Social Media And Russian Territorial Irredentism: Some Facts and a Conjecture

Jesse Driscoll and Zachary C. Steinert-Threlkeld

Word Count: 9,364 (inclusive of footnotes and references)

After Kremlin policymakers decided to incorporate the territory of Crimea into Russia, updates on public attitudes in Russian-speaking communities elsewhere in Ukraine would have been in high demand. Because social media users produce content in order to communicate ideas to their social networks, online political discourse can provide important clues about the political dispositions of communities. We map the evolution of Russian-speakers' attitudes, expressed on social media, across the course of the conflict as Russian analysts might have observed them at the time. Results suggests that the Russian-Ukrainian interstate border only moved as far as their military could have advanced while incurring no occupation costs – Crimea, and no further.

Revise and Resubmit, *Post-Soviet Affairs*

Epigraph:

"I would like to remind you that what was called Novorossiya, back in the tsarist days, Kharkiv, Luhans'k, Donets'k, Kherson, [M/N]ikolayev, and Odes[s]a were not part of Ukraine back then. The territories were given to Ukraine in the 1920s by the Soviet government. … Why? God only knows!..."

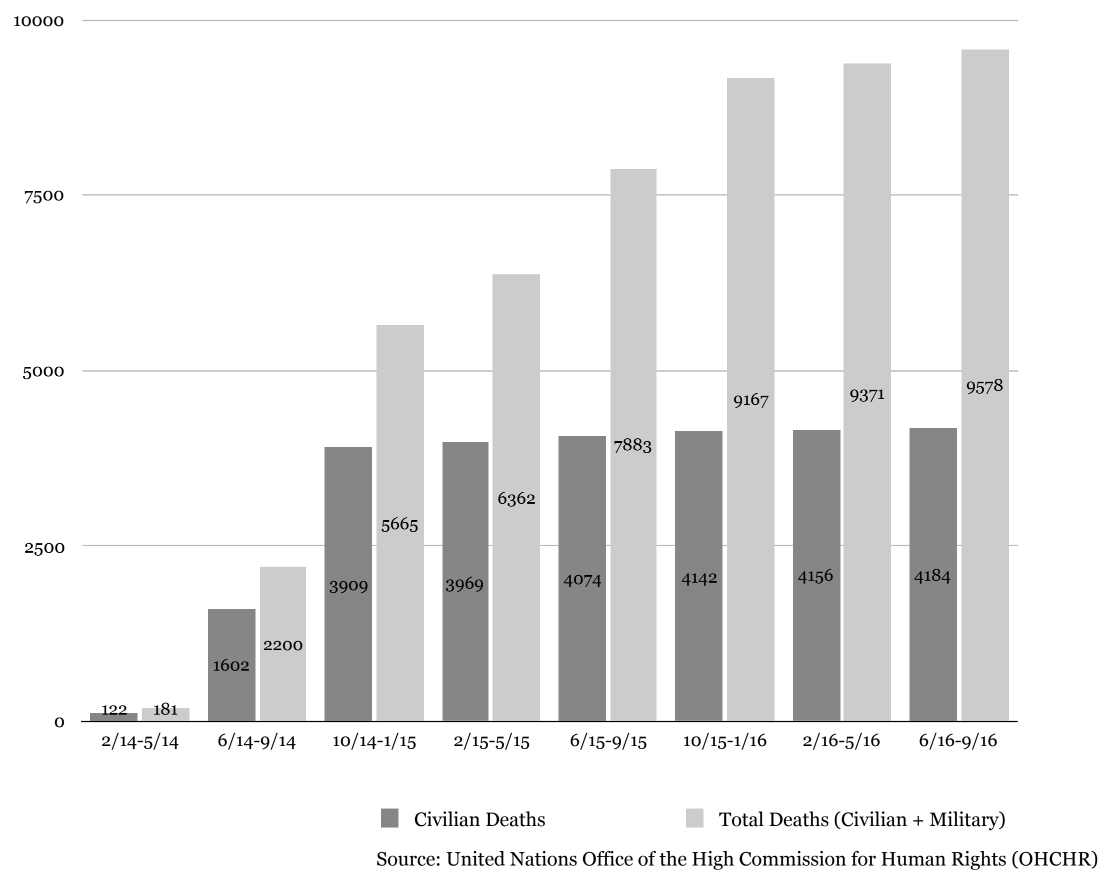
-- Vladimir Putin, April 2014

1. Introduction: Some Facts

On November 21, 2013, protests in Kyiv started against President Victor Yanukovych. On the night of February 21, 2014, Yanukovych fled Ukraine. On or around February 27, 2014, Russian special forces entered Crimea. Russia's interstate borders seemed to be expanding, and Ukraine's contracting, under a "regime phase" of territorial realignment (Lustick 1993, 123). With the map already re-drawn, and no Ukrainian authority to make arrests, newly-formed pro- and anti-government militias acted on their own accord. Throughout 2014 these militias clashed with each other and brutalized civilians. Few of the anti-government militias had much success seizing or holding government buildings or other symbols of power. The exception to this general rule was the eastern Donbas region of Ukraine, where indigenous insurgents captured the regional apparatus of the state in two regions (oblasts), Donets'k and Luhans'k.

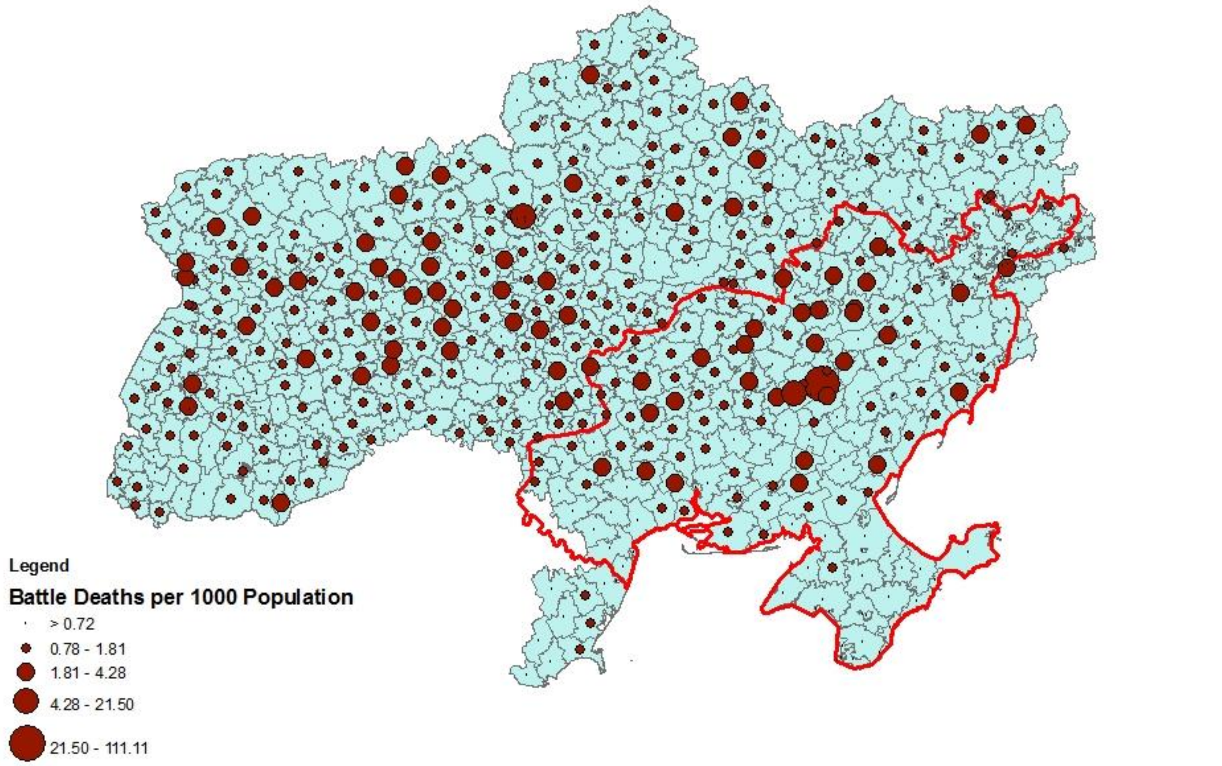
Militias clashed, then consolidated, and, eventually, formed stable coalitions with hierarchical chains of command. Those coalitions are today referred to as "the Ukrainian army" and "the secessionist rebels" in both academic and policy shorthand. In the first year of fighting approximately 4,000 civilians were killed. More than one million individuals fled their homes as refugees or IDPs. Property and industry damage is estimated in the tens of billions of dollars. Zones of fighting calcified into stable frontlines in the winter of 2015 after Russia sent regular troops to tip the scales at two critical junctures, the Battles of Ilovaisk and Debaltseve. Territory has not changed hands significantly since those battles. As the war conventionalized along a territorially-fixed line of contact, the brutalization of civilians slowed (see **Figure 1**).

Figure 1



CAPTION: Cumulative civilian and military deaths in Ukraine between February 2014 and September 2016. The darker column is civilian deaths and the lighter column is the sum of civilian and military deaths. Data from the United Nations Office of the High Commission for Human Rights (OHCHR).

Figure 2



CAPTION: Military deaths as a fraction of total oblast population on the Ukrainian side since the spring of 2014. Data on the birthplace of the deceased is from Ukrainian Memorial. Data on Oblast populations is from the 2001 Ukrainian census. A thick line surrounds the territory of historical Novorossiya, which in defiance of early predictions is producing anti-invasion/anti-Putin martyrs at a rate that consummate with other parts of Ukraine. The largest dot represents the city of Dnipro. Crimea is shown on this map and others in this paper as part of Ukraine in order to reflect the plasticity of interstate borders that would have been felt at the time.

We are now five years in. What has emerged is not the ethnic bloodbath some experts feared. There has been no Srebrenica. Ethnic identities have hardened, perhaps, but not in simple ways pitting ethicized Russians against ethicized Ukrainians. Civilizational and religious fault lines have not been weaponized as some pessimists predicted. Most Russian-speakers living in Ukraine in 2014 rejected calls for rebellion against the post-Yanukovych government.

Many born in historical *Novorossiya* went further, actively performing their Ukrainian patriotism by volunteering to fight off an invasion. Consider **Figure 2**. It is a map of Ukrainian martyrs per capita by oblast. Each dot represents the fraction of an oblast's population that have died on the frontlines, usually victims of shells fired from the territory of the self-declared 'Donets'k people's republic' (DNR) or 'Luhans'k people's republic' (LNR). It has been well-documented that many pro- and anti-Putin fighters on both sides have pilgrimaged from distant lands, but it is also well-documented that many fighters are locally recruited. Russian is the lingua franca on both sides of the line of contact. Many soldiers imagine themselves to be fighting for their homes. Ukraine's war pits Russian-speakers that accept the premises of the Russian state narrative against Russian-speakers that are inoculated against that narrative.

This paper uses social media data to reconstruct how the Russian-state narrative was received by Russian speakers living in Ukraine during the critical period between February 2014 (when Yanukovych fled Kyiv) and the Battle of Ilovaisk (when the Russian military intervened directly and froze the territorial front-lines). Our conjecture is that during that time, policy elites in Moscow would have been considering using their conventional military to move the undeclared front-lines of the war further West. These planners would have been hungry for information on the social attitudes of Ukrainian Russian-speakers (*russkoiazychnoe naselenie*).

Russian planners would have wanted to know if they were interested in opting out of the Ukrainian polity.

We show that Russian-language social media traffic could have been one new source of military intelligence. Since the prevalence of overtly political behaviors on social media provides important clues about the political dispositions within communities, a growing body of scholarship has taken advantage of these data to understand contentious action in Ukraine (Onuch 2015, Metzger et al. 2015, Wilson 2017, Metzger and Tucker 2017). Our departure from previous studies is emphasizing the potential for these data to be repurposed for crisis decision-making. As proof of concept, we reconstruct a number of different maps of social attitudes shared by users of Russian-speakers active on social media. Our dataset contains approximately 7 million online user entries (tweets), all generated within the territorial borders of Ukraine. Aggregated patters in the data we analyze provide a measure – noisy, but informative – of how many self-identified Russians living in Ukraine would have favored border revision. Most did not.

Our supposition is that if Russian strategists were considering expansion beyond Crimea, they would have been able to use social media information to assess, with a great deal of precision and in real time, the reception that they would likely receive. Since interstate border changes are rare events, the re-purposing of public data for military reconnaissance has not yet been considered despite excellent studies of how polarized media bubbles allow conflicting coverage of the same events (Warren 2014, Baum and Zhukov 2015, Peisakhin and Rozenas 2018), how internet connectivity enables cyber-operations (Gartzke 2013, Kostyuk and Zhukov 2017), and the advantages that some states seek by deliberately muddying the historical record (Beissinger 2015, Laurelle 2015, Hopf 2016, Snyder 2018, Hale et al 2018).

2. Background: Divergent Narratives

Violence between self-identified Russians living in Eastern Ukraine and their self-identified Ukrainian neighbors was not an issue after the disintegration of the Soviet Union. In a study comparing the characteristics of four Russian-speaking "beached diasporas" – communities that found themselves living on parts of what they construed as their homeland, but divided among new post-Soviet states – Laitin (1998) attributes peaceful interethnic relations in Ukraine to a combination of deterrence and the ambiguity of political identity boundaries:

> The major mechanism holding back interethnic violence in Ukraine … is the feeling by Russians … that if they were ever terrorized (*qua* Russians) by the [Ukrainians], the Russian Federation would come to their aid. […] But another mechanism reducing the likelihood of interethnic violence … is the embarrassing fact (for both sides) that *the boundaries of opposition are not at all clear*.[1]

The breakdown of this peaceful equilibrium began in the fall of 2013. On November 21, 2013, Ukrainian President Viktor Yanukovych declined to sign an association agreement with the European Union (EU) in order to explore membership in Russia's Eurasian Economic Union (EEU). This reversal was seen as the culmination of years of friction and competition between Russia and Western Europe (Colton 2016, Charap and Colton 2017). Social forces mobilized. Maidan Square in central Kyiv became a focal point for "Euromaidan" protests which, as the weeks passed, took on an all-or-nothing anti-regime flavor. Clashes between state security forces and armed protesters gradually produced martyrs. Ukraine's government imploded on February 21, 2014. Yanukovych fled Kyiv that night.

---

[1] Laitin (1998), 185, emphasis added.

Russian special forces seized Crimea a few days later. The popular understanding within Russia was that its military was "coming to the defense" of ethnic Russians at risk.[2] Over the next few months, military drills along the border provided cover for 50,000 Russian soldiers to mass, signaling that Russia could invade "at a moment's notice."[3] The areas that eventually became the front-lines of the conventional war – parts of Donets'k and Luhans'k – were areas directly adjacent to Russian territory, where secessionist militias could anticipate the possibility of easy military resupply. Between February and August, on the Western side of the gradually solidifying conventional front lines, there were sporadic attempts by provocateurs sympathetic to the Russian cause to provoke general uprising in Eastern Ukraine (historical *Novorossiya*). Most failed. The Russian military did not send aid to militias outside of Crimea or the Donbas. Though widespread speculation of clandestine Russian assistance persists, and is made plausible by few prominent pro-Kremlin volunteers – and more facts may yet come to light – direct Russian intervention did not occur until July (four months after Crimea and ten weeks after the Ukrainian government began its "Anti-Terrorist Operation" (ATO) to forcibly re-incorporate the East). In the end, Russia only sent conventional ground forces to assist secessionist militias in the Donbas that had already demonstrated capacity to hold government buildings for months.

Though Russia did not engage in overt kinetic military activity outside of Crimea, Russian-language media broadcasts during the time represent an exemplar information warfare

[2] The entire local government institutional apparatus, which were all remnants of the suddenly-defunct Party of Regions, accepted Russian rule almost immediately. That said, how much popular support there was for Russian military actions within the permeant population of Crimea will continue to be disputed. See, for example Suslov (2015) and Faizullaev and Cornut (2017).

[3] Charap and Colton (2017), 132.

campaign.[4] One goal was to solidify Russian domestic opinion. Another was to encourage Russian-speakers within Ukraine to take advantage of the temporary window of Ukrainian state incapacitation and rise up. Petersen (2001) identifies three analytically distinct triggering mechanisms that can impel leaderless resistance: (1) the amplification of emotions of resentment, especially caused by prospective status reversals for one's ethnic group and subordination to another ethnic group (especially a hated one); (2) coordination on a few focal points and infusing them with special symbolism; and (3) valorization of heroic resistance, assuring citizens that incurring small risks of martyrdom will be accompanied by large community status rewards.

All three triggering mechanisms were prominent in the content of Russian television coverage of post-Maidan Ukraine. Emphasis on status reversals for Russians was overt. A constant barrage of news stories – including fabricated stories about Russian boys being crucified by Ukrainian far-right groups and staged photographs of soldiers proudly displaying flags of the *Azov* paramilitary group alongside the NATO flag and a Nazi flag – left no doubt that Russians, stranded in Ukraine, were potential hostages and under imminent threat.[5] The subordination of Russians in a new status hierarchy below Ukrainians was a reoccurring theme.[6] Valorization of heroic resistance to Ukrainian fascism was accompanied by promises of status rewards to patriotic volunteers from across the Former Soviet Union. The reciprocal decision by

[4] Reisinger and Golts (2014), 5, 3-8, Darczewska (2014), Lucas and Pomerantsev (2016), and Chivvis (2017). Gerasimov (2013) merits a close read, as does Beissinger (2015b), Pomerantsev (2015), Romanets (2017), and Snyder (2018) Chapter 5. See also McFaul (2018), 430-40.

[5] See Peisakhin and Rozenas (2018). For evidence of saturation of Ukraine-related stories on Russia's Channel One News, see especially their Figure 1.

[6] Russia's narrative reinforced analogies to World War II (e.g., by "NATO" with "Nazis" using consonant repetition, substituting "Germany" for "the EU," the explicit claim that Maidan was a CIA coup, etc.) and rabid anti-Americanism. See for example Cottiero et al. (2015).

the interim, post-Maidan Ukrainian government to respond to uprisings in Donbas with a national counterinsurgency policy called an "Anti-Terrorist Operation" (ATO) was also obviously strategic messaging, meant to resonate in NATO capitals and in the imaginations of Ukrainian patriots.[7]

This gloss is not meant as a comprehensive history, simply an *amuse-bouche* to whet the appetite for empirical exposition. After an unexpected regime change in Kyiv, Russian-speakers were provided two competing narratives to make sense of the tectonic political shift. Different anchoring keywords – one promulgated by the Kremlin and one the other promulgated by the new government Kyiv – resulted in bifurcated narratives. These narratives containing well-understood focal points (coup, fascist, terrorist, invasion, etc.) calibrated to exile one's political enemies from respectable coalition politics. **Table 1** summarizes the competing narratives.

[7] Boyd-Barrett (2017), Scholz (2016).

TABLE 1: COMMON COMPONENTS OF COMPETING NARRATIVES

	Pro-Kremlin Narrative	**Anti-Kremlin Narrative**
The appropriate Russia-Ukraine relationship, taking the relevant historical facts into account, ought to be one of …	...natural hierarchy. Borders are gifts from Soviet times.	...diplomacy between sovereign equals.
Future historians, writing about the Maidan events, will describe them as...	...a coup by far-right social forces, emboldened by material and moral support of the NATO alliance and Western intelligence agencies.	...a broad-based social movement against an illegitimate government.
Future historians, writing about Putin's responses to the Maidan events and their aftermath— including the seizure of Crimea — will describe them as...	…heroic.	…criminal.
The proximate cause of the violence in East Ukraine is...	...the CIA coup which brought fascists to power.	...Putin's illegal seizure of the Crimean peninsula, leading some in Ukraine's east calculated that if they organized militias, Russia might assist them too.
Any account of the violence in East Ukraine is incomplete if it does not reference deeper structural causes, such as…	…decades of Western policies to encircle Russia, expanding NATO and aggressively pushing regime change in post-Soviet states under the aegis of democracy promotion.	… the basic incompatibility of values between Putin's regime and the European Security Community.
Soldiers fighting to secede from Eastern Ukraine are best described as...	…Russian patriots.	…terrorist insurgents.
Main keyword for narrative track (by revealed user preferences, aggregated from the word-clouds found in supplementary materials):	fascist	terrorist

3. Methods: Mapping Divergent Narratives

Starting on August 26, 2013, we connected to Twitter's streaming application programming interface (API), requesting only tweets with GPS coordinates. We first filtered for time, focusing on the 188 days from February 22, 2014 to August 28, 2014. This filter generated a sample of roughly 940,000,000 geotagged tweets, which we then reduced to 6,880,623 tweets originating within the territorial borders of Ukraine (Crimea inclusive).[8]

We divide the sample into three periods: Crimea (dated from the flight of Viktor Yanukovych on February 22 until the March 15 voting referendum in Crimea); the post-Crimea period (in which local forces organized for secession knowing that the Ukrainian-Russian interstate border was in flux, which ended with the election of Petro Poroshenko via a mass-participation voting exercise on May 26); and the subsequent conventionalized artillery war in the Donbas region (May 27 until August 28). When Russian armor intervened directly in late August at the Battle of Ilovaisk, while at the same time official Russian diplomats were denying that they were doing so, it was clear that ease of seizing Crimea would not be repeated.

Two primary considerations caused us to prefer Twitter to Facebook or *VKontakte*. One concern was minimizing platform bias. Platform choice was itself a signal of political preferences: *VKontakte* features heavily pro-Russian users and Facebook, more pro-Maidan ones

---

[8] We foresee two distinct potential methodological objections to this methodology: (a) Twitter-users are systematically different from non-users and, (b) that Twitter users who geotag tweets are different than other Twitter users. The first concern will be addressed in the text below. Though we do not have an empirical strategy to address the second concern, other research teams employ geotagged Tweets in studies of behavior in Ukraine and find patterns consistent with behavioral expectations (Wilson 2017). More research on geotagged tweet bias, however, is needed. In the United States, geo-located accounts are more likely to be from smartphones, residents of cities, certain minorities, and higher income U.S. census tracts (Malik et al. 2015).

(Gruzd and Tsyganova 2015). Twitter, by contrast, was new enough at the time that it had little reputation beyond being a popular social media platform that did not censor, and it contained a sufficient population of both pro-Kremlin and anti-Kremlin users for analysis.

Multiple studies have shown that Twitter contained, and perhaps still does, substantial numbers of pro-Kremlin and anti-Kremlin Russian language accounts and tweets. An analysis of Russian Twitter users from 2010-2011 shows that many users focused on Ukraine, including a sizable group with a positive attitude towards Russia (Kelly et al. 2012). Pro-Putin attitudes were popular on Twitter in Russia around the 2011 Duma and 2012 presidential elections and continued to track offline events (Spaiser et al. 2017).

Second, was is more practical for our team to acquire and work with large quantities of Twitter data than *VKontakte* or Facebook data in 2014. *VKontakte*'s API is neither well-documented in English nor reliably uncensored. Using Facebook profile data requires working with internal researchers, providing the company veto power at every stage of research.

To make responsible inferences about public opinion expressed on social media, it is necessary to contrast earnest reproduction of keywords that signal support for the Kremlin narrative against the prevalence of users overtly rejecting those arguments. For this we applied language filters. Metzger et al. (2015) demonstrate that many multilingual Twitter users performed solidarity with Maidan protesters by activating their Ukrainian identity by communicating on Twitter in Ukrainian, but they switched *back into Russian* after the success of Maidan. An interpretation is that they switched back in order to participate in online information warfare – the replication (or self-production) of pro- or anti-Kremlin propaganda. Our study restricts the investigation to Russian-language content since the Russian-speaking populations (*russkoiazychnoe naselenie*) are the subsample of Ukrainian citizens whose beliefs would have been most salient to Kremlin strategists. Though the master cleavage of this war (Kalyvas 2006: 14, 389) is the East-West (Russophone-Ukrainophone) division, our design is calibrated to illuminate intra-Russian-speaking discordant politics. Our tests assume that Russian-speaking civilians were purposive agents competing with each other to explain the many unexpected off-line political upheavals that took place within Ukraine during the study period.

We employ a dictionary of keywords for parsimony and interpretability. The six months in this study were marked by a series of dramatic, contentious events with their own vocabulary. A few weeks after the voting exercise in Crimea, coordinated protesters occupied government buildings in the eastern Ukrainian city of Kharkiv in April demanding a referendum on independence. There was an attempt to storm a police station in Mauripol to seize heavy weapons. There was a military siege on the city of Sloviansk that would last through early July (live-webcast, with constant YouTube updates). In early May, violent clashes in Odessa left 42 people dead when a building caught fire. The framing language of "fascism" and "terrorism" was prevalent in descriptions of all these events. **Table 2** presents the dictionaries.[9]

Two steps code the tweets. First, a Python script filters all tweets from Ukraine so that each tweet in the sample contains at least one keyword from a narrative's dictionary. We considered complicating our bag-of-words approach with cases and declensions but, based on initial visual inspection of the sample, opted for a dictionary including only nouns and adjectives

---

[9] Though complex syntax and subtle reasoning can stymie dictionary classifiers (Schwartz and Ungar 2015), we chose a bag of words approach because 140 character tweets are direct and short. A similar "bag of words" method has been used in other event-based studies relying on tweets (Ramakrishnan et al. 2014; Ritter, Etzioni, and Clark 2012).

in the nominative case for initial filtering. The irregular use of declensions (and irregular spellings generally) on Twitter may be a confound, but we opt for clear coding of meanings in the initial dictionary (knowing we can use these accounts to build a supervised model that will pull other important variants, including the same words with different declensions – see below). Given the much longer and larger pro-Kremlin dictionary we have no reason to believe the decision to search only in the nominative biases inferences systematically.

After manually screening for automated accounts ("bots"), this process yielded 5,328 tweets from 1,339 individual accounts. Second, teams of Russian speakers – four native Ukrainians and three fluent Russian-speaking residents of North America – read each tweet and coded it as pro-Kremlin or anti-Kremlin. This second step is necessary both because irony confuses unsupervised computer classifiers and sometimes has poor inter-coder reliability and because visual inspection was the most reliable way to spot automated accounts. To understand the demographics and professions of these users, we searched Google, Facebook, and *VKontakte* for each user in our sample. Tentative results suggests that the sample skews slightly young (16-36) and male, but with a bulk of accounts unidentifiable on these characteristics.[10]

Having identified the 5,328 tweets containing at least one of the keywords from **Table 2**, we built two separate supervised models to identify pro-Kremlin and anti-Kremlin tweets that our dictionaries might have missed. Hand-coded tweets were used as a training set on which we built each model. Through processes described in our Supplementary Materials we stemmed all

[10] Only 24 accounts, and 196 tweets, are from bots, based on manual inspection. All hand-coded analysis in the paper's main results excludes them.

words, removed stopwords, and dropped all *Foursquare*-account generated data.[11] The resulting classifiers identify 58,689 tweets as pro-Kremlin and 107,041 as anti-Kremlin, a total of 166,454 tweets. Training the supervised models provided us a higher proportion of pro-Russia narrative than the dictionary, bringing our results more in line with other studies of the Ukrainian and Russian Twittersphere (Kelly et al. 2012, Spaiser et al. 2017, Wilson 2017).

[11] The first output of the classifier identifies 204,189 tweets as concerning either narrative, 144,776 of which are anti-Kremlin. Validating the output showed many were of the form of "I'm at [place]", indicating that the tweet was created on the app *Foursquare*. Since no *Foursquare* tweets had appeared in the hand-coded data, this confound was not discovered until the referee process. Simply removing all tweets that start with "I'm at" (a total of 37,735) from the study did not change results substantively, but we rely on the smaller dataset excluding *Foursquare* data for all analysis in this paper. See Supplementary Materials for more details.

TABLE 2: KEYWORDS DICTIONARIES FOR INITIAL CODING

Anti-Kremlin Dictionary (English)	**Anti-Kremlin Dictionary (Russian)**
Terrorist	tеррорист
Terrorists	террористы
Terrorism	терроризм
Pro-Kremlin Dictionary (English)	**Pro-Kremlin Dictionary (Russian)**
Radical, radicals	радикальные, радикалы
Right-Wing Radical (adj.)	праворадикальные
Nationalist Radical (adj.)	национал-радикальный
Right-Wing Extremist (adj., fem.)	правоэкстремистская
Right Terrorism	правый терроризм
Extremist (adj.), Extremism	экстремистский, экстремизм
Neo-Nazism, Neo-Nazi	неонацизм, неонацистский
Nazis, Nazism, Nazi (fem.)	нацисты, нацизм, нацистская
Nationalist, Nationalist (adj.)	националист, националистическое
Nationalist-Radical (adj.)	национал-радикальный
National Minorities	нацменьшинства
Ultra-nationalist (adj.)	ультранационалистические
Fascism, Fascist (adj.)	фашизм, фашистский
Mercenaries, Fighters	наемники, боевики
Antisemites	антисемиты
Russophobes	русофобы

4. Results

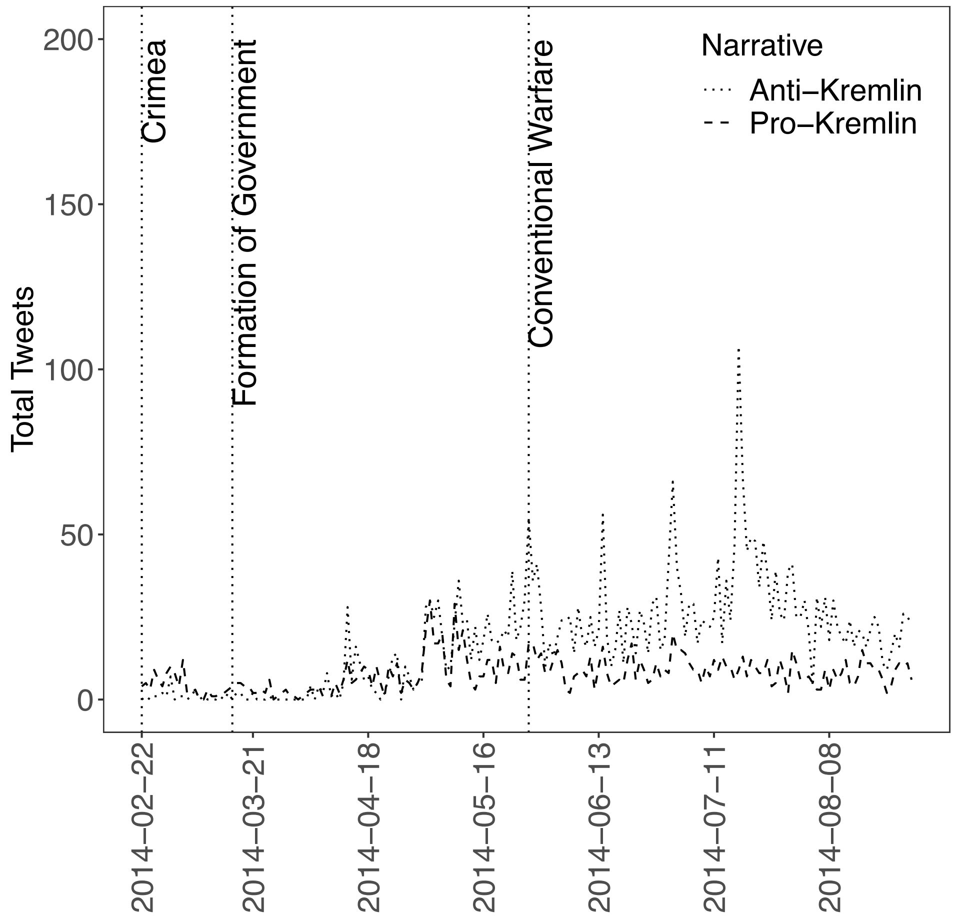
4a. Results

The keyword filter finds a large majority of the tweets (85%) to be anti-Kremlin. This result surprises, given that the pro-Kremlin selection dictionary is much larger than the anti-Kremlin dictionary.[12] **Figure 3** is a time plot of the raw data, organized by narrative track. The dotted lines in this figure divide the sample into three periods discussed above. Until mid-May, the two narratives peak on the same days, suggesting an online clash of narratives as locals used their accounts to narrate the same off-line events competitively (signaling solidarity by performing the pro-Kremlin/anti-Maidan line or performing anti-Kremlin/pro-Maidan line for an audience in the social network).[13]

[12] The proportion our sample may be higher than 85%. Hand-coded tweets from the dictionary method contain 273 clearly pro-Kremlin, 4,338 clearly anti-Kremlin, and another 543 with disputable content (e.g., with inter-coder variation across the options of “neither” or “both”).

[13] The Supplementary Materials contains evidence of spatial and temporal correlations between social media behaviors that earnestly reproduce the anti-Ukraine (“fascist”) narrative and ironic “trolling” behaviors (using the fascist keywords, but intending to mock that position).

Figure 3



CAPTION: Vertical dotted lines divide the sample into three periods: Crimea (dated from the flight of Viktor Yanukovych until the March 15 referendum); the post-Crimea period in which local forces organized for secession (March 16 to May 26); and the subsequent conventionalized artillery war in the Donbas region (May 27 until August 28). Cauterized uprisings by Russian-speakers in various parts of Novorossiya occur in late April and early May. The Ukrainian Government's ATO ("Anti-Terrorist Operation") is initiated in May. The visible outlier in anti-Russian Twitter activity on July 17 is descriptions of the downing of Malaysian Airlines Flight 17 as terrorism.

The greatest density of pro-Kremlin tweets occurred in April and May. During this period, the Russian military had consolidated control of Crimea, but it was unclear whether the Kremlin would come to the assistance of militias who seized territory and advertised their desire to secede. The anti-Kremlin narrative did not emerge as dominant until the government's dedicated counterinsurgency policy (the ATO), which transformed "terrorism" into the focal point for resistance. By early July, the anti-Kremlin narrative had a clear advantage on Russian-language social media.

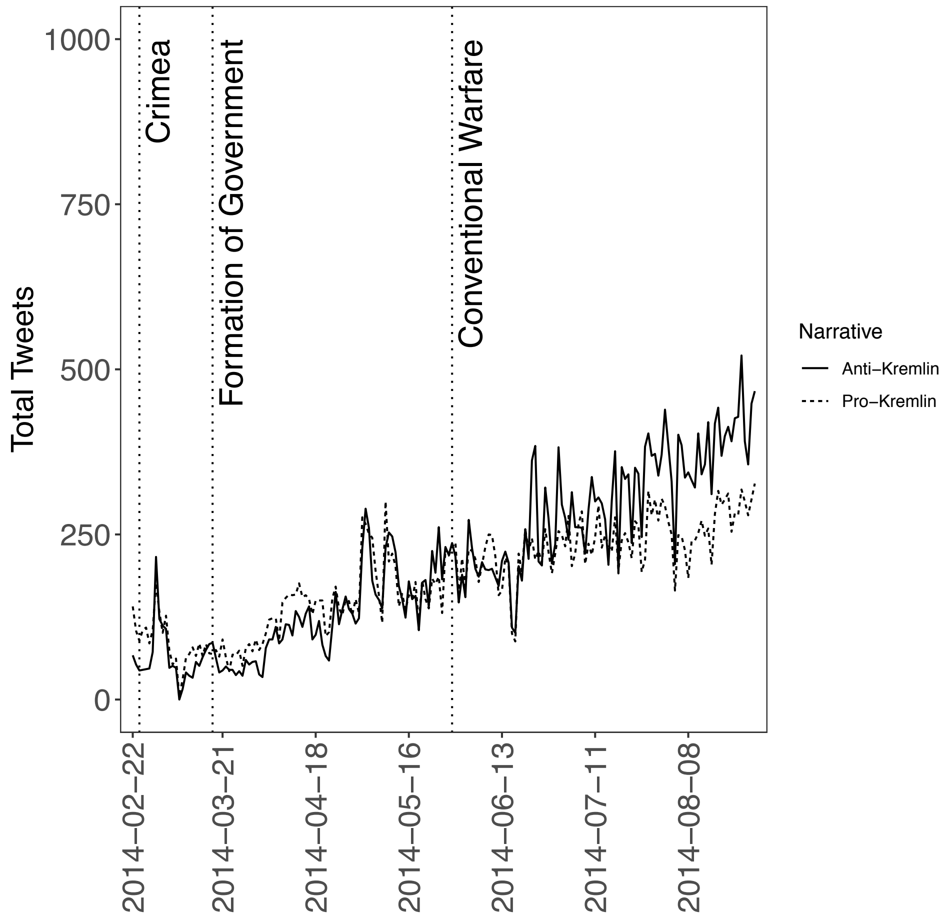
Just as in all highly-polarized news coverage, in some cases these competitive narratives "latched on" to the same offline events, providing different frames for their description, and in other cases the narratives "change the subject" and simply focus on different events. The Twitter Streaming API provides only 1% of tweets and thus prevents the reconstruction of full threads. We must make inferences about offline events based on what amount to conversation snippets. It is still straightforward to see observe narratives in tandem. In the immediate aftermath of the downing of Malaysian Airlines Flight 17 (MH17) on July 17, 2014, Pro-Kremlin tweets promulgated the narrative that the Ukrainian military had shot down its own plane. Anti-Kremlin tweets promulgated the more standard narrative outside of Russia (that a BUK missile had been fired by separatists and that Russia was a state supporter of terrorism). We manually analyzed the subsample of 567 tweets from our human-coded sample from the day before to the week after and found that even in this week, when the airline crash provided a clear focal point for news coverage, a plurality of tweets in our sample referred to other aspects of the conflict (e.g., movement of weapons across the Russian border, POWs captured and exchanged, SBU operations, stockpiled weapons, battlefronts in Luhansk, individuals tweeting the location of separatists to authorities, etc.) or employed generic name-calling. Approximately twice as many anti-Kremlin tweets as pro-Kremlin tweets (41% compared to 20%) in our sample referenced the event, suggesting it was more common for the pro-Russian narrative to dwell on other themes.

The front lines, where Russian military intervention would have been operationally feasible at low cost, are of special interest. **Figure 4** replicates **Figure 3,** but using the larger machine-learning dataset and only examining geotagged tweets generated in *Novorossiya*. This region is where uprisings of ethnic Russians never occurred, but if they *had,* might have been joined with logistical support from the Russian military.[14] Note that the two narratives track each other relatively evenly for the first two periods, suggesting an ongoing battle on social media for hearts and minds in Russian-speaking communities in *Novorossiya* in the early months of the conflict followed by dominance of anti-Kremlin messaging in the third period once conventional frontlines solidified.

The Supplementary Materials show the same graph using data from all of Ukraine, and, not surprisingly, the Anti-Kremlin discourse clearly dominates the entire period. We also undertake two additional robustness checks to ensure that automated accounts ("bots") do not drive the results from the machine learning models. First, we submit every user to Botometer, a service that produces a probability estimate that an account is a bot and drop tweets from any account with a probability of being a bot greater than or equal to .4 (Varol et al. 2017). The Anti-Kremlin narrative dominates the Pro-Kremlin one even more once these tweets are removed. Second, we drop tweets from any account at or above the 95th percentile of the tweet frequency or friend:follower distribution, as these behaviors are common features of bots (Bessi and Ferrara 2016). Dropping tweets on these criteria does not substantively alter our inferences.

[14] The quote from Vladimir Putin in the epigraph, openly questioning the legitimacy of the border between Ukraine and Russia, was delivered with scripted sincerity and references these oblasts.

FIGURE 4



CAPTION: The subset of the raw data from the machine-learning dataset (N=166,454) after dropping , replicating Figure 3, using only data generated from oblasts in historical Novorossiya. In March and early April, the pro-Kremlin narrative ebbed and flowed but was generally dominant in the Novorossiya sample. As the conventional warfare phase gets underway, the anti-Kremlin narrative begins to dominate.

To enable spatial comparisons across Ukraine, we exploit the variation in the prevalence of each narrative across oblasts as a fraction of overall Russian Twitter behavior. We first calculate the percentage of all Russian-language geotagged tweets originating within an oblast.[15]

[15] We did not aggregate to a lower geographic level, such as city or raion, because of lack of tweets available in our hand-coded sample, especially if we also wanted to subdivide the data into smaller bins by time period.

This number is the denominator. To calculate the numerator, we repeat the oblast-level calculation for the population of tweets that contain keywords from either narrative track. The percent of all tweets in Russian from each *oblast* (A_i), the percent of all pro-Kremlin tweets in Russian from each *oblast* (B_i), and the percent of all anti-Kremlin tweets from each *oblast* (C_i) can be used to compare the percent difference between B_i and A_i, and the percent difference between C_i and A_i. This quantity measures the over- or under-production of pro-Kremlin (B_i) or anti-Kremlin (C_i) tweets, against an oblast-specific production baseline. So long as we observe *some* pro-Kremlin behavior in every oblast there is enough data to make comparisons across space.[16]

The non-parametric nature of this operationalization generates two advantages. First, it mechanically controls for an oblast's Russian-speaking Twitter population and any omitted demographic, social, economic, technological, or political variables that might correlate with the overall percentage of Russian-language Twitter users (since the denominator of each oblast is the total number of Russian-language geotagged tweets).Second, it avoids mechanically capturing the East-West (Rusian-Ukrainian) cleavage since it incorporates oblast-specific amounts of Russian production. For example, it is not surprising that Crimea would produce a lot of pro-Russia tweets. What is surprising is that it produces more than would be expected given its baseline production of tweets in Russian. This measure therefore captures the residual "cultural

[16] For example, in one period, Crimea produced 4.3% of all Russian-language tweets, but it produced 10.93% of all Russian-language tweets reproducing the pro-Kremlin narrative. Our method would calculate that it produced 154.19% (10.93-4.30)/(4.30) more pro-Kremlin tweets than the baseline expectation. Results are robust to an alternative model specification in which the denominator is the percentage of all geotagged tweets originating from within each *oblast* regardless of language.

package" of pro-Kremlin attitudes that outlived the institutional implosion of the Party of Regions.

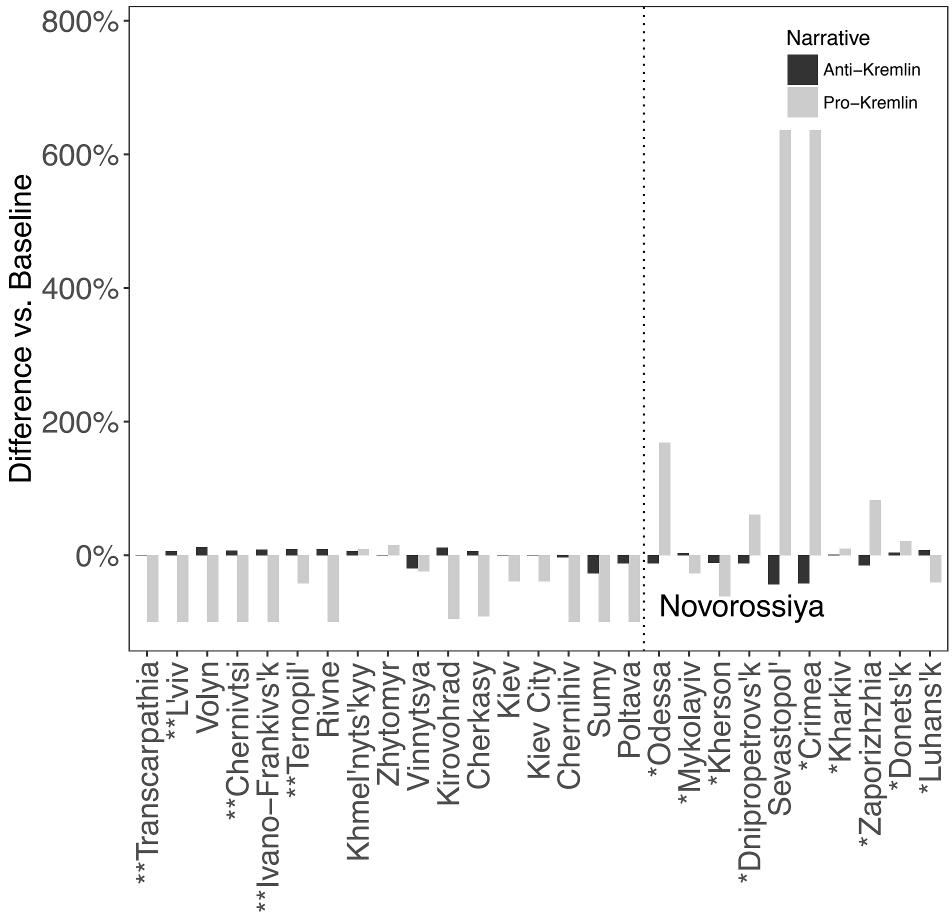
Figure 5 displays relative over- or underproduction of narrative track keywords by oblast using the full sample of our hand-coded data (N=5,328). Oblasts are arrayed from west to east. The clear outlier is Crimea and its capital Sevastopol'.[17] The other oblasts where social media users reproduced the pro-Russia ("fascism") narrative were Transcarpathia, Khmel'nyts'kyy, Zhytomyr, Odessa, Dnipropetrovs'k, Kharkiv, Zaporizhahia, and Donets'k. Over-production of the anti-Kremlin narrative is more common in the West, but also, crucially, in Mykolayiv, Donets'k, and Luhans'k. These areas are precisely where it would have been easiest for Russia to expand if it wanted to. The Spearman rank correlation between this relative production measure and the proportion of pro-Kremlin tweets by oblast is .58. That many people in these supposedly pro-Russia oblasts were against Russia may have given pause to military planners.

Figure 6 replicates **Figure 5** with the machine-coded dataset. Broad trends are similar, but two differences deserve noticing. First, the Crimean oblasts are no longer extreme outliers. Second, the East-West dimension of the data is now much more pronounced. Russian-language Twitter behaviors in the 7 oblasts that were formerly part of the Habsburg Empire systematically

[17] We speculate the relative over-production of pro-Russia discourse in Crimea is explained jointly by a few causal processes: (a) over-production as a reflection of an authentic broad-based outpouring of support for rejoining the homeland; (b) information operations conducted by pro-Kremlin agents that were not indigenous citizens of the peninsula; (c) the Russian military presence deterred the production of the "anti-Kremlin" narrative by indigenous citizens, creating strategic self-censorship by citizens who left the Twitter-sphere once military occupation was a fait accompli; (d) the linked claim that in other parts of Ukraine, post-Maidan residual state capacity deterred the irredentist Russian narrative, but those constraints ceased to be present in Crimea very early in our study. Our design cannot disentangle these mechanisms. Since Twitter users are anyway not representative of the entire population, these trends in our data should not be interpreted as evidence that Russian annexation was overwhelmingly popular with the population, an "authentic" victory for national self-determination, or anything of the sort.

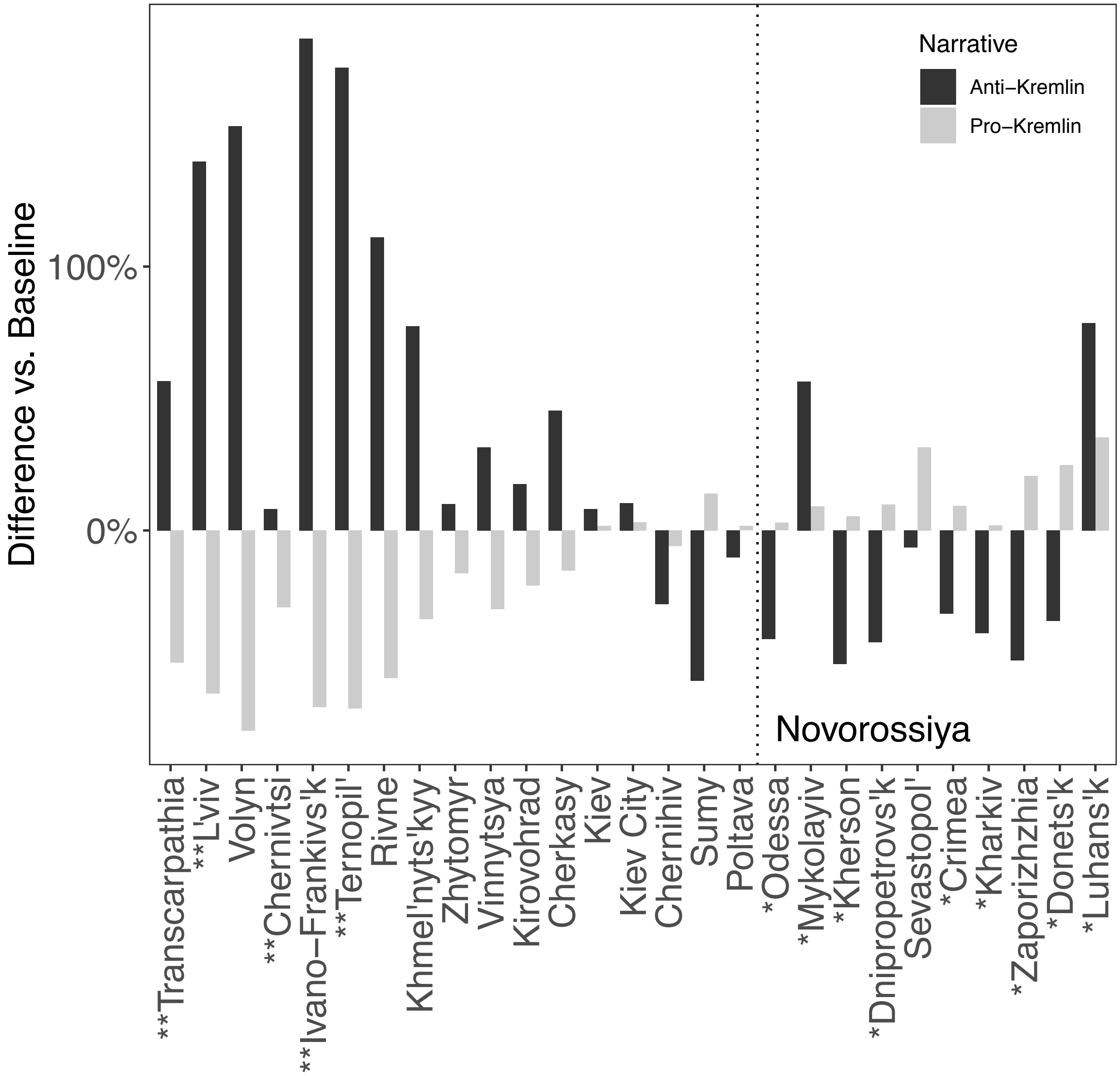
under-produce the focal point keywords in the pro-Kremlin dictionary and over-produce keywords from the anti-Kremlin dictionary. The opposite trend occurs in eight of the nine eastern-most oblasts, and also in Odessa. The most active “front lines” of the social media conflict were Mykolayiv and Luhans’k. The Spearman rank correlation between this relative production measure and the proportion of pro-Kremlin tweets by oblast is .62. **Figure 7** includes two paired visual maps of the data in **Figure 6,** with oblasts shaded corresponding to their relative narrative production.

FIGURE 5



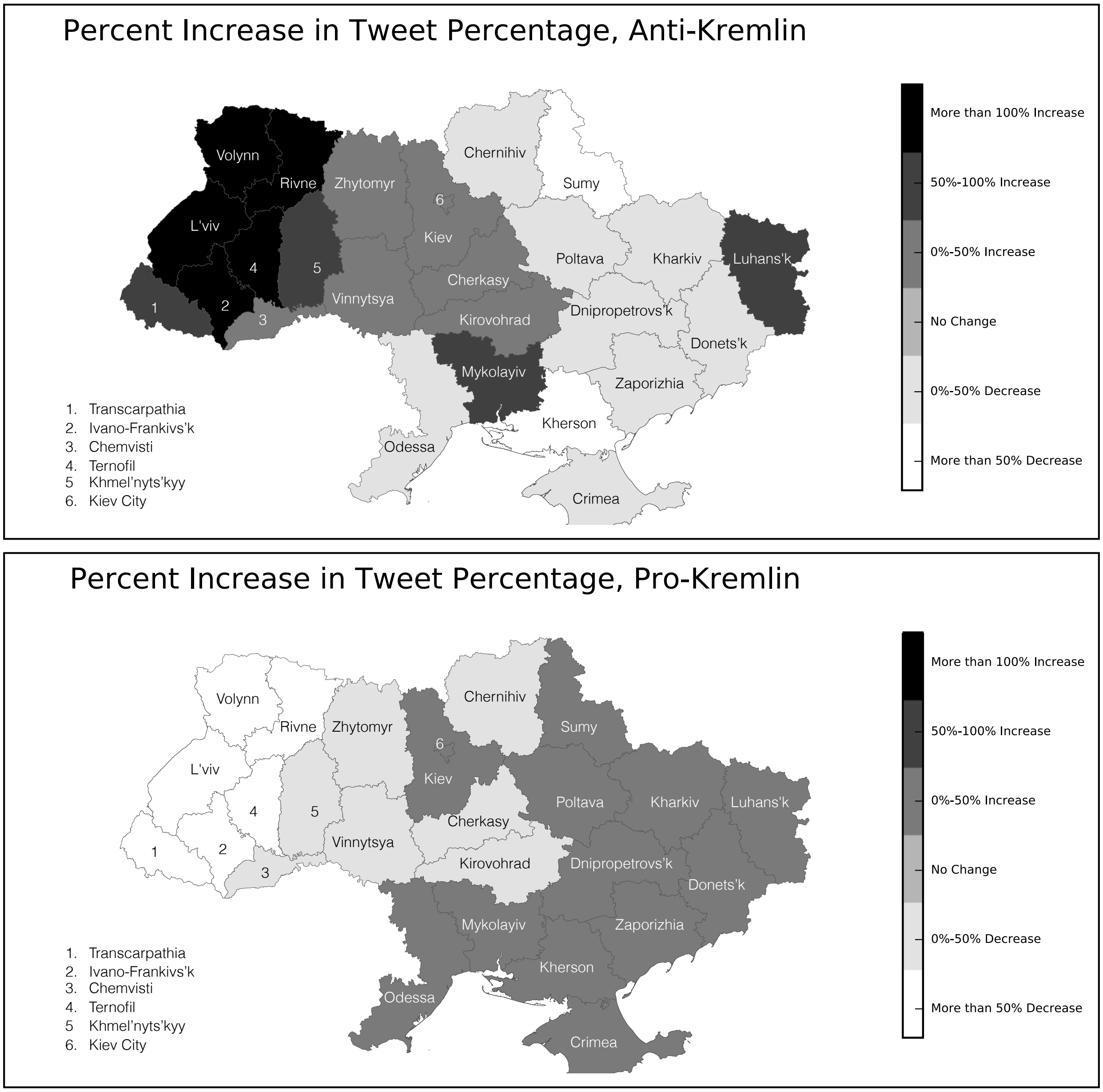
CAPTION: Relative over- or underproduction of narrative track keywords, by oblast, using the full sample of hand-coded data. Oblasts are arrayed roughly from west to east. The Pro- Kremlin outliers are Crimea and its capital city, Sevastopol. Over- production of Anti-Kremlin narratives takes place in the west, in the center Kyiv, and on the military frontline of the Donbas.

FIGURE 6



CAPTION: Relative over- or underproduction of narrative track keywords, by oblast, using the full sample of machine-coded tweets. Oblasts are arrayed roughly from west to east. Historical Novorossiya over-produces pro-Kremlin narratives and under-produces anti-Kremlin narratives. Mykolayiv, Kyiv, Sevastopol', and Luhans'k are notable as oblast where both

FIGURE 7



CAPTION: Relative over- or underproduction of narrative track keywords, by oblast, using the full sample of machine-coded tweets. To visualize, we bin the results: whether the oblast produced 50-100% fewer tweets than expected, 0-50% fewer, no change, 0-50% more, 50-100% more, or more than double the number of tweets expected. Darker colors indicate a relative surge in tweets containing target keywords relative to overall Twitter traffic in the district.

4b. Interpretation

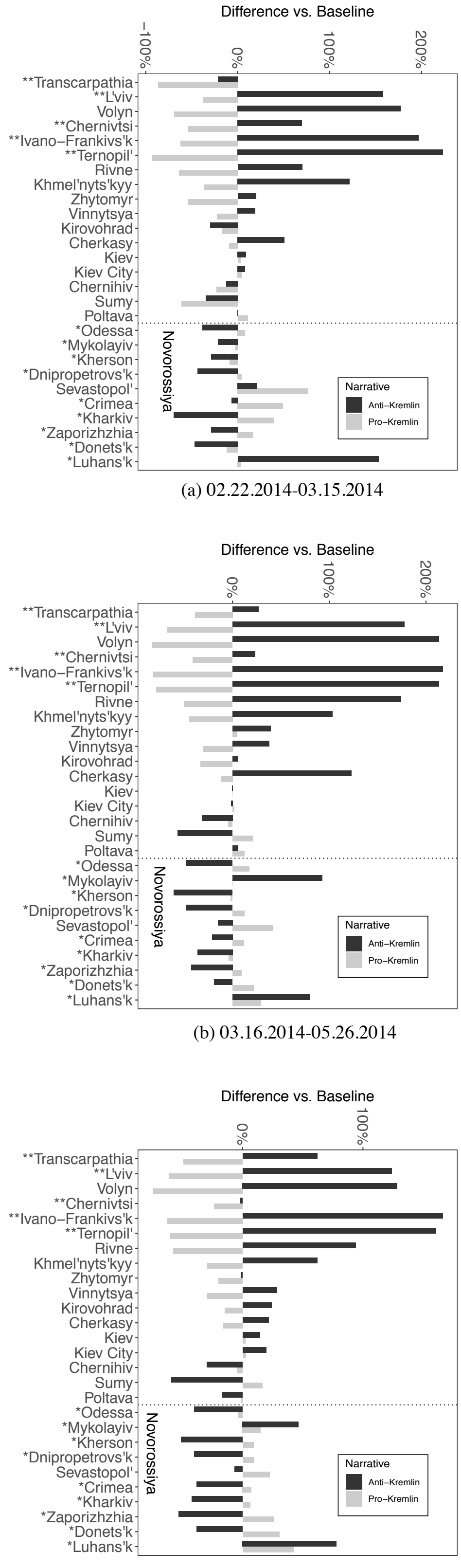
Our supposition is that spatiotemporal trends in these online social behaviors would have correlated with the offline social behaviors that would have been easily visible to civilians, journalists, or embedded observers reporting to Russian intelligence. The behaviors described spatially in **Figure 2**, and the main results in **Figure 3** and **Figure 4** reveal the *limits* of the Kremlin's capacity to compete with the West in soft power projection. The Kremlin's narrative of events seems to have found limited reception in the *Russkii Mir,* even in Russia's historical sphere of influence, even for a population historically sympathetic to its message (such as those living in parts of *Novoroissiya* that had reliably delivered votes to the Party of Regions*)*, even when Kremlin-influenced producers monopolized the airwaves (Peisakhin and Rozenas 2018). All factors suggested, *a priori*, hegemonic dominance of the pro-Kremlin narrative. *Ex-post* analysis of outcomes reveals a decidedly mixed picture.

These data suggest that a key point of failure for a Russian "social tip" towards wide-spread pro-Kremlin sedition against the post-Maidan Ukrainian political regime occurred in Luhans'k. Eventually this oblast emerged as the front line of conventional warfare. In **Figure 7,** while the patterns of production in Luhans'k are not strongly differentiated from the rest of *Novorossiya* in terms of *pro-Kremlin* production**,** this oblast was a focal point for *anti-Kremlin* social media activity. The conventional frontline was a front line in a war of ideas first: the influx of military activity brought volunteer journalists with Twitter accounts. This altered the sample to reflect a different set of social dynamics (and social media dynamics) than elsewhere.[18]

Figures 8A, **8B, and 8C** allow visual inspection of the 166,454 tweets in a rough time series. The progression in the variation in attitudes by Russian-speakers in Ukrainians oblasts emerges in these snapshots, as a Russian intelligence analyst might have seen them. Data from the first period, February 22 through March 15 (**Figure 8A)**, suggests there was support for the Russian narrative, and a relative dearth of anti-Kremlin pushback, in territories near Crimea. Occupied Sevastopol' was, to our surprise, a site of contestation according to these data. Kharkiv especially, but also Zaporizhzhia, Dnipropetrovs'k, and Odessa, might have been tempting targets for annexation in March and April. Between mid-March and late May (**Figure 8B**), there were a few of attempts by pro-Russia forces to engineer uprisings. The anti-Kremlin tweets in Mykolayiv in period 2 (**8B**) probably reflect sentiments by residents, after uprisings in Odessa, expressing fears that Russian planners might be tempted to create a land bridge linking Crimea to Transniestria. After the election of Petro Poroshenko, the consolidation of the post-Maidan Ukrainian state, and the intensification of artillery war (**Figure 8C**), Russian military intervention would have been more difficult. Outside of Mykolayiv and Luhans'k Russian-speaking social media users in Novorossiya continued to be relatively receptive to the Kremlin's point of view in this period. This interpretative exercise is not meant to be the final word on public sentiment by Russian-speaking communities in Novorossiya – just an exposition of how these tools could have been used by a computer-literate observer monitoring social media trends.

[18] Luhans'k is not the only oblast that over-produces both narratives. Mykolayiv, Kyiv, Kyiv City, and occupied Sevastopol' are also sites of contestation.

FIGURE 8 ARRAY



(a) 02.22.2014-03.15.2014

(b) 03.16.2014-05.26.2014

(c) 05.27.2014-08.28.2014

4c. Caveats and Speculations

Twitter was essentially an anti-Kremlin platform in 2014 during our study period everywhere in Ukraine except Crimea. Our supposition is that spatiotemporal trends in online behaviors, which we can measure from a distance, correlate with the offline signaling behaviors that would have been taking in Russian-speaking communities, but it is worth re-emphasizing that extrapolating wider trends in public opinion from these data is fraught. There may well have been a hidden density of pro-Kremlin Russian-speakers that decamped from Twitter and continued to communicate on Tor, in chat rooms, on forums, or using platforms beyond the scope of our analysis. That said, social media communications clearly can be used to estimate levels of support for seditious political attitudes such as secession. Since border revisions are very rare, and since social media is very new, inferential limitations should be made explicit.

Our decisions targeted the opinions of a single population (Russian-speaking Twitter users residing within Ukraine) during a complicated period of institutional collapse and state weakness (February-August 2014). Context-specific variables matter. We make no claim to external validity. Twitter is not a perfect substitute for representative public opinion sampling.[19] Different kinds of people do not use social media for different reasons. Even for those that enthusiastically “opt-in” to online politics, every 140-character tweet (or status on Facebook) is not an authentic political act. There is exciting behavioral work to be done, but until there is academic convergence on best practices for how to interpret spontaneous performances on social media, frontier-mapping exercises like ours should be treated with cautious care.

---

[19] Yet we are aware of no public polling during this period in Crimea, Luhans’k, or Donets'k. The shortcomings of social media should be weighed against the ability to conduct studies that danger would otherwise forbid.

Three assumptions must hold to justify online observation as measurement for offline behavior. First, it must be the case that individuals in our sample do not maintain a performance identity on social media that promulgates information contrary to offline beliefs. The same people should share the same ideas on Twitter as on other platforms – and, more importantly, as around kitchen tables or on soccer fields. This assumption is plausible but contestable.[20] Context-specific research is needed to sort extremist cheap talk on social media from sincerely-held extremist beliefs, especially for populations flirting with radicalization in active war zones.

Second, any study of community signaling behaviors tacitly assumes that the imagined audiences for tweets are local friend and family networks. Spatial comparisons of production patterns across oblasts may provide a window into public sentiment if density of production is related to latent characteristics of the communities Twitter users are trying to influence (and have private information on, as community members). This assumption is plausible, and consistent with academic understandings (McGee et all 2011), but more site-specific research is needed.

Third, the native population of an area must produce the bulk of the data in a sample. It would be a huge problem if the majority of data coming from non-community members such as journalists or mercenaries. Strategic efforts to create a false impression of local support through the use of bots or clandestine operatives (which is why we laboriously coded user characteristics) could also contaminate inferences. Manual inspection of the 5,328 tweets and 1,339 accounts convinced us that this study contains relatively few accounts originating from outside an oblast,

[20] See Hill et all (2016), Malik et al. (2015). Because behaviors on Twitter replicate known offline phenomena such as Dunbar's Number (Dunbar et al. 2015) and diurnal patterns of activity (Golder and Macy 2011) we are cautiously comfortable with this assumption. For a more thorough defense of measuring offline data with online sources, see Steinert-Threlkeld (2018).

but we admit caution on this point. A sophisticated information operation, if prepared years in advance, could foil visual inspection of the sort that we employ in this paper.

Weighed against these concerns are certain advantages of analyzing data from social media platforms. Unstructured data from populations that would be otherwise impossible to reach (in this case Crimea or behind the lines of control in the Donbas) can be analyzed. Unlike surveys, there is no attempt to claim population representativeness, so neither social desirability bias nor strategic non-response confound inferences. Unlike ethnographic observation, which is limited by the range of the researcher's own sensory equipment, research designs that employ social media data can compare patterns of production that occur at the same time in many places.

Perhaps the most salient objection to these results is that they are not novel. The East-West split has defined Ukrainian politics since independence (Arel 2002, Barrington and Herron 2004, Darden and Grzymala-Busse 2006, Clem and Craumer 2008, Constant 2011, Constant 2012, Frye 2015, Zhukov 2016). Using new social media data to draw costly maps that reproduce old maps (such as the second map in **Figure 7**) may be criticized as old wine in new bottles. There are three reasons not to dismiss this paper's methodology or results so quickly.

First, unlike a cross-sectional survey, these data mirror the series of updates that would have arrived, in real time, to Russian military personnel during a period of crisis bargaining. New information would have been at a premium for Kremlin policymakers. Maidan, the implosion of the Party of Regions, and Russia's seizure of Crimea were major events. Old understandings of public opinion would have been held up under close scrutiny. In that moment of crisis, no party, academic or military, would have had time to collect or analyze survey data.

Second, this research *has* generated new knowledge. The outline of Novorossiya is included in **Figure 2** as a reminder that old maps are not all useful guides to high-stakes behavior by Russian-speaking communities. There were many surprises among our research team as we conducted this study. The dominant narrative used by political elites in Kyiv describing this period is one of Russian agents sowing discord (which complements the dominant narrative in the United States is that Russia is an innovator in the information warfare domain). As such, we anticipated finding widespread geographic support for Russia (expressed in the "fascism" narrative) and extensive evidence of astroturfing (bots or dubious accounts reproducing Russian talking points). Neither appeared. Only 24 accounts, responsible for 196 tweets, were from bots. Outside of occupied Crimea, most Russian speakers did not use Twitter as a forum to voice support for Russia. The facts were surprising to our team but stubbornly clear.

Our supposition is that the failure of the pro-Kremlin narrative to catch on would have been an important source of military intelligence for Russian planners in 2014. Recall that having begun the process of redrawing the post-Soviet territorial map, it was not clear where Russia would define the natural end-point to its irredentism. Russian mechanized units could have moved quickly to establish facts on the ground if they had expected to find a population ready to greet them as liberators. The frontlines of Ukraine's conflict could easily be many kilometers further west. Some claim that Russia did not go further because its leadership feared international censure, but Russian diplomats could have easily justified the action, much as they justified Crimea, by invoking familiar "Responsibility to Protect" and "self-determination" arguments. That works if and only if many Russians call for help, however. The information that military planners needed for a more ambitious policy, but did not have, is whether they were likely to encounter resistance. If the Kremlin had access to data like ours, they would have known that they were unlikely to be greeted as liberators by many Russian-speaking communities – even in the Eastern Donbas. The Russian-Ukrainian interstate border moved only as far as Russian forces could advance while incurring no occupation costs – Crimea, and no further.

Third, the question of whether new kinds of technologies – in this case social media – enable irredentist mobilization is intrinsically worthy of study. If we are correct, social media has under-appreciated implications for revisionist powers trying to assess occupation costs prospectively. This is analytically separate from other well-analyzed applications of social media (e.g., lowering the costs of collective/connective action, lowering the costs for state actors surveillance of dissident networks, real-time source-checking of "fake news", etc.). When war weaponized radio and film, states had a comparative advantage in what might be called "memetic supply" (the production and dissemination of narrative embedded in memorable slogans, catchy songs, and viral images). Until the recent proliferation of inexpensive smartphones, states did not have the capability to reliably and systematically measure "memetic demand" in real-time. Our empirical results suggest that this capability probably already exists.

5. Conclusion

Social media behaviors are public signals analogous to scrawling graffiti, whistling a patriotic tune on a bus, talking loudly about politics in a public setting, or flying a flag. Since social media users add content to platforms in order to communicate ideas to their social network, the prevalence of overtly political behaviors can provide important clues about the political dispositions of the community that is the imagined audience for those messages.

We present no evidence supporting the claim that Russian military actions in 2014 were altered as a result of re-purposing social media trends for military intelligence – merely a variety of evidence consistent with our conjecture that such re-purposing is now possible. Social media data are straightforward to analyze systematically and can be collected at relatively low cost. Following Kostyuk and Zhukov (2017: 3), we favor the analogy between information warfare techniques and airplanes at the start of the First World War. Recall that planes were used primarily for reconnaissance before they were used to drop bombs. Conventional militaries are just beginning to explore the ways that emergent information technologies can shape battlefields. As techniques for real-time data mining become commodified, they will be integrated into best practices for counterinsurgency (Berman, Felter, and Shapiro 2018) and, more generally, into military planning. This paper has shown one way in which they could have been useful.

Bibliography

Baum, Matthew and Yuri M. Zhukov. 2015. "Filtering Revolution: Reporting Bias in International Newspaper Coverage of the Libyan Civil War." *Journal of Peace Research* 52 (3): 384-400.

Barrington, Lowell and Erik Herron, 2004. "One Ukraine or Many? Regionalism in Ukraine and its Political Consequences." *Nationalities Papers*, Volume 32, Number 1: 53-86.

Beissinger, Mark R. 2015. "Self-determination as a Technology of Imperialism: The Soviet and Russian Experiences." *Ethnopolitics* 14, no. 5: 479-487.

Beissinger, Mark. 2015b. "Russia's War in Ukraine and Putin's Propaganda State," *Comparative Politics Newsletter*. 25 (2): 26-29. https://bit.ly/2HLyCBl

Berman, Eli, Joseph H. Felter, and Jacob N. Shapiro. 2018. *Small Wars, Big Data: The Information Revolution in Modern Conflict*. Princeton University Press.

Bessi, Alessandro, and Emilio Ferrara. "Social bots distort the 2016 US Presidential election online discussion." *First Monday* 21, no. 11-7 (2016).

Boyd-Barrett, Oliver. 2017. "Ukraine, Mainstream Media and Conflict Propaganda," *Journalism Studies* 18 (8): 1016-1034.

Charap, Samuel and Timothy Colton. 2017. *Everyone Loses*. International Institute for Strategic Studies.

Chivvis, Christopher S. 2017. "Understanding Russian 'Hybrid Warfare' And What Can Be Done About It." Testimony Before The Committee on Armed Services, United States House of Representatives. *The RAND Corporation*. March 22, 2017.

Clem, Ralph S., and Peter R. Craumer. 2008. "Orange, Blue and White, and Blonde: The Electoral Geography of Ukraine's 2006 and 2007 Rada Elections." *Eurasian Geography and Economics* 49 (2): 127-151.

Colton, Timothy J. 2016. *Russia: What Everyone Needs to Know*. Oxford University Press.

Cottiero, Christina, Katherine Kucharski, Evgenia Olimpieva, and Robert W. Orttung. 2015. "War of Words: The Impact of Russian State Television on the Russian Internet." *Nationalities Papers* 43 (4): 1–23.

Constant, Amelie. Martin F. Kahanec, and Klaus F. Zimmerman. 2011. "The Russian-Ukrainian Political Divide." *Eastern European Economics* 49 (6): 97-109.

Constant, Amelie. Martin F. Kahanec, and Klaus F. Zimmerman. 2012. "The Russian-Ukrainian Earnings Divide." *Economics of Transition* 20 (1): 1-35.

Darczewska, Jolanta. 2014. *The Anatomy of Russian Information Warfare: The Crimean Operation, A Case Study*. Warsaw: Center for Eastern Studies (OSW).

Darden, Keith, and Anna Grzymala-Busse. 2006. "The Great Divide: Literacy, Nationalism, and the Communist Collapse." *World Politics* 59(01): 83–115.

Dunbar, R.I.M., Valerio Arnaboldi, Marco Conti, and Andrea Passarella. 2015. "The Structure of Online Social Networks Mirrors Those in the Offline World." *Social Networks* 43: 39–47.

Faizullaev, Alisher and Jérémie Cornut. 2017. "Narrative Practice in International Politics and Diplomacy: The Case of the Crimean Crisis," *Journal of International Relations and Development* 20: 578-604.

Frye, Timothy. 2015. "What Do Voters in Ukraine Want?: A Survey Experiment on Candidate Ethnicity, Language, and Policy Orientation." *Problems of Post-Communism* 62, (5): 247-257.

Gartzke, Erik. 2013. "The Myth of Cyberwar: Bringing War in Cyberspace Back Down To Earth." *International Security*. 38 (2). 41-73.

Gentile, Michael. 2015. "West oriented in the East-oriented Donbas: a political stratigraphy of geopolitical identity in Luhansk, Ukraine." *Post-Soviet Affairs* 31(3):201-223.

Gerasimov, Valery. 2013. Translated as: "The Value of Science Is in the Foresight: New Challenges Demand Rethinking the Forms and Methods of Carrying out Combat Operations". *Military Review*, January-February (February 27) 2016.

Giles, Kier. 2016. "Russia's 'New' Tools For Confronting The West: Continuity and Innovation in Moscow's Exercise of Power." *Chatam House Research Paper*. The Royal Institute of International Affairs. 1-69.

Golder, Scott a, and Michael W Macy. 2011. "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures." *Science* 333(6051): 1878–81.

Gramsci, Antonio. 1933. *Selections from the Prison Notebooks of Antonio Gramsci*: Ed. and Transl. by Quintin Hoare and Geoffrey Nowell Smith. eds. Geoffrey Nowell-Smith and Quintin Hoare. International Publishers. First Printing 1971, Sixth Printing 1980.

Grimmer, J., and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*. 21(3): 267–97.

Gruzd, Anatoliy and Ksenia Tsyganova. 2015. "Information Wars and Online Activism During The 2013/2014 Crisis in Ukraine: Examining the Social Structures of Pro- and Anti-Maidan

Groups," *Policy and Internet*. 7(2): 121-158.

Hale, Henry, Oxana Shevel and Olga Onuch. 2018. "Believing Facts in the Fog of War: Identity, Media and Hot Cognition in Ukraine's 2014 Odesa Tragedy." *Geopolitics*, DOI: 10.1080/14650045.2018.1465044

Herman, Edward S., and Noam Chomsky. 2010. *Manufacturing Consent: The Political Economy of the Mass Media*. Random House.

Hill, Shawndra, Han Zhang, and David Rothschild. 2016. *Geolocated Twitter Panels to Study the Impact of Events*. In *2016 AAAI Spring Symposium Series*. 2016.

Hopf, Ted. 2016. "'Crimea is Ours': A Discursive History," *International Relations*. 30 (2): 227-55.

Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War*. Cambridge University Press, 2006.

Kelly, John, Vladimir Barash, Karina Alexanyan, Bruce Etling, Robert Faris, Urs Gasser, and John Palfrey. 2012. "Mapping Russian Twitter." *Berkman Center for Internet & Society, 2012-2013* : 1-16.

Kostyuk, Nadiya and Yuri Zhukov. 2017. "Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events?" *Journal of Conflict Resolution*, November 10, 2017. https://doi.org/10.1177/0022002717737138

Laitin, David. 1998. *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Cornell University Press.

Laruelle, Marlène. 2015. "The Three Colors of Novorossiya, or the Russian Nationalist Mythmaking of the Ukrainian Crisis," *Post-Soviet Affairs*. 32 (1): 55-74.

Lucas, Edward and Peter Pomerantsev. 2016. *Winning the Information War: Techniques and Counter-Strategies to Russian Propaganda in Central and Eastern Europe*. Center for European Policy Analysis (CEPA) Information Warfare Project, Legatum Institute.

Lustick, I. 1993. *Unsettled States, Disputed Land: Britain and Ireland, France and Algeria, Israel and the West Bank-Gaza*. Ithaca: Cornell University Press.

Malik, Momin M, Hemank Lamba, Constantine Nakos, and Jurgen Pfeffer. 2015. "Population Bias in Geotagged Tweets." *9th International AAAI Conference on Weblogs and Social Media*, 18–27.

McGee, Jeffrey, James Caverlee, and Zhiyuan Cheng. 2011. “A Geographic Study of Tie Strength in Social Media.” *International Conference on Information and Knowledge Management,* Glasgow, 2333–36.

McFaul, Michael. 2018. *From Cold War to Hot Peace*. Houghton Mifflin Harcourt.

Metzger, Megan, Jonathan Nagler, and Joshua A. Tucker. 2015. “Tweeting Identity? Ukrainian, Russian, and #Euromaidan.” *Journal of Comparative Economics* 44(1): 16-40.

Metzger, Megan and Joshua A. Tucker. 2017. "Social Media and EuroMaidan: A Review Essay," *Slavic Review*, 76 (1). 169-191.

Onuch, Olga. 2015. "EuroMaidan Protests in Ukraine: Social Media Versus Social Networks." *Problems of Post-Communism*, 62(4):217-235.

Peisakhin, Leonid and Arturas Rozenas. "Electoral Efforts of Biased Media: Russian Television in Ukraine." *American Journal of Political Science*. 2018 (forthcoming).

Petersen, Roger D. 2001. *Resistance and Rebellion: Lessons from Eastern Europe*. Cambridge University Press.

Pomerantsev, Peter. 2015. “Unplugging Putin TV,” *Foreign Affairs*.

Ramakrishnan, Naren et al. 2014. “‘Beating the News’ with EMBERS: Forecasting Civil Unrest Using Open Source Indicators.” In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, , 1799–1808.

Reisinger, Heidi and Aleksandr Golts, "Russia's Hybrid Warfate: Waging War Below The Radar of Traditional Collective Defense." Research Paper. *Research Division, NATO Defense College*, Rome. No. 105. November 2014.

Ritter, Alan, Oren Etzioni, and Sam Clark. 2012. “Open Domain Event Extraction from Twitter.” *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’12*: 1104–12. http://dl.acm.org/citation.cfm?doid=2339530.2339704.

Romanets, Maryna. 2017. “Virtual Warfare: Masculinity, Sexuality, and Propaganda in the Russo-Ukrainian War,” *East/West: Journal of Ukrainian Studies*. 4 (1): 159-177.

Scholz, Fabian. 2016. “When Journalists and Readers Diverge: Contestation to Der Spiegel’s Framing of the Russia-Ukraine Crisis,” International Journal of Media and Cultural Politics 12 (2), 249-256.

Schwartz, H. a., and L. H. Ungar. 2015. “Data-Driven Content Analysis of Social Media: A Systematic Overview of Automated Methods.” *The ANNALS of the American Academy of*

Political and Social Science. 659(1): 78–94. http://ann.sagepub.com/cgi/doi/10.1177/0002716215569197 (April 11, 2015).

Spaiser, Viktoria, Thomas Chadefaux, Karsten Donnay, Fabian Russman & Dirk Helbing. 2017. "Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests." *Journal of Information Technology & Politics*. (14)2:132-152.

Snyder, Timothy. 2018. “Truth or Lies (2015),” in *The Road to Unfreedom: Russia, Europe, America*, New York: Tim Duggan Books. Chapter 5, 159-215.

Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. Cambridge: Cambridge University Press.

Suslov, Mikhail D. 2015. “‘Crimea Is Ours!’ Russian Popular Geopolitics in the New Media Age,” *Eurasian Geography and Economics*. 55 (6): 588-609.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. "Online human-bot interactions: Detection, estimation, and characterization." In *Eleventh International AAAI Conference on Web and Social Media*.

Warren, T. Camber. 2014. "Not by the Sword Alone: Soft Power, Mass Media, and the Production of State Sovereignty." *International Organization* 68(1):111-141.

Wilson, Steven. L. 2017. "Detecting Mass Protest through Social Media." *The Journal of Social Media in Society*, *6*(2), 5-25.

Zhukov, Yuri. 2016. “Trading Hard Hats for Combat Helmets: The Economics of Rebellion in Eastern Ukraine,” *Journal of Comparative Economics*. 44 (1): 1-15.

Post-Soviet Affairs: Supplementary Materials

1. Hand-Coding the Tweets

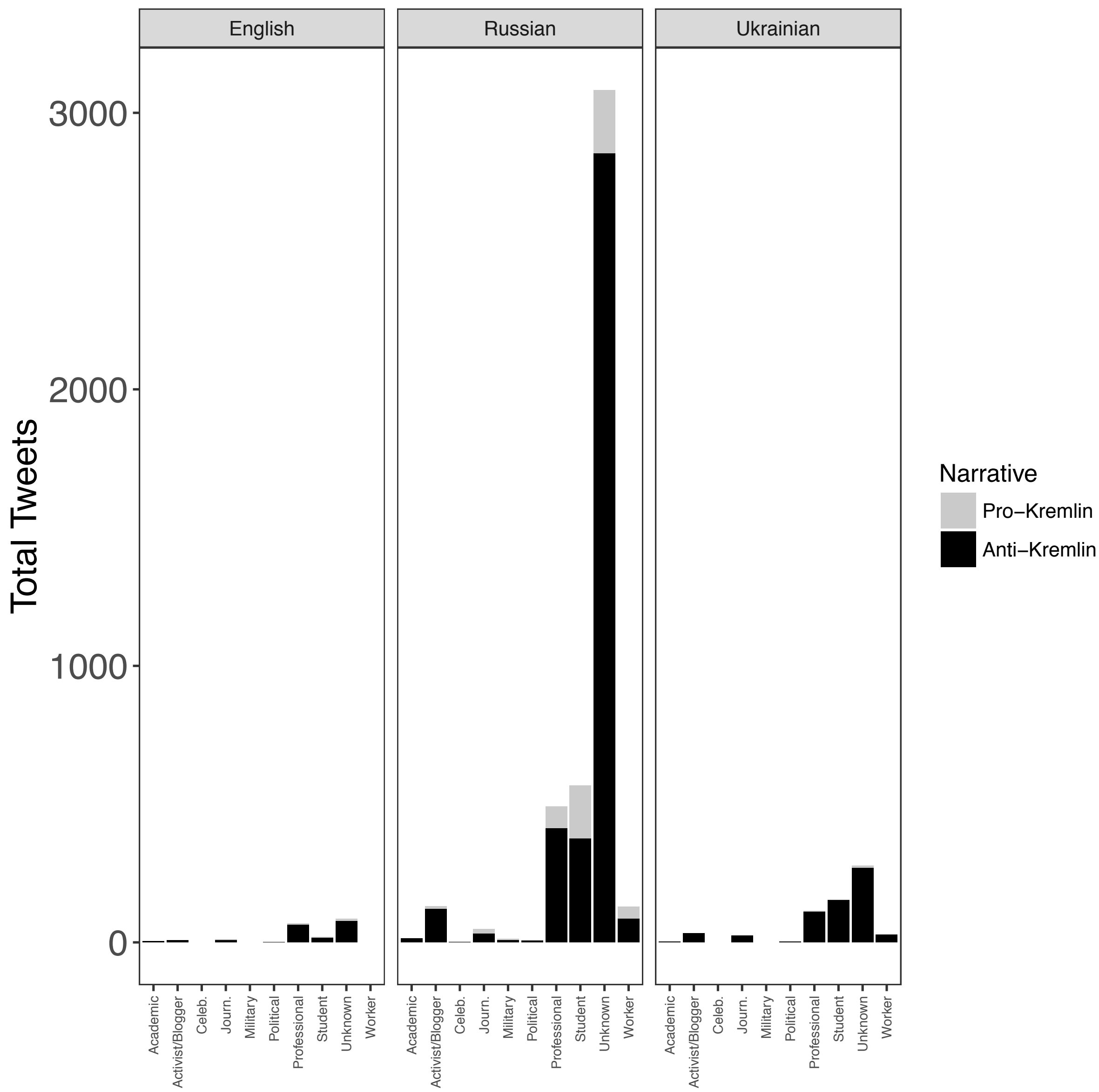
We manually coded each of the 1,535 accounts which authored at least one geotagged tweet in our sample. A Kyiv-based team of four used a combination of public searches on Google, Facebook, and *VKontakte*. As explained in the main paper, the initial purpose of this exercise was to ensure that we were collecting information that would be meaningful evidence of social attitudes within our population, rather than artifacts of sophisticated efforts to use social media to "plant evidence" of attitudes as part of the information warfare campaign. We did not record names of users out of respect to human subjects, but coded respondents on profession, sex, age, and primary language. As **Table SM-1.1** shows, the sample is dominated by people whose ages are indeterminate from publicly searchable online data. Other information was easily discernable. Automated accounts were particularly easy to identify – conditional on someone taking the time to search the user profile, "bots" are distinguishable. Contrary to our prior beliefs based on knowledge of Russia's information warfare strategy, few bots (196) enter our sample. Also surprising was that bot accounts were not all pro-Kremlin. As can be seen in the summary statistics **Table SM-1.5**, bot accounts were somewhat more prevalent in Kiev City (which we expected), Khmel'nyts'kyy, Poltava, Zaporizhzhia and Dnipropetrovs'k (which we did not expect, though in most of these cases the high percentage is due to the low denominator). Our going theory is that providing GPS coordinates to a tweet was, and is, not considered important for the bot's effectiveness, eliminating them mechanically by our geotagging filter.

TABLE SM-1.1: SOCIAL MEDIA BEHAVIORS BY IDENTITY

	Individuals	Tweets
Academic	7	20
Blogger/Activist	41	172
Celebrity	3	3
Journalist	49	85
Military	5	14
Political	7	12
Professional	303	675
Student	529	743
Unknown	332	3446
Worker	63	158
Total	1,535	5,328

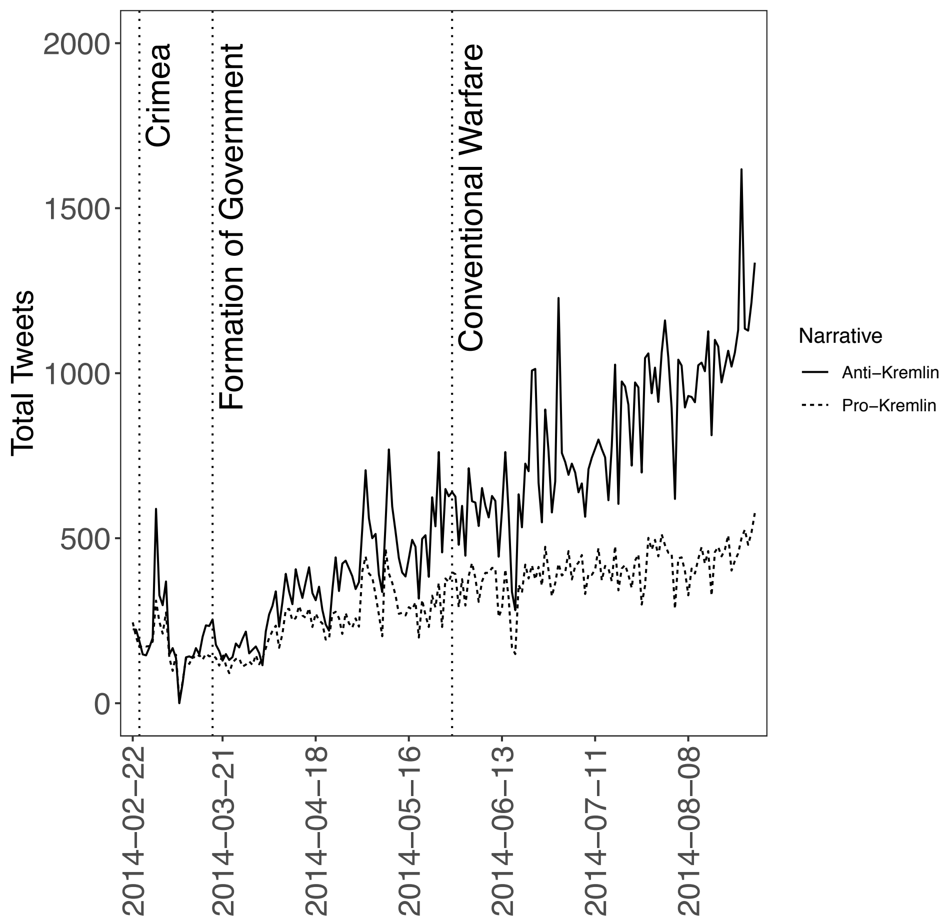
If Russian intelligence had been interested in populations' receptiveness to irredentism, they would presumably have eliminated bots as a source of noise. In order to paint a picture most consistent with what they might have seen, we drop all bots from the descriptive statistics in the paper (**Figure 3** and **Figure 5**) and in these supplementary materials. Importantly, we found that bots have low values for median followers. Since automated accounts do not have many followers, whatever they are saying cannot possibly be seen by many people. We opted to include tweets originating in automated accounts when we later built our ML classifier.

FIGURE SM-1



CAPTION: Most, but not all, tweets in our sample were in Russian. A tweet originating in the territory of Ukraine in English or Ukrainian that also contained a keyword was almost guaranteed to be identified using the anti-Kremlin keyword dictionary. The dominance of the pro-West narrative, in spite of the larger pro-Kremlin selection dictionary, is very clear in this particular visualization of the data.

FIGURE SM-2



CAPTION: This figure is identical to Figure 4 except using data from all of Ukraine rather than just Novorossiya. The Anti-Kremlin narrative more clearly dominates the full sample.

Demographic characteristics of the sample and summaries of how characteristics map onto the two narratives are presented in **Figure SM-1, Table SM1.1** and **Table SM1.2**. Languages other than Russian seem to have been pre-filtered by users to signal anti-Russian attitudes. Our filters searched exclusively for Russian words, but some returned tweets written primarily in English or Ukrainian. Only 1.72% of Ukrainian tweets in our sample were pro-Kremlin, compared to 8.54% of English tweets and 12.17% of Russian-language tweets in our sample. In the course of cleaning the data we eliminated approximately 300 garbled or incoherent messages, messages written primarily in non-study languages (Romanian, Spanish), and separated automated accounts (bots), which explains the slightly lower number of tweets in tables below. For 34.39% of the accounts (528/1535), it was not possible to identify the user's profession, but, again, our team was confident the account belonged to a person, not a bot. These "unknown" individuals also often produced many different tweets and thus comprised the bulk of the data, as **Figure SM-1** makes clear. An even more dramatic visualization that **Figure 3** in the main paper of just how dominant the anti-Russian narrative is in these overall data. The same story is clear in **Figure SM-2**, which replicates the main paper result in **Figure 4** but for the entire territory of Ukraine.

The three groups that tweet the most are students, professionals, and unidentified accounts. **Figure SM1** displays tweet production as it varies across language, profession, and narrative track. Though the majority of the tweets were anti-Kremlin, certain professional groups, at least when tweeting in Russian, were systematically more likely than other groups to voice opinions consistent with the pro-Kremlin narrative. The three groups with the greatest concentration of pro-Kremlin narrative were the military (35.71%), celebrities (33.33%), and workers (27.22%). The three groups with the greatest concentration of anti-Kremlin narrative were activists and bloggers (94.18%), politicians (91.67%), and unknown (92.95%), with professionals not far behind at 87.41%. Our favored interpretation is that young professionals and students engaged with state-backed actors engaged in psychological operations. Some celebrities chased the story, seeking controversy. Exploring these suppositions would require access to the entire data, rather than just the 1% sample made available to academic researchers.

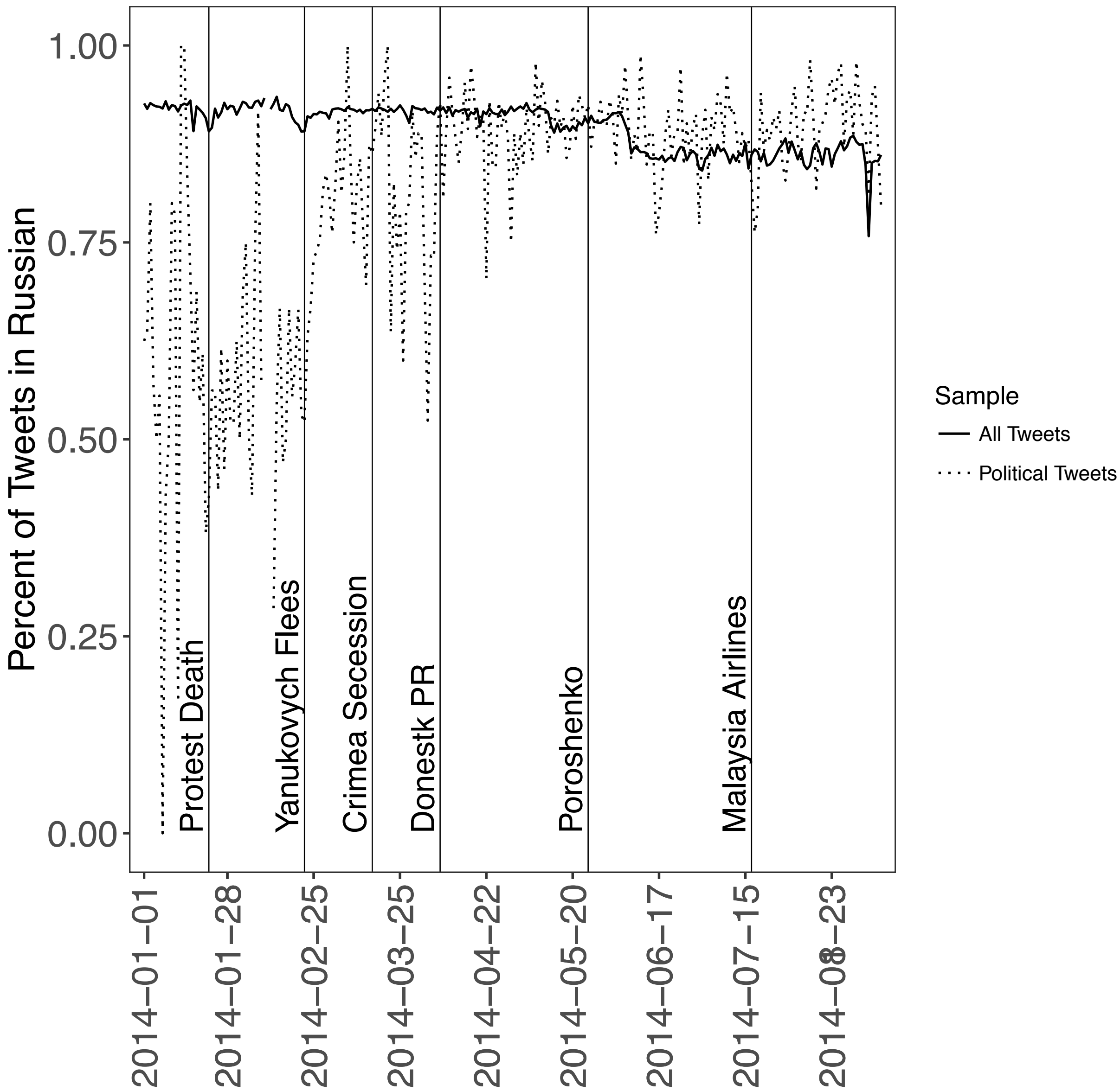
TABLE SM-1.2: SOCIAL MEDIA BEHAVIORS BY LANGUAGE, AGE GROUP

By Language			
	Pro-Kremlin	**Anti-Kremlin**	**Total**
English	17	182	199
Russian	575	3916	4725
Ukrainian	11	627	638
Total	603	4725	5,328
By Age			
0-15	45	99	144
16-24	247	786	1033
25-36	114	1058	1172
37-60	57	293	350
61+	2	29	31
Unknown	138	2460	2598
Total	603	4725	5,328

The sample skews young: only 31 tweets are from individuals over the age of 60 and the bulk of tweets are from users aged 16-36. Very few accounts (7) belong to self-identified government officials or state bureaus. Only 49 belong to news organizations. Twenty- four accounts, producing 196 tweets, were automated accounts (bots). Though it was a source of discussion between coauthors, we ultimately opted to keep these few accounts in our regression models of human-coded tweets (reported below) and in the training set we used to code the larger dataset to avoid "throwing away data," once we confirmed (a) excluding bots did not alter core results (see regression robustness checks below), and (b) that only our human coders could convincingly discern intent to overcome the irony confound (discussed below).

The most important previous work on the use of Twitter in Ukraine for our study at the time of initial analysis was a pre-publication version of Metzger et al. (2015). This paper analyzed behaviors during the same period and reported users were likely to tweet politically in Russian, even if their default language on Twitter was Ukrainian, after Crimea. Their study used self-reported language of origin as the decision criteria for whether to include tweets in the study rather than geotagging. As discussed in the main text, we had some residual anxiety about whether a sample of only geotagged Tweets might somehow be non-representative. In order to assuage these concerns among ourselves, we replicated one of their core results using our geotagged database. **Figure SM-3** extends the periods of observation back a full month before the Maidan protests turned violent. Our hand-coded evidence provides important confirmation of Metzger et al. (2015)'s theory: In our data, the switch to Russian occurred among politically-activated individuals engaging in online political contention, a phenomenon distinct from the rest of the online communication occurring on Twitter (such as social mobilization for Maidan, sharing music videos, etc.). Before they began rebutting claims embedded in Russia's irredentist information warfare campaign, many were communicating in Ukrainian on Twitter. **Figure SM-3** suggests to us that the Ukrainian Twitter-sphere switched to Russian to tell competing stories.

FIGURE SM-3

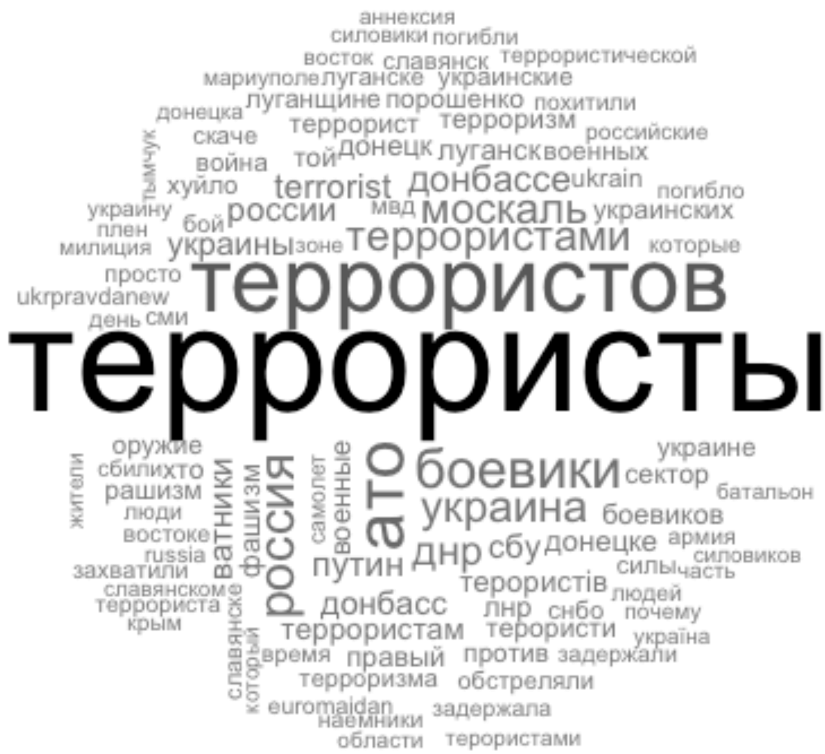


CAPTION: The frequency of tweeting in Russian increases after Viktor Yanukovych flees the country. The solid line is all accounts in our sample; the dotted line, for only those accounts that engage in pro-Kremlin or anti-Kremlin narrative construction. While accounts generating tweets that use politically-charged keywords clearly switch into Russian after the Maidan events, a similar change is not seen in the full sample

Figure SM-4 shows the 100 most common words based on the content of initial tweets selected by the dictionary sorting. The results are shown in word clouds scaled by size to reflect prevalence in the overall hand-coded sample. Russian speakers will be able to quickly parse the two polarized narratives. The anti-Kremlin cloud (above) has "Terrorists" as the most prevalent word. The pro-Kremlin cloud below) has "Right" (a reference to the *Right Sektor*, ultra-nationalist right-wing party that played a key role in the Maidan events) as the most prevalent word. Careful readers will also that, even though it is not one of our keywords, "terrorists" is quite prominent in the pro-Kremlin word cloud, as well.

Though the data passed a basic face validity test, we struggled with user intent – and in particular with what we call amongst ourselves the *irony confound*. Consider the layers of irony embedded the phrase: "This is a moment that you do not want to be seen celebrating, because people celebrating victory over fascism are dangerous, as fascists." Assuming that references to "fascist" imply a pro-Kremlin bias would completely misread user intent. Consensus on how to sort signal from noise in the shadow of the irony confound has not yet emerged. Analysis of large quantities of social media data usually depends on keyword string searches and machine-learning algorithms. These methods do not reliably identify irony, double-entendres, or sarcasm.

FIGURE SM-4


аннексия
силовики погибли
восток славянск террористической
мариуполе луганске украинские
луганщине порошенко похитили
донецка террорист терроризм
скаче донецк луганск военных российские
война той
хуйло terrorist донбассе ukrain погибло
украину бой россии мвд москаль украинских
плен
милиция украины зоне террористами которые
просто
террористов
ukrpravdanew
день сми
террористы
оружие украине
сбили хто боевики сектор
рашизм ато батальон
люди украина боевиков
востоке армия
russia днр сбу донецке силовиков
захватили ватники фашизм россия самолет военные путин силы часть
славянском терористів людей
террориста донбасс лнр снбо почему
крым терористам терористи україна
время правый против задержали
терроризма обстреляли
euromaidan задержала
наемники
области терористами


террористам страны
сми нацизм убивают новороссия
euromaidan antiterrorista националист
terrorista день называют славянск
война радикалы наемники майдан
odikov
вообще terrorist украине нужно
донбассе
novostidnua нацисты террористами войны
россия
dbnmjr этих
город люди украина фашизма
ещё
террористы одесса
стране
ukrpravdanew
города
вчера que которые той ukrain слова
жить
говорят
правый
армии нашей
новости сектор
своих
фашизм
сепаратисты
крыма
kiev
майдане
какие людей
одессе нам боевики
хотят
сделали террористов просто
стреляют крым националисты москаль
говорит
фашизмом националистов мирные мирных майдана
ватники жителей украины днр kadirovrussia
терроризма пиздец
правыйсектор донецк
фашистский
оружие операción жители
duringstorm
украинцев

Before proceeding to analysis of the full data, therefore, we checked to see whether a different picture of social opinions among Russian-speaking Ukrainians would have emerged if we employed fluent Russian-speakers to try to parse tweets for user intent. This second step found that 983 tweets – about one out of six – used the dictionary language ironically. Common sources of miscoding requiring human correction arose from ironic usage, anti-Kremlin Ukrainians describing Putin as a fascist, expressions of outrage at police tactics at Maidan, and (rarely) sympathetic descriptions of the *Pravyi Sektor* movement. Tables **SM-1.3** and **SM-1.4** show the results of manual re-coding. These manual codings are employed in all of the data visualizations in the paper that reference hand-coded data.

TABLE SM-1.3: PRO-KREMLIN TWEETS AND IRONY

		Irony		
		No	Yes	
Pro-Kremlin	No	3255	588	3843
	Yes	1090	395	1485
		4345	983	5,328

TABLE SM-1.4: ANTI-KREMLIN TWEETS AND IRONY

		Irony		
		No	Yes	
Anti-Kremlin	No	722	339	1061
	Yes	3623	644	4267
		4345	983	5,328

Table SM-1.5: Summary Statistics For All Hand-Coded Tweets

	Oblast	Tweets	Russian	English	Ukrainian	Anti-Kremlin	Pro-Kremlin	Users	Account Age	Median Followers	% Male	% Bots	Tweets per 100k
Not Novorossiya	Transcarpathia	19	14	1	4	14	5	15	754	266	72.22	0	1.51
	L'viv	149	73	1	75	119	28	80	828	202	50.91	5.71	5.88
	Volyn	15	5	0	10	13	1	14	654	156	71.43	0	1.44
	Chernivtsi	22	10	0	12	17	2	15	605	160.5	9.52	0	2.42
	Ivano-Frankivs'k	28	10	3	15	21	7	22	675	74.5	63.64	0	2.02
	Ternopil'	34	6	2	26	33	1	15	1343	2667	70.97	0	3.19
	Rivne	36	16	1	19	30	8	12	707.5	62	73.53	0	3.1
	Khmel'nyts'kyy	28	24	0	4	19	2	17	725	42	50	35.71	2.16
	Zhytomyr	17	10	2	5	10	5	12	374	177	33.33	0	1.36
	Vinnytsya	52	30	4	18	31	16	29	734	72.5	51.28	0	3.24
	Kirovohrad	427	411	3	13	376	80	17	1619	1464	23.53	0.47	43.81
	Cherkasy	237	208	4	25	177	87	33	697	406	87.5	0.88	19.02
	Kiev	254	199	2	53	195	60	104	409.5	74.5	62.05	28.63	14.67
	Kiev City	838	596	89	153	580	233	406	1019	98	51.65	2.05	28.89
	Chernihiv	38	31	0	7	24	9	18	942	537.5	50	0	3.63
	Sumy	15	12	0	3	7	6	12	895	58	41.67	7.14	1.35
	Poltava	61	53	1	7	35	24	31	864	136	52.63	18.52	4.23
	Odessa	190	168	12	10	112	82	97	854	175	65	0	7.96
Novorossiya	Mykolayiv	191	181	1	9	159	37	42	1584	1298	39.06	0.54	16.47
	Kherson	52	45	2	5	34	20	26	909.5	202.5	45.24	0	4.89
	Dnipropetrovs'k	384	289	36	59	238	128	181	742	116.5	42.75	11.64	11.78
	Sevastopol'	39	39	0	0	8	26	24	1267	189	58.82	0	10.22
	Crimea	151	140	5	6	56	67	70	1140	775	61.82	1.41	7.69
	Kharkiv	364	305	5	54	255	122	114	920	475.5	50.6	1.13	13.38
	Zaporizhzhia	142	124	4	14	78	52	67	596.5	97.5	41.05	17.97	8.09
	Donets'k	1212	1178	10	24	1006	292	157	1602	1374	54.81	0.6	27.62
	Luhans'k	333	314	11	8	292	85	27	1615	1451	46.51	0.6	14.71

NB: Percent Male and Percent Bots are for tweets, not accounts. Account Age is the median age, in days, of a Twitter account.

Table SM-1.6: Summary Statistics For All Machine-Coded Tweets

	Oblast	Tweets	Russian	English	Ukrainian	Anti-Kremlin	Pro-Kremlin	Users	Account Age	Median Followers	Tweets per 100k
Not Novorossiya	Transcarpathia	17359	8281	1662	4447	966	126	1281	671	88	1378.25
	L'viv	146749	64119	8936	61324	11115	836	5427	602	232	5787.83
	Volyn	35921	13375	944	18432	2984	133	962	519	117	3444.49
	Chernivtsi	52493	35336	1695	10842	2109	536	1171	486	144	5768.46
	Ivano-Frankivs'k	30068	10317	2502	14108	2724	149	1566	550	108	2174.55
	Ternopil'	18410	6605	1294	8836	1713	94	994	745	99	1726.17
	Rivne	50580	23712	1669	21191	3202	336	1298	473	97	4352.66
	Khmel'nyts'kyy	40133	22641	1360	13177	2193	412	1439	376	109	3096.44
	Zhytomyr	31443	21761	1236	6374	1130	400	1582	373	53	2517
	Vinnytsya	109695	66663	3599	31330	4565	1198	2509	414	73	6837.69
	Kirovohrad	69183	44682	11971	4307	2470	851	1858	410	115	7097.7
	Cherkasy	105786	70223	3497	24402	4960	1388	2950	554	99	8488.92
	Kiev	473579	366492	19399	51476	16008	7165	14881	638	101	27348.06
	Kiev City	522716	376911	32293	63578	25090	7356	18989	822	86	18018.97
	Chernihiv	66051	51396	1853	7282	1722	983	1832	485	65	6308.46
	Sumy	28221	23642	989	1129	437	499	879	389	64	2530.92
	Poltava	67679	51458	2212	8438	2079	1076	2167	535	83	4697.7
	Odessa	356523	289294	17853	11639	9250	5512	8593	592	92	14934.26
Novorossiya	Mykolayiv	89091	76025	1624	5033	4289	1504	2361	497	88	7682.68
	Kherson	90998	77878	2359	3491	2857	1449	2473	489	85	8554.03
	Dnipropetrovs'k	644973	545817	16837	31571	18541	10193	10099	572	77	19792.31
	Sevastopol'	40863	35063	1139	1303	1629	777	2324	908	68	10705.95
	Crimea	181050	149008	9052	5271	5453	3018	6755	564	116	9219.51
	Kharkiv	241548	190845	15555	11962	6285	3694	5916	593	85	8879.32
	Zaporizhzhia	172294	147333	4436	7122	3384	3212	4676	553	77	9813.61
	Donets'k	266715	230295	9887	5251	5599	5084	5439	539	73	6078.69
	Luhans'k	34127	28204	2154	812	2022	708	1250	470	63	1507.59

NB: Account Age is the median age, in days, of a Twitter account. Tweets column is all tweets, not just political ones.

2. Machine-Coding the Tweets & Curating The Sample

Building each model followed the same process. First, we removed stopwords and tokenized the remaining ones in each tweet. Second, we made a training set from 80% of the tweets. Third, we made a term frequency-inverse document frequency matrix for the training and test tweets. Fourth, we generated bagged estimators for each narrative: a support vector machine, Bernoulli Naïve Bayes, Gaussian Naïve Bayes, and a multinomial Naïve Bayes. "Bagged" means that for each classifier, we generated it on *k* random subsets of the training data, generating *k* predictions for each tweet; the predictions were averaged, ensuring that results were not driven by a specific part of the parameter space. We varied the number of features (variables) each classifier could have; whether or not a variable could consist of 1, 2, or 3 words (it is an n-gram); and how many bags to use for each classifier. For each combination of these parameters, we recorded the classifier's precision, recall, accuracy, and F1 score.[1]

To determine which classifier to use for which narrative, we chose a combination of parameters to maximize F1 or precision.[2] For the pro-Russia tweets, we chose the classifier with the highest F1. For the pro-Ukraine tweets, however, we chose the model with the highest precision, as models with high F1 scores tended to have too many false positives for our comfort. The pro-Russia classifier is Bernoulli Naïve Bayes with 45 bags, an n-gram of 1,800 features,

---

[1] For an explanation of these steps and metrics, see Grimmer and Stewart (2013) and Lucas et al. (2015).

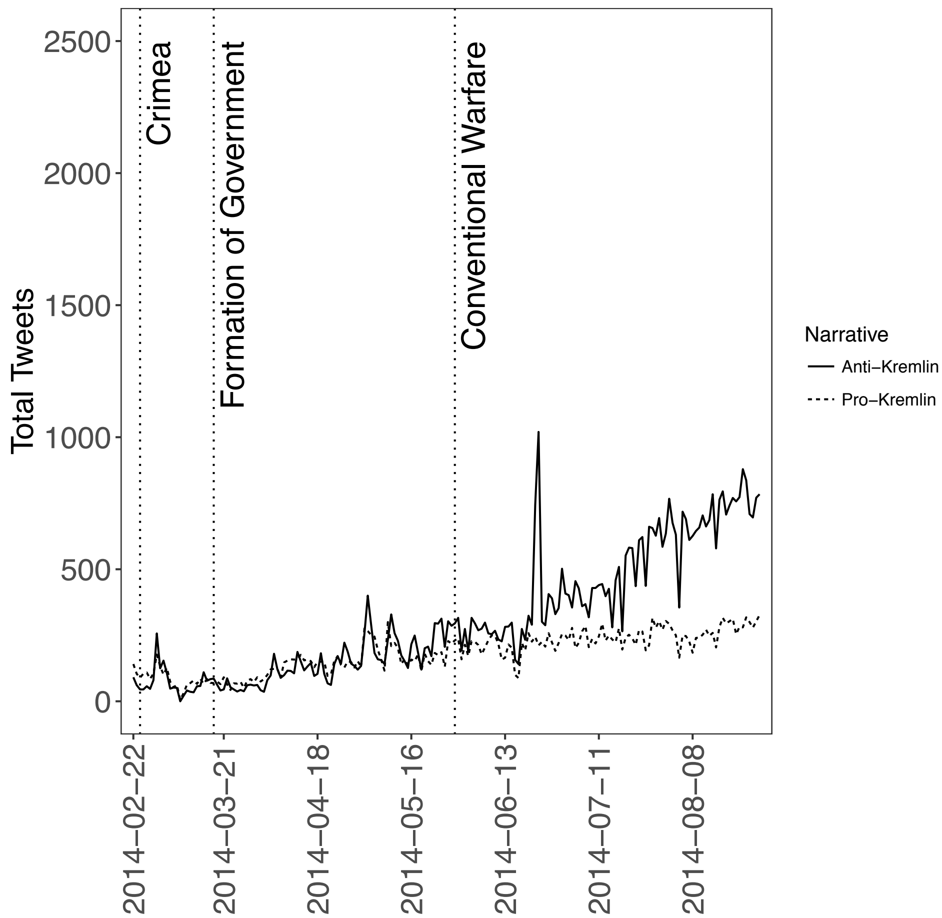
[2] Precision here means the percent of all tweets from the test set the classifier labels as "Pro-Russian" or "Pro-Ukraine" that actually are. *Recall* is the percent of all tweets in the test set that were manually labeled as "Pro-Russia" or "Pro-Ukraine" that the classifier correctly labels. F1 is the weighted combination of the two.

precision of .62, and an F1 score of .58. The pro-Ukraine classifier is Gaussian Naïve Bayes with 10 bags, an n-gram of 1, 50 features, a precision of 1, and an F1 score of .56.

One interesting and unexpected result of this process, completely unrelated to this paper's theory, is that the language used for pro-Kremlin narrative appears to be simpler than the anti-Kremlin language. The pro-Kremlin classifier has more features (words) than the anti-Kremlin one (800 to 50) and the pro-Russia tweets constitute a greater percentage of the 204,189 machine-coded tweets than they do of the dictionary-coded ones. This increase is also notable because we initially seeded the dictionary with more pro-Kremlin words. Not until training a classifier to recognize co-occurrences with dictionary words did we recover the volume of pro-Russia tweets expected. During the referee process we discovered another confound in the data: tweets originating from the *Foursquare* app. *Foursquare* is a mobile app where users indicate they are in specific places and are connected with nearby app users. If the user has not disabled certain settings, and has connected their Twitter account, when they check-in to a location tweets are sent. Since no *Foursquare* tweets had appeared in the hand-coded data, this confound was not discovered until the referee process, but the *Foursquare* relaunch of its app during our study period seems to have unexpectedly yielded certain days with "tweet dumps."

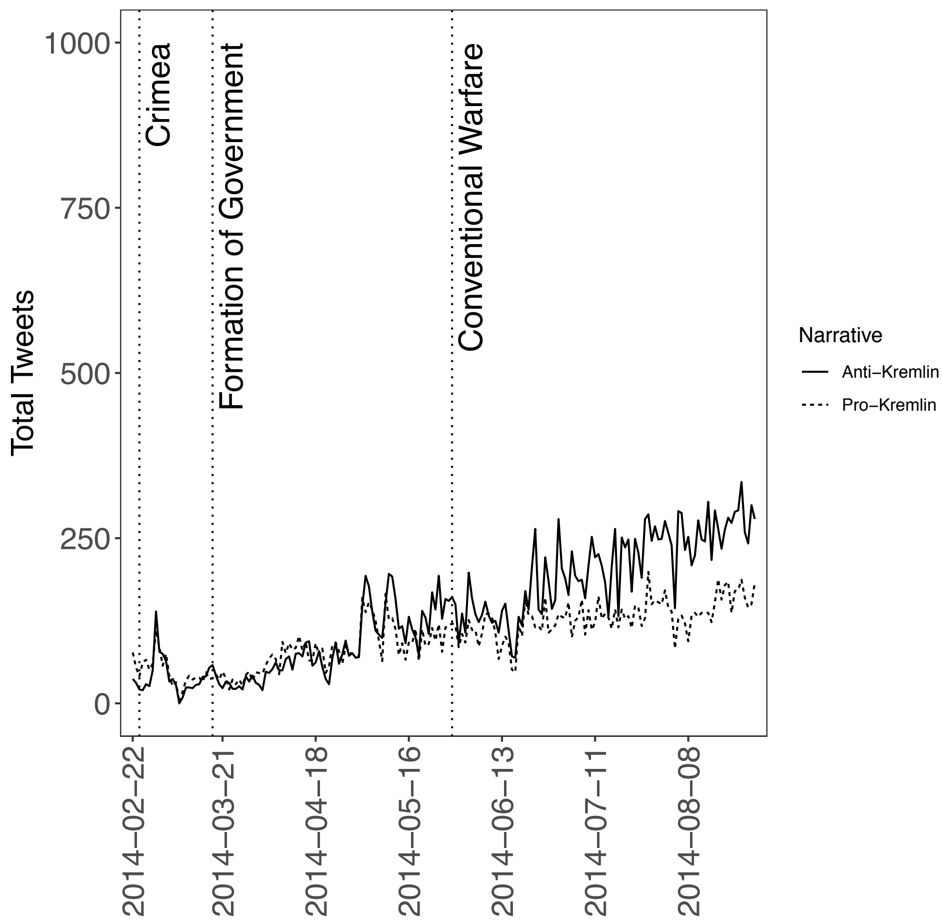
We described our methods for assuring results were not drive by bots in the main text, but readers may also be curious to know how these robustness checks altered the sample. We therefore reproduce **Figure 4** from the main text in a few variants (recall that this is a visualization of the time trends of the two narratives in the larger dataset, for the subsample residing in historical *Novorossiya,* the site of an anticipated uprising). **Figure SM-5, SM-6,** and **SM-7** replicate the form of this figure in order to demonstrate the negligible substantive effect of removing the *Foursquare* tweets or suspected bots or influencers from the data.

Figure SM-5



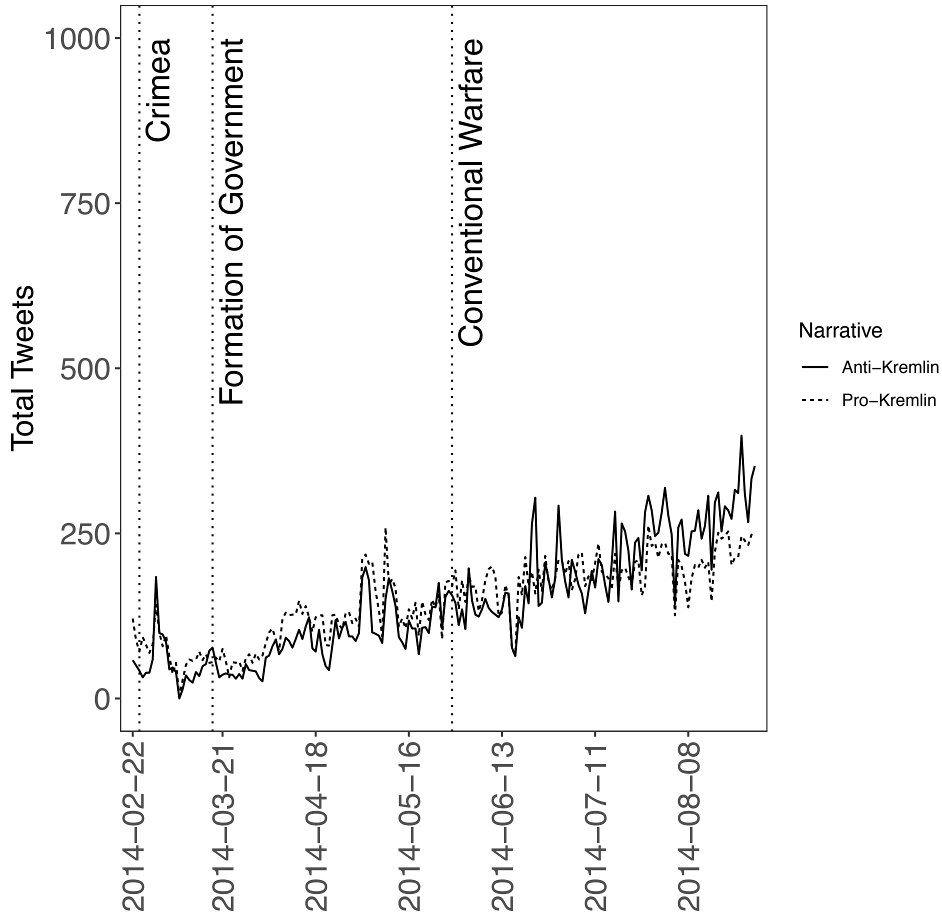
CAPTION: The subset of the raw data from the machine-learning dataset (N=204,189), replicating Figure 4, using only data generated from oblasts in historical Novorossiya on the full sample. Note the huge spike between June 13 and July 11. This was not due to any particular offline event: only due to a surge in Foursquare account activity. This was our first clue that it would be important to systematically eliminate these accounts. Luckily, the fact that all such tweets open with the phrase "I'm at [place]" made the identification and elimination of this confound easy. The N of our sample shrunk from 204,189 to 166,454 as a result of dropping all Foursquare tweets.

Figure SM-6



CAPTION: This figure reproduces Figure 4 except removing tweets from accounts that the Botometer service determines are likely to be from bots. The Anti-Kremlin narrative more strongly dominates once these tweets are removed.

Figure SM-7



This figure is the same as Figure 4, except with tweets removed if an account in the top 5 percent of the tweet distribution produced them. In this specification the pro-Kremlin narrative seems to have a slight upper-hand in Novorossiya until quite late in the conventional warfare phase, though the two narratives track together. The analytic take-away (for us) is that results can depend a great deal on filtering assumptions (and anyway does not indict the core conjecture of the paper, which is that outside of Crimea the pro-Russia narrative did not have a decisive upper-hand in Russian-speaking communities). We have performed the same analysis on all tweets from Ukraine but could discern no meaningful difference from ***Figure SM-2****, so we do not show that figure.*

3. Distribution of Tweets by Oblast

To determine if tweets were well-distributed across users and oblasts, we normalized the number of accounts per oblast by population. We normalized both the hand-coded accounts and the accounts identified using the topic model. We also report the mean and median number of tweets per person. **Table SM-5.1** shows these results. The correlation between the number of hand-coded accounts and the density of those accounts is .9069; for the topic model accounts, .827. The distribution of the account densities, regardless of identification method, follows a log-normal distribution (histograms provided upon request). The distribution of average number of tweets follows a normal distribution. Note that each oblast has a much higher mean than median number of tweets per user, suggesting that there are some users who are simply more active than others. This a common and well-documented feature of social networks.

TABLE SM-5.1

Oblast	Pop. (Millions)	Accounts (Hand)	Density (Hand)	Accounts (NLP)	Density (NLP)	Mean Tweets (NLP)	Median Tweets (NLP)
Cherkasy	1.25	33	26.48	867	695.73	7.34	2
Chernihiv	1.05	18	17.19	494	471.81	5.48	2
Chernivtsi	0.91	15	16.48	354	389.01	7.49	2.5
Crimea	1.96	70	35.65	1729	880.45	4.93	2
Dnipropetrovs'k	3.26	181	55.54	3568	1094.91	8.08	3

Donets'k	4.39	157	35.78	1667	379.93	6.45	2
Ivano-Frankivs'k	1.38	22	15.91	437	316.04	6.58	2
Kharkiv	2.72	114	41.91	1630	599.19	6.15	2
Kherson	1.06	26	24.44	675	634.52	6.39	2
Khmel'nyts'kyy	1.3	17	13.12	407	314.02	6.4	2
Kiev	1.73	104	60.06	4044	2335.31	5.75	2
Kiev City	2.9	406	139.96	5281	1820.46	6.17	2
Kirovohrad	0.97	17	17.44	457	468.85	7.28	2
Luhans'k	2.26	27	11.93	421	185.98	6.54	2
Lviv	2.54	80	31.55	1489	587.27	8.04	2
Mykolayiv	1.16	42	36.22	618	532.93	9.39	2
Odessa	2.39	97	40.63	2321	972.24	6.38	2
Poltava	1.44	31	21.52	563	390.79	5.63	2
Rivne	1.16	12	10.33	409	351.96	8.67	2
Sevastopol'	0.38	24	62.88	557	1459.32	4.34	2
Sumy	1.12	12	10.76	220	197.3	4.26	2

Ternopil'	1.07	15	14.06	259	242.85	6.98	2
Vinnytsya	1.6	29	18.08	801	499.29	7.22	2
Volyn	1.04	14	13.42	337	323.15	9.26	2
Zakarpattia	1.26	15	11.91	267	211.99	4.1	1
Zaporizhzhia	1.76	67	38.16	1356	772.36	4.89	2
Zhytomyr	1.25	12	9.61	406	325	3.77	2

4. Suggestive Mechanism Evidence: A War of Position (Co-Production of Narratives)

Visual inspection of time trends in the hand-coded and machine-coded data suggests that both anti-Kremlin and pro-Kremlin narratives often peak on the same day. Throughout our study, the inference we draw from this trend, reinforced by discussions with the coders, is that partisans on both sides were commenting on the same events in the media cycle, on the same day, reading and responding to each other's commentary, mutually raising the political temperature. We cannot test this directly because of the structure of our data, of course, but multivariate statistical models are a straightforward method see whether social media behaviors receptive to Russia's narrative and behaviors that are oppositional correlated temporally and spatially. This is an imperfect test of whether polarized Russian-speaking communities were "shouting" at each other on social media about the same online events.

We model the production of narratives by country-day, oblast-day, and oblast-week using the coded tweets from the two classifiers. **Table SM-4.1** presents the results from five models. In all models, the outcome variable is the number of anti-Kremlin tweets. All models include numerous temporal and socioeconomic controls. Model 1 takes the country-day as the unit of analysis. Models 2 and 3 aggregate to the oblast-week, and Models 4 and 5 employ the oblast-day as the unit of analysis. Since we have sufficient data to analyze oblast-days, Model 5 is our preferred model. In all models, each discourse is contemporaneous with the other with very small p-values. The only consistently significant socioeconomic variable is rural population, which positively correlates with the production of anti-Kremlin tweets. The large sample size means that some other control variables are statistically significant in some models but not others, but not in a way that lends itself to easy interpretation or theorization (which is anyway beyond the scope of our ambition).

Also notable is that while anti-Kremlin behavior is contemporaneous with its pro-Kremlin counterpart, previous pro-Kremlin content has no correlation with future anti-Kremlin content. We view this as cautious evidence that users were competing in real-time. Past anti-Kremlin tweets positively correlate with future anti-Kremlin tweets in all model specifications.

Table SM-4.2 revisits trends in the smaller, hand-coded dataset. While reading carefully for false positives, we had our team code the production of *ironic keyword use* -- when a user employs words from one narrative's keyword dictionary, but is clearly doing so in order to draw readers attention, then expose the absurdity of the entire line of argument. In common-use parlance, the online behavior of interest is "trolling." In these two models, the outcome is the ironic use of the anti-Kremlin discourse (e.g., using anti-Kremlin words in a way that is pro-Kremlin). Models are estimated using a negative binomial model. These models show that ironic anti-Kremlin and non-ironic anti-Kremlin narratives tend to co-occur, regardless of whether the unit of analysis is oblast-day or oblast-week.[3] Suffice to say that a variety of additional model specifications are possible (e.g., using only the hand-coded dataset, the hand-coded dataset but treating bots differently, oblast fixed effects, etc.), but since there is no logical end-point to this sort of a-theoretical fishing expedition, rather than sprawl this Supplementary Materials needlessly, we invite future scholars to explore the replication data themselves.

---

[3] We kept the bots in this analysis, since automated accounts are presumably also responding to offline events through unmolded processes.

TABLE SM-4.1: PRO-KREMLIN AND ANTI-KREMLIN NARRATIVES OCCUR IN THE SAME OBLASTS AT THE SAME TIME

	DV: Count of Anti-Kremlin Tweets						
	Country-day	Oblast-Week	Oblast-Week	Oblast-Day	Oblast-Day	Oblast-Day	Oblast-Day
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Pro-Kremlin)$_t$	1.448***	.677***	.386***	.607***	.175***	.288***	.604***
	(.058)	(.028)	(.033)	(.014)	(.014)	(.015)	(.013)
Log(Anti-Kremlin)$_{t-1}$	.535***						
	(.064)						
Log(Pro-Kremlin)$_{t-1}$	-.682***						
	(.110)						
Log(Anti-Kremlin)$_{i,t-1}$			.845***		.739***	.606***	
			(.020)		(.010)	(.012)	
Log(Pro-Kremlin)$_{i,t-1}$			-.285***		.0002	.124***	
			(.034)		(.014)	(.016)	
Avg. Monthly Wage$_i$		.00005	-.00001	.0001***	.00002	-.001***	.0003***
		(.0001)	(.00004)	(.00003)	(.00002)	(.0001)	(.00004)
Urban Pop.$_i$		-.0001	0.00000	-.00002	-.00001	.00003*	-.00003**
		(.00004)	(.00002)	(.00002)	(.00001)	(.00002)	(.00002)
Rural Pop.$_i$		.001***	.0001***	.001***	.0002***	-.001***	.001***
		(.0001)	(.00004)	(.00003)	(.00002)	(.0001)	(.00004)
Internal Net Migration Rate$_i$		.033***	.006	.030***	.007**	-.070***	-.009
		(.012)	(.006)	(.005)	(.003)	(.008)	(.006)
International Net Migration Rate$_i$		-.132***	-.012	-.083***	-.024***	-.071***	.079***
		(.031)	(.016)	(.012)	(.008)	(.020)	(.021)
Unemployment Rate$_i$		.012	.007	.040***	.008	-.359***	-.099***
		(.037)	(.019)	(.015)	(.010)	(.047)	(.021)
Russian Speakers (%)							-.006***
							(.001)
Intercept	-.596***	.476**	.123	-.075	.007	2.948***	-.123
	(.130)	(.233)	(.121)	(.093)	(.061)	(.262)	(.092)
Fixed Effect	N	N	N	N	N	Ntv. R. Spkng.	N
Observations	187	702	676	4,737	4,711	4,711	4,737
Adjusted R^2	.922	.610	.897	.495	.779	.794	.504
Note:	*p<0.1; **p<0.05; ***p<0.01						

TABLE SM-4.2: EARNEST PRO-KREMLIN & IRONIC "PRO-KREMLIN" TROLLING OCCURS IN THE SAME OBLASTS AT THE SAME TIME

	DV: Count of Ironic Anti-Kremlin Tweets	
	Country-day	Oblast-Week
	(1)	(2)
Pro-Kremlin$_{t}$	.042**	.031
	(.021)	(.029)
Anti-Kremlin$_{t}$	.015***	.038***
	(.003)	(.005)
Pro-Kremlin$_{t-1}$	-.001	.021
	(.022)	(.028)
Anti-Kremlin$_{t-1}$	.003	-.025***
	(.003)	(.006)
Ironic Anti-Kremlin$_{t-1}$	.007	.208***
	(.016)	(.018)
Intercept	.913***	-.273***
	(.094)	(.065)
Observations	185	501
Log Likelihood	-422.059	-760.241
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Negative binomial model	