \beta_1 + \beta_2 \mathrm{VK}_j + \beta_{4,ns} \mathrm{NationalistSupport}_j + \eta_j. \tag{5}$$

We present the results of this estimation in Table 7.[33] As one can see, the results are largely consistent with the results at the individual level (Table 6), though the coefficients in Table 7 are slightly larger in terms of magnitudes, potentially due to some spillover effects underestimated by the specification used at the individual level. The relationship from this table, in the form of bin scatter plot, is also illustrated in Figure B6.

[33]Note that the number of observations is different for different specifications, as not all cities have a large enough number of respondents to compute the difference between the mean number of options chosen by the members of treatment and control in each subgroup. For robustness, we also report the results for the sample of cities with non missing estimates for every category in Table B9.

Overall, the results in Tables 6 and 7 indicate that social media penetration had a positive effect on the share of people who have xenophobic attitudes, and more so among the groups of respondents likely to be involved in hate crimes (and, in the case of younger and low-educated individuals, groups that are arguably likely to be persuadable). These findings speak in favor of the persuasive effect of social media on xenophobic attitudes.

3.4 Self-reported hostility

We also examine the effect of social media on direct, self-reported xenophobic attitudes. This is estimated at the individual level using the following specification:

$$\text{SelfReportedHate}_{ij} = \beta_0 + \beta_1 \text{VK}_j + \beta_2 \text{X}_{ij} + \varepsilon_{ij}. \tag{6}$$

The results of this estimation are reported in Table 8, Panel A. The coefficient of interest, β_1, is generally not statistically significant and has a negative sign. For one particular specification in which we look at the subset of younger respondents (column 6), the 95% weak-instrument-robust confidence set lies entirely below zero.

To make sure that the lack of an effect on self-reported xenophobic attitudes is not a consequence of the timing of the survey (almost twelve years after VK was founded) or the number of respondents, we replicate this analysis using data from a much larger survey, conducted in February 2011.[34] This survey contained a direct question on dislike toward other ethnicities with exactly the same wording as the question we used to measure self-reported hostility in our survey.

The results of estimating Equation (6) based on this sample are presented in Table 8, Panel B. These results indicate that, as in the case of the 2018 survey, there is no significant relation between social media penetration and self-reported xenophobic attitudes. This null result holds

[34]This MegaFOM opinion poll, conducted by FOM, has a regionally representative sample of 54,388 respondents in 79 regions, of which 29,780 respondents come from the 519 cities in our sample.

regardless of the initial level of nationalism in a city. Weak instrument robust confidence sets are, again, too large to claim that these are indeed zero results, though for the direct effect we can rule out a more than 25% increase in reported xenophobic attitudes following a 10% increase in social media penetration.

Overall, our survey analysis implies that in cities with higher social media penetration respondents are more likely to have xenophobic attitudes, but at the same time are not more likely to express them openly to a stranger, such as a surveyor.

By contrasting the effects on hostility inferred from list experiments and on self-reported hostility, we see no evidence that social media reduces stigma associated with expression of hateful opinions to surveyors.[35]

4 Conclusion

We study the longer-term, causal effects of exposure to social media on ethnic hate crimes and xenophobic attitudes in Russia, using exogenous variation in city-level initial penetration of social media. We find that higher penetration of social media increases ethnic hate crime. This effect is stronger in cities with a higher baseline level of nationalist sentiment as well as for crimes with multiple perpetrators. These findings suggest a role of social media as mechanism of coordination of hate crime. We also use a national survey to show that social media penetration also had a persuasive effect and increased xenophobic attitudes, especially for young and male individuals, as well as those with lower levels of education.

These findings contribute to a growing body of evidence indicating that social media is a complex phenomenon that has both positive and negative effects on the welfare of people (see also

[35] In Table B10 we report the effect of social media on self-reported intolerance when the elicited level of hostility is controlled for. Unfortunately, here we hit the limits of our identification approach, with weak instrument robust confidence sets being very imprecise and most of them including the entire grid. The results for the city-level estimation of Equation (6) are qualitatively similar and are presented in Table B11.

Allcott et al., 2020). These effects need to be taken into account when discussing policy implications of the recent changes in media technologies, as well as possible government regulation or self-regulation by social media platforms.

Our paper also hints at promising directions for future research. One direction is finding more direct evidence on the effect of social media on polarization, which might help understand whether and when social media may contribute to moderation. More generally, it would be interesting to understand the factors that determine opinion formation in social networks. Finally, one could analyze direct evidence on how social media facilitates coordination in practice, by analyzing text content in social media forums and understanding how online discussions lead to offline interactions.

References

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun, "The Power of the Street: Evidence from Egypt's Arab Spring," *Review of Financial Studies*, 2018, *31* (1), 1–42.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of the Nazis in Prewar Germany," *Quarterly Journal of Economics*, Nov 2015, *130* (4), 1885–1939.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari, "The European Trust Crisis and the Rise of Populism," *Brookings Papers on Economic Activity*, 2017, *Fall*, 309–382.

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–36.

_ , Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, March 2020, *110* (3), 629–76.

Andrews, Donald W. K., "Identification-Robust Subvector Inference," Working Paper 2017.

Andrews, Isaiah, James Stock, and Liyang Sun, "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*, 2019 2019, *11*, 727–753.

Blair, Graeme and Kosuke Imai, "Statistical Analysis of List Experiments," *Political Analysis*, 2012, *20* (1), 47–77.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler, "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, September 2012, *489* (7415), 295–298.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro, "Greater Internet Use is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups," *Proceedings of the National Academy of Sciences*, 2017, *114* (40), 10612–10617.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin, "From Extreme to Mainstream: The Erosion of Social Norms," working paper 2019.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio, "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation," *Journal of the European Economic Association*, aug 2018, *16* (4), 1094–1136.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *Quarterly Journal of Economics*, may 2019, *134* (2), 1021–1077.

Chaudhuri, Saraswata and Eric Zivot, "A New Method of Projection-based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 2011, *164* (2), 239 – 251.

Chiang, Chun-Fang and Brian Knight, "Media Bias and Influence: Evidence from Newspaper Endorsements," *The Review of Economic Studies*, 02 2011, *78* (3), 795–820.

Coutts, Elisabeth and Ben Jann, "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)," *Sociological Methods & Research*, 2011, *40* (1), 169–193.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova, "Social Media and Protest Participation: Evidence from Russia," *Econometrica*, forthcoming.

_ , _ , _ , and Leonid Polishchuk, "Social Image, Networks, and Protest Participation," Technical Report 2017.

_ , Maria Petrova, and Konstantin Sonin, "Social Media and Corruption," *American Economic Journal: Applied Economics*, 2018, *10* (1), 150–74.

Enke, Benjamin, "Moral Values and Voting," working paper 2019.

Fergusson, Leopoldo and Carlos Molina, "Facebook Causes Protests," *Working paper*, 2020.

Gentzkow, Matthew and Jesse M. Shapiro, "Ideological Segregation Online and Offline," *Quarterly Journal of Economics*, 11 2011, *126* (4), 1799–1839.

Glynn, Adam N., "What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment," *Public Opinion Quarterly*, 01 2013, *77* (S1), 159–172.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya, "3G Internet and Confidence in Government," Technical Report, mimeo, Paris School of Economics 2019.

Halberstam, Yosh and Brian Knight, "Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter," *Journal of Public Economics*, 2016, *143*, 73 – 88.

Hatte, Etienne Madinier Sophie and Ekaterina Zhuravskaya, "Reading Twitter in the Newsroom: How Social Media Affects Traditional-Media Reporting?," *working paper*, 2020.

Imai, Kosuke, "Multivariate Regression Analysis for the Item Count Technique," *Journal of the American Statistical Association*, 2011, *106* (494), 407–416.

Manacorda, Marco and Andrea Tesei, "Liberation Technology: Mobile Phones and Political Mobilization in Africa," *Econometrica*, 2020, *88* (2), 533–567.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie, "The economic effects of Facebook," *Experimental Economics*, 2020, *23* (2), 575–602.

Müller, Karsten and Carlo Rasmus Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime," *Available at SSRN: https://ssrn.com/abstract=3082972 or http://dx.doi.org/10.2139/ssrn.3082972*, 2018.

_ and _ , "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment," *Available at SSRN: https://ssrn.com/abstract=3149103*, 2019.

Qin, Bei, David Strömberg, and Yanhui Wu, "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda," *Journal of Economic Perspectives*, February 2017, *31* (1), 117–40.

_ , _ , and _ , "Social Media, Information Networks, and Protests in China," working paper 2019.

Raghavarao, Damaraju and Walter T. Federer, "Block Total Response as an Alternative to the Randomized Response Method in Surveys," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1979, *41* (1), 40–45.

Sen, Ananya and Pinar Yildirim, "Clicks Bias in Editorial Decisions: How Does Popularity Shape Online News Coverage?," *working paper*, 2016.

Settle, Jaime E., *Frenemies: How Social Media Polarizes America*, Cambridge University Press, Aug 2018.

Steinert-Threlkeld, Zachary C., Delia Mocanu, Alessandro Vespignani, and James Fowler, "Online Social Networks and Offline Protest," *EPJ Data Science*, Nov 2015, *4* (1), 19.

Sun, Liyang, "Implementing Valid Two-Step Identification-Robust Confidence Sets for Linear Instrumental-Variables Models," *The Stata Journal*, 2018, *18* (4), 803–825.

Sunstein, Cass R., *Republic.Com*, Princeton, NJ, USA: Princeton University Press, 2001.

_ , *#Republic: Divided Democracy in the Age of Social Media*, Princeton, NJ, USA: Princeton University Press, 2017.

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 11 2014, *129* (4), 1947–1994.

_ , Maria Petrova, and Ruben Enikolopov, "Echo Chambers: Does Online Network Structure Affect Political Polarization?," mimeo 2019.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political Effects of the Internet and Social Media," *Annual Review of Economics*, 2020, *12* (1), null.

Figures and Tables

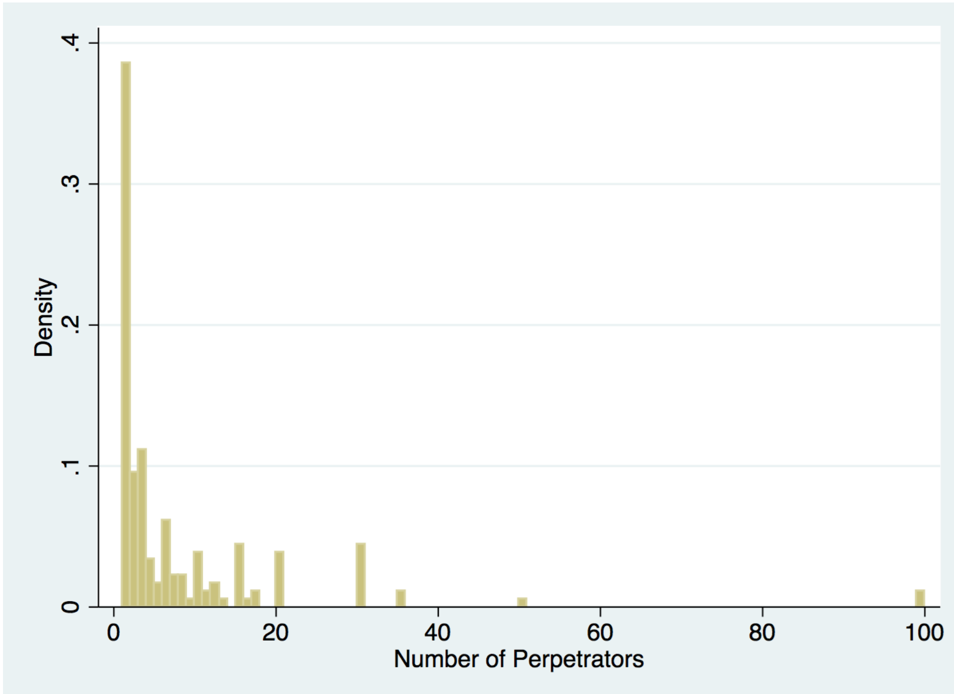
Figure 1: Total Number of Hate Crimes by City Across Russia (2007-2015)

Notes: The bubble map shows the total number of hate crimes by city across Russia from 2007 to 2015. Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis.

Figure 2: Total Number of Ethnic Hate Crimes by City Across Russia (2007-2015)

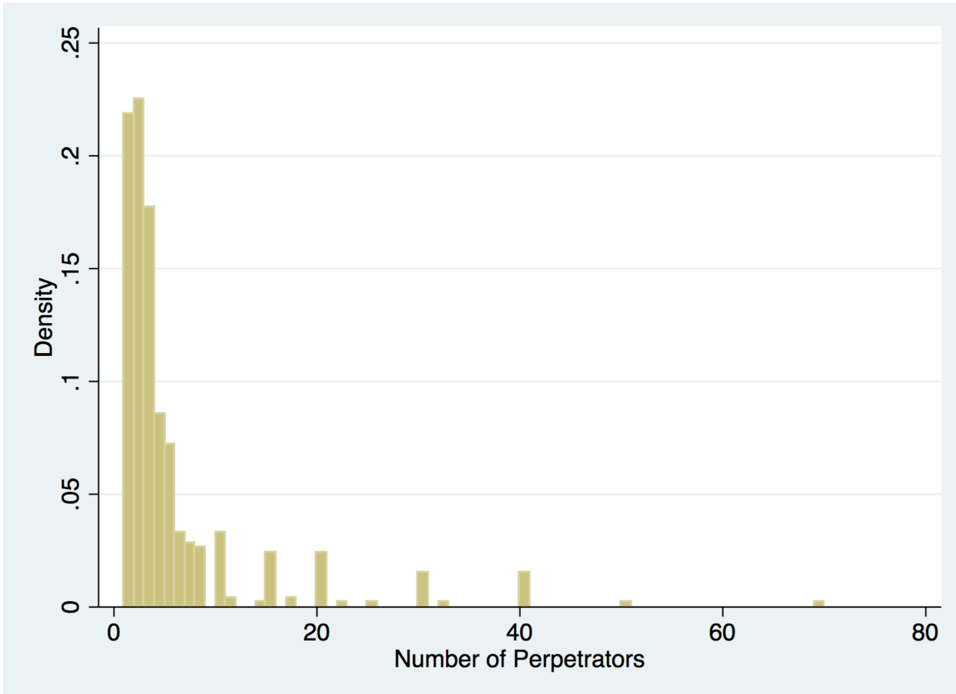
Notes: The bubble map shows the total number of ethnic hate crimes by city across Russia from 2007 to 2015. Data on ethnic hate crimes comes from the database compiled by SOVA Center for Information and Analysis.

Figure 3: Histogram for the Number of Perpetrators of Non-Ethnic Hate Crime



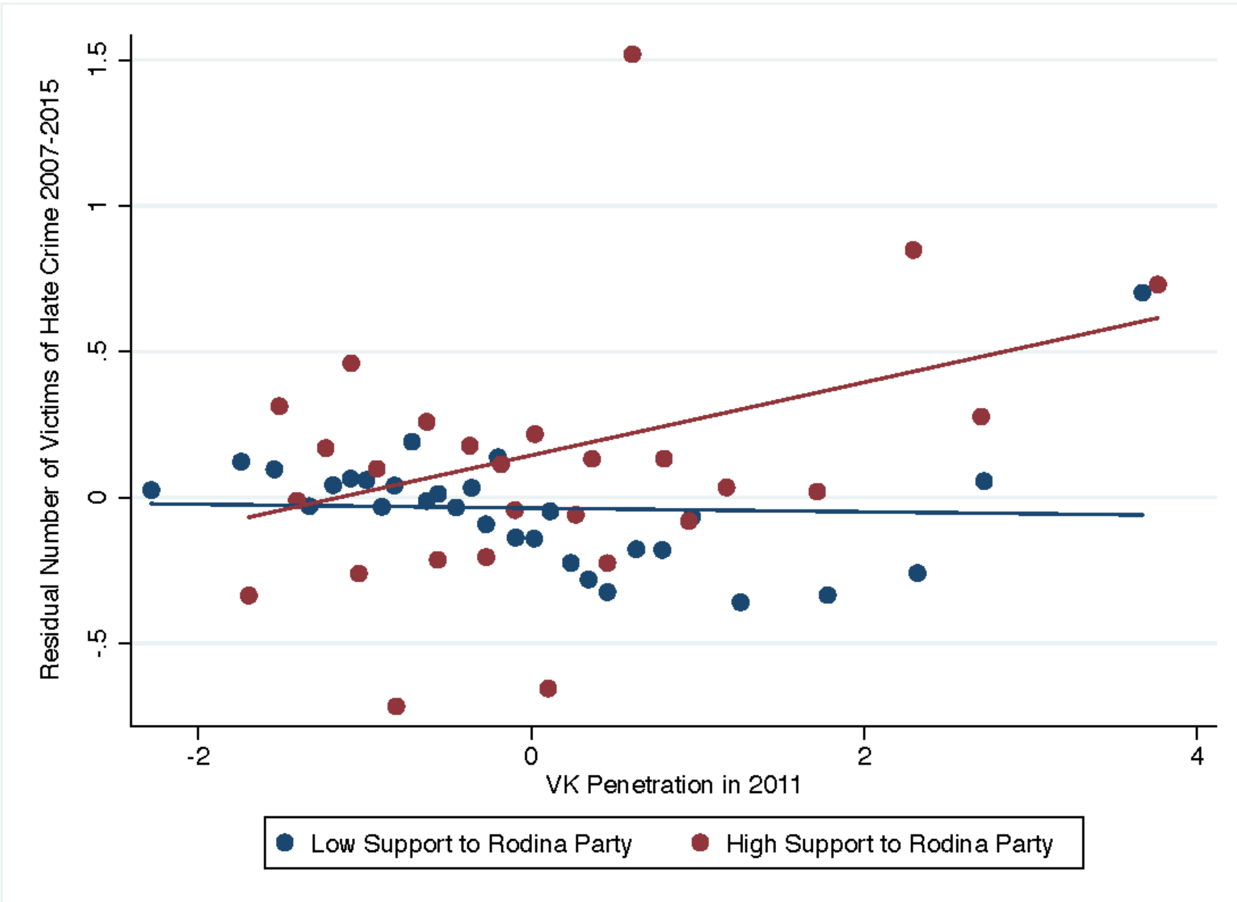
Notes: Histogram for the number of perpetrators of non-ethnic hate crimes. Data on non-ethnic hate crimes comes from the database compiled by SOVA Center for Information and Analysis, 2007-2015.

Figure 4: Histogram for the Number of Perpetrators of Ethnic Hate Crime



Notes: Histogram for the number of perpetrators of ethnic hate crimes. Data on ethnic hate crimes comes from the database compiled by SOVA Center for Information and Analysis, 2007-2015.

Figure 5: Number of Hate Crime Victims and VK Penetration, by Level of Support for Nationalist Party



Notes: Binned scatter plot for the relationship between the number of victims of hate crimes and VK penetration separately by the level of support of the nationalist party. Both the number of victims of hate crimes and VK penetration are measured as the residuals of the respective measures in the linear regression on the control variables from Equation (2). The threshold for the high level of support of the nationalist party is set at the 85th percentile of the number of votes for the party. The number of bins is 30.

Table 1: VK Penetration, SPbSU Student Cohorts, and Nationalistic Party Support

	Log(Number of VK users, 2011)	Nationalistic party support in 2003	
	(1)	(2)	(3)
Log (SPbSU students), same 5-year cohort as VK founder	0.142***	-0.002	
	[0.039]	[0.001]	
Log (SPbSU students), one cohort younger than VK founder	-0.024	-0.002	-0.002
	[0.042]	[0.001]	[0.002]
Log (SPbSU students), one cohort older than VK founder	0.051	-0.002*	-0.002
	[0.044]	[0.001]	[0.002]
Nationalistic party support in 2003	4.602***		
	[1.178]		
Log(Number of VK users, 2011)			-0.016
			[0.011]
Socioeconomic city-level controls	Yes	Yes	
Observations	625	625	

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Number of Victims by Type

Victims	Freq.	Percent
Ethnic		
Central Asia	325	23.81%
Caucasus	265	19.41%
Blacks	74	5.42%
Russians	63	4.62%
Arabs	33	2.42%
Jews	10	0.73%
Other "non-slavic"	209	15.31%
Other Asians	108	7.91%
Other Ethnicity	85	6.23%
Total Ethnic	770	56.41%
Non-Ethnic		
Youth groups and left-wing groups	402	29.45%
Religious Groups	106	7.77%
Homeless	42	3.08%
LGBT	32	2.34%
Unknown	13	0.95%
Total Non-Ethnic	595	43.59%
Total	1 365	100%

Notes: Number of hate crime victims by ethnic and non-ethnic characteristics. Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis, 2004-2015.

Table 3: Social Media and Hate Crime. Period: 2007-2015

Panel A. OLS.	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011	0.103*	0.072**	0.057	0.072	0.071***	0.023	0.04	0.019	0.037
	[0.061]	[0.032]	[0.059]	[0.053]	[0.026]	[0.050]	[0.047]	[0.019]	[0.045]
Nationalist Party Support in 2003	1.378	0.607	0.942	1.558	0.424	1.36	0.192	0.006	-0.123
	[1.509]	[0.744]	[1.464]	[1.380]	[0.604]	[1.311]	[1.125]	[0.526]	[1.131]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
R-squared	0.589	0.425	0.548	0.555	0.339	0.514	0.45	0.24	0.389

Panel B. IV.	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011	-0.130	0.238	-0.229	-0.211	0.348*	-0.417	0.350	0.007	0.502
Weak Instrument Robust Confidence 95% Sets	(-1.169; .660)	(-.262; .787)	(-1.265; .558)	(-1.167; .516)	(-.036; .853)	(-1.460; .316)	(-.423; 1.196)	(-.403; .381)	(-.177; 1.393)
	[0.420]	[0.241]	[0.418]	[0.386]	[0.204]	[0.390]	[0.372]	[0.180]	[0.360]
Nationalist Party Support in 2003	2.407	-0.127	2.208	2.810	-0.801	3.307	-1.178	0.057	-2.176
	[2.492]	[1.250]	[2.477]	[2.301]	[0.956]	[2.341]	[2.069]	[0.945]	[2.042]
Log (SPbSU students, one cohort younger)	-0.084	-0.060	-0.060	-0.155**	-0.058*	-0.122*	0.075	-0.015	0.087
	[0.068]	[0.039]	[0.070]	[0.063]	[0.033]	[0.065]	[0.059]	[0.031]	[0.063]
Log (SPbSU students, one cohort older)	0.101	0.065	0.089	0.099	0.008	0.113*	0.014	0.057*	-0.035
	[0.077]	[0.040]	[0.076]	[0.067]	[0.035]	[0.068]	[0.066]	[0.031]	[0.067]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	13.150	13.150	13.150	13.150	13.150	13.150	13.150	13.150	13.150
Effective F-statistics (Montiel Olea and Pflueger 2013)	13.571	13.571	13.571	13.571	13.571	13.571	13.571	13.571	13.571
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Social Media and Hate Crime. Specification with Interaction. Period: 2007-2015

Panel A. OLS.	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	4.488***	1.768**	4.006**	4.823***	1.903***	4.599***	1.276	0.088	0.914
	[1.467]	[0.781]	[1.624]	[1.405]	[0.690]	[1.433]	[1.306]	[0.693]	[1.420]
Log (number of VK users), 2011	0.150**	0.100***	0.098	0.114**	0.091***	0.064	0.070	0.029	0.060
	[0.062]	[0.032]	[0.061]	[0.056]	[0.026]	[0.053]	[0.049]	[0.021]	[0.048]
Nationalist Party Support in 2003	1.507	0.459	1.063	1.974	0.473	1.695	-0.268	-0.206	-0.504
	[1.623]	[0.745]	[1.591]	[1.521]	[0.611]	[1.452]	[1.184]	[0.539]	[1.191]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
R-squared	0.601	0.427	0.559	0.569	0.356	0.530	0.445	0.232	0.383
Full Effect at min level of Nationalist Party Support	-0.065	0.015	-0.094	-0.118*	0.000	-0.156***	0.009	0.025	0.016
p-value for the effect at minimum	.339	0.699	.186	.056	.994	.01	.89	.389	0.816
Full Effect at max level of Nationalist Party Support	1.097***	0.473***	0.943**	1.131***	0.493***	1.034***	0.340	0.048	0.253
p-value for the effect at maximum	.001	.009	.013	.001	.002	.002	.247	.762	.428

Panel B. IV.	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	12.002***	6.349***	11.605***	10.578***	5.056**	10.282***	10.365***	1.823	9.125**
Weak Instrument Robust Confidence 95% Sets	(4.537; 23.199)	(1.588; 13.491)	(4.120; 22.833)	(3.701; 20.895)	(1.114; 10.971)	(3.304; 20.749)	(3.004; 21.407)	(-1.623; 6.991)	(1.983; 19.839)
	[4.570]	[2.915]	[4.583]	[4.211]	[2.414]	[4.272]	[4.507]	[2.110]	[4.373]
Log (number of VK users), 2011	0.053	0.362	-0.055	-0.046	0.446**	-0.276	0.529	0.051	0.667*
Weak Instrument Robust Confidence 95% Sets	(-.976; .740)	(-.105; 1.062)	(-1.081; .629)	(-.984; .578)	(.050; 1.041)	(-1.215; .351)	(-.201; 1.624)	(-.410; .359)	(-.036; 1.720)
	[0.420]	[0.286]	[0.419]	[0.383]	[0.243]	[0.383]	[0.447]	[0.188]	[0.430]
Nationalist Party Support in 2003	5.384	1.168	5.534*	4.978*	0.180	5.633*	2.214	0.509	1.137
	[3.298]	[1.527]	[3.260]	[2.930]	[1.281]	[3.006]	[2.557]	[1.096]	[2.504]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger + their interaction with Nationalistic Party Support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Full Effect at min level of Nationalist Party Support	-0.522	0.057	-0.611	-0.554	0.204	-0.769*	0.032	-0.036	0.229
p-value for the effect at minimum	.255	0.831	.176	.173	.35	.062	0.939	.862	0.573
Full Effect at max level of Nationalist Party Support	2.584**	1.701**	2.392**	2.184**	1.512**	1.893*	2.715**	0.436	2.591**
p-value for the effect at maximum	0.017	0.027	0.028	0.032	0.021	0.064	0.023	0.380	0.024

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Social Media and Hate Crime. Specification with Interaction. Period: 2004-2006

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethnic hate crime)
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	-1.898	-0.165	-1.732	-1.321	-0.165	-1.156	-0.576
Weak Instrument Robust Confidence 95% Sets	(-5.311; .378)	(-.749; .127)	(-5.065; .489)	(-4.583; .853)	(-.749; .127)	(-4.326; .957)	(-2.094; .436)
	[1.393]	[0.179]	[1.360]	[1.331]	[0.179]	[1.294]	[0.619]
Log (number of VK users), 2011	0.018	0.014	0.005	0.145	0.014	0.132	-0.127
Weak Instrument Robust Confidence 95% Sets	(-.258; .433)	(-.021; .066)	[-.271; .280)	(-.075; .476)	(-.021; .066)	(-.086; .459)	(-.376; .039)
	[0.169]	[0.021]	[0.168]	[0.135]	[0.021]	[0.133]	[0.101]
Nationalist Party Support in 2003	-0.954	-0.217	-0.737	-1.307*	-0.217	-1.091	0.353
	[0.772]	[0.213]	[0.717]	[0.769]	[0.213]	[0.695]	[0.370]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Full Effect at minimal level of Nationalist Party Support	0.109	0.022	0.088	0.209	0.022	0.187	-0.100
p-value for the effect at minimum	.471	.423	.554	.116	.423	.14	.234
Full Effect at maximum of Nationalist Party Support	-0.382	-0.021	-0.361	-0.133	-0.021	-0.112	-0.249
p-value for the effect at maximum	0.340	.504	.364	0.698	.504	.743	.245

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. There were no non-ethnic hate crimes, committed by a single perpetrator, during this period, thus we are not able to compute the results for non-ethnic hate crime separately by the number of perpetrators. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Social Media and Ethnic Hostility, Elicited from List Experiment

	Number of options in List Experiment						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dislike Other Ethnicities Option in List Experiment (LE) x Log (Number of VK users, 2011)	0.075**	0.109*	0.043	0.164***	-0.008	0.105***	0.050
Weak Instrument Robust Confidence 95% Sets	(.009; .208)	(-.005; .222)	(-.039; .247)	(.071; .257)	(-.091; .075)	(.026; .224)	(-.055; .207)
	[0.041]	[0.069]	[0.050]	[0.057]	[0.051]	[0.049]	[0.064]
Log (Number of VK users, 2011)	-0.053	-0.001	-0.080	0.017	-0.085	0.066	-0.067
	[0.167]	[0.277]	[0.189]	[0.228]	[0.220]	[0.191]	[0.253]
Dislike Other Ethnicities Option in LE	0.203**	0.110	0.293**	-0.019	0.422***	0.087	0.310**
	[0.101]	[0.173]	[0.123]	[0.131]	[0.130]	[0.119]	[0.157]
Nationalistic Party Support, 2003	-0.832	-1.227	-0.363	-1.390	-0.045	0.120	-1.477
	[1.037]	[1.399]	[1.492]	[1.716]	[1.310]	[1.299]	[1.555]
Dislike Other Ethnicities Option in LE	1.040	0.680	1.032	0.526	0.762	0.061	2.087
x Vote share of nationalistic party, 2003	[1.195]	[2.177]	[1.431]	[1.748]	[1.355]	[1.501]	[1.989]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,447	2,118	2,329	1,954	2,493	2,164	2,283
Kleibergen-Paap F-statistics	4.541	4.366	4.507	4.469	4.445	4.559	4.012
Number of cities	125	125	124	125	124	125	123

Notes: Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Social Media and Ethnic Hostility, Inferred from List Experiment. City Level

Subsample:	List Experiment elicited hostility						
	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	0.123***	0.158**	0.050	0.204***	-0.008	0.210***	0.099
Weak Instrument Robust Confidence 95% Sets	(.045, .208)	(.026, .290)	(-.042, .151)	(.099, .334)	(-.136, .107)	(.080, .353)	(-.022, .219)
	[0.041]	[0.070]	[0.049]	[0.062]	[0.062]	[0.069]	[0.064]
Nationalistic Party Support, 2003	1.486	1.058	2.725	1.695	-0.362	1.444	2.912
	[1.522]	[2.700]	[1.896]	[1.953]	[2.178]	[2.500]	[2.188]
Observations	124	116	122	124	111	121	116
Kleibergen-Paap F-statistics	78.994	74.394	81.499	78.994	56.186	73.944	67.506
Effective F-statistics (Montiel Olea and Pflueger 2013)	105.021	98.222	103.035	105.021	75.060	97.275	92.187
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Social Media and Self-Reported Ethnic Hostility

	Self-reported hostility to other ethnicities						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. VK and self-reported hate. 2018 Own Survey.							
Log (Number of VK users, 2011)	-0.114	-0.214	-0.106	-0.158	-0.114	-0.333*	0.139
Weak Instrument Robust Confidence 95% Sets	(-.526; .237)	(-.731; .226)	(-.662; .296)	(-.737; .221)	(-.638; .409)	(-.990;-.014)	(-.366; 1.028)
	[0.154]	[0.193]	[0.194]	[0.202]	[0.212]	[0.190]	[0.243]
Nationalistic Party Support, 2003	0.191	1.054	0.018	0.252	0.170	1.375	-1.265
	[0.837]	[1.279]	[1.141]	[1.431]	[1.116]	[1.179]	[1.265]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,927	927	1,000	853	1,074	943	984
Kleibergen-Paap F-statistics	8.875	8.982	8.600	9.709	8.205	9.858	7.105
Montiel Olea-Pflueger Effective F-stat	8.942	9.046	8.666	9.776	8.270	9.929	7.159
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Number of cities	125	122	117	125	114	122	118
Panel B. VK and self-reported hate. 2011 FOM survey.							
Log (Number of VK users, 2011)	0.003	-0.065	0.053	0.015	-0.069	-0.091	0.094
Weak Instrument Robust Confidence 95% Sets	(-.248; .254)	(-.442; .199)	(-.150; .373)	(-.242; .313)	(-.513; .197)	(-.448; .173)	(-.122; .453)
	[0.094]	[0.116]	[0.098]	[0.104]	[0.128]	[0.116]	[0.104]
Nationalistic Party Support, 2003	0.487	0.610	0.421	0.455	0.683	0.894	0.134
	[0.522]	[0.615]	[0.559]	[0.568]	[0.596]	[0.653]	[0.537]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,696	12,285	15,411	20,766	6,930	13,571	14,125
Kleibergen-Paap F-statistics	7.897	7.849	7.876	7.798	7.566	8.136	7.368
Montiel Olea-Pflueger Effective F-stat	6.326	6.063	6.500	6.197	6.200	6.179	6.245
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Number of cities	512	512	511	512	503	512	511

Notes: Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts. Individual-level controls include gender, education categories, dummy for pilot (for Panel A) and age categories. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Supplementary Appendix

(Not For Publication)

A Student Cohort Data Collection and Odnoklassniki Data

One possible concern with the approach that we are using is that we do not have administrative records for SPbSU students. Instead, we use information from the profiles of Odnoklassniki users to infer the number of students in each university at each point in time. This concern is partially alleviated by the fact that 80% of Internet users in Russia had profiles in Odnoklassniki in 2014 (in times of data collection). This number was likely to be higher for recent students, e.g. people born after 1978 (older cohort), which further strengthens representativeness of our data. In order to correct for a possible measurement error bias due to the non-random variation in Odnoklassniki penetration, we control for the number of Odnoklassniki users in each city in all of our specifications. We should also note that the Odnoklassniki platform had no specific relationship to SPbSU, to Saint Petersburg, or Durov's age cohort. The founder of Odnoklassniki, Albert Popkov, was born in Yuzno-Sakhalinsk on Sakhalin island, studied in Moscow at a technical college in the early 90's, and founded the network while living in London.

Nevertheless, it is possible people could be more likely to have an Odnoklassniki account in cities with higher VK penetration, and, possibly, this effect could be stronger in places with a greater number of SPbSU students in Durov's cohort. To address this concern, we conduct two additional tests.[36] First, we check whether the number of Odnoklassniki users is correlated with the number of VK users in a city at different stages of VK diffusion. The results in Columns (1)-(3) of Table A1 show that early VK penetration (the number of users in a city among the first 5,000, 50,000, or 100,000 users of the network) is not positively related to the subsequent

[36] This part of the analysis repeats Table A12 from Enikolopov et al. (forthcoming).

penetration of Odnoklassniki (if anything, the corresponding coefficients are significantly negative for column (2) and (3)). This result is consistent with assumption that the initial diffusion of VK was not driven by general preferences for social media and that there might have been a substitution effect between different social networks. At the same time, VK penetration in 2011 is positively related to Odnoklassniki penetration in 2014, although this effect is also not statistically significant (column 4). Second, we test whether Odnoklassniki penetration was related to the student flows from Russian cities to Saint Petersburg State University. The results in column (5) indicate that there is no such association, with the standard errors being substantially larger than the coefficients for the VK founder's cohort. We conclude that the potential selection, introduced by our data collection process, is unlikely to bias our results.

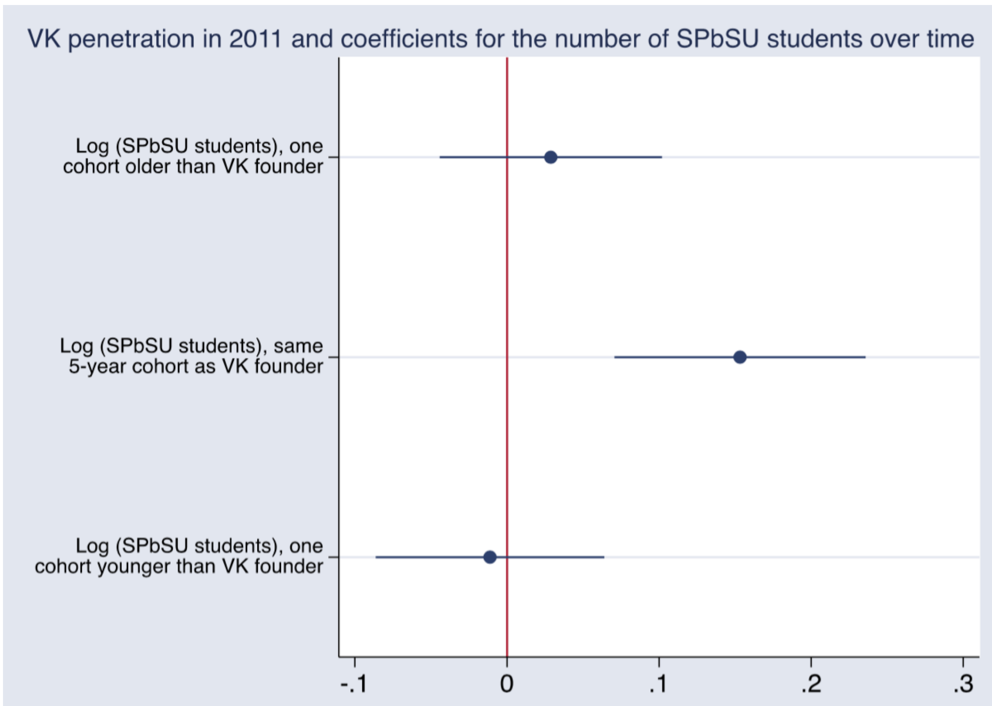
Table A1: VK Penetration and Odnoklassniki.

	Log (Number of Odnoklassniki users), 2014				
	(1)	(2)	(3)	(4)	(5)
Log (early VK users, from first 5,000 users)	-0.005				
	[0.042]				
Log (early VK users, from first 50,000 users)		-0.090**			
		[0.037]			
Log (early VK users, from first 100,000 users)			-0.084***		
			[0.031]		
Log (number of VK users), 2011				-0.030	
				[0.053]	
Log (SPbSU students), same 5-year cohort as VK founder					-0.034
					[0.046]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes
Other SPbSU student cohorts					Yes
Observations	625	625	625	625	625

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1.

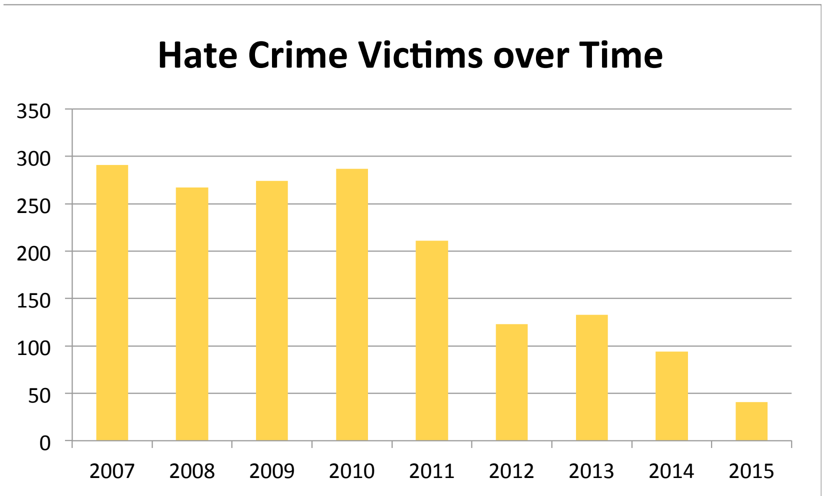
B Appendix Figures and Tables

Figure B1: Coefficients for student cohorts and VK penetration



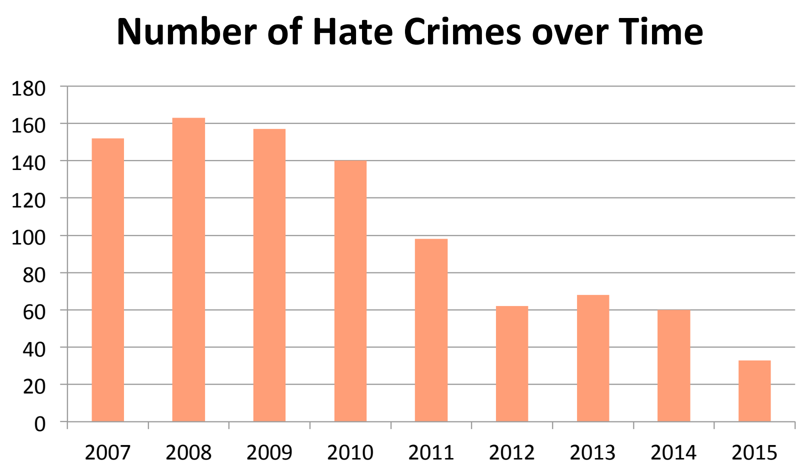
Notes: Coefficients for the first stage regression from Table 1. Reported coefficients come from the specification with all baseline controls included, namely logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, ethnic fractionalization according to 2010 Russian Census.

Figure B2: Hate Crime Victims Over Time



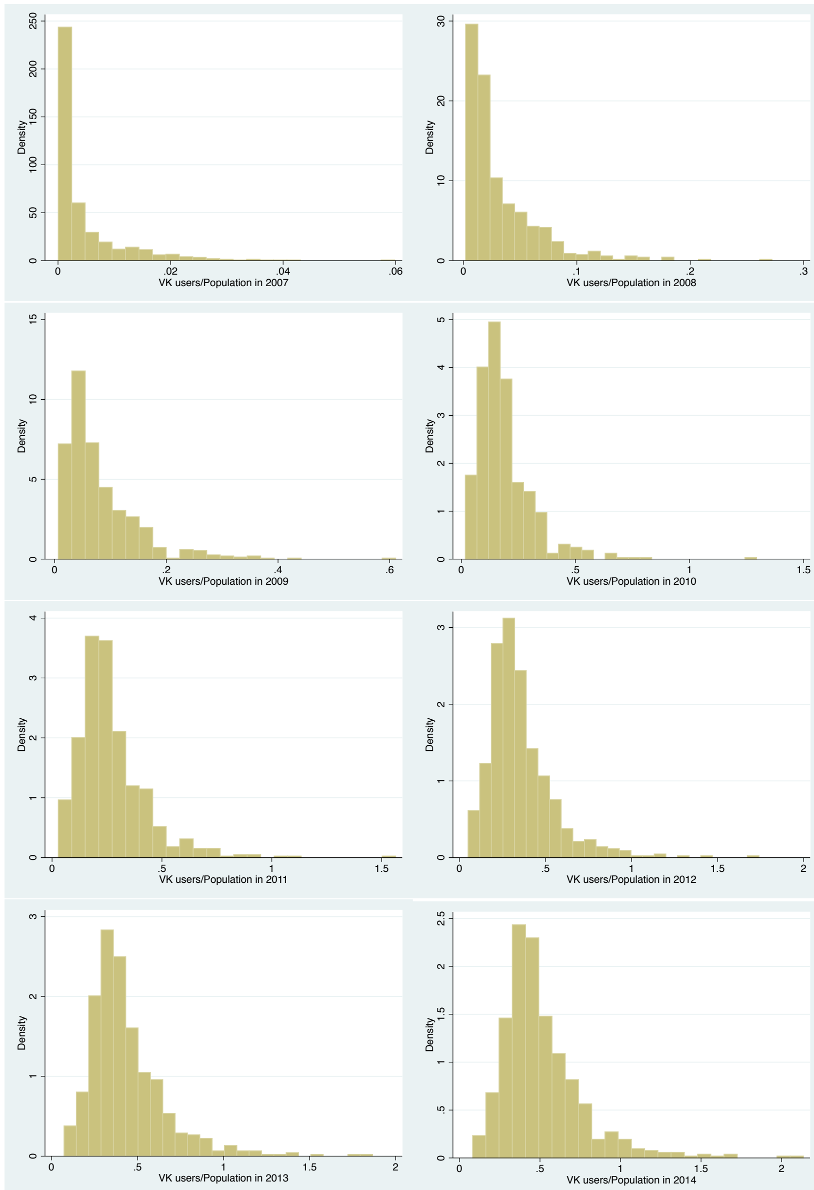
Notes: Number of recorded hate crimes from 2007 to 2015. Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis.

Figure B3: Number of Hate Crimes Over Time



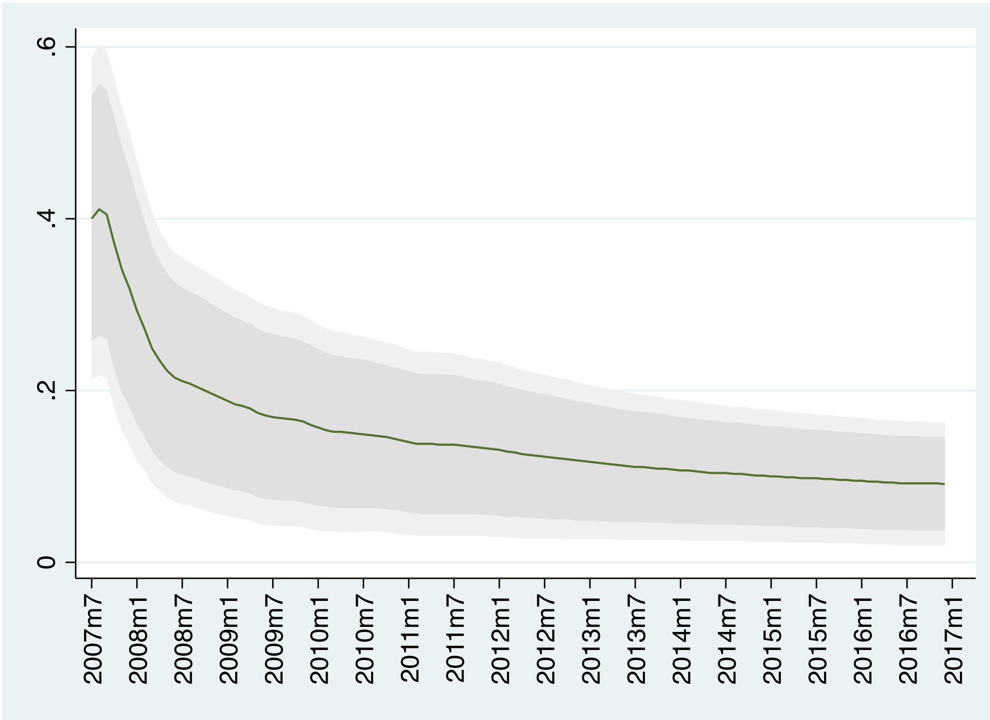
Notes: Number of recorded hate crime victims from 2007 to 2015. Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis.

Figure B4: VK Penetration over Time for 2007-2014



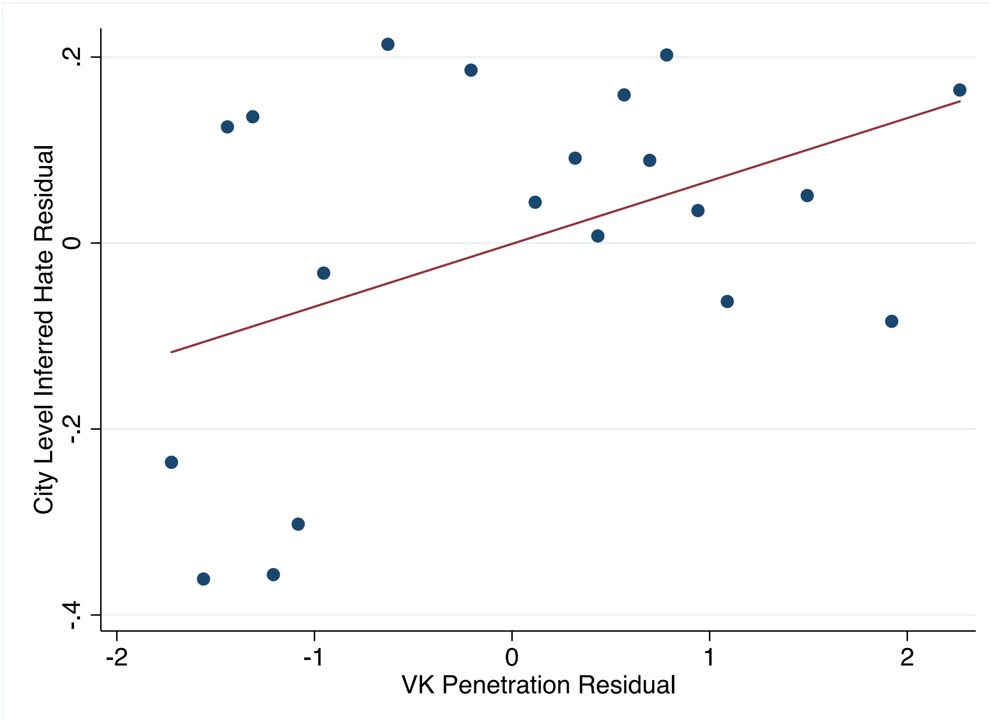
Notes: Histograms of VK user population from 2007 to 2014. Data on VK penetration comes from Enikolopov et al. (forthcoming)

Figure B5: Changes in the coefficient for Durov's cohorts and VK penetration



Notes: Coefficients for the first stage regression from Table 1 for penetration measured at different points in time. Reported coefficients come from the specification with all baseline controls included, namely logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, ethnic fractionalization according to 2010 Russian Census.

Figure B6: VK penetration and Inferred Hate



Notes: This figure presents binned scatter plot for the OLS version of the relationship between VK penetration and city-level inferred hate, as reported in column (1) of Table 7. City-level inferred hate is computed as the difference in means in the number of options chosen by the respondents in treatment and control groups.

Table B1: Nationalistic Party Support and Measures of Xenophobia

	Log (# of hate crimes)	Log (# of ethnic hate crimes)	Log (# of non-ethninc hate crimes)	Self-reported hostility to other ethnicities
	(1)	(2)	(3)	(4)
Nationalist Party Support in 2003	6.132***	6.126***	0.727	0.507***
	[1.672]	[1.532]	[0.927]	[0.187]
Ln(Population)	0.686***	0.552***	0.428***	-0.005
	[0.042]	[0.040]	[0.041]	[0.006]
Observations	625	625	625	27,696
R-squared	0.493	0.439	0.382	0.003

Notes: Unit of observation is a city in columns (1)-(3) and an individual respondent in column (4). Robust standard errors in brackets in columns (1)-(3) and standard errors clustered at the city level in column (4). Self-reported hostility from FOM survey. *** p<0.01, ** p<0.05, * p<0.1.

Table B2: Social Media and Hate Crime. Specification with Interactions. Period: 2007-2015

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	12.002***	6.349***	11.605***	10.578***	5.056**	10.282***	10.365***	1.823	9.125**
Weak Instrument Robust Confidence 95% Sets	(4.537; 23.199)	(1.588; 13.491)	(4.120; 22.833)	(3.701; 20.895)	(1.114; 10.971)	(3.304; 20.749)	(3.004; 21.407)	(-1.623; 6.991)	(1.983; 19.839)
	[4.570]	[2.915]	[4.583]	[4.211]	[2.414]	[4.272]	[4.507]	[2.110]	[4.373]
Log (number of VK users), 2011	0.053	0.362	-0.055	-0.046	0.446**	-0.276	0.529	0.051	0.667*
Weak Instrument Robust Confidence 95% Sets	(-.976; .740)	(-.105; 1.062)	(-1.081; .629)	(-.984; .578)	(.050; 1.041)	(-1.215; .351)	(-.201; 1.624)	(-.410; .359)	(-.036; 1.720)
	[0.420]	[0.286]	[0.419]	[0.383]	[0.243]	[0.383]	[0.447]	[0.188]	[0.430]
Nationalist Party Support in 2003	5.384	1.168	5.534*	4.978*	0.180	5.633*	2.214	0.509	1.137
	[3.298]	[1.527]	[3.260]	[2.930]	[1.281]	[3.006]	[2.557]	[1.096]	[2.504]
Log (# students one cohort younger than Durov) x Nationalist Party Support in 2003	-8.330	-7.265*	-7.336	-8.079	-5.820*	-5.874	-8.412	-2.867	-7.818
	[6.110]	[3.978]	[6.167]	[5.711]	[3.326]	[5.726]	[6.330]	[2.674]	[6.205]
Log (# students one cohort older than Durov) x Nationalist Party Support in 2003	-3.061	1.234	-5.058	-0.557	1.186	-2.609	-4.868	0.561	-5.216
	[3.535]	[2.252]	[3.525]	[3.182]	[1.851]	[3.220]	[3.424]	[1.568]	[3.428]
Log (# students one cohort younger than Durov)	-0.080	-0.060	-0.051	-0.155**	-0.058*	-0.120*	0.086	-0.014	0.099
	[0.066]	[0.039]	[0.068]	[0.060]	[0.034]	[0.063]	[0.059]	[0.031]	[0.065]
Log (# students one cohort older than Durov)	0.118	0.072*	0.105	0.113*	0.014	0.128*	0.027	0.059*	-0.024
	[0.077]	[0.043]	[0.076]	[0.067]	[0.038]	[0.067]	[0.070]	[0.032]	[0.071]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Full Effect at min level of Nationalist Party Support	-0.522	0.057	-0.611	-0.554	0.204	-0.769*	0.032	-0.036	0.229
p-value for the effect at minimum	.255	0.831	.176	.173	.35	.062	0.939	.862	0.573
Full Effect at max level of Nationalist Party Support	2.584**	1.701**	2.392**	2.184**	1.512**	1.893*	2.715**	0.436	2.591**
p-value for the effect at maximum	0.017	0.027	0.028	0.032	0.021	0.064	0.023	0.380	0.024

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B3: Social Media, Hate Crime, and Nationalistic Party Support. Reduced Form Estimates. Period: 2007-2015

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (# students in Durov's cohort)	8.340***	5.003***	7.875**	7.185**	4.265***	6.575**	8.058***	1.343	7.451***
x Nationalist Party Support in 2003	[3.223]	[1.665]	[3.231]	[2.910]	[1.333]	[2.927]	[2.497]	[1.322]	[2.435]
Log (# students one cohort older than Durov)	-2.330	1.012	-4.156	0.257	0.721	-1.397	-5.117**	0.593	-5.802**
x Nationalist Party Support in 2003	[3.306]	[1.707]	[3.313]	[2.984]	[1.367]	[3.002]	[2.561]	[1.355]	[2.498]
Log (# students one cohort younger than Durov)	-0.565	-2.360	-0.084	-1.459	-1.536	0.009	-0.552	-1.584	-0.418
x Nationalist Party Support in 2003	[3.453]	[1.784]	[3.461]	[3.118]	[1.428]	[3.136]	[2.676]	[1.416]	[2.609]
Log (# students in Durov's cohort)	-0.027	0.031	-0.041	-0.037	0.046*	-0.067	0.042	0.002	0.065
	[0.058]	[0.030]	[0.058]	[0.052]	[0.024]	[0.053]	[0.045]	[0.024]	[0.044]
Log (# students one cohort younger than Durov)	0.097	0.078**	0.079	0.090	0.027	0.093*	0.034	0.058**	-0.007
	[0.061]	[0.031]	[0.061]	[0.055]	[0.025]	[0.055]	[0.047]	[0.025]	[0.046]
Log (# students one cohort older than Durov)	-0.096	-0.078**	-0.065	-0.167***	-0.077***	-0.126**	0.059	-0.018	0.070
	[0.062]	[0.032]	[0.062]	[0.056]	[0.026]	[0.056]	[0.048]	[0.025]	[0.047]
Nationalistic Party Support in 2003	0.535	-0.020	0.404	0.318	-0.112	0.151	0.021	-0.048	0.037
	[1.567]	[0.809]	[1.570]	[1.415]	[0.648]	[1.423]	[1.214]	[0.642]	[1.184]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
R-squared	0.595	0.438	0.553	0.564	0.359	0.523	0.462	0.242	0.402

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B4: Social Media and the Number of Hate Crimes. Specification with Interactions. Period: 2007-2015

	Log (# of hate crimes)			Log (# of ethnic hate crimes)			Log (# of non-ethninc hate crimes)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	11.379***	5.516***	10.971***	10.567***	4.605***	10.188***	6.388**	1.289	5.369**
Weak Instrument Robust Confidence 95% Sets	(5.129; 20.755)	(1.445; 11.624)	(4.89; 20.09)	(4.600; 19.516)	(1.215; 9.690)	(4.315; 18.99)	(1.395; 13.88)	(-1.598; 5.620)	(.839; 12.163)
	[3.827]	[2.493]	[3.723]	[3.653]	[2.076]	[3.596]	[3.057]	[1.767]	[2.773]
Log (number of VK users), 2011	0.081	0.286	0.025	0.005*	0.308	-0.120	0.391	0.101	0.438*
Weak Instrument Robust Confidence 95% Sets	(-.781; .655)	(-.112; .884)	(-.817; .586)	(-.812; .550)	(-.027; .811)	(-.930; .420)	(-.101; 1.129)	(-.155; .358)	(-.002; 1.097)
	[0.352]	[0.244]	[0.344]	[0.334]	[0.205]	[0.331]	[0.301]	[0.157]	[0.269]
Nationalist Party Support in 2003	4.754*	1.182	4.707*	4.761*	0.550	5.038**	0.863	0.177	0.170
	[2.657]	[1.359]	[2.562]	[2.475]	[1.133]	[2.471]	[1.759]	[0.988]	[1.611]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Full Effect at minimal level of Nationalist Party	-0.465	0.022	-0.501	-0.501	0.087	-0.609*	0.084	0.040	0.181
p-value for the effect at minimum	.201	.926	.152	.133	.645	.064	.776	0.825	.497
Full Effect at maximum of Nationalist Party Support	2.480***	1.449**	2.338**	2.234**	1.279**	2.028**	1.738**	0.373	1.570**
p-value for the effect at maximum	.008	.026	.012	.015	.02	.025	.026	.354	.026

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B5: Social Media and Hate Crimes. Specification with Interactions. Specification By Period

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
2007-2009	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	8.226**	2.553*	8.071**	9.202***	3.146**	8.311***	4.538	-0.330	4.743
Weak Instrument Robust Confidence 95% Sets	(1.443; 15.008)	(-.165; 6.631)	(1.331; 14.810)	(2.957; 15.447)	(.644; 6.899)]	(2.273; 14.348)	(-1.605; 13.754)	(-2.424, 1.066)	(-1.382; 13.931)
	[4.152]	[1.664]	[4.126]	[3.824]	[1.532]	[3.696]	[3.761]	[0.855]	[3.750]
Log (number of VK users), 2011	0.133	0.171	0.202	-0.089	0.102	-0.054	0.778***	0.111*	0.740**
Weak Instrument Robust Confidence 95% Sets	(-.551; 1.158)	(-.117; .604)	(-.487; 1.234)	(-.736; .557)	(-.169; .508)	(-.690; .583)	(.201; 1.644)	(-.013; .296)	(.166; 1.600)
	[0.418]	[0.177]	[0.421]	[0.396]	[0.166]	[0.390]	[0.354]	[0.076]	[0.351]
Nationalist Party Support in 2003	0.854	-0.016	0.593	2.540	0.543	2.080	-2.280	-0.685	-0.685
	[1.976]	[0.908]	[1.990]	[1.958]	[0.888]	[1.851]	[2.012]	[0.453]	[1.958]
Full Effect at minimal level of Nationalist Party Support	-0.262	0.049	-0.185	-0.530	-0.049	-0.452	0.561*	0.126	0.512
p-value for the effect at minimum	.462	.767	.609	.108	.755	.161	.098	.108	.122
Full Effect at maximum of Nationalist Party Support	1.867	0.710	1.904*	1.851*	0.765*	1.699	1.735*	0.041	1.740*
p-value for the effect at maximum	.105	.112	.096	.086	.061	.107	.074	.844	.074
2010-2012									
Log (number of VK users), 2011 x Nationalist Party Support in 2003	10.671***	3.394*	9.319***	7.002**	3.490**	4.249	5.775**	-0.250	6.135**
Weak Instrument Robust Confidence 95% Sets	(4.360; 20.136)	(-.085; 6.873)	[(3.195; 18.504)	[(1.251; 12.753)	(.391; 8.138)	-1.564; 10.063)	(.563; 13.593)	(-2.603; 2.103)	(1.248; 13.466)
	[3.864]	[2.130]	[3.749]	[3.521]	[1.897]	[3.559]	[3.191]	[1.441]	[2.992]
Log (number of VK users), 2011	0.167	0.199	0.111	0.286	0.263*	0.060	-0.132	-0.031	-0.053
Weak Instrument Robust Confidence 95% Sets	(-.766; .789)	(-.133; .696)	(-.795; .715)	(-.276; 1.128)	(-.028; .700)	(-.483; .602)	(-.940; .407)	(-.333; .170)	(-.808; .451)
	[0.381]	[0.203]	[0.370]	[0.344]	[0.178]	[0.332]	[0.330]	[0.123]	[0.308]
Nationalist Party Support in 2003	5.365*	1.312	4.320	2.592	0.947	1.743	4.014*	0.190	3.685*
	[2.884]	[1.125]	[2.715]	[2.200]	[0.938]	[2.074]	[2.136]	[0.694]	[2.053]
Full Effect at minimal level of Nationalist Party Support	-0.344	0.036	-0.336	-0.050	0.096	-0.144	-0.409	-0.019	-0.347
p-value for the effect at minimum	.387	.857	.388	.88	.537	.658	.211	.892	.257
Full Effect at maximum of Nationalist Party Support	2.418**	0.914*	2.076**	1.762*	0.999*	0.956	1.086	-0.084	1.241
p-value for the effect at maximum	.011	.093	.023	.053	.051	.292	.185	0.800	.106
2013-2015								\	
Log (number of VK users), 2011 x Nationalist Party Support in 2003	3.476	1.733	2.677	1.954	-0.077	2.720	2.876	1.937	1.200
Weak Instrument Robust Confidence 95% Sets	(-3.944; 8.423)	(-1.073; 5.942)	(-4.320; 7.341)	(-4.861; 6.498)	(-2.826; 1.756)	(-3.824; 7.082)	(-.252; 7.568)	(-.369; 5.397)	(-.975; 4.463)
	[3.029]	[1.718]	[2.856]	[2.782]	[1.122]	[2.671]	[1.915]	[1.412]	[1.332]
Log (number of VK users), 2011	-0.066	0.195	-0.215	-0.075	0.210**	-0.205	0.057	0.018	-0.006
Weak Instrument Robust Confidence 95% Sets	(-.710; .363)	(-.062; .581)	(-.824; .191)	(-.634; .297)	(.064; .427)	(-.752; .160)	(-.406; .366)	(-.302; .232)	(-.343; .219)
	[0.263]	[0.158]	[0.249]	[0.228]	[0.089]	[0.223]	[0.189]	[0.131]	[0.138]
Nationalist Party Support in 2003	1.920	-0.489	2.456	1.380	-1.263	2.500*	0.805	0.663	0.446
	[1.697]	[1.105]	[1.552]	[1.385]	[0.819]	[1.338]	[1.178]	[0.860]	[0.898]
Full Effect at minimal level of Nationalist Party Support	-0.233	0.112	-0.343	-0.169	0.213*	-0.335*	-0.081	-0.075	-0.064
p-value for the effect at minimum	.396	.556	.158	.438	.066	.095	.704	.635	0.683
Full Effect at maximum of Nationalist Party Support	0.667	0.561	0.349	0.337	0.193	0.369	0.664	0.427	0.247
p-value for the effect at maximum	.365	.129	.627	.632	.397	.595	.132	0.160	.417

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B6: Social Media and Hate Crimes. Reduced Form with Interactions. Specification By Period

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
2007-2009	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (# students in Durov's cohort)	5.886**	2.058**	5.902**	6.163***	2.342***	5.614**	4.497**	-0.031	4.570**
x Nationalist Party Support in 2003	[2.603]	[1.000]	[2.622]	[2.319]	[0.867]	[2.329]	[2.027]	[0.563]	[2.018]
Log (# students one cohort older than Durov)	0.398	1.705*	-0.730	0.994	1.013	0.303	-0.923	0.899	-1.755
x Nationalist Party Support in 2003	[2.670]	[1.025]	[2.689]	[2.378]	[0.889]	[2.388]	[2.079]	[0.577]	[2.069]
Log (# students one cohort younger than Durov)	0.332	-1.056	0.286	0.606	-0.959	1.280	-1.982	-0.584	-1.727
x Nationalist Party Support in 2003	[2.789]	[1.071]	[2.809]	[2.484]	[0.929]	[2.495]	[2.172]	[0.603]	[2.162]
Nationalist Party Support in 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPbSU Students Cohorts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2010-2012									
Log (# students in Durov's cohort)	7.628***	2.683**	6.598***	5.317**	2.865***	3.025	3.734**	-0.227	4.121**
x Nationalist Party Support in 2003	[2.602]	[1.322]	[2.521]	[2.340]	[0.997]	[2.262]	[1.893]	[0.986]	[1.758]
Log (# students one cohort older than Durov)	-3.056	0.981	-4.879*	-1.514	-0.621	-1.924	-2.712	1.515	-4.384**
x Nationalist Party Support in 2003	[2.668]	[1.355]	[2.586]	[2.400]	[1.022]	[2.320]	[1.942]	[1.011]	[1.803]
Log (# students one cohort younger than Durov)	-1.621	-0.827	-0.653	-3.823	-0.940	-2.261	0.689	0.025	0.486
x Nationalist Party Support in 2003	[2.787]	[1.416]	[2.701]	[2.507]	[1.068]	[2.424]	[2.029]	[1.056]	[1.884]
Nationalist Party Support in 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPbSU Students Cohorts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2013-2015									
Log (# students in Durov's cohort)	2.271	1.537	1.458	1.209	0.319	1.505	2.077*	1.363	0.813
x Nationalist Party Support in 2003	[1.981]	[1.065]	[1.855]	[1.760]	[0.695]	[1.690]	[1.236]	[0.890]	[0.974]
Log (# students one cohort older than Durov)	-1.540	-1.140	-0.634	0.128	0.553	-0.422	-2.086	-1.736*	-0.495
x Nationalist Party Support in 2003	[2.032]	[1.092]	[1.902]	[1.805]	[0.713]	[1.733]	[1.267]	[0.912]	[0.999]
Log (# students one cohort younger than Durov)	-0.082	-1.072	0.507	0.508	-0.280	0.554	-1.028	-0.893	-0.178
x Nationalist Party Support in 2003	[2.123]	[1.141]	[1.987]	[1.886]	[0.744]	[1.811]	[1.324]	[0.953]	[1.044]
Nationalist Party Support in 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SPbSU Students Cohorts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
p-value for 2007-2009 equals 2010-2012	0.637	0.734	0.850	0.809	0.752	0.474	0.816	0.853	0.891
p-value for 2007-2009 equals 2013-2015	0.347	0.742	0.230	0.159	0.0943	0.237	0.327	0.176	0.119
p-value for 2010-2012 equals 2013-2015	0.0901	0.553	0.0946	0.145	0.0767	0.592	0.488	0.232	0.114

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B7: Social Media and Hate Crime. Specification with Interaction. Period: 2004-2015

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethnic hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003 x Dummy for the 2004-2006 period	-1.961	0.143	-2.398	-2.416	0.152	-2.891	-0.072	0.052	0.078
	[2.047]	[0.682]	[2.031]	[1.890]	[0.597]	[1.858]	[1.058]	[0.419]	[1.053]
Log (number of VK users), 2011 x Nationalist Party Support in 2003 x Dummy for the 2007-2015 period	7.433**	3.160**	6.717*	7.843***	2.601**	7.879***	3.296	0.574	2.016
	[3.461]	[1.599]	[3.594]	[3.008]	[1.300]	[3.009]	[2.753]	[1.337]	[2.863]
Log (number of VK users), 2011 x Dummy for the 2004-2006 period	-0.271	0.008	-0.279	-0.193	0.075	-0.243	-0.062	-0.040	0.011
	[0.181]	[0.082]	[0.176]	[0.161]	[0.073]	[0.158]	[0.137]	[0.056]	[0.135]
Log (number of VK users), 2011 x Dummy for the 2007-2015 period	0.229	0.272	0.110	0.163	0.311**	-0.016	0.338	0.051	0.411
	[0.284]	[0.174]	[0.282]	[0.260]	[0.145]	[0.260]	[0.264]	[0.125]	[0.253]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger, their interaction with Nationalistic Party Support, 2003, intereacted with period dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,250	1,250	1,250	1,250	1,250	1,250	1,250	1,250	1,250
p-value for the equality of interaction effects of Log (number of VK users) and Nationalist Party Support for two periods	0.053	0.153	0.068	0.017	0.148	0.012	0.335	0.755	0.586
p-value for the equality of direct effects of Log (number of VK users) for two periods	0.006	0.013	0.033	0.034	0.009	0.185	0.011	0.221	0.010

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B8: Social Media and Hate Crimes. Specification with Interactions. Only Cities with Population above 50k Included

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	30.821**	17.622*	26.674**	29.040**	11.554	28.426**	19.431	4.586	12.581
Weak Instrument Robust Confidence 95% Sets	(6.061; 129.863)	(-6.670, 66.205)	(3.112; 109.14)	(6.639; 141.04)	(-18.83; 41.94)	(5.30; 144.04)	(-61.58; 73.44)	(-22.19; 31.37)	(-98.79; 68.27)
	[15.160]	[9.915]	[14.426]	[13.715]	[7.442]	[14.157]	[16.534]	[8.198]	[17.047]
Log (number of VK users), 2011	0.378	0.853	0.325	0.106	0.888	-0.134	1.651	0.193	1.839*
Weak Instrument Robust Confidence 95% Sets	(-2.130; 3.721)	(-.325; 3.798)	(-2.086; 3.540)	(-2.248; 3.245)	(-.051; 3.707)	(-3.200; 2.932)	(-.0505; 6.754)	(-.8220; 1.207)	(.210; 6.724)
	[1.024]	[0.721]	[0.984]	[0.961]	[0.575]	[0.939]	[1.041]	[0.414]	[0.997]
Nationalist Party Support in 2003	-0.648	-3.067	-0.882	0.406	-3.024	0.614	-6.937	-1.014	-7.245
	[5.006]	[3.368]	[4.787]	[4.585]	[2.581]	[4.444]	[5.476]	[1.957]	[5.341]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	323	323	323	323	323	323	323	323	323
Kleibergen-Paap F-statistics	3.134	3.134	3.134	3.134	3.134	3.134	3.134	3.134	3.134
Full Effect at minimal level of Nationalist Party Support	-1.097	0.010	-0.951	-1.283	0.335	-1.494	0.721	-0.027	1.237
p-value for the effect at minimum	0.330	0.990	0.390	0.207	0.593	0.145	0.555	0.960	0.313
Full Effect at maximum of Nationalist Party Support	6.877*	4.569*	5.950*	6.230*	3.325*	5.860*	5.748	1.160	4.492
p-value for the effect at maximum	0.053	0.054	0.076	0.057	0.058	0.078	0.127	0.527	0.240

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B9: Social Media and Ethnic Hostility, Inferred from List Experiment. City Level. Only Cities with Non Missing Estimates for All Categories Included

Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	List Experiment elicited hostility						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	0.085*	0.157**	0.012	0.168**	-0.013	0.175**	0.021
Weak Instrument Robust Confidence 95% Sets	(-.010; .191)	(.007; .308)	(-.093; .127)	(.041; .309)	(-.133; .107)	(.052; .327)	(-.098; .15159)
	[0.051]	[0.080]	[0.055]	[0.068]	[0.064]	[0.073]	[0.063]
Nationalistic Party Support, 2003	1.186	2.771	0.744	0.833	0.078	0.659	2.424
	[1.624]	[2.732]	[1.857]	[2.085]	[2.202]	[2.282]	[2.099]
Observations	104	104	104	104	104	104	104
Kleibergen-Paap F-statistics	57.095	57.095	57.095	57.095	57.095	57.095	57.095
Effective F-statistics (Montiel Olea and Pflueger 2013)	57.095	57.095	57.095	57.095	57.095	57.095	57.095
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B10: Social Media, Self-Reported Ethnic Hostility, and Inferred City Level Hate (cities with at least 40 respondents)

	Self-reported hostility to other ethnicities						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	-0.058	-0.377	0.009	0.031	-0.136	-0.175	0.150
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	(-1.357; 1.364)	entire grid	entire grid	entire grid
	[0.235]	[0.439]	[0.310]	[0.275]	[0.398]	[0.380]	[0.287]
Nationalistic Party Support, 2003	-0.472	1.092	-0.089	-1.566	0.013	0.378	-1.827
	[1.257]	[2.343]	[1.721]	[2.265]	[1.416]	[1.972]	[1.827]
City-level hate to other ethnicities, inferred from LE	0.033	-0.006	0.060	0.186*	-0.072	-0.052	0.078
	[0.060]	[0.112]	[0.099]	[0.098]	[0.075]	[0.082]	[0.077]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2.147	1.800	2.591	4.189	1.626	2.297	2.467
Kleibergen-Paap F-statistics	2.181	1.827	2.631	4.251	1.651	2.331	2.505
Montiel Olea-Pflueger Effective F-stat	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Number of cities	61	61	61	61	61	61	61

Notes: Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts. Individual-level controls include gender, education categories, dummy for pilot, and age categories. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

Table B11: Social Media and Self-Reported Ethnic Hostility. City Level

Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	-0.134	-0.270	-0.146	-0.349	0.110	-0.205	-0.207
Weak Instrument Robust Confidence 95% Sets	(-.002; 2.733)	(-1.845; .954)	(-.019; 2.650)	(.005; 3.935)	(-.883; 1.028)	(.264; 3.250)	(-.884; 2.048)
	[0.181]	[0.227]	[0.208]	[0.225]	[0.212]	[0.212]	[0.208]
Nationalistic Party Support, 2003	-0.339	1.110	0.218	0.627	-1.450	0.892	0.040
	[1.093]	[1.459]	[1.199]	[1.207]	[1.243]	[1.333]	[1.172]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	125	125	114	122	118	122	117
Kleibergen-Paap F-statistics	8.390	7.340	6.840	7.084	7.408	7.220	7.054
Montiel Olea-Pflueger Effective F-stat	7.691	8.896	7.647	8.402	8.589	8.653	8.384
Montiel Olea-Pflueger threshold for 10% worst cas	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1.

C Translated Survey Script

Questionnaire

Q0. Which city have you been living in for the last 6 months?
List of cities

Q1. How often do you use social networks?
One answer

1	Not at all **[skip to question 3]**
2	Once a month or less
3	Once a week
4	Every day or almost every day
5	Several times a day
6	I'm using social networks nonstop

Q2. Which of the following social networks do you use?
Several answers possible + rotation

1	VKontakte
2	Facebook
3	Odnoklassniki.ru
4	LiveJournal
5	Twitter
98	Other **(please specify)**

Q3. Which websites do you visit most often?
One answer

1	News and analytics websites
2	Social networks
3	Games and entertainment websites
4	Online stores
5	Search engines
98	Other
99	Unsure

Q4. On social networks, do you use your real name or an alias?
One answer

1	Real name
2	An alias, for privacy concerns
3	An alias, but for a reason other than privacy concerns

Q5. How many friends/followers do you have in social networks?
One answer

1	Less than 10
2	10-100
3	100-250
4	250-500
5	500-1000
6	More than 1000

Q6. Do you agree with the statement "I get a lot of important news from social networks"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q7. Do you agree with the statement "Social networks help me find people with similar interests"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q8. Do you agree with the statement "In social networks, people are more sincere than in real life"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q9. To what extent do you trust information in social networks?
One answer

1	Completely trust [skip to question 10]
2	Somewhat trust [skip to question 10]
3	Somewhat distrust [skip to question 11]
4	Completely distrust [skip to question 11]

Q10. Why do you trust information in social networks?
Several answers possible

1	People are more sincere in social networks than in real life
2	In social networks one can find a variety of opinions
3	Certain information is only available in social networks
98	Other reason (please specify)

Q11. Why do you distrust information in social networks?
Several answers possible

1	Many users deliberately spread incorrect information
2	Many users unwittingly spread incorrect information
3	Many users play the fool and write rubbish
98	Other reason (please specify)

Q12. In social networks, how often do you encounter:
[scale: A. Very often, B Often, C Occasionally, D Rarely, E Never]
Rotation of statements, one answer

1	Personal insults
2	Obviously incorrect information
3	Extremist statements
4	Propaganda of violence
5	Religious propaganda
6	Pornography

Q13. Which modern technology do you use to organize gatherings with friends or acquaintances?
Several answers possible + rotation

1	Yes, video calls (e.g., Skype)
2	Yes, messengers embedded in social networks (VKontakte, Facebook, etc)
3	Yes, standalone messengers (WhatsApp, Telegram, ICQ, etc)
4	Yes, blogs or public posts in social networks
5	Yes, SMS (short text messages sent over the phone)
6	Yes, phone calls

THERE ARE TWO RANDOMIZED CELLS.

CELL 1 [QUESTION Q14_1]

Q14_1. **Please think, which of the following statements you agree with. Without telling which particular statements you agree with, please specify the number of statements you agree with.**

THE ANSWER IS A NUMBER BETWEEN 0 AND 5, ROTATION

1	Each week I usually read at least one newspaper or magazine
2	I want Russia to be a country with high living standard
3	I know the name of the Chairman of the Constitutional Court of the Russian Federation
4	I feel annoyance or dislike toward some ethnicities
5	Retirement benefits in our country are sufficiently high

CELL 2 [QUESTIONS Q14_2, 15, 16, 17]

Q14_2. **Please think, which of the following statements you agree with. Without telling which particular statements you agree with, please specify the number of statements you agree with.**

THE ANSWER IS A NUMBER BETWEEN 0 AND 4, ROTATION

1	Each week I usually read at least one newspaper or magazine
2	I want Russia to be a country with high living standard
3	I know the name of the Chairman of the Constitutional Court of the Russian Federation
4	Retirement benefits in our country are sufficiently high

Q15. **Do you feel annoyance or dislike toward some ethnicities?**
One answer

1	Yes
2	No

Q16. **In your opinion, which percentage of the survey participants from your city answered "Yes" to the previous question? If your answer is the most accurate, you will get an additional 100 rubles.**
Enter a number with a percentage sign – restrict from 0 to 100

Q17. **How certain are you in your answer to the previous question?**
SLIDER FROM 0 (COMPLETELY UNCERTAIN) TO 10 (COMPLETELY SURE)

QUESTIONS ON GENDER AND AGE ARE ASKED ON THE TECHNICAL PAGE "CIRCLE", SURVEY RESTRICTED TO PEOPLE 18 – 55 YEARS OF AGE

S3. Please specify your education.
One answer

1	Incomplete secondary
2	Secondary
3	Vocational
4	Incomplete higher
5	Higher
6	Doctorate
99	Not sure

S4. Please specify your occupation (your position).
One answer

1	Director, deputy director
2	Division head (of a branch, shift, department)
3	Specialist with a higher education (medical doctor, teacher, sales manager, engineer, etc)
4	Mid-level employee (secretary, salesperson, security, driver, etc)
5	Creative work (photographer, artist, actor, etc)
6	Small business (owner of a business or individual entrepreneur)
7	Technical or service personnel
8	Worker
9	Military
10	Student
98	Other **(please specify)**

S5. How would you describe your family's current financial well-being?
One answer

1	Not enough money even for food
2	Enough money for food, but purchasing clothes is problematic
3	Enough money for food and clothes, but purchasing a TV, a fridge or a washer would be difficult
4	Enough money for major appliances, but we would not be able to buy a new car
5	Enough money for everything except expensive purchases like a country house or an apartment
6	No material difficulties. Can afford to buy a country house or an apartment if necessary