Under the Radar: Analyzing Recent Twitter Information Operations to Improve Detection and Removal of Malicious Actors, Part 1

Cody J. Wilson

August 12, 2022

Executive Summary

Russian information operations (IOs) targeting the U.S. political system have generated a significant amount of attention in recent years.[1] Yet a recently published report that I co-authored argues that this focus on a single threat actor leaves the West open to being blindsided by other, emerging threats. While analysts in the West focused on Russia, other actors like Iran, North Korea, and China were actively conducting their own IOs on social media platforms such as Twitter. In fact, Twitter regularly makes announcements about IO takedowns. The company also makes datasets of tweets posted by accounts tied to these IOs available for researchers. This report analyzes some of the key similarities and differences across three tweet datasets collected from recent IOs in the hope of shedding some light on the possible reach and return on investment gained by malign actors conducting these IOs. Each of the three Twitter datasets focus on a separate country, Russia, China, and Iran, all disrupted by Twitter in 2021. This analysis found that Iran's IO had the widest reach. This was contrary to, or perhaps because of, the growing attention placed on Russian IOs and, to a lesser extent, Chinese IOs. This analysis also revealed that more accounts do not necessarily yield a greater overall reach for the IO and that a more generalized response mechanism for countering IOs, regardless of country of origin, remains imperative. Lastly, these datasets lend themselves well to further analysis, which I am hopeful will prompt a series of follow-on reports aimed at improving detection and disruption of these operations.

---

[1] An information operation is a form of warfare that uses information to target a specific population with a tailored message aimed at furthering some larger goal. These operations can take the form of mis/disinformation campaigns, efforts to mobilize individuals in furtherance of a goal, or even to prevent people from doing something.

Introduction

On July 23, 2020, a handful of anonymous Twitter users began tweeting material containing the hashtag #RamaphosaResigns, referring to the purported resignation of South African President Cyril Ramaphosa. These accounts flooded the South African Twittersphere with the hashtag in the early morning hours, when few other users were online. Upon waking, South African Twitter users saw the hashtag trending and started talking about the surprise resignation, eventually leading the media to pick up the story. The only problem was Ramaphosa had not resigned—legitimate users were tricked into perpetuating a false narrative by malign actors (Jean Le Roux, 2020).

In recent years, Twitter and other social media platforms have been prominent targets of activity such as this by online trolls as well as coordinated, well-resourced information operations. Despite their best efforts, social media companies have struggled to find a way to stem the tide of false and deliberately misleading information, commonly known as misinformation or disinformation, respectively. At times, these efforts have even faced strong backlash, some of which has been justified. A recent Lawfare article I contributed to digs into some of these issues, including the backlash and the challenges involved in combatting disinformation on social media in light of growing concerns about online censorship (Daveed Gartenstein-Ross et al., 2022).[2]

However, one of the biggest challenges for social media companies is that information operations are constantly evolving as part of a perpetual cat and mouse game where malign actors continuously innovate to better evade detection. Worse yet, malign actors learn from each other in order to develop and adapt new tactics. Drawing attention to this learning mechanism was one of the main contributions my co-authors and I put forth in a recent report titled *Blind Sided: A Reconceptualization of the Role of Emerging Technologies in Shaping the Tactics, Techniques and Procedures of Information Operations in the Grey Zone*. The Blind Sided report also proposed examining IO threats from a tactic-oriented perspective rather than an actor-oriented perspective. The main downside of the actor-oriented perspective is that it leads analysts into being blind-sided by emerging threats while focusing on threats from yesterday (Ashley A. Mattheis et al., 2022). That key takeaway from the Blind Sided report laid bare to me the need for an actor-neutral IO detection and response mechanism. The goal of this series of analyses is to analyze real-world IO data that can contribute to and support the creation of this mechanism in the future.

In 2021, Twitter reported takedowns of several information operations on its Transparency Center website, including a Chinese operation focusing on denying and covering the systematic repression of the Uyghur population in Xinjiang Province, an Iranian operation targeting the West, and a Russian Internet Research Agency operation also targeting the West. Using datasets provided by Twitter containing tweets from these aforementioned operations, this series of reports examine some of the key similarities and differences across these three information operations with the goal of finding overlapping indicators that an information operation is underway on a social media platform like Twitter. This particular study (the first of hopefully several) asks, what was the estimated reach of these IOs? The purpose of this report is to shed some light on the return on investment that malign actors get from running an IO targeting a specific audience in order to highlight the importance of disrupting IOs that seek to undermine political stability. The broader goal of the entire series of reports, once they are completed, is to provide insights that can improve

[2] The full report by Daveed Gartenstein-Ross, Madison Urban, and myself that this article was based on can be found on Valens Global's website here.

the detection and removal of IO accounts from social media platforms like Twitter. What emerged from this analysis was a story that reflected what my co-authors and I warned about in the Blind Sided report—the threats most closely observed, Russia and to a lesser extent China, were not actually the IOs with the farthest reach and greatest potential of breaking into the mainstream.

Methodology

To prepare for this analysis, several datasets were acquired. The three datasets of interest were acquired from Twitter's Transparency Center in July 2022 (Twitter Transparency Center, 2022).

The first Twitter dataset contained over 15,000 tweets from a suspected Chinese information operation focusing on covering up systematic repression of Uyghur Muslims in China's Xinjiang Province. Twitter announced the takedown of these accounts in December 2021 (Twitter Safety 2021b). The second Twitter dataset contained almost 70,000 tweets from a suspected Russian Internet Research Agency (IRA) operation focusing on the United States and European Union. Twitter announced this takedown in February 2021. As part of this same announcement in February 2021, Twitter also announced a separate takedown of tweets from an Iranian information operation. This dataset consisted of over 560,000 tweets from suspected Iranian accounts that had originally tried to influence the 2020 U.S. Presidential election (Twitter Safety 2021a).

These three datasets from China, Russia, and Iran represent recent examples of suspected information operations conducted by actors that represent significant security challenges for Western democracies, particularly the United States. One operation from each country was chosen to compare and contrast the overall structures and characteristics of each operation from the others. This selection criterion helps to determine if there is any indication of overlapping indicators or tactics. With the exclusion of Russia, the datasets were the most recently announced takedowns conducted by Twitter, suggesting usage of the most up-to-date tactics. With regard to Russia, a December 2021 operation by the IRA targeting parts central Africa was available (Twitter Safety 2021b). However, the dataset contained only 16 accounts, leading to concerns that the dataset would not contain as much variety of tweets as the other two datasets. The Russia dataset contained in this analysis is also smaller than its counterparts are, but it is not quite as small as the dataset from December 2021.

After selecting the datasets, the data cleaning and preparation phase started. All data cleaning, transformation, and analysis steps utilized the R statistical programming language and the RStudio IDE. The R scripts used in this analysis and the Quarto markdown file used to produce this report reside at https://github.com/CWilson01.

Cleaning and transformation of the Twitter datasets took place in the following ways. The variables *tweetid, in_reply_to_userid, in_reply_to_tweetid, quoted_tweet_tweetid, retweet_userid, retweet_tweetid, user_mentions* were converted from numbers to characters to prevent inadvertent inclusion of these unique identifiers in descriptive statistics calculations.[3] Tweet hashtags were converted to all lowercase for a future analysis project, brackets and single quotes were stripped from the hashtags

[3] It is important to note that Twitter hashed many of these unique identifiers for accounts with under 5,000 followers to protect the anonymity of users that may have been included in the dataset erroneously. These hashed values cannot currently identify the user ID or account screen name of these accounts, so they will not appear in this report. The unhashed account names were previously available upon written request to Twitter but are not currently available.

and user_mentions variables, and NAs were introduced where data was missing entirely. The cleaned datasets were saved to new files for easy access during descriptive analysis.

For each of the three Twitter datasets, the following questions guided the descriptive analysis of the data. "How many unique accounts are present?" "How many unique tweets are present?" "What is the average number of tweets per account?" "What is the distribution of the dates of the tweets?" "What is the distribution of account creation date?" "How many followers does each account have?" "How many other accounts does each account follow?" "What is the proportion of retweets?" "What are the amplification metrics for each dataset?" This last question covered the number of quotes, replies, retweets, and likes for the tweets in each dataset. Such amplification metrics provide a sense of how many people may have interacted with the tweets produced by each information operation.

Results

On February 23, 2021, Twitter announced the takedown of a Russian IRA information operation, comprised of 31 accounts, as well as an Iranian information operation comprised of 238 accounts (Twitter Safety, 2021a). Later that year, on December 2, 2021, Twitter also announced the takedown of a large Chinese information operation. Twitter released a dataset containing what it calls a "representative sample of 2,048 accounts" from the operation (Twitter Safety, 2021b). Table 1 below shows an overview of the scale of each operation based on the datasets made available by Twitter.

Table 1: Overview of Russian, Iranian, and Chinese Twitter IO Activity

Metrics	Russia	Iran	China
Accounts	24	209	1247
Tweets	68914	560541	15635
Tweets per account	2871	2682	12.54

One important thing immediately stands out about this table—the number of accounts for each dataset is lower than the numbers reported by Twitter. This suggests that some accounts simply did not tweet but nonetheless aroused Twitter's suspicions for any number of reasons, such as having the same IP address or exhibiting patterns of behavior similar to other suspended accounts. Next, the number of tweets and the average tweets per account varies widely across the three operations. While the Chinese operation had an order of magnitude more accounts in its network, each account tweeted significantly less than the other operations, resulting in the lowest number of tweets produced. The Russian IRA operation was the smallest, but its highly active accounts produced more content per account than both of the other operations did. Iran's operation produced the most tweets overall, with well over half a million tweets posted.

Moving on to the distribution of tweets for each information operation, seen in Figure 1 below, the majority of tweets took place in 2020 for both the Russian and Iranian IOs, with the Russian operation having a large amount of tweets also in 2018 and 2019, though to a lesser extent. The length of time that these Russian and Iranian accounts were tweeting is also of particular interest. The graphs in Figure 1 show that the accounts were tweeting for quite some time, dating back to 2009 in some cases. Either this suggests a very long-term strategy or acquisition of the accounts, presumably through nefarious means, sometime after their creation. The Chinese operation differs starkly from the Iranian and Russian operations by being exclusively active in the first four months

of 2021. It is not clear if this difference is due to sampling by Twitter or if the Chinese IO was operational only in early 2021.

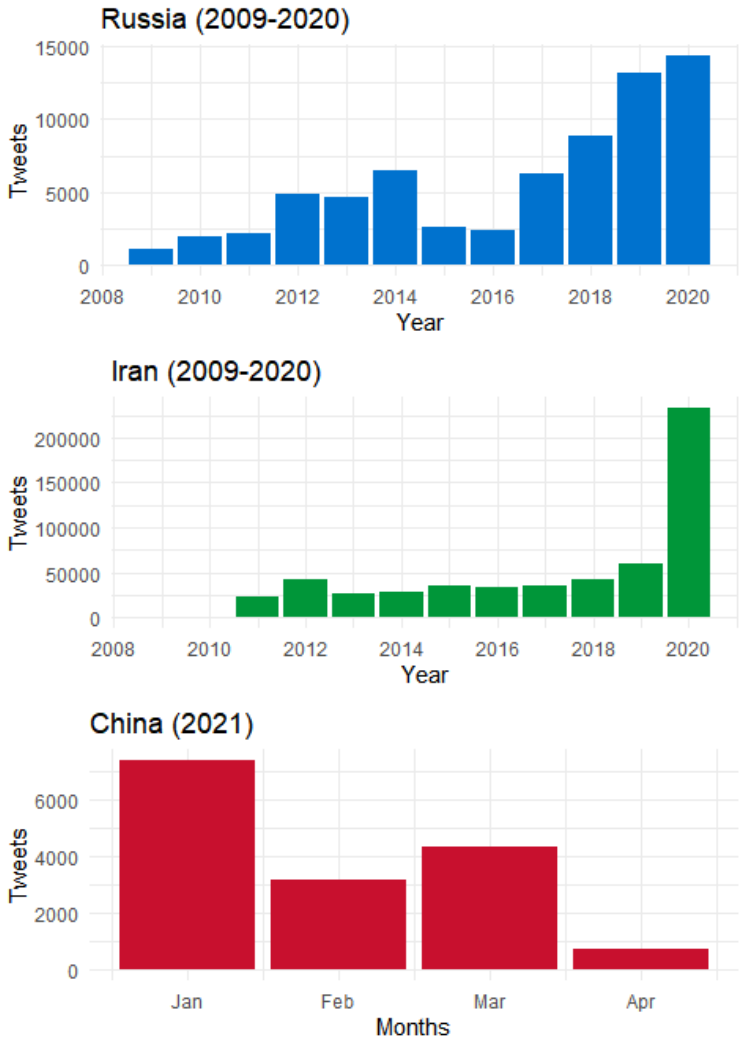

Figure 1: Distribution of Tweets by Information Operation

Figure 2 below, which depicts the distribution of account creation dates, suggests that the Chinese IO's sudden flurry of activity in 2021 was intentional. Based on account creation dates for the Chinese operation, it appears that 85 percent of the accounts had creation dates prior to 2021. However, all the tweets in the dataset started at the beginning of 2021 and continued until Twitter removed the accounts from its platform, presumably in April 2021. By contrast, the Russian operation steadily created 2 to 3 accounts per year throughout most of the period spanning 2009 to 2020. Iran's account creation pattern is a bit of a mix between China's approach and Russia's. The Iranian operation acquired the majority of its accounts in 2019 and 2020, but it also maintained a small but growing presence on Twitter dating back to 2009.

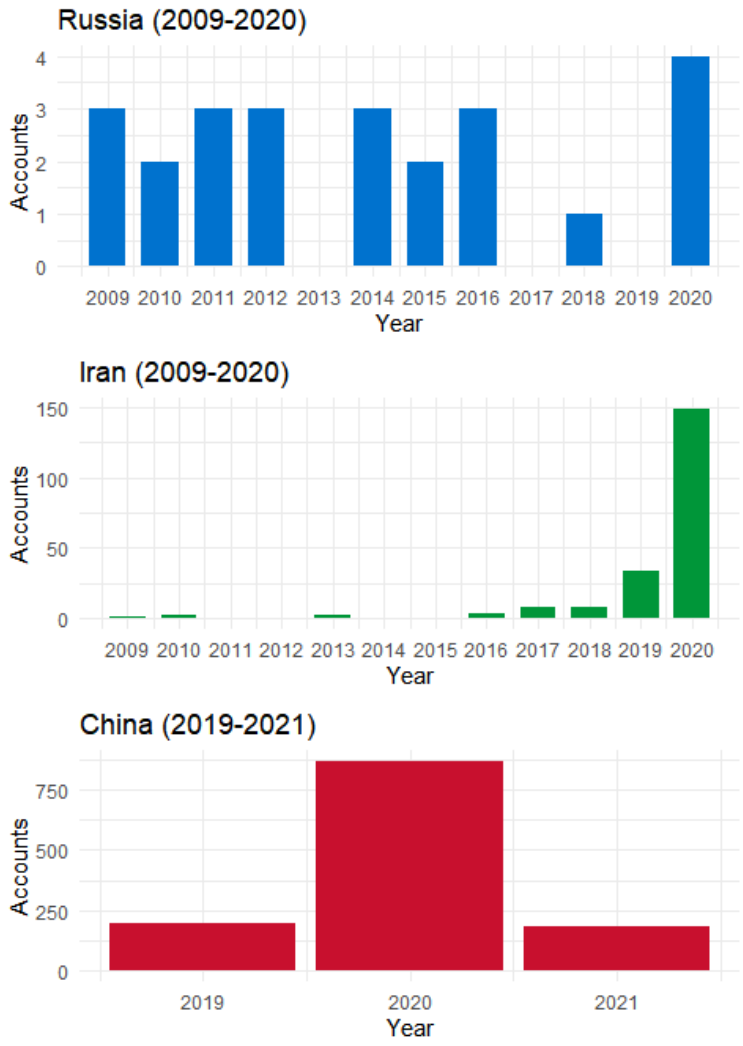

Figure 2: Distribution of Account Creation Dates by Information Operation

Turning to analysis of follower counts, the average number of followers for accounts in the Russian IO was 1,815 while the median was 178.5, suggesting a strongly right-skewed distribution. The histograms in Figure 3 below confirm the presence of right-skewness. The skewness found in the Iranian and Chinese IOs is also very pronounced. The Iranian IO had an average of 1,452 followers per account but a median of only 118. The Chinese IO by contrast had an average number of 2.7 followers per account and a median value of zero. Many of the accounts in each dataset had hashed usernames, as noted in the methodology section. The top three accounts in the Russian IO dataset were not among the accounts that had hashed usernames. Their usernames were *ValdaiClub*, *bluevvelvet*, and *uvidata*. Together, these three accounts had over 29,000 followers. The top seven accounts in the Iranian IO had unhashed usernames. The top account, *Hispantv*, had 161,512 followers. Accounts *ParadisMireille* and *Atrumphater* were a distance second and third with 11,036 and 6,528 followers, respectively. In the Chinese IO dataset, the top three accounts were all hashed values, but their follower counts were 485, 385, and 342, respectively. These follower counts for the top Chinese accounts were much lower than either the Russian or the Iranian datasets were. Indeed, the entire Chinese IO had fewer overall followers; over three quarters of the other accounts in the

Chinese IO had less than two followers each. In each of the three IOs, a handful of accounts dominated the others in follower count.

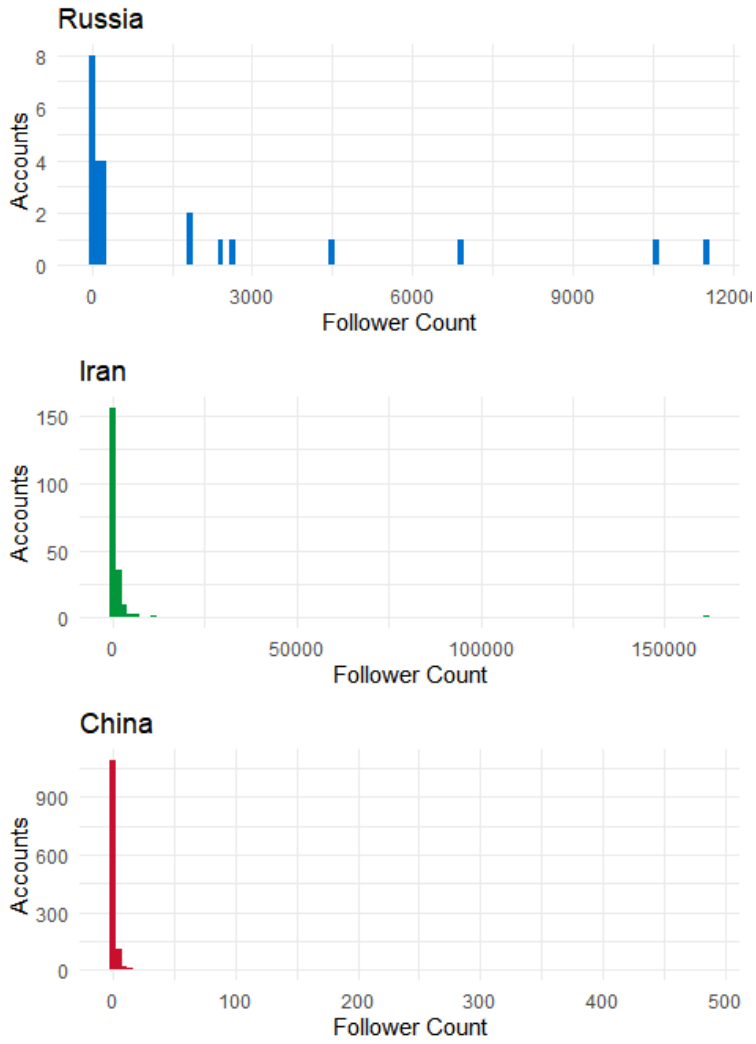

Figure 3: Distribution of Twitter Followers by Information Operation

With the distribution of accounts followed, there is again strong right-skewness present in each of the three IOs, as seen below in Figure 4. The Russian accounts followed 305 other accounts on average, though the median value was only 169.5. Iranian accounts followed 857.5 accounts on average, with a lower median of 250. Chinese accounts followed on 15.6 accounts on average, with a median of seven. The top nine Russian accounts were usernames with hashed values, though number ten was once again *bluevelvet*. The top three accounts followed 2,642, 765, and 613 other accounts, respectively. Three quarters of the Russian accounts followed 267 or fewer other accounts. Two of the top Iranian accounts were also familiar names. Account *ParadisMireille* followed 11,652 accounts while *MariequMoi*, appearing for the first time thus far, and *Atrumphater* followed 6,122 and 5,932 accounts, respectively. Three quarters of the Iranian accounts followed 1,035 or fewer other accounts, making Iranian accounts the most active in terms of both followers and accounts followed. China was again the least active on these metrics. The top three accounts for China had

their usernames hashed, but each followed 749, 571, and 445 other accounts, respectively. Three quarters of the accounts followed 14 or fewer other accounts.

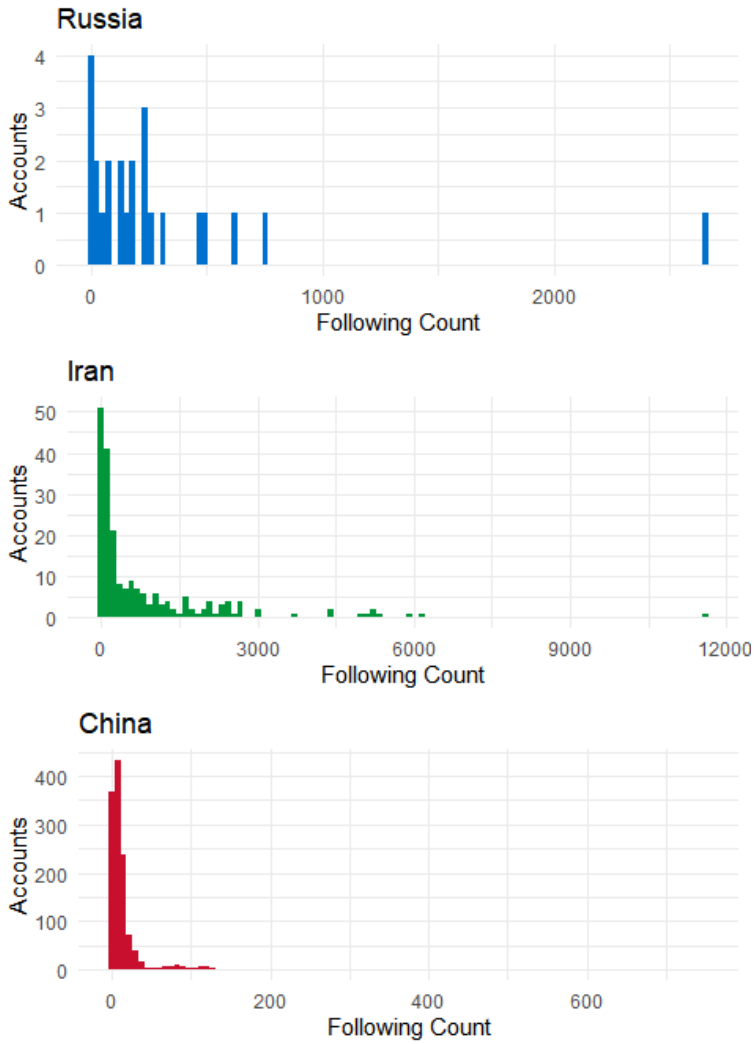

Figure 4: Distribution of Accounts Followed by the Accounts in Each Information Operation

The analysis now turns to the proportion of retweets in each IO's dataset. As can be seen in Table 2 below, the majority of the tweets in each dataset were not retweets. The Chinese IO contained the lowest number of retweets, with only 10.23 percent of the tweets consisting of retweets. Meanwhile, the Russian IO retweeted the most, with almost one third of the tweets in the dataset consisting of retweets. The Iranian IO was in the middle with just under 18 percent retweets.

Table 2: Retweet Statistics of Each Operation

	Retweet	Not Retweet
China	10.23	89.77
Iran	17.92	82.08
Russia	31.60	68.40

The remainder of this section is dedicated analysis of amplification metrics. In this sense, amplification metrics are the count, mean, and median of quote tweets, replies, retweets, and likes for the tweets in each IO's dataset. These metrics give a sense of how many other Twitter users knowingly or unknowingly amplified a particular tweet by sharing it with others or interacting with the tweet in some way, such as liking it. Tweets with high interactions are more likely to be served to other users. The visuals of the distributions in this section do not appear here due to their extreme right skewness, which persisted even after normalization attempts, making it nearly impossible to discern any meaningful characteristics visually. The code to reproduce the skewed visuals is, however, present in the R scripts for this project, available at https://github.com/CWilson01.

The first amplification metric is the number of quote tweets. The Russian IO was the least often quoted of the three IOs, with an overwhelming majority of tweets in the Russian IO having zero quotes. This led to a median of zero and a mean of 0.038, suggesting an extreme right skew. The most popular tweet received 389 quotes tweets. The second and third most popular were quoted 45 and 29 times, respectively. Accounts *ValdaiClub*, *bluevvelvet*, and *uwidata* appeared several times in the top ten most quoted tweets, suggesting they were more highly amplified according to this metric than other accounts in the IO. The Iranian IO followed a similar pattern with over three quarters of the tweets in the dataset having no quotes, leading to a median of zero and a mean of 0.24. The most popular tweet, however, received 1,155 quotes. The second and third most popular were quoted 1,109 and 617 times, respectively. Accounts *Hispantv* and *ParadisMireille* appeared several times in the top ten most quoted tweets, again suggesting they were more highly amplified accounts according to this metric than others in the network were. The Chinese IO had only slightly more quotes on average than the Russian IO, with over three quarters of the dataset having zero quotes, leading to a median of zero and a mean of 0.039. The top three most popular accounts produced tweets that received 88, 84, and 67 quotes, respectively. While the top quoted account names all had their usernames hashed, they do not appear to be the same accounts that were highly followed, suggesting more diffuse user interaction with the Chinese IO.

The second amplification metric is the number of replies to tweets in each dataset. In terms of replies, the Russian IO was in the middle of the pack, with a mean of 0.098 replies per tweet and a median of zero, again suggesting an extreme right skew. Over three-quarters of the tweets had no replies, further supporting the presence of a right skew. An account with a hashed username produced two of the most replied to tweets. One tweet from this account had 190 replies and another tweet had 85 replies. Account *uwidata* was third with a tweet that had 68 replies. The Iranian IO was again the top network overall on this amplification metric. Tweets in the Iranian dataset had 0.42 replies per tweet on average and a median of zero. The distribution had a right skew, with three quarters of the tweets having no replies. The most replied to tweet was made by account *78williamjones* and had 2,222 replies. Second and third were hashed user names with 1,385 and 644 replies, respectively. Accounts *ParadisMireille* and *Hispantv* also appeared in the top ten. The Chinese IO was the least replied to network with an average of only 0.006 replies per tweet and a median of zero. The distribution had a right skew, with three quarters of the tweets having zero replies. The top three most replied to tweets were all from accounts with hashed usernames, each having only three replies.

The third amplification metric is the number of retweets to each tweet in the datasets. The Russian IO was again in the middle of the pack, having 0.48 retweets per tweet on average and a median of zero. The data again contained a right skew, with three quarters of the tweets having no retweets. The top three most retweeted tweets garnered 264, 221, and 143 retweets, respectively. Account *uwidata* appeared multiple times in the top ten and *ValdaiClub* appeared once. The Iranian IO was

again the top with the most retweets per tweet at 5.6 on average and a median of zero. The dataset had a right skew but the third quartile value was four retweets, instead of zero. The top three tweets, each from hashed username accounts, received 26,269; 9,967; and 4,503 retweets, respectively. Accounts *Hispantv* and *ParadisMireille* also appeared in the top ten two and three times, respectively. The Chinese IO was again the least interacted with in terms of retweets, receiving an average of only 0.004 retweets per tweet and a median of zero. The third quartile value was zero, and the top tweets only garnered 10, 7, and 6 retweets, respectively, each from the same hashed username account.

The last amplification metric is the number of likes to each tweet in the datasets. The Russian IO was again middle of pack. The tweets in the dataset received on average 0.54 likes per tweet with a median of zero. The third quartile value was zero, suggesting another right skew. Three accounts dominated the top ten most liked tweets. One hashed username account appeared seven times alongside *uwidata* and *ValdaiClub* in the top ten. The top three most liked tweets had 923, 331, and 250 likes, respectively. The Iranian IO was again the most interacted with in terms of likes. On average, the tweets received 5.19 likes per tweet with a median of zero. The third quartile value was two, suggesting right skewness. Each of the top ten most liked tweets were hashed username accounts. The top three tweets respectively received 63,863; 18,145; and 12,618 likes. The Chinese IO had the fewest likes with an average of 0.02 likes per tweet and a median of zero. A right skew appeared in the data, with the third quartile having a value of zero. The top three tweets received 43, 24, and 19 likes, respectively.

Analysis and Implications

The analysis of these three IO networks produced several key takeaways:

- With the exception of China, which is somewhat newer to the externally facing IO space, accounts used by the information operations were created and actively tweeting long before their detection. This increased the time horizon that other users could encounter material from these accounts.
    - This behavior is in keeping with behaviors observed in other IOs. Many IO accounts go about building a following by sharing material that people want to believe is true or through so-called "seeds of truth," which are small pieces of accurate information amid inaccurate information intended to increase a message's credibility. These tactics make the account appear genuine and appealing, before switching to more overt mis/disinformation (Cailin O'Connor & James Owen Weatherall, 2019). This can be a part of a long-term strategy by the actor operating the account. In other instances, bot accounts on social media have been repurposed from one information operation to another (Persily & Tucker, 2020).
- Tweet virality appears to be a critical component to both the Iranian and Russian operations. Most accounts in these operations saw little engagement, but a few accounts enjoyed some amount of viral success, which appears to have produced some scaling behavior. The result of this is that these few accounts attracted almost all of the attention while the rest languished. This power law-like behavior produced extreme skews across numerous metrics where a handful of accounts enjoyed all the success. This likely fed further interactions with those accounts, bolstering the metrics of subsequent tweets from those select few accounts.
- Having more accounts in an operation does not always translate to a greater chance of virality or success. China fielded 66 times more accounts than Russia and over eight times as

many accounts as Iran. Despite this, all of the Chinese accounts combined likely reached fewer people than just the top Iranian account.

- The Iranian IO likely reached the largest amount of people of the three IOs. It is plausible that the Iranian network may have reached hundreds of thousands of people. The fact that the most followed account operated by the Iranian IO had over 100,000 followers by itself and tweets from top Iranian accounts amassed well over 10,000 retweets on multiple occasions further supports this conclusion.
    - Moreover, it seems that the Iranian IO positioned itself well to capitalize on the viral potential of social media. While most of the accounts still saw little interaction, the accounts that did capture attention captured it in a big way and translated that into high follower counts, ensuring continued exposure of the propaganda.
    - Further research is needed to elucidate more fully the mechanism behind this selective virality.
- Unlike Iran, the Chinese IO was much more diffuse in terms of how Twitter users interacted with it. Accounts with the highest followers, were often not the accounts with the highest amplification metrics. Accounts that were high in one amplification metric were often not in the top ten in others. This diffusion may have contributed to the overall lack of virality potential present across the IO's content. Based on these factors, it appears likely that the Chinese IO only reached users in the thousands.
- The Russian IO likely yielded a low-to-moderate return on investment for the Kremlin. From extensive research by Thomas Rid, it is known that prior IRA operations operated around the clock with high costs and reached relatively few Americans. Based on Rid's analysis, the so-called "Russia investigation" likely brought more eyes to IRA propaganda than the operation itself did (Rid, 2020). Thus, Russia likely only fielded the 24 accounts present in this dataset because they were expensive to operate. This may also explain why the Russian network retweeted content more often than the other two IOs—as a cost saving measure to avoid having to create original content. Based off follower counts topping out in the low thousands and amplification metrics in the hundreds, it is likely that the Russian operation only reached users in the low tens of thousands.
- The IO threats receiving the most media attention, Russia and to a much lesser extent China, were the least effective, despite, or perhaps because of, the closer attention paid to them. While this could be interpreted to mean that keeping very close attention on these threats keeps them somewhat suppressed, this is likely unsustainable as other threat actors enter the fray. There are finite resources available to watch for threats, meaning as new threats emerge, there will not be the same level of resources available to deal with them. Thus, a generalizable threat detection and response model is necessary.

Conclusion

While most of the accounts present in the IO networks analyzed here had relatively limited reach, the scalability present in a few select accounts highlights the importance of quickly detecting and removing malign actors from social media platforms. Accounts or tweets that go viral at the scale seen in the Iranian dataset could easily be amplified even further by the other accounts in the network, pushing that material over the edge of what is known as the "breakout scale." Ben Nimmo proposed the breakout scale in a 2020 report published by the Brookings Institution. In the report, Nimmo lays out a framework for analyzing an IO's reach, which culminates in the IO escaping its original platform to spread to other platforms and ultimately across mediums, eventually reaching

mainstream audiences(Ben Nimmo, 2020). If the Iranian IO had its tweets seen by a prominent influencer who then retweeted it, the IO could easily have catapulted into the mainstream view.

This analysis is only the first of many possible avenues of further research. From just the datasets used in this report, four additional research projects are currently planned:

- First, are there similarities or differences between each operation when conducting a social network analysis? Furthermore, if data from a baseline network of general, legitimate Twitter conversations about some random topic were introduced to the analysis, would there be statistically significant differences discernible between the IO networks and the baseline?
- Second, using the same networks, what would happen if the three IO networks were combined together and the two baseline networks were combined together? Would the comparison change? How would the network structure of the unified IO network evolve? Would overlap or common nodes make themselves apparent? Would there be important nodes that would stand out as connections between the three IOs? Would there be similar statistical results on the networked metrics?
- Third, what would a text mining and sentiment analysis of the hashtags and tweet text in the IOs reveal? If a text-based analysis of IO networks were introduced, would that enhance possible detection of IO networks? Are there similar tactics, techniques, and procedures in terms of how language and hashtags are used across the IOs?
- Fourth, is it possible to split each IO and baseline dataset in half, train an algorithm on the first half, and then test the algorithm against the other halves of the datasets combined together into one large network? Could the model detect the IO sub-networks amid noise from the baseline datasets?

Beyond these future projects, there is the question of if the results discussed in this report hold across other IOs. Moreover, can evolving TTPs interfere with the ability to produce generalizable results? Additionally, while conducting the descriptive analysis of each of the IO networks, one additional question was considered that could not be answered without access to internal Twitter data—"with regard to overall reach of each information operation, how many people were these tweets served to?" Such a question could be an area of further study for researchers that can gain access to such impression data through a partnership with Twitter, perhaps.

References

Ashley A. Mattheis, Daveed Gartenstein-Ross, Cody J. Wilson, & Varsha Koduvayur. (2022). *Blind Sided: A Reconceptualization of the Role of Emerging Technologies in Shaping the Tactics, Techniques and Procedures of Information Operations in the Grey Zone* (pp. 1–102). Valens Global. https://valensglobal.com/blind-sided/

Ben Nimmo. (2020). *The Breakout Scale: Measuring the Impact of Influence Operations*. Brookings Institution. https://www.brookings.edu/research/the-breakout-scale-measuring-the-impact-of-influence-operations/

Cailin O'Connor & James Owen Weatherall. (2019, September). How Misinformation Spreads—And Why We Trust It. *Scientific American*, *3*(321), 54–61.

Daveed Gartenstein-Ross, Madison Urban, & Cody J. Wilson. (2022, July 27). Anti-Censorship Legislation: A Flawed Attempt to Address a Legitimate Problem. *Lawfare*. https://www.lawfareblog.com/anti-censorship-legislation-flawed-attempt-address-legitimate-problem

Jean Le Roux. (2020, August 19). South African Twitter accounts gamed trending algorithms to promote prank political hashtags. *Digital Forensic Research Lab*. https://medium.com/dfrlab/south-african-twitter-accounts-gamed-trending-algorithms-to-promote-prank-political-hashtags-7ad1c6cb0622

Persily, N., & Tucker, J. A. (Eds.). (2020). *Social Media and Democracy: The State of the Field, Prospects for Reform* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108890960

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.

Twitter Safety. (2021a, February 23). Disclosing networks of state-linked information operations. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2021/disclosing-networks-of-state-linked-information-operations-

Twitter Safety. (2021b, December 2). Disclosing state-linked information operations we've removed. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed

Twitter Transparency Center. (2022). *Information Operations*. Twitter. https://transparency.twitter.com/en/reports/information-operations.html

About the Author

Cody Wilson is a researcher focused on solving societal problems, particularly those created or exacerbated by modern and emerging technologies, through the application of research and data analysis. He was a student in Google's Data Analytics Professional Certification program, which concludes with a capstone analytics project. This report serves to fulfill that certification requirement. He has worked as a full-time national security analyst and consultant at Valens Global and a volunteer research analyst focusing on terrorism and radicalization with the virtual think tank NextGen 5.0. He holds a master's degree in global studies and international relations, with a concentration in conflict resolution, from Northeastern University. Cody previously earned a bachelor's degree in political science with a concentration in international relations from the University of California, Los Angeles.

As an undergraduate, he completed a capstone research project focused on Iran's nuclear program, and in graduate school, he completed two additional significant research projects, one on the recruitment of women by ISIS and the other on the drivers of, and policy responses to, the decades-long conflict in Somalia. After graduate school, Cody produced a study examining the relationship between online terrorist propaganda and real-world terrorist activity. During his time at Valens Global, Cody contributed to reports used in federal litigation, helped execute eight successful wargames, designed a cybersecurity-focused tabletop exercise, produced technical and analytical reports for the U.S. government, and co-authored a study putting forth a framework for better understanding grey zone information operations.