**Journal article:** Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. (2019). "For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia." SAGE Open, January-March 2019: 1–16.

# Supplemental Appendix

## Appendix A: Robustness Checks

	Precision	Recall
Pro-Kremlin	0.96	0.99
Neutral	0.98	0.96
Pro-opposition	1.0	0.98
Pro-Kiev	1.0	1.0

Table 1: Dropout: Classifier Performance.

Note: Entries are performance metrics for the test set.  $\begin{aligned} Precision &= Pr(A|\widehat{A});\\ Recall &= Pr(\widehat{A}|A), \text{ where } A \text{ is a given cat-}\\ egory of bots, \text{ and } \widehat{A} \text{ is category } A \text{ predicted.} \end{aligned}$ 

Table 2: Orientation Classifier Confusion Matrix.

	Predicted orientation			
	Pro-Kremlin	Neutral	Pro-opposition	Pro-Kiev
True orientation				
Pro-Kremlin	66	1	0	0
Neutral	1	45	0	0
Pro-opposition	1	3	48	1
Pro-Kiev	0	0	0	28

*Note:* Entries are frequencies on the test set.





Histograms represent predicted probabilities for the most probable class for every observation. As one can see, the MLP model makes high-confident predictions for most observations.

Neutral bots

regnum.ru tvzvezda.ru www.kp.ru www.rbc.ru www.youtube.com rg.ru www.vipservicegas.com vk.cc ria.ru dlvr.itvk.com lenta.ru twitter.Comow.ly **afs-gasservice.com** russian.rt.combit.lyfb.me life.ru lifenews.ru goo.gl youtu.be www.portnoff.od.ua j.mp www.gazeta.ru news.mail.ru www.interfax.ru www.vesti.ru www.interfax.ru www.vesti.ru www.interfax.ru www.vesti.ru

Pro-Kiev bots

www.ukrinform.ua zloy-odessit.livejournal.com www.pravda.com.ua gordonua.com news.liga.net inforesist.org replyua.net fb.mebit.ly www.dialog.ua tvrain.ru nv.uagoo.gl zn.ua tvrain.ru nv.uagoo.gl censor.net.ua echo.msk.rutwitter.com vk.com politolog.netyOutu.be In.is espreso.tv www.youtube.comjoinfo.ua www.facebook.com newsdaily.com.uadumskaya.net obozrevatel.com 24today.net topnews.odessa.ua apostrophe.com.ua

Figure 2: Top 32 urls that appeared in the highest number of (different) bot accounts by each group of predicted bots (predicted probability  $\geq 0.7$ ). Pro-Kremlin wesbsites in red; pro-opposition in green; pro-Kiev in blue; content-sharing platforms, news aggregators, and politically neutral websites are not color-coded (black). Gazeta.ru, which switched from being pro-opposition to pro-regime, and Lenta.ru, which switched from being neutral to pro-regime, are coded in orange. See endnote 10 for details and caveats of color-coding.

Pro-Kremlin bots

www.facebook.com www.vesti.ru rusnovosti.ru WWW.gazeta.ru www.interfax.rulifenews.ru dlvr.it www.youtube.com vesti.ru goo.gl ria.ru vz.ru fb.melife.ru bit.ly ren.tv ru.rt.comtwitter.com www.ntv.ru russian.rt.com izvestia.ru youtu.be ruposters.ru tvzvezda.ru vk.com tass.ruln.is lenta.ru www.kp.ru www.ridus.ru amp.twimg.com

Pro-opposition bots www.vedomosti.ru gordonua.com www.znak.com www.rosbalt.ru www.novayagazeta.ru r.rbc.ru www.facebook.com zona.media navalny.com ria.ru fb.me goo.gl slon.ru theins.ruWWW.youtube.com vk.com twitter,com vk.cc www.mk.ru youtu.bewww.rbc.ru In.is bit.lytvrain.ru top.rbc.ru meduza.io openrussia.org echo.msk.ru eta.ru www.svoboda.org www.interfax.ru www.newsru.com

Neutral bots

Pro-Kremlin bots

@mayskolpino @volkova\_v@youtube @iliabelous@lavrovmuesli @tass\_agency@rpgrpgs@rentvchannel @rishikesh\_news @probubnist@zvezdanews @vrsoloviev@spacelordrock @us\_progress@rianru@dolg132004 @ruposters@dimsmirnov175 @imerkouri@rt\_russian@gazetaru @lifenews\_ru@gazetaru @lifenews

©looqqyzoox25199 ©iwiceexino02786 @pravitelstvo\_rf @senhovestu1984 @dimindband @tass\_agency @vesti\_news @onlinekpru @dimsmirnov175 @ru\_rbc@gazetaru @vedersuven1978 @kremlinrussia @lifenews\_ru @youtube @downsillootham1 @vedersuven1978 @kremlinrussia @lifenews\_ru @youtube @downsillootham1 @zvezdanews @rjanru @hapkotnkn @svobodaradio @rt\_russian @townduartaren19 @navalny@xop\_xop @lentaruofficial @150157ocaqenoob @medvedevrussia @highway\_to\_heli @868354wogeroqoo @59747liryjecooh

Pro-opposition bots

@maryel2002 @pani\_walewska @elshad\_babaev @Indcalling @mix\_voronezh @victorkvert2008 @leonidvolkov @svobodaradio @granitweet @shulz @aavst @lyapunovs @kermlinrussia @echomskru @fake\_midrf @oldlentach @navalny @youtube @stalingulag @tvrain @anakoyher @atalingulag @tvrain @anakoyher @adagamov @pavel\_xii @vvp2\_0 @fakemorf @cynicarea @nashacanada @meduzaproject @golub @openrussia\_org @dw\_russian @sandy\_mustache Pro-Kiev bots

@obozrevatel\_ua @victorkvert2008 @a\_vitauskas @nashacanada @radiosvoboda @crimeaua1 @belamova@mh17\_rus @golub@moscow\_advokat @golub@elshad\_babaev @robertbik@anakoyher@igorz\_ua @inforesist@fake\_midrf@ale4kko @krymrealii@odessahistory @mark\_feygin@shulz @ryabtchuk\_roman @realuran@vvp2\_0@ian\_tanya @stalingulag@youtube @odessaunews@cynicarea @evgeniamer @svobodaradio @demonte\_cristo

Figure 3: Top 32 mentions that appeared in the highest number of (different) bot accounts by each group of predicted bots (predicted probability  $\geq 0.9$ ). Pro-Kremlin accounts in red; pro-opposition in green; pro-Kiev in blue; accounts which do not clearly belong to either category are not color-coded (black). Gazeta.ru, which switched from being pro-opposition to pro-regime, and Lenta.ru, which switched from being neutral to pro-regime, are coded in orange. See endnote 10 for details and caveats of color-coding.

## Appendix B: Keywords and Hashtags Used for Collecting Twitter Data

This list uses transliteration. A Cyrillic equivalent was uses when "Cyril." is given in parentheses.

- medvedev (Cyril.)
- dukhovniyeskrepy (Cyril.)
- putinvor (Cyril.)
- putinakh (Cyril.)
- pzhiv (Cyril.)
- Strategiya31 (Cyril.)
- triumfalnaya (Cyril.)
- bolotnaya (Cyril.)
- oppozicia (Cyril.)
- gorozhaneprotiv (Cyril.)
- surkovskayapropaganda (Cyril.)
- navalniy (Cyril.)
- zanavalnogo (Cyril.)
- komandanavalnogo (Cyril.)
- suvkirove (Cyril.)
- $\bullet$  PussyRiot
- PussyRiot (Cyril.)

- tolokonnikova (Cyril.)
- narodniyskhod (Cyril.)
- sdnempobedi (Cyril.)
- #sochi2014
- #sochi
- #putinsgames
- #sochi (Cyril.)
- #vitishko (Cyril.)
- #schitaemvmeste (Cyril.)
- #sochifail
- #sochi2014problems
- golodovka (Cyril.)
- MinutaNeMolchaniya (Cyril.)
- zhalkiy (Cyril.)
- puti (Cyril.)
- spasiboputinuzaeto (Cyril.)
- priamayaliniya (Cyril.)
- partiyazhulikovivorov (Cyril.)
- edro (Cyril.)
- 6maya (Cyril.)
- sobyaninnashmer (Cyril.)

- marshmillionov (Cyril.)
- zachestniyevibory (Cyril.)
- bolotnoyedelo (Cyril.)
- 6may
- svobodupolitzaklyuchennym (Cyril.)
- svoboduuznikam6maya (Cyril.)
- rosuznik (Cyril.)
- odinzavsekh (Cyril.)
- vsezaodnogo (Cyril.)
- rasserzhennye (Cyril.)
- chestniyevybory (Cyril.)
- udaltsov (Cyril.)
- vysurkovskayapropaganda (Cyril.)
- 37godvernulsya (Cyril.)
- DMP (Cyril.)
- privet37god (Cyril.)
- krovaviyrezhim (Cyril.)
- kirovles (Cyril.)
- tolokno (Cyril.)
- biryulevo (Cyril.)
- khvatitkormitkavkaz (Cyril.)

- khvatitvinitkavkaz (Cyril.)
- russkiymarsh (Cyril.)
- Sochi2014 (Cyril.)
- #olimpiada (Cyril.)
- #olimpiyskayazachistka (Cyril.)
- #sochiproblems
- Odessa (Cyril.)
- $\bullet \ \# {\rm Nemtsov}$
- #Nemtsov (Cyril.)
- nemtsov (Cyril.)
- savchenko (Cyril.)
- #FreeSavchenko
- #Putinkiller
- maidan (Cyril.)
- maidaner (Cyril.)
- maidanutiy (Cyril.)
- #PutinUmer (Cyril.)
- #MinutaNeMolchaniya (Cyril.)
- Su24
- Su-24 (Cyril.)
- Su24 (Cyril.)

- #samolet (Cyril.)
- #RussianJet
- #ExpelTurkeyFromNATO
- #Russianplane
- #Erdogan
- #Latakia

### Appendix C: SMaPP Lab (NYU) Russian Twitter Project Coding Instructions

### Section 1. Introduction

### Dear project participants!

The goal of our research project is to study the character (patterns, connections, sequences, etc.) of the activity of bots (automated programs that, to a varying degree, imitate real people) in the political segment of the Russian Twitter. To this end, we have collected data on a large number (over 1.3 million) of Twitter accounts that have tweeted about Russian politics in the last couple of years. However, we have no foreknowledge about which of these accounts belong to human beings and which are bots. Obviously, classifying these accounts by hand isn't just impossible (it would require an unthinkable number of man-hours of work), but also it's not in line with the methodology of modern social sciences. Computer science and statistics are now used to solve tasks like that. We also have developed an algorithm that classifies accounts as either real human beings or automated bots, based on data drawn from a small sample of accounts that have already been classified. These already classified accounts are called the *training set*.<sup>1</sup>

Your job in this project is to help create the training set. This task could be split into two subtasks:

- *classifying accounts on Twitter* so that a computer could correctly classify a much larger number of accounts using your decision – which category an account belongs to – as a template
- *identifying accounts' political orientation*

The main goal of our classification is to separate *real human beings* writing on Twitter from *bots*, i.e. programs that automatically post information lifted from elsewhere or composed by the program itself (the latter is quite rare, though). In some cases, differentiating real people from bots is fairly easy. For example, compare the account of someone you know with an account full of odd spam. But in other cases, it could be hard.

To make the process of classification *reliable* (meaning that two people working independently and not communicating to each other would classify the account the exact same way), we need a schema

<sup>&</sup>lt;sup>1</sup> Methodological note: To understand what "training" means, think of fiting a linear regression. When we fit a linear regression, our goal is to predict, in an optimal way, the value of the dependent variable (y), based on one or many explanatory variables (x's). To solve this problem, you need values of both x's and y. The algorithm (for example, the least squares approach) uses the values of y from the training set to train itself in predicting y based on the values of x's that correspond to y in the training set.

Classification is a similar task: to learn (to train) how to predict the category the object belongs to based on a number of its characteristics.

There are tasks that don't require a training set. For instance, the cluster analysis simply identifies groups (clusters) of similar objects based on their characteristics. The results of cluster analysis could be quite different from what you expected. It could help, though, for exploratory purposes.

that makes it easy to figure out the category an account fits the best. At the same time, we need to keep the number of categories relatively small, otherwise the process of classification would be difficult and, more importantly, the analysis would be uninformative. We developed a *typology of accounts* (aka the classification scheme), which strikes a balance between accuracy and functionality.

Based on this typology of accounts (we describe it in detail below) you will classify an account snapshot we created based on the tweets from this account in our collection. Why will you use static snapshots instead of actual live Twitter accounts? The computer algorithm that we will train (using the training set you are going to create) will work with tweets from our collection and their metadata, not live Twitter accounts. To be confident that both you and the algorithm will be working with the same data, we created snapshots using the same tweets and metadata.

### Section 2. How to use this instruction

Although it's always a good idea to read this whole document again, it can also be used as a guide where you can find answers to specific questions that you might encounter.

Section 3 contains the *typology of Twitter accounts* in a condensed form. It gives you an understanding of the logic we applied to classifying accounts. In some simple cases (like spam), this will be enough to classify accounts correctly.

In most cases though, you will need to refer to **Section 4** that contains a *detailed description of each classification category*. It's conveniently divided into subsections, so if, for instance, you're convinced that an account is not a bot but you are not sure whether it's a community or a transmitter, you only need to consult subsection 4.6, "Humans". In the beginning of Section 4, we present an algorithm that can help you identify the category an account belongs to even in the most complicated cases. In addition, you can find a picture example for each category in our typology in a separate file called **Library.pdf**, and compare how much an account looks like the examples we provide.

Our classification of political orientation is found in Section 5.

**Section 6** "The recommended coding procedure" describes how to organize your work in the most efficient way, not to make mistakes when filling the tables, and correctly save and send out the file.

Finally, **Section 7** contains important instructions on what is allowed and what is not during the coding process, as well as the FAQ. **Make sure you read this section before you get started!** 

### Section 3. General typology of accounts on Twitter

(The numbers after the dash correspond to the code that we assign to each category of accounts. It is this code (number) that should be entered into the Excel files.)

- Official institutional accounts 1
- Bots

0

- Single-content:
  - Retweets 2
  - Pictures 3
  - Videos 4
  - Text:
    - News headlines
      - $\circ$  Text with links 5
      - $\circ$  Text without links 6
    - Other text -7
- $\circ$  Diverse content 8
- Cyborgs -9
- Humans
  - $\circ$  Transmitters 10
  - $\circ$  Personal accounts 11
  - $\circ$  Communities 12
- **Spam** 13
- $\Box$  *Deleted* -0
- $\Box$  Suspended 99
- $\Box$  Protected 999
- $\Box$  Never twitted 9999

### Section 4. Classification categories

Below, we present detailed descriptions for each classification category. In the description, we use the same reference codes as in the general typology found in Section 3. You will need to enter these digital codes into the Excel files (for details, see Section 6).

See picture example for each category in "Library.pdf".

### 4.1 <u>Unavailable accounts (CODES 0, 99, 999, 9999)</u>

THE DESCRIPTION OF THIS CATEGORY IS RETAINED FOR THE PURPOSES OF COMPLETENESS OF THE INSTRUCTIONS AS WELL AS FOR HISTORICAL REASONS. AS YOU WILL BE WORKING WITH STATIC SNAPSHOTS YOU WILL NOT ENCOUNTER ACCOUNTS OF THIS CATEGORY! So, the first thing you will need to code is whether you have access to the account, and, most importantly, if not — then why.

An account could be **suspended**. In this case, you will get: <u>https://twitter.com/account/suspended</u> in the address bar of the browser where you enter the account address. The code for these accounts is 99.

Another option is a deleted account (if you try to visit its webpage page, the account address will remain in the address bar, but the page will say "Sorry, that page does not exist!"). The code for these accounts is 0.

The third option is that an account is **protected** from viewing, that is, only its followers can view it. This is indicated as follows: "@ Dunyasha\_A11's Tweets are protected". The code for these accounts is 999.

Finally, the fourth option is an account that **hasn't tweeted yet** (Twitter tells you this much: "@absametov has not tweeted yet"). The code for these accounts is 9999.

### 4.2 Official accounts (CODE 1) and spam (CODE 13)

After opening an account, you should first evaluate whether there is any meaningful content at all. If all you see are links to online gadget and game stores (sometimes with hashtags or pictures) or gibberish, then you found a **spam** account that you should label with code 13 (look at an example in the Library.pdf). Sometimes, accounts switch to spam after posting more meaningful content. It is worth, therefore, to scroll through all available tweets from the account.

If there is meaningful content, check to see if it's an **official** account. We consider any account held by an <u>organization</u> to be official, be it a state organization, a private firm, or a political party. Many accounts of this type (<u>but not all</u>) have a blue checkmark. The code for these accounts is 1.

In order for the account to be considered official, it must satisfy two key criteria. *First*, it must represent an organization and not a person. *Second*, it must be maintained by the organization itself (i.e. by one of its employees), and not by someone else. Your task is to <u>indirectly assess</u> the account's compliance with these two criteria. For example, a link to the organization's website can serve as evidence on the second criterion, but it is not necessary for this criterion to be met.

The most frequent type of official accounts that you will encounter are media accounts (for instance, the accounts of newspaper *Kommersant*). At the same time, the accounts of *Kommersant* journalists should be classified as personal accounts (codes 10 or 11). Numerous bots that simply post *Kommersant* headlines but are not connected with it in any way should be classified as bots (codes 2 - 8).

A few additional examples. Accounts of famous people, for example, Alexei Venediktov, or even Dmitry Medvedev, should be classified as personal accounts and not as official ones. At the same time the Kremlin's Twitter, for example, should be classified as an official account, because Vladimir Putin does not write these tweets, and it only contains news headlines from *Kremlin.ru*. Similarly, accounts of Echo of Moscow and the Russian Government would belong to official accounts.

### **4.3** Single-content bots (CODES 2, 3, 4, 5, 6, 7)

If an account belongs to neither spam nor official categories, you should first look at the type of content that it posts: are all tweets similar to each other? If all the tweets are so homogenous and similar that there is no doubt that they were written by a robot (for example, only news headlines like "Putin to visit China" are posted, or all tweets contain pictures with captions, or every tweet is a link to Youtube vidoe without any comment), and not by a human being, the account belongs to one of the following categories:

- Only retweets code 2
- Only **pictures** code 3
- Only video code 4
- Text:
  - If it is **news headlines** or something similar
    - and it contains a link to the news story code 5
    - and it **does not contain a link** to the news story code 6
  - If this is any other text, from jokes to quotes from Brodsky, with or without links code 7

Within this category, the main pitfalls for correct classification are as follows.

**First**. While the attribution of an account to categories 3 and 4 ("pictures only" and "video only" bots) seems obvious at the first glance, "retweets only" can be tricky because the retweets can contain text, links, pictures, or videos, in various combinations. The distinctive feature here is that the owner of the account has not written any tweets of her own (nor any replies to other users). To confirm this, it is worth **scrolling several screens down** to see whether there are any tweets with this user's profile picture, and not someone else's. If you do find such tweets, look below to discern which category (8, 9 or 10) this account belongs to.

**Second**. Continuing on "retweets only": what if it's a human being rather than a bot who is retweeting? Generally, this problem can apply to any type of bots, for example "pictures only" and "video only" bots. But in those cases, it is usually easy to see that the account has no signs of a human present, but simply is an RSS feed put on Twitter. In the case of retweets, it could be more complicated. After all, there are people who do not write their own tweets, and retweet strangers. In this case, it is worth paying attention to **other characteristics of the account**. Does the (nick)name appear to belong to a human being? Does the profile picture look like a real photo or a picture chosen by a human being? Is there a link to a profile on VKontakte, Facebook, etc.? Does the account have many followers and do they look like human beings? How much and how often does the account retweet? The attached "Library.pdf" will be especially helpful here.

Third. Which text is considered news? Headlines like "Shoygu briefed Putin on the readiness of the Northern Military District" are easy. It's harder with things like "Vladimir hosted of Moscow honoring City Day": some bots post news headlines with missing words. You will therefore need to discern that something is a news headline or a quote, even if it is distorted.

**Fourth**. Bots with news headlines that post links to news stories can, for example, post links to *Kommersant*'s official website, but they can also post links to an unknown website with lots of ads and articles that were re-published from *Kommersant*. In this case, it does not matter: both fall in the category of news headlines with links. It's important not to confuse <u>an unauthorized bot</u> that posts links to an official authorized Twitter account (for example, *Kommersant*'s account).

### 4.4 Bots with heterogeneous content (CODE 8)

If the content of tweets and user information clearly indicate that the account **does not belong to a human being, but the tweets vary in terms of the content they post**, then the account most likely falls in category 8, "bots with diverse content". Usually, they consist of a combination of news headlines, retweets, and pictures.

What could these accounts be confused with most easily? With cyborgs. Bots with diverse content post a variety of different things, and among the stream of automatically generated tweets (see below on cyborgs) you might miss some original text clearly written and posted by a human being. If you doubt the originality of the text – even something mundane, like musing about the weather – you can google it: an original text will not be found anywhere except for the account you are looking at. If an original text exists (but is surrounded by a stream of tweets clearly generated by a bot), you are dealing with a cyborg. If there is no original text at all, it is a bot with diverse content.

### 4.5 Cyborgs (CODE 9)

Cyborgs are accounts where content that is typical for bots (either single-content or diverse) is occasionally interspersed with original text purposefully written by human beings (for example, meaningful answers to other users).

It is important not to confuse cyborgs with the so-called trolls. The latter are human beings who write tweets to provoke others, for money, or fun, or out of conviction, whereas the former are primarily bots, with only **a small part of their tweets** containing real human interaction or writing. The code for cyborgs is 9.

### **4.6** Human beings (CODES 10, 11, 12)

Finally, if you do not see any signs that it is a bot, perhaps it is really a **human being**? A human being on Twitter would usually have an informative profile which says something about herself, a moderate amount of meaningful and varied tweets, and other signs that you will surely recognize without additional explanation. Most of real human beings on Twitter belong to category 11 (**personal account**) in our typology. Two categories are exceptions.

First, there are accounts that only feature links to the content posted on other social media platforms (for example, check-ins on Foursquare, or pictures on Instagram). These are **transmitters**, their code is 10.

Second, although it is not as common on Twitter as it is on other social media platforms, there are community-like pages. Often, they feature tweets about some city or a city district. They usually mix original tweets with news headlines about the city from the press. If an account looks like a news and information board about a city maintained by one or more moderators, this is a **community**, code 12.

### Section 5. Political orientation

We are interested in the political orientation of all accounts. To encode the political orientation, we use *Latin* letters:

### 1. Pro-Kremlin: K

- 2. Pro-opposition (Russian domestic opposition): O
- 3. Pro-Kiev: U
- 4. Neutral / Non-political / Uncertain (tweets represent several different orientations): N

### Spam is automatically classified as N

<u>What NOT to do:</u> read every last tweet in the snapshot to determine if the author supports Putin based on cat memes and other irrelevant content.

<u>What to do:</u> if an account is 1) primarily about politics and 2) takes very strong, consistent and explicit position on political matters, then your task is to define and label this position. If the position is not strong, is hard to pinpoint, or the account is non-political altogether, it falls in the neutral category (N).

If you are convinced that an account has a strong and consistent political orientation, be careful to choose the right classification for it.

For example, accounts that criticize the Kremlin might be doing this primarily in connection to its activities in Ukraine, and then they belong to the **pro-Kiev category U** (this includes Ukrainian patriots of all stripes). Accounts that support or cover events from the point of view of the Donetsk and / or Lugansk People's Republics do NOT belong to the pro-Kiev category, since they represent separatists fighting against the central government in Kiev.

Other accounts also criticize the current Russian leadership, but focus on Russian domestic politics (for example, corruption in the government or repressions against the opposition). They fall in the category of **pro-opposition accounts (O)**. This also includes supporters of various Russian opposition politicians from Navalny to Khodorkovsky, and so on.

All accounts that support Putin, Medvedev, United Russia, or that are against the opposition and support the Donetsk People's Republic, the Lugansk People's Republic, the National Guard, "the faith, the tsar, and the fatherland" belong to the **pro-Kremlin (K)** category.

Politically colored accounts that do not fall under any of these types (for example, the accounts of the Belarusian opposition), should be classified as neutral (N).

### Section 6. Recommended coding procedure

1. Create a folder called "Twitter\_project" on your computer. If this folder already exists, proceed to the next step.

- 2. Download the Excel file that we emailed you with words "coding\_form" in its name (at different stages of the project, a stage identifier will be added to it). Save this file into the "Twitter\_project" folder. Download and unzip the static snapshots (html pages that artificially recreate accounts using data in our collection).
- 3. Open the Excel file. It contains a tab with 3 columns: Account, Code, Orientation. The Account column stores account IDs for the accounts you'll need to classify. The remaining columns are empty, and you will enter the codes you assign there:
  - Code: you enter the numerical code indicating the account type here;
  - Orientation: you enter the Latin code indicating the political orientation of the account here.
- 4. Print or open on your computer two classification charts (see attached) and keep them around, as they will help you figure out accounts' category and orientation more quickly.
- 5. Open the snapshot corresponding to the first account ID in your Excel file.
- 6. After determining the type of account and its political orientation, enter the required codes in the Excel file.
- 7. Save the Excel file.
- 8. Navigate to the next account.
- 9. After you are done coding all the accounts, save the Excel file again and send it to two addresses: denis.stukal@nyu.edu and sanovich@nyu.edu, indicating in the body of your email how much time you spent coding this portion of accounts in hours.

### Section 7. Important rules and FAQ

 If you see unreadable gibberish when you open html files with snapshots, this might be due to conflicting encoding settings in our browsers. You need to change the encoding to Unicode (UTF-8).

How to do it?

- Chrome: https://support.google.com/chrome/answer/95290?hl=ru
- Safari: https://support.apple.com/kb/PH21424?locale=ru RU&viewlocale=ru RU
- Opera: http://forums.opera.com/discussion/1789662/kak-pomenyat-kodirovku-veb-stranicyv-opera18/p1
- 2. Some tweets do not look like tweets. Instead, you just see a plain text. This means that these tweets or the entire account have been deleted. You should not ignore them under any circumstances! Here are two examples of how to spot and code them.

In the first example, it is easy to spot two tweets with news headlines between tweets containing only hashtags: one with a link (underlined in red), the other without. (This account thus belongs to bots with diverse content).

GazetaRu #Nemtsov #ДНР #ЛНР #Новоросс #АТО #Украина #Донбасс #Донецк #Лутанск #Немцов Татьяна Бунина (@a45955148) <u>2015-03-01 06:19:46</u> ЕchoMskNews #Nemtsov #ДНР #ЛНР #Новоро #АТО #Украина #Донбасс #Донецк #Лутанск #Немцов Татьяна Бунина (@a45955148) <u>2015-03-01 06:19:47</u> Донецк #Лутанск #Харьков #Одесса #мосийчук Этот своей смертью не умрет http://t.co/MsTwEMRKYj.. Татьяна Бунина (@a45955148) <u>2015-03-01 13:34:29</u> Путин требует автономии для Донбасса http://t.co/CqkrccD6ku... #Киев #Донецк Татьяна Бунина (@a45955148) <u>2015-03-01 13:34:30</u> Путину плевать на судьбу русских в Донецке...Борис Немцов Татьяна Бунина (@a45955148) <u>2015-03-01 15:35:19</u>

Ukraine #Україна #Украина #Донбасс #россия #рф #путин #ОБСЕ #НАТО #США #терроризм

In the second example, a bot is retweeting both accounts that are still available and those that are already deleted. In this case, both deleted and available tweets point to the same category of a bot with retweets, although they look different.

Соборная Украина!         ▼ Follow           @soborna_com         ▼ Follow           Порошенко поздравили лидеры многих государств мира, а Путин – нет tinyurl.com/lwh66pv         11:24 AM - 24 Aug 2014           ↑         1 3 9         1	
<ul> <li>Валера Кончалов @ValeraKonchalov</li> <li>Да и Навальный - это человек Путина. Потом вспомните мои слова.</li> <li>11:37 AM - 24 Aug 2014</li> <li>▲ ▲ 3 ♥ 2</li> </ul>	
RT @GOSDEP1USA: Путин "забыл" поздравить Украину с /           http://t.co/wH0CSGDMvw           ну и ну (@Hy_u_Hyyy) 2014-08-24 15:42:11 <u>RT @ReggaeMortis1:</u> Придумал конспирологическую версинчтобы вату по обе стороны г           ну и ну (@Hy_u_Hyyy) 2014-08-24 20:54:14	Цнем Независимости http://t.co/IHIGiM5VUH о, что Путин так слабенько и бесполезно ввел войска, просто
<ul> <li>Alex Doppelkorn</li></ul>	
<u>RT @kozakmamai:</u> История начинает повторяться. То как бы срочников от 20 до 25 ле ну и ну (@Hy_u_Hyyy) <u>2014-08-24 21:16:54</u>	ло в Чечне. Путин и Ко начинает бросать в пекло призывников-

- 3. Before sending out your Excel file, please make sure that all the cells are filled. Additionally, please make sure that you had not accidentally put a non-existent code in the political orientation (for example, by hitting the adjacent P instead of O on the keyboard).
- 4. If you are encountering an account you already coded, please code it again and do it **independently of the previous attempts**. You can encounter the same account twice for two main reasons. First, to check whether the coding is consistent.<sup>2</sup> Second, to recode those accounts that caused the most disagreement between coders at the previous stage of the coding.
- 5. We urge all participants to work independently and not discuss their coding with other participants. Methods for measuring intercoder reliability rely on the assumption that coding decisions are

 $<sup>^{2}</sup>$  In social science, we distinguish between three types of reliability: the weakest one (aka consistency) demands the same results when the object is coded multiple times by the same coder; the medium one (intercoder reliability) requires the results to be identical for different coders coding the same object; and the strongest one (precision) requires the coding results to concur with a exogenously determined gold standard.

made independently. If coders discuss their work with each other, this assumption is violated and we get biased intercoder reliability estimates.

- 6. While working on the project, do not communicate in any way on Twitter with either people or cyborgs or bots that you were assigned to code. Any study involving interaction with people requires the approval of the ethics committee (called IRB<sup>3</sup>). We did not request such approval as we do not expect or need to interact with the owners of the accounts we study. Therefore, any communication with them will put us in violation of research ethics rules.
- 7. **Do not change in any way letter and number codes you use.** For example, Latin (!) letters N, U, O, K are used to code political orientation. Please do not replace them with Russian letters. This severely complicates the processing your results.
- 8. **Do not put more than one code in each cell.** If you have doubts about the code you assign, please indicate that in your email to us.

<sup>&</sup>lt;sup>3</sup> <u>https://en.wikipedia.org/wiki/Institutional review board.</u>



Does the account contain anything other than spam, ads, or



No, the account consists almost exclusively of spam,





Appendix F: Library

August 7, 2016

## 0. Deleted



## 99. Suspended



### 999. Protected



ill Joined January 2012

K Tweet to Imma Star

### 9999. Never twitted



## 1. Official accounts

 Belong to known institutions: party, media outlet, etc. (Regional Vesti in Sochi)



## 1.1 Official accounts

- Link to a professionally-maintained community, not your average Livejournal community
- Note also a clear pro-Kremlin orientation



## 2. Bots: RTs

Retweets of any content: text, pictures, videos, replies. Most consistently politically charged bots are RTs



## 3. Bots: pictures

### Labor-intensive bots: require uploading pictures directly to twitter

FOLLOWERS LIKES





## 4. Bots: videos

### Usually, embeded vidoes from Youtube



## 5. Bots: news headlines with links

These links go directly to RIA; others go to news agregators



## 6. Bots: news headlines without links

#### Pure text news

	Theers FOLLOWERS	
рмолай	Tweets Tweets & replies	
pa52028212 extarea p ass=ProfileHeaderCardEdiling-	Ермолай @a52028212 - 2h За рост цен после антитурецких санкций ответят Ткачев и Мантуров	0
ieasuringField/p /div div ass=ProfileHeaderCardEditing-location	45 £3 ¥ ···	
rofileHeaderCardEditing-item	Ермолай ⊚₅52028212- 2h Президент наградил Хазанова орденом «За заслуги перед Отечеством»	0
🦯 Tweet to Ермолай	45 E3 V ···	
	Ермолай @s52028212 - 2h Законопроект о запрете сухого алкоголя внесут в ГД до Нового года	0
	45 123 V ····	
	Ермолай @в52028212 - 2h Руководству Спецстроя вызовут «доктора»	6
	6 ±3 ♥ ···	
	Ермолай @в52028212 · 2h Между Сциллой и Харибдой	6
	6 ±3 ♥ ···	
	Ермолай @в52028212 - 5h Медведее посетит Китай в рамках заседания глав правительств ШОС	3
	45 123 V ···	
	Ермолай ⊚а52028212 - 5h Мэр Краснодара объяснил увольнение всех своих замов «перезагрузкой»	0

## 7. Bots: other text

## News interspersed with unrelated content, such as famous aphorisms



## 8. Bots: diverse content

### Mix of different content type: RTs, videos, text, etc.



## 8.1 Example: Bot with diverse content, not a human

Livejournal link goes to generic news account, nothing personal. Twitter content – Moscow region governor Vorobyev promotion (=pro-Kremlin), nothing personal



## 9. Cyborgs

### Usually: same as bots, but post replies clearely written by humans



### 10. Humans: translations

#### Links to author's posts elsewhere, e.g. on Facebook



## 11. Humans: personal accounts

#### "Regular" twitter account



## 11.1 Example: Human, not cyborg

## Tweets look like cyborg, but link in the profile goes to a real Facebook account



### 12. Humans: communities

### Curated collections of content on a subject: city, politician, etc.



## 13. Spam

### Mostly links to ads or stores



## 13.1 Example: Bot, not spam

#### **Political hashtags**



#### Столярова Юлия

@syv\_od
 Crostsposa
 2013 гол. 1
 Одесса
 2013 гол. 1
 Столтярова
 Ш Дага регистрации: Магсh 08, 2014
 Вот и ивст
 Столтярова
 Столтарова
 Столтярова
 Столтярова

Твиты Ноавится Списки 414 0 "Встреча в тракт 🖗 http://t.co/X0rAGErcrO #путин Столярова Юлия (@svv od) 2014-03-18 20:05:53 Xax, вот скоро но http://t.co/HcinY40dov #путин Столярова Юлия (@syv od) 2014-03-18 19:42:12 2013 год. Решил зав http://t.co/iOdRpfkT3B #путин Столярова Юлия (@svv od) 2014-03-18 19:40:09 Вот и наступила � http://t.co/4pIa8rW2Nu #путин Столярова Юлия (@syv od) 2014-03-18 19:26:34 Я маленькая колд http://t.co/galgLx7YOw #путин Столярова Юлия (@syv\_od) 2014-03-18 19:25:44 Обустраиваюсь н http://t.co/7pBgoi3ifh #путин Столярова Юлия (@svv od) 2014-03-18 19:25:43 Это мой новый дно http://t.co/m539eMd3Vt #путин Столярова Юлия (@syv od) 2014-03-18 19:22:13 Официальный ано♠ http://t.co/Z8DIbA021b #путин Столярова Юлия (@svv od) 2014-03-18 19:08:27 Есть у меня одна http://t.co/UzC3pW7Zq2 #путин Столярова Юлия (@syv od) 2014-03-18 19:06:03 http://t.co/Zw1halZ4IJ #путин Столярова Юлия (@svv\_od) 2014-03-18 18:58:54

## 13.2 Example: Spam, not bot

Hashtags are not necessarily political, or not political at all



## Political orientation

## U: Pro-Ukrainian

## Critical of Putin, but not pro-opposition: all tweets related to Ukraine





Follow

Україна Online

@OnlineUkraine

## O: Pro-opposition

## Some critique of Putin and Kremlin is related to Ukraine, but plenty of other topics present

G	Твиты Читаемые Читатели Нравится Списки 92852 1087 719 161 15
людмила надточии @ljuka0312 Ф	Суд вернул авропорт <u>Сдесса в</u> собственность города. biz liga net/all/transport/ 537 PM - 13 Jul 2015 ♠ 13 - 22 ♥ 2
⊞ Дата регистрации: June 24, 2015	Маrk Feygin         ✔ Foltow           Дети мой Оказывается мне надо получать пропуск в ФСБ, чтобы принять участие в суде по Савенкю в Донецке, посолкиу это пограничная зона 6:02 РМ - 13 Jul 2015         € \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$
	Правлюст         ▼ Гойом           Карельский город не дохдался газа, торжаственным пуском которого руководил Динтрий Модведее Павлюст Солтическ/каневку         14.9 № 1.3 Jul 2015           -         5 № 1.0 %         5 %

## K: Pro-Kremlin

P

### Either praise of Putin or mocking of his opponents



Дмитрий Смирнов 🥯 @dimsmirnov175

Follow

Путин перечислил с чем будет бороться БРИКС: терроризм, экстремизм, возрождение нацизма. наркотики, пиратство

## N: Neutral

### News mention Putin, but no explicit message about his virtues

