Published on *USC Center on Public Diplomacy* (https://uscpublicdiplomacy.org)

Thumbnail Image:

Feb 26, 2019 by Ali Fisher

Mapping Russian & Iranian Cyber Networks

[1]

As the sources and instruments of power have adapted to a new information age, the extent to which different groups have access to influence has also shifted.

Public diplomacy in the 21st century has to navigate the complex architecture of multi-hub, multi-directional networks that exist between communities around the world. As a result, calibrating public diplomacy and strategic communications to face specific challenges requires fine-tuning the specific approach to engagement with foreign publics as well as an understanding of the tactics employed by any adversaries who wish to influence those same publics.

This complexity is not a new phenomenon, but the information age has added an additional layer of methods to communicate or share information, and additional opportunities to analyze available data. This is true whether discussing how the Media Mujahidin seek to spread their theology, Russian attempts to influence political discourse, or attempts by political

cyberhackers to influence elections.

Influencing the Information Environment

There are many ways nations seek to influence the information that individuals use to shape their individual perspectives. In addition to the usual public diplomacy and cultural relations infrastructure in the UK, the 77th Brigade was created to contest information wars. As Carl Miller wrote, "They are soldiers, but the 77th Brigade edit videos, record podcasts and write viral posts. Welcome to the age of information warfare." This decision is "an unorthodox one, but in the eyes of the British Army also a necessary innovation; simply reflecting the world in which we all now live and the new kind of warfare that happens within it."

For public diplomacy, the ability to identify opportunities for influence—and the ability identify the influence others exert—within an information environment requires both tools and the diplomats able to use them.

For the UK military, information operations play a key part in the non-lethal approach to targeting, but they are not alone. There has been much discussion of influence operations targeting Facebook which has sparked interest in developing new insight into the impact of social media on elections and the subsequent data release by Facebook.

Similarly, Twitter has noted that "it is clear that information operations and coordinated inauthentic behavior will not cease." As part of their effort to protect what they term "a healthy public conversation," they have released data relating to their investigation "into foreign interference in political conversations on Twitter." This disclosure includes details of activity "potentially connected to a propaganda effort by a Russian government-linked organization known as the Internet Research Agency."

This is an example of how those responsible for public diplomacy, strategic communications and information operations can utilize newer communication methods to exert influence. The specific *Options for Influence* that they choose may be different, but they are all operating in the same spaces and on the same platforms. This data release also highlights that for public diplomacy scholars and practitioners there are greater opportunities to analyze how others are using similar spaces and platforms to exert influence.

Analyzing an Information Environment

The recent Twitter data release relating to Russia and Iran demonstrates opportunities to gain insight into how platforms are being exploited by those conducting information operations, but only if public diplomacy organizations can handle the data appropriately.

This means there is a need to extend the level of data analysis which is taught during degree level or postgraduate courses relating to public diplomacy. Equally the culture within organizations conducting public diplomacy may have to increase the values they place on the ability to use the host of programming languages freely available.

Those interested may want to try Jupyter and follow guides such as Mining the Social Web or Python for Data Analysis (Jeffrey Stanton's Introduction to Data Science provides a similar introduction for anyone who would prefer to focus on R).

Beginning with Jupyter, users can take advantage of a single start point and expand out to experiment with a wide range of kernels including Python, R, Bash, Perl, Lua, Java or MATLAB and unlock the analytical capabilities of libraries and packages such as Pandas and Networkx. No matter the specific language someone chooses, it should allow students, scholars and practitioners to develop the skills and greater flexibility in handling data to get beyond the common overreliance on Excel.

To demonstrate the opportunity to analyze the information environment, the following analysis of information operations on Twitter was conducted using only freely available and open source resources, including Jupyter, run on a desktop PC that is between three and four years old.

The Data

According to the statement on the Twitter blog:

> *These large datasets comprise 3,841 accounts affiliated with the IRA, originating in Russia, and 770 other accounts, potentially originating in Iran. They include more than 10 million Tweets and more than 2 million images, GIFs, videos, and Periscope broadcasts…*

This is a large amount of data in public diplomacy terms: 297 GB and 65 GB relating to Russian and Iranian operations, respectively.

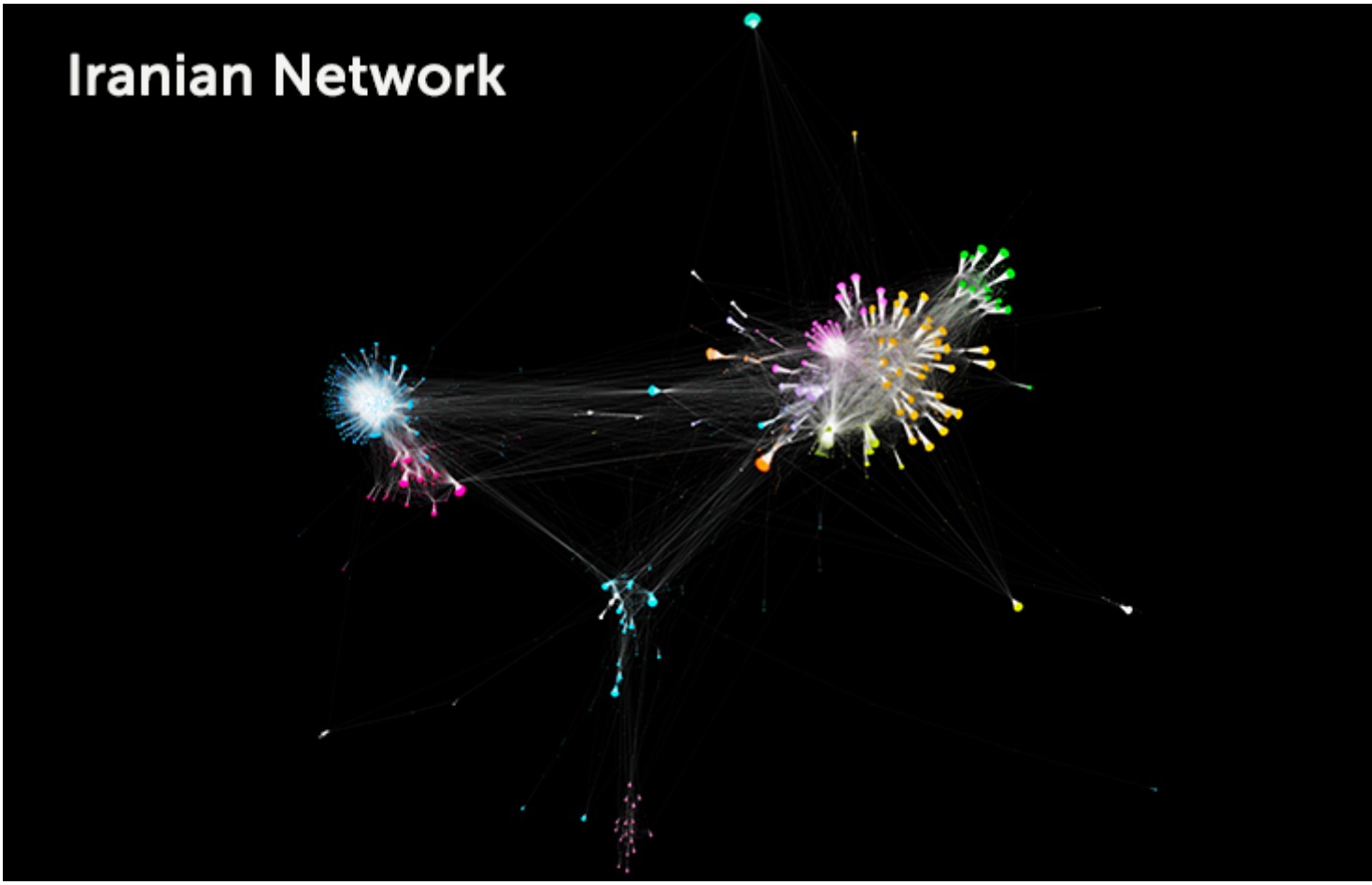
Using Python and Pandas within Jupyter, one can read the data provided by Twitter and locate all the retweets. As Excel only allows 1,048,576 rows per worksheet and the Russian data alone contains approximately 3,333,000 retweets, the need for access to alternative data handling solutions becomes clear.

The 3.3 million retweets result in a network of approximately 205,000 accounts made up of 844,000 tweet and retweet connections. Once you can create the network, it can be either analyzed within Jupyter—options include Python package Networkx, or igraph if using R. For those with a preference for a more visual form of analysis, data can be exported for analysis in Gephi.

Important note for those who analyze the network visually: ensure the layout has run for long enough. If it is still basically square, you almost certainly haven't let it run long enough. Visual analysis based on an incomplete network layout are invariably baseless nonsense. This type of error can be seen in the ISIS Twitter Census (p.57).

The Networks

Using Gephi, we find in the “Russian” network that 99.9 percent of nodes are connected within a single “giant component” but that some parts of that network are more interconnected than others, as the giant component contains 19 statistically distinct communities. Using the same method on the “Iranian” data, there are 232,000 retweets, connecting 32,700 accounts and 64,500 tweet and retweet connections. 99.8 percent of the accounts connect to make up a single “giant component.” This time, the giant component contains 23 statistically distinct communities (again visualized using Gephi).



This overview shows that in both Russian and Iranian cases:

- There are many more accounts in the network than those accounts identified by the original data release.
- There are a number of different hubs of activity with little connection between them. This allows them to reach out in different directions.

Collectively the overview of these information environments shows that both Russian and Iranian information operations are drawing on content produced by other accounts, but for various reasons these accounts serve their purpose. With control of the data, public diplomacy organizations could find information about the accounts producing content which is subsequently exploited by the Iranian or Russian information operation using a simple call to the Twitter API.

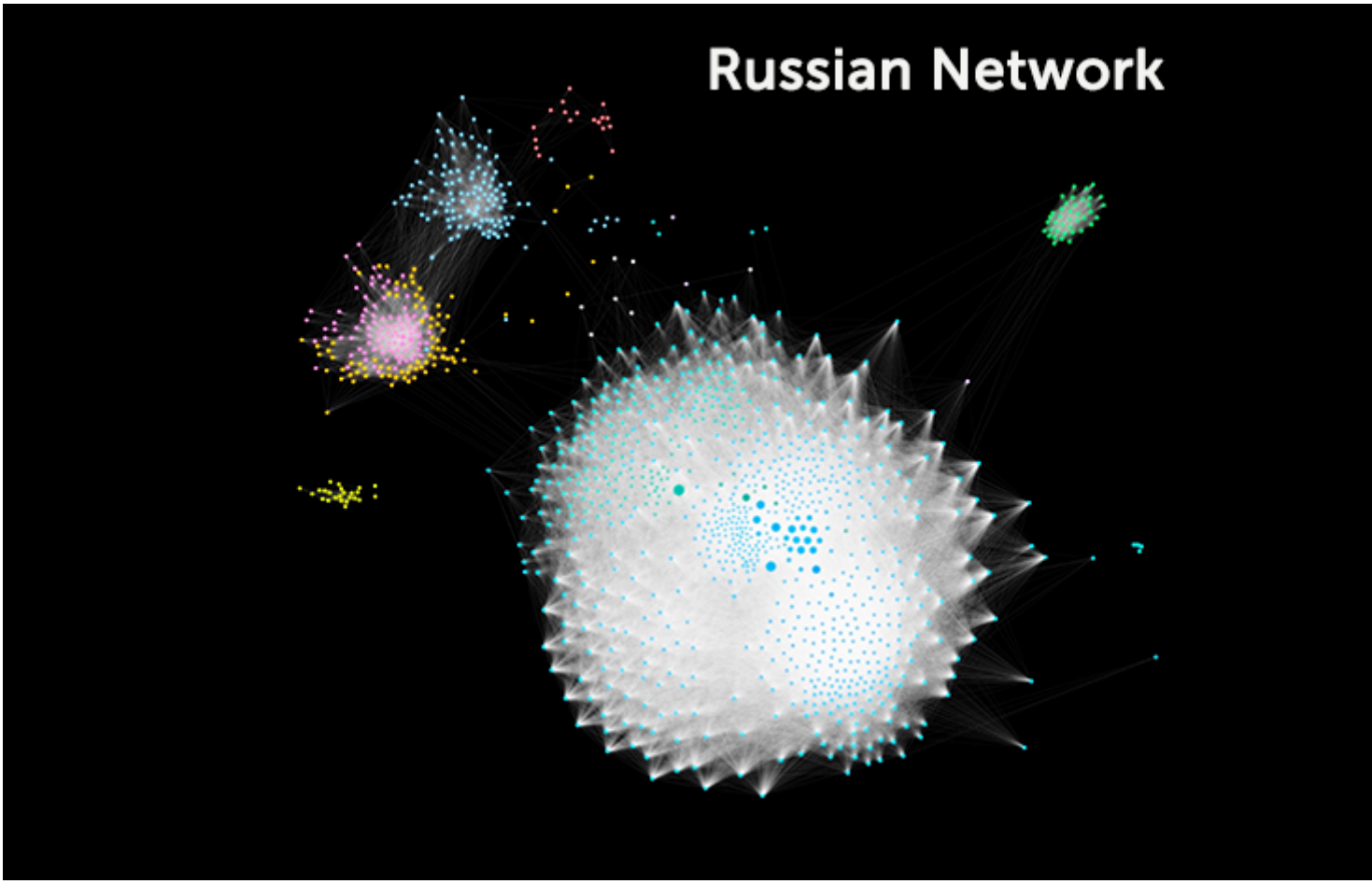
This information would provide another perspective on the different online communities which were being specifically targeted and would help triangulate findings from other methods. For example, community targeting was also uncovered by the Atlantic Council's Digital Forensic Research Lab in their analysis of the #tags and URLs being shared. The methods applied by

the Digital Forensic Research Lab to find the #tags or URLs commonly used could be applied at the level of specific communities within the network, once you know which accounts are in which cluster just that content can be analyzed.

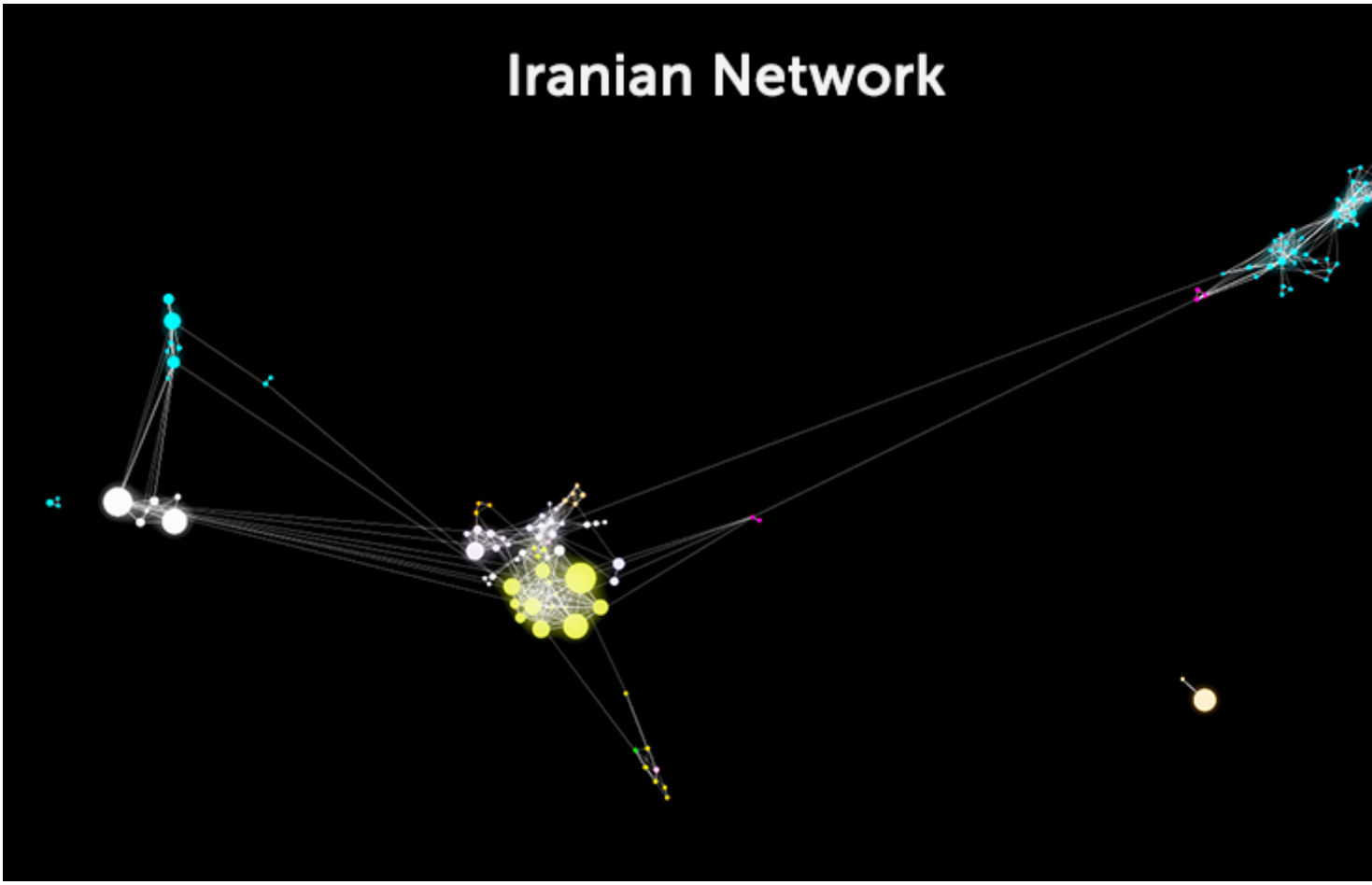
Working Together

One benefit of analyzing content based on the specific clusters is the greater level of granularity in the analysis. For example, are accounts alleged to be part of the information operation working together, or merely tweeting the same content, URL and #tags?

Analysis of the overall network provides some insight, which can be extended if just the accounts which have a mutual degree (two accounts that both retweeted each other) are analyzed. This means both are thought to be part of the information operation. The Russian network, when filtered using mutual degree, contains 1,198 accounts. This is 0.58 percent of the total accounts in the network but includes 13 percent of the edges (lines representing retweet relationships) present in the total network.



In the Iranian network, the number of accounts drops to 219, 0.67 percent of the original nodes and only 1.59 percent of the edges. As one would intuit from the images, the statistics indicate that the accounts which are believed to be part of the Russian information operation worked together much more closely than those in the Iranian version. The Iranian version engaged more frequently with accounts not identified as part of the information operation.



Conclusion

The tools for developing innovative strategies for public diplomacy in the big data era have been evolving for many years as commercial tools and freely available programming languages. Some influence and information activity will be conducted in the shadows, whether this is by GCHQ, who like the UK military also had an information warfare unit, with their tools having code names like “Badger,” “Gateway,” “Burlesque” and “Clean Sweep,” or other actors. But much more will happen in the open.

For public diplomacy, the ability to identify opportunities for influence—and the ability identify the influence others exert—within an information environment requires both tools and the diplomats able to use them. Some will come from “off-the-shelf” commercially available tools. However, as shown here, there are many instances where data could be analyzed and visualized by diplomats with freely available tools.

It is a question of organizational culture and training. *The Economist* recently reported that Python is becoming the most popular coding language, but how many diplomats could use it or any of the many other programming languages freely available to deliver insight and influence?

Public diplomacy may not use information operations methods alleged by Twitter in their data release. However, having the skills to develop an appropriate situational awareness of what others are doing in the same place or space continues to be vital to the successful planning and practice of public diplomacy.