asd
G|M|F
Brief
2017 | No. 001

The Methodology of the Hamilton 68 Dashboard

by J.M. Berger

The Hamilton 68 Dashboard, hosted by the Alliance for Securing Democracy, tracks Russian influence operations on Twitter. The following document explains exactly what the dashboard shows and the methodology used to construct it.

Dashboard Overview

The dashboard monitors the activities of 600 Twitter accounts linked to Russian influence efforts online. Accounts were selected for their clear connection to Russian influence, but not all of the accounts are directly controlled by Russia. The method is focused on understanding the behavior of the aggregate network rather than the behavior of individual users.

The list includes samples from three different networks:

- A network derived from openly pro-Russian users.
- A network derived from users who tweeted as part of a disinformation campaign linked to openly attributed Russian media.
- A network of accounts that engage in automated behavior (bots) on behalf of other accounts reflecting Russian messaging priorities.

The list includes the following types of users:

- Accounts likely controlled by Russian government influence operations.
- Accounts for "patriotic" pro-Russia users that are loosely connected or unconnected to the Russian government, but which amplify themes promoted by Russian government media.

Accounts for users who have been influenced by the first two groups and who are extremely active in amplifying Russian media themes. These users may or may not understand themselves to be part of a pro-Russian social network.

The dataset is properly understood as a network of accounts linked to and participating in Russian influence campaigns. Modes of participation include both knowing and unknowing participation. The composition of the list is discussed in more detail below.

The 600 accounts are monitored in real time, and the output from their accounts is analyzed to produce the dashboard. There are three primary categories of data, each of which is analyzed in two different ways. The categories are:

1. Hashtags
2. Topics (terms used in tweets excluding hashtags)
3. Links (including top-level domains and specific URLs)

4. Each of these categories are then analyzed in the following ways:

    a. Top items (most-tweeted over the last 48 hours)

    b. Trending items (content with the highest percentage increase over the last 48 hours)

Understanding the Content

While the users in the network generally serve to promote Russian influence themes, the content within the network is complex and should be understood in a nuanced way. There are three broad categories of content documented by the dashboard:

1. Content generated by attributable Russian media and influence operations. This is a relatively small proportion of the network's content. It includes, for example, content generated by Russia Today and Sputnik.

2. Content amplified to reflect Russian influence themes. This content is typically produced by third parties, including but not limited to mainstream media, hyperpartisan sites and so-called "fake news" sites. Third-party content is sometimes amplified because it complements Russian influence themes. At other times, it is amplified for the opposite reason, meaning that users in the network are seeking to attack or discredit the content.

3. Less relevant content. This includes popular hashtags, which the users employ to increase the reach of their messages, developing news stories, and (less commonly) random social dynamics in the network.

We cannot at this time offer a breakdown of the relative amount of content from each of these categories, although we may pursue this question in future analysis. Nevertheless, three important caveats contained in the description above are worth emphasizing:

1. Not all content in this network is "created" by Russia. A significant amount—probably a majority—of content is created by third parties and then amplified by the network because it is relevant to Russian messaging themes.

2. Not all content amplified by this network is pro-Russian. The network frequently mobilizes to criticize or attack individuals or news reports that it wishes to discredit.

3. Because of the two points above, we emphasize it is NOT CORRECT to describe sites linked by this network as Russian propaganda sites. We are not claiming that content producers linked by this network are Russian propaganda sites. Rather, content linked by this network is RELEVANT to Russian messaging themes.

How the Monitoring List Was Assembled

The monitoring list is based on analysis conducted over the course of roughly three years, with the specific networks identified over the last year. The networks were revisited, updated and the list finalized in the summer of 2017. The three networks are based on the following criteria:

1. We tracked disinformation campaigns that synchronized with overt Russian propaganda outlets like Sputnik and RT (Russia Today). We analyzed the social networks of users who were promoting this disinformation to identify which users were centrally involved, and to remove users who tweeted disinformation casually, after encountering it online.

2. We identified a group of users online who openly professed to be pro-Russian and tweeted primarily in support of Russian government policies and themes. We analyzed followers of these accounts to identify a large and interconnected social network that tweeted the same themes and content.

3. We identified accounts that appear to use automation (bots) to boost the signal of other accounts linked to Russian influence operations, or to be the beneficiaries of such Each of the above activity.

Analytical techniques

Each of the above networks consisted of thousands of accounts. In order to identify the most relevant accounts for each, we employed social network analytical techniques largely developed by J.M. Berger and Jonathon Morgan. These techniques have been previously published, with a detailed description of the relevant methodologies in the following papers, both of which are freely available online:

- Who Matters Online: Measuring influence, Evaluating Content and Countering Violent Extremism in Online Social Networks, J.M. Berger and Bill Strathearn, International Centre for the Study of Radicalisation (March 2013)

- The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter, J.M. Berger and Jonathon Morgan, The Brookings Institution (March 2015)

A brief summary of the methodologies follows. After a network was identified based on the criteria above (for instance, "accounts that tweeted a specific piece of disinformation"), the users were analyzed using weighted metrics. Each of these metrics is a score that correlates to how an account in the network is similar to the accounts that seeded the network. The metrics were tested by extensively coding results, as detailed in the papers linked above. The metrics employed to create the monitoring lists included the following criteria:

- Influence: A metric that measures how many interactions a Twitter account inspires from other users in the same network, weighted by types of interactions and the number of followers.

- Exposure: A metric that measures how many interactions a Twitter account directs to other users in the same network, weighted by types of interactions and the number of followers.

- Engagement: The sum of Influence and Exposure, reflecting how engaged a user is with the network and how much the account reflects the overall interests of the network

- In-Groupness, or IQI: This is a more advanced metric that examines interactions and relationships inside the network with those outside the network, to identify accounts that are most relevant to the network's interests. In other words, accounts with high in-groupness are most similar to the accounts used to "seed" the network collection.

In addition to the metrics above, which are described in detail in the linked papers, we also employed a Fast In-Groupness/IQI metric devised by Morgan and Berger subsequent to the previous publications. Because the full IQI requires lengthy analysis and data collection (weeks and sometimes months), we employed a machine learning algorithm based on existing IQI scores and coded results. The resulting metric can be calculated in considerably less time. The fast IQI was extensively tested against the full IQI and produced very comparable results.

In addition to the metrics described here, we also manually reviewed the final list in an effort to further eliminate any possible noise. As part of this review, we removed accounts using one specific social media application for automated tweeting. In our evaluation, this network requires further study before we can fully assess its relationship to Russian influence operations.

Bot analysis

Identifying bots and cyborgs (bots which feature periodic human intervention) is a constantly evolving process. There are many different types of bots. Some are relatively simple, others are more sophisticated. No single approach is adequate to identify all of the different types.

We identified bots in this network using a new Influence-Outlier metric recently developed by Berger. Within Russian influence networks, we had frequently observed accounts which were disproportionately engaged with other accounts in the network relative to their follower counts. So for example, a user with 5,000 or 10,000 followers might be ranked among the top 10 influencers in the network, alongside accounts with 100,000 or millions of followers.

A formula was developed to isolate these accounts, which we initially suspected were being amplified by retweets sent by bots.[1] On further examination, we found this to be true, but we also found that many of these outliers appeared to be bots themselves, with some tweeting hundreds of times per day, primarily retweets, which were then further amplified by their followers. Several variations of the influence-outlier metric were employed to sift bot accounts from other networks related to Russian influence, and then the bots themselves were used to seed a network, which was subsequently scored using the fast IQI metric.

Further identification of users

We expect that we would have faced criticism for identifying the 600 accounts, and we expected (correctly) that we would face criticism for not identifying the accounts. We choose not to identify the accounts for the following reasons:

1. As noted above, our metrics are very accurate, but not 100 percent accurate. We believe based on the performance of the metrics described above and the subsequent manual review, that the monitoring list is at least 98 percent accurate, but that leaves as many as 12 accounts that may not be relevant. We are not willing to publicly attribute even one specific account incorrectly.

2. A 98 percent accuracy rate is, however, more than adequate to perform analysis on the aggregate set.

3. If we identified the users in the dataset, they would certainly change their behavior and render the dashboard essentially useless.

[1] A bot is a piece of computer code that tweets automatically based on pre-determined rules. When we refer to bots here, the category includes "cyborgs," meaning accounts that feature automated activity, but which are also manually supplemented by human users.

About the Author:

J.M. Berger is a non-resident fellow with the Alliance for Securing Democracy at GMF.

G | M | F The German Marshall Fund of the United States
STRENGTHENING TRANSATLANTIC COOPERATION

Washington • Ankara • Belgrade • Berlin
Brussels • Bucharest • Paris • Warsaw

www.gmfus.org