PONARS Eurasia
NEW APPROACHES TO RESEARCH AND SECURITY IN EURASIA

The Use of Twitter Bots in Russian Political Communication

Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker[1]
New York University

The question of how regimes respond to online opposition has become a central topic of politics in the digital age. In previous work, we have identified three main options for regimes: offline response, online restriction of access to content, and online engagement. In order to better understand the latter category—efforts to shape the online conversation—in Russia, we study the political use of Russian language Twitter bots using a large collection of tweets about Russian politics from 2014–2018. To do so, we have developed machine learning methods to both identify whether a Twitter account is likely to be a bot—that is, an account that automatically produces content based on algorithmic guidelines—as well as the political orientation of that bot in the Russian context (pro-Kremlin, pro-opposition, pro-Kyiv, or neutral).

We find that despite public discussion that has largely focused on the actions of pro-Kremlin bots, the other three categories are also quite active. Interestingly, we find that pro-Kremlin bots are slightly younger than either pro-opposition or pro-Kyiv bots, and that they were more active than the other types of bots during the period of high Russian involvement in the Ukrainian crisis in 2014. We also characterize the activity of these bots, finding that all of the political bots are much more likely to retweet content produced by other accounts than the neutral bots. However, neutral bots are more likely to produce tweets that have identical content to those produced by other bots. Finally, we use network analysis to illustrate that the sources of retweets from Russian political bots are mass media and active Twitter users whose political leanings correspond to bots' political orientation. This provides additional evidence in support of the claim that bots are mostly used as amplifiers for political messages.

Methodology

To study the political use of Russian language Twitter bots in Russia, we rely on a large collection of tweets about Russian politics from 2014–2018. We collected over 36 million tweets (short messages) posted by over 1.88 million Russian users on Twitter[2] that contain at least one of 86 keywords that we put together to capture major aspects of Russian politics.

We use this dataset to accomplish two sequential tasks. First, we develop a new methodology for detecting Twitter bots that allows us to perform both real-time and retrospective analysis.[3] Second, we develop another new methodology to identify the political orientation of detected bots.[4] Both methodologies rely on a variety of modern machine-learning methods and belong to the overarching class of supervised approaches, i.e., they require "ground truth" as an input and use patterns available in the provided "ground truth" to label the rest of the data.

In our analysis, the "ground truth" (also called the *labeled set*) encompasses over 1,000 Twitter accounts labeled as bots (or not) by human coders; and another set of almost 2,000 Twitter bots labeled as pro-Kremlin, neutral, pro-opposition, or pro-Kyiv. We define pro-Kremlin bots as automated accounts that actively advocate the Kremlin political agenda; pro-Kyiv bots are those that feature critical tweets about the Russian involvement in the crisis in Crimea and Eastern Ukraine only, whereas pro-opposition bots by definition produce anti-Kremlin tweets on a broader range of topics including both domestic and international dimensions. Neutral bots are defined as a residual category for any account that was not coded as belonging to one of the other three categories.

We hired a large group of native Russian-speakers who study social sciences in a major Russian university to perform the coding. Each account was labeled by at least five coders, and we used in the labeled set only those Twitter accounts that had high values of inter-coder reliability (over 0.77).[5]

---

[2] Since most Twitter users are anonymous, there is no way to know for sure that a given user is Russian. However, the meta-data that comes with tweets via Twitter API shows the language of the author's Twitter account. We define an account to be a Russian one if at least 75 percent of all its political tweets that we are able to collect are posted while the account has the Russian-language interface.

[3] Other popular methods of bot detection do not allow for retrospective, and therefore replicable, classification. For more details see Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker, "Detecting Political Bots on Russian Twitter," *Big Data*, Vol. 5, No. 4, 2017, pp. 310–324.

[4] For details, see Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker, "For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia," Unpublished manuscript, forthcoming at *Sage Open* in 2018.

[5] A detailed account of the coding procedure is available in: Stukal et al. 2017 and Sergey Sanovich, Denis Stukal, and Joshua A. Tucker, "Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia," *Comparative Politics*, Vol. 50, No. 3, 2018, pp. 435–482.

Our bot-detection methodology uses over four dozens of accounts' tweeting characteristics (the devise used for tweeting, the entropy of time intervals between consecutive tweets, proportion of hashtags, etc.) and other meta-data (followers-to-friends ratio, the use of default Twitter account photo, etc.). In case of orientation detection, we use over 30,000 textual features the include the most common words and word pairs, hashtags, and URL domains included in tweets.

Results

Having put together the labeled set and estimated the machine-learning model for bot-detection, we applied it to our collection of over 1.88 million Russian Twitter users and found that over 450,000 accounts were predicted to be bots. However, many of these predicted bots have only a few tweets in our collection, which makes it hard to predict their political orientation. In order to avoid any substantive analysis based on estimates from these accounts, we restrict our further analysis to the 120,000 predicted bots that either had over 10 tweets after 2015 or had over 100 tweets before 2016. As a robustness check, we also impose an additional restriction on the quality of our model performance and focus on almost 93,000 bots for which we were had a prediction of political orientation with a probability over 0.9.

Who Dominates the Political Twittersphere in Russia?

Although the dominant mass media narrative focuses on "Putin's bots," our analysis shows that the political landscape of Russian-language Twittersphere is very diverse. We do find tens of thousands of pro-Kremlin bots. However, as **Figure 1** reveals, these are far from the only type of Twitter bots in Russia. What we label as "neutral" bots actually make up the modal category (although, if we restrict ourselves to a smaller subset of long-living accounts with at least 10 tweets each year in 2015-2017, pro-Kremlin bots are the most frequent category, followed by neutral, pro-opposition, and pro-Kyiv bots). These are a very heterogeneous group of bots that include accounts that automatically tweet news headlines, or whose political orientation is not overtly pro- or anti-Kremlin.

**Figure 1. Distribution of Predicted Political Orientation**
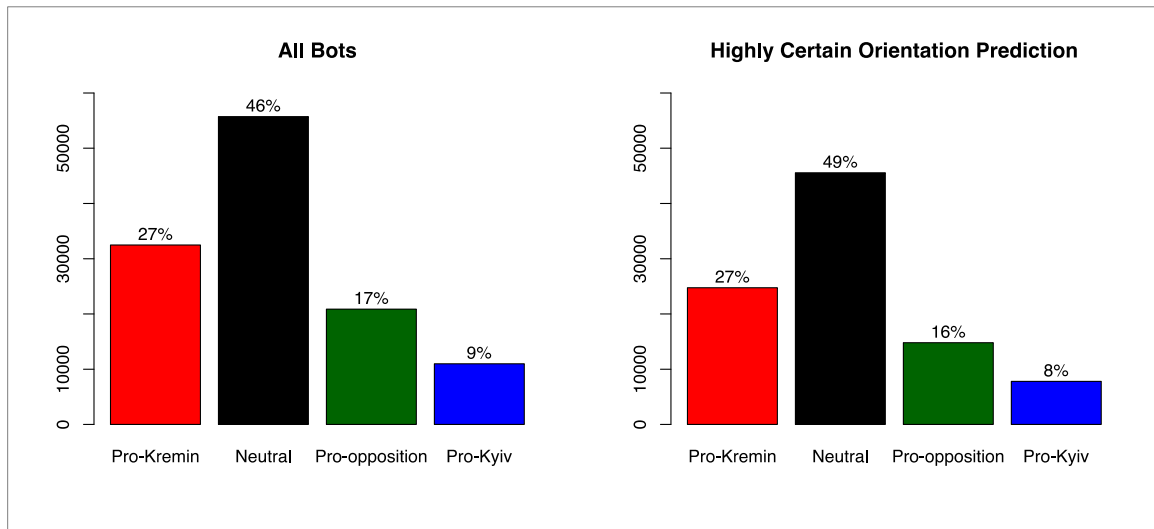
Bar heights represent absolute numbers of detected bots with different political orientations. Percentages are computed out of all 120,000 bots under consideration (on the left panel), and out of 93,000 bots for which we were able to predict the correct political orientation with probability over 0.9 (on the right panel).

Figure 2 provides a more nuanced representation of bot distribution by plotting orientation proportions over time on a monthly basis. We observe that neutral accounts were the most frequent and accounted for nearly half of all bots up until late 2017, when there was an increase in the proportion of pro-Kremlin and pro-opposition bots.

Overall, as **Figure 2** demonstrates, the proportion of pro-opposition bots has been growing and gradually narrowing the gap with pro-Kremlin ones since late 2016. In the meantime, there has been little change in the proportion of pro-Kyiv bots, except for a spike in late 2014, during one of the most heated periods of the military crisis in eastern Ukraine.

Figure 2. Proportion of Bots with Different Political Orientation Over Time
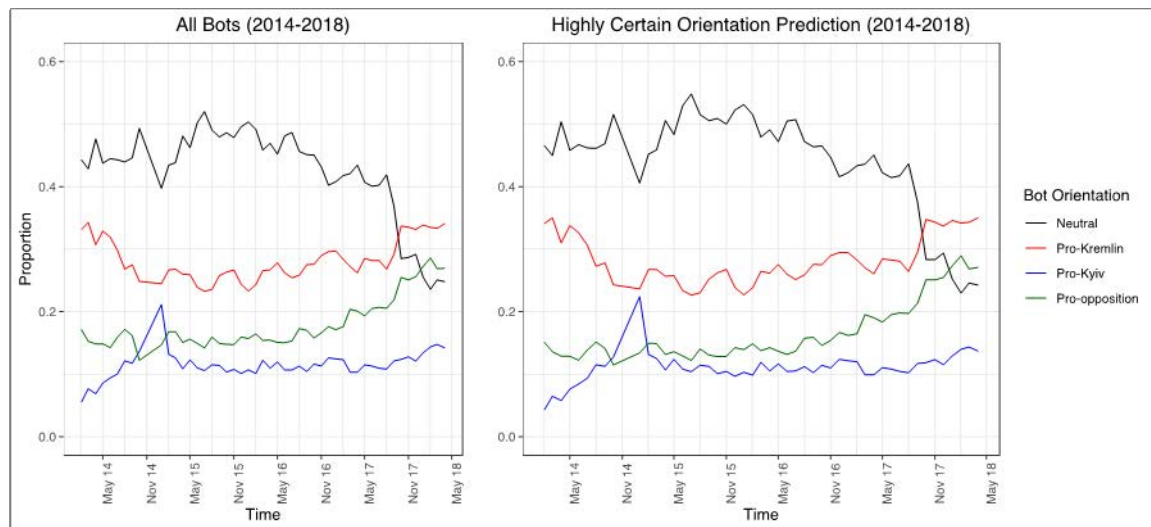
Sequence lines represent the proportion of bots with a given type of political orientation among all bots that tweeted in a given month. The left panel uses data on all 120,000 bots; the right panel only uses data on 93,000 bots for which we were able to predict the correct political orientation with probability over 0.9.

We suspect that these diverging patterns reveal different strategies used to disseminate information and pedal partisan agenda on Twitter. Many duplicative neutral bots use hashtags to spread unfiltered streams of news that tend to be quite diverse even coming out of partisan media. Selecting news stories and other messages to advance a particular agenda is left to leaders of public opinion, whose tweets are then retweeted by networks of properly attuned non-neutral bots.

When Did All of This Start?

Space limitations do not allow us to provide an in-depth account of how bots interact dynamically within and across orientation groups. However, we can gain some insights about the history of Russian Twitter bot wars from **Figure 3**. As the figure demonstrates, pro-Kremlin bots are slightly younger than pro-opposition and neutral ones. This serves as an additional piece of evidence supporting the claim we make elsewhere that the Kremlin's active engagement with social media was motivated by the need to catch up with democratic forces using social media as "liberation technology."[6]

---

[6] See: Sergey Sanovich, "Computational Propaganda in Russia: The Origins of Digital Disinformation," Report for the Project on Computational Propaganda, Oxford Internet Institute, 2017; Joshua A. Tucker, Yannis Theocharis, Margaret Roberts, and Pablo Barberá, "From Liberation to Turmoil: Social Media and Democracy," *The Journal of Democracy*, 28(4), 2017, pp. 46-59

Figure 3. Median Age of Tweeting Bots
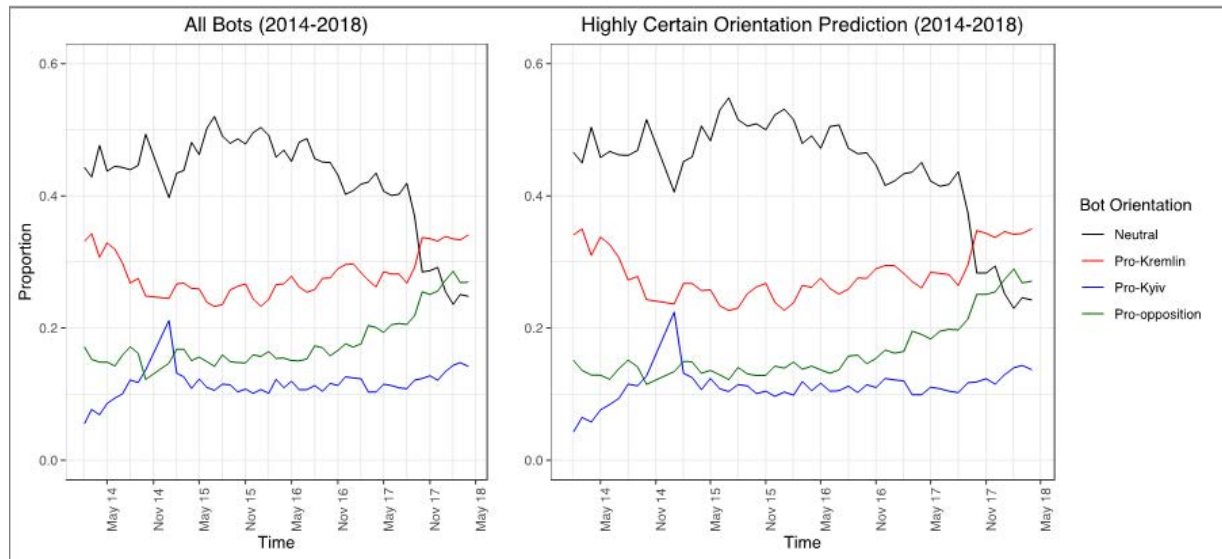
Solid lines represent the median age of bots that tweeted in a given month for each political orientation separately. The left panel uses data on all 120,000 bots; the right panel only uses data on 93,000 bots for which we were able to predict political orientation with probability over 0.9.

Another important aspect revealed from **Figure 3** concerns when bots were created. Apparently, the earliest bots were created around 2011 (around 1,000 days before February 2014), but most of them were created in 2012–2013.

Interestingly, there were a lot of changes in the average age of bots after 2014. New bots were created and replaced older ones, as reflected in the drops of the median lines in **Figure 3**. However, starting in mid-2015, the population of bots stabilizes somewhat, and we can see a nearly linear increase in the median bots' age going forward from this point.

What Do Different Types of Bots Do?

Bots with different political orientations show dissimilar patterns of activity, both in terms of tweeting frequency and types of tweets they post.

Figure 4. Number of Bot Tweets Over Time
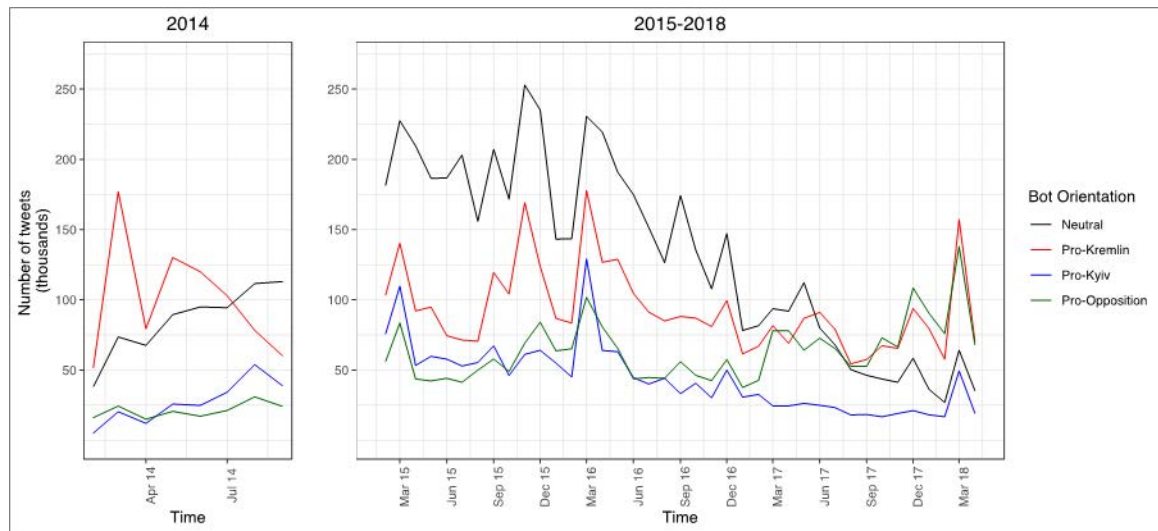
Year 2014 is presented separately for the clarity of exposition. Computations made using all 120,000 bots.

As **Figure 4** reveals, the numerical prevalence of neutral bots helped them dominate also in terms of the number of tweets. However, this prevalence dissipated in 2017 with an increase in both the number of pro-Kremlin and pro-opposition bots, and their activity. Figure 4 also shows how different the activity of bots were during the most active phase of the Russian involvement in the Ukrainian crisis in 2014, when pro-Kremlin bots clearly dominated in their tweeting activity.

As one can infer from **Figure 4**, there is some correlation between the number of tweets that bots of different types post over time. It is particularly large for neutral and pro-Kyiv bots (Spearman's rank correlation $\rho = 0.76$) and neutral and pro-Kremlin ($\rho = 0.60$). This is most likely due to bots' reaction to similar political events.

Although bots react to similar events, they do this in different ways, as **Table 1** illustrates. The most drastic difference is between neutral bots on the one hand, and the rest on the other. Whereas both pro- and anti-Kremlin bots retweet a lot (over two thirds of their tweets are retweets), neutral bots basically do not retweet other accounts. They also do not post tweets mentioning other users (i.e., containing the @ symbol), although mentions are rare across all types of bots. Another peculiarity of neutral bots is that they are much more active than other bots in using hashtags.

Table 1. Descriptive Statistics: Retweets, Hashtags, and Mentions

Orientation	Number of accounts	Number of tweets	% RT	% #	% @
Pro-Kremlin	32,479	4.4 million	66	9	5
Neutral	55,730	5.9 million	7	16	1
Pro-opposition	20,881	2.7 million	73	4	6
Pro-Kyiv	10,993	1.9 million	71	4	4

Percentages are out of all tweets posted by bots of a given political orientation.

Another interesting aspect of bot activity is how "creative" bots are. **Table 2** characterizes bot creativity by looking at repetitive tweets. As one can see from the table, neither pro- nor anti-Kremlin bots engage in active posting of identical tweets (except some outliers, as the "max" column shows). On the other hand, tweeting identical text is quite common for neutral bots.

Table 2. Descriptive Statistics on Repetitive Tweets

Orientation	Mean	Median	Max
Pro-Kremlin	0.965	0	10,210
Neutral	6.661	3	2,592
Pro-opposition	0.272	0	1,143
Pro-Kyiv	0.036	0	147

For each repetitive tweet that is not a retweet, we computed the number of times it was tweeted by bots of different political orientations. The mean, median, and max statistics were computed on those frequencies.

Non-neutral bots also show different patterns in terms of who they retweet. **Figure 5** shows the retweet network for account pairs with at least 1,000 retweets from 2014–2018. As one can see, anti-Kremlin accounts reveal a great deal of concentration in their retweet strategies. In particular, pro-Kyiv bots mostly retweet "newsdaily_ukr," whereas pro-opposition accounts retweet either Navalny or "VVP2_0," which tweets pro-oppositional comments and jokes on the Russian political agenda. Pro-Kremlin bots show much more diversity in their retweet network, and we give examples of just two frequently retweeted accounts: the pro-Kremlin mass-media outlet *Vzglyad*, and "BOLLIK12"—an account that is currently suspended by Twitter.

Figure 5. Bots Retweet Network for Retweet Pairs With at Least 1,000 retweets 2014–18
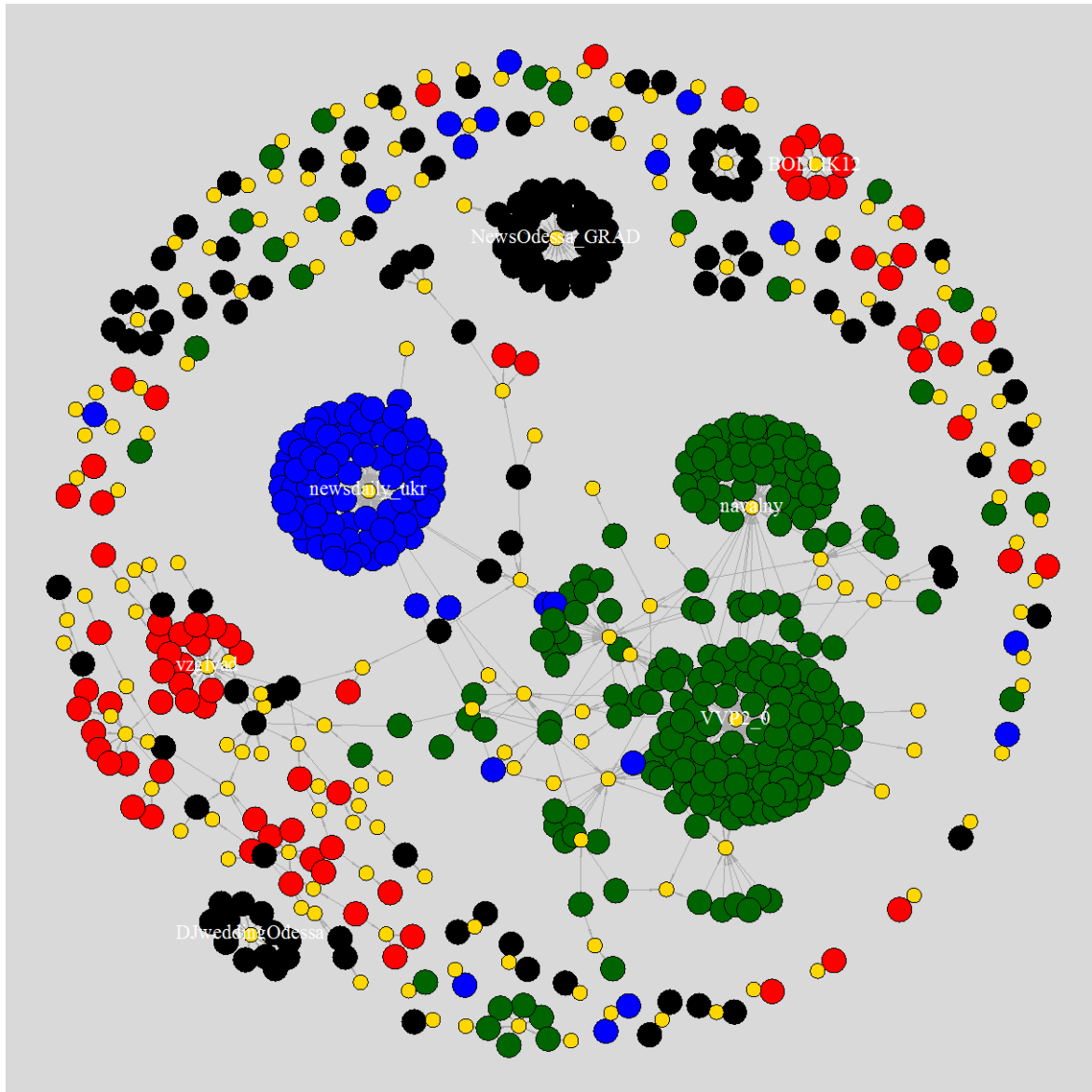
Golden circles are retweeted accounts. Pro-Kremlin bots are in red, neutral bots are in black, pro-opposition bots are in green, whereas pro-Kyiv bots are in blue. Gray arrows represent retweets.

Conclusion

Our analysis of millions of Russian tweets over 2014-2018 reveals that bots make up a large proportion of the Russian political Twittersphere. However, an important lesson from our region is that one cannot assume that simply because there are bots present in the Russian political Twittersphere that they are pro-Kremlin. Instead, as it turns out, pro-opposition, pro-Kiev, and neutral bots proliferate as well. We therefore also developed machine learning models that allow us to distinguish between three major

groups of political bots in Russia at scale, including pro-Kremlin, pro-opposition, and pro-Kyiv bots. It is worth noting, though, that the fourth residual category of bots that we call neutral actually make up a plurality of these bot-orientation types.

Our preliminary analysis of bot activity shows that across the entire data set, bots mainly seem to be being used to amplify political messages. In the case of neutral bots, amplification is conducted via tweeting repetitive texts, whereas non-neutral bots achieve this via retweeting. It appears that the sources of retweets from Russian political bots are either mass media with strong political orientation or prominent political figures. Exciting topics for future research would include more deeply diving into the topics of the messages shared by bots, better understanding whether the target audience for these shared messages are humans or other computer algorithms (e.g., to influence search rankings), and testing hypotheses related to over-time variation in the use of political bots, both in Russia and beyond.