Matt Sellman

SOCIAL MEDIA, CROWDSOURCING AND CITIZEN SENSING

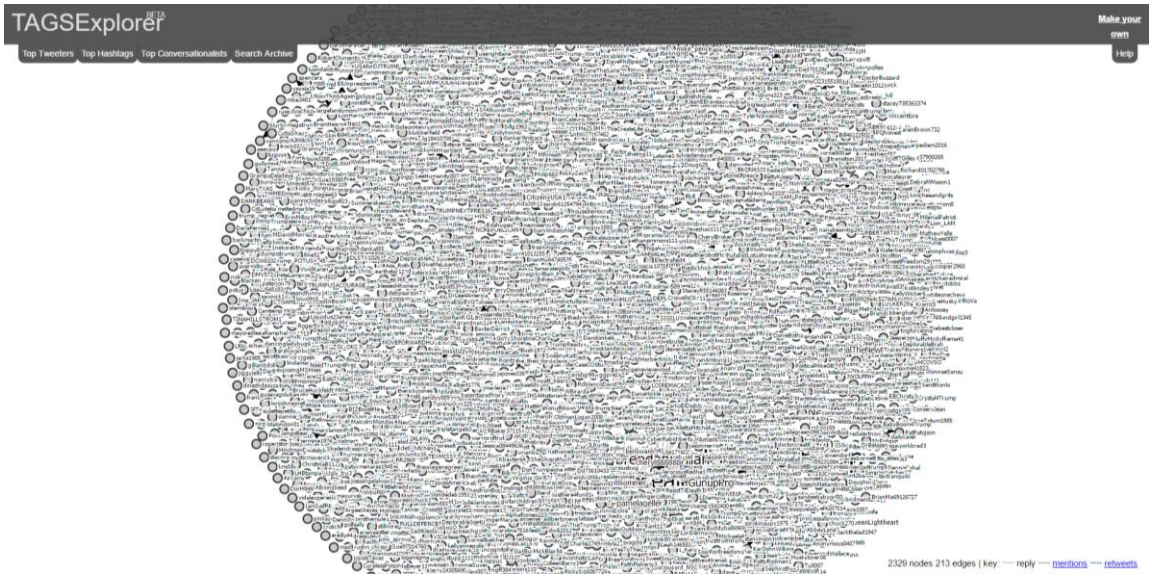
Finding Russian Pro-Trump bots on twitter and verifying their origin.

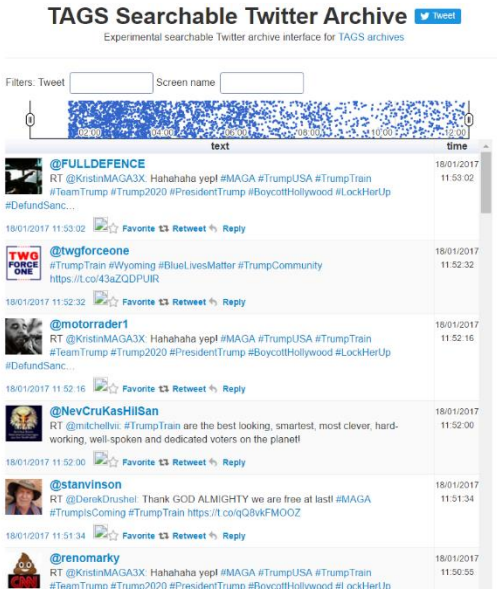
Starting evidence: -

We are all aware of trolls and bots in some form or another, but in my practical assessment, I wanted to see to what extent the Russians had interfered in the US presidential election, to do this I wanted to carry out an analysis of twitter bots, to see where these operated and how linked the accounts truly were. What I wanted to find was evidence of Russian bot activity and can find at least five bot accounts that supported Donald.

Methods:

The first thing I needed to do was to find a phrase that was synonymous with all the bots used by the kremlin, to do this I took a quick look through twitter and the wider internet for an event that all Russian bots would be tweeting about, in the end I chose the US election, since that was an event that had been influenced by the Russian ministry. The first thing I looked at was the hashtag #trumptrain, this led me to a large map of a large number of users. The

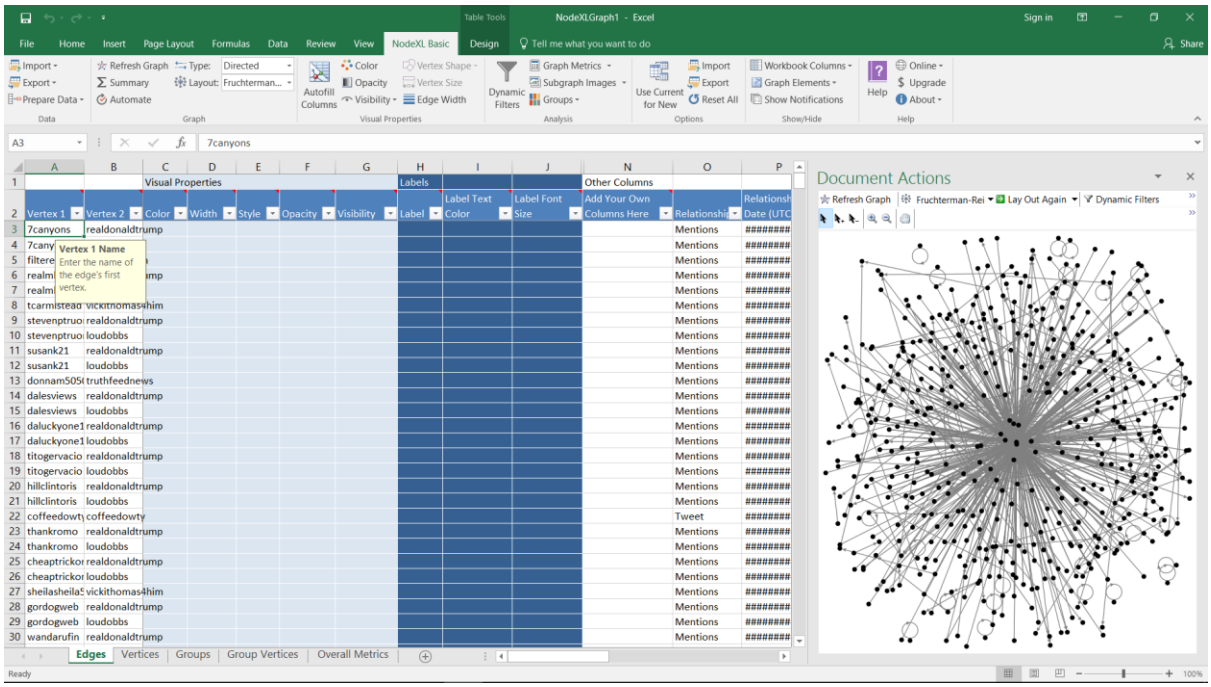




network graph above shows many unconnected tweets with @realdonaldtrump lurking in the bottom right corner. Although there are several smaller players with a larger than average number of connections within the tags graph. The twitter archive shows an in depth detailing of each tweet which allowed me see how positive or negative the general hashtag trend was towards the Donald.

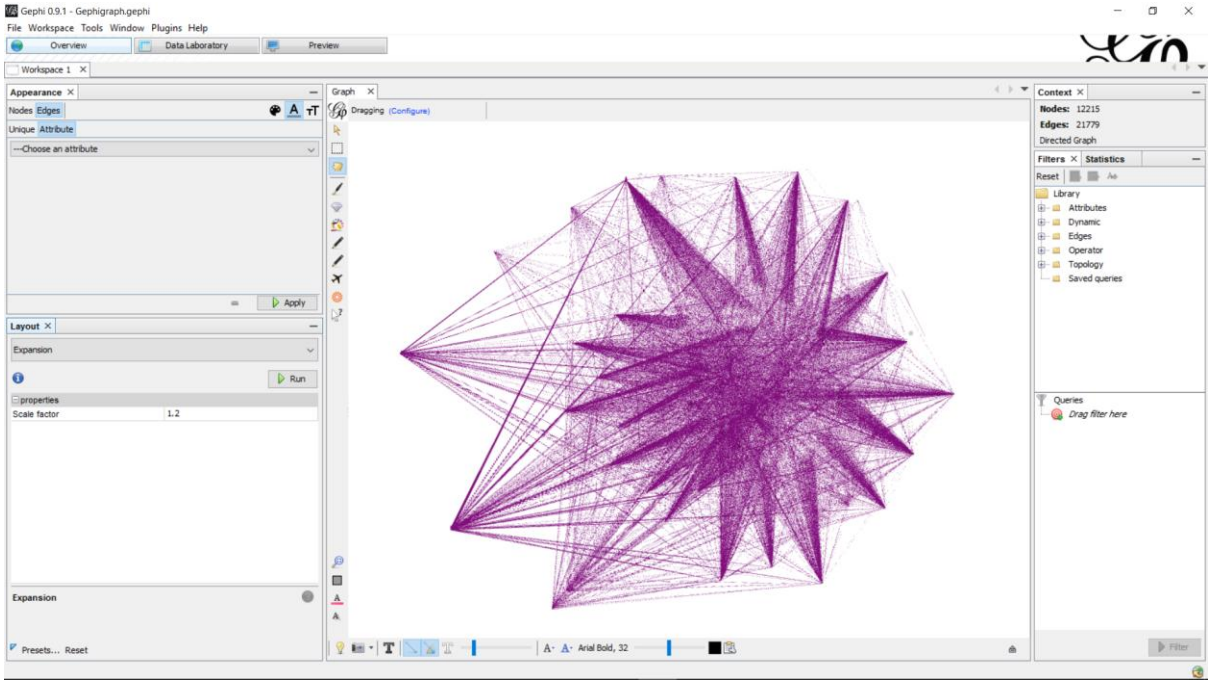
The next step was separating out these users, finding the bots from the humans and the Russians from the Americans. To do this I needed to find a few bot accounts, by doing this I would have at least a small number of Russian bots to work with. To find the Russian bots, I needed to either find a hashtag that went against the grain, for example, a hashtag that calls any accusations (and there are many) ‘lies’ or ‘unsupported’, etc. etc.

I went for the #trumptrain hashtag and using an open source online tool called NodeXL that I discovered from one of the articles from one of the slides, using this I imported a 3000 #trumptrain tweets and I made a graph that showed the relationship between them, what I came up with was a graph that was much more linked than the tags network above.



What this showed me was the relationships between the different accounts, which was interesting given the simplicity of the parameters I had set. The two centre accounts are called robertw08604367 and realdonaldtrump, these are mixed with much smaller accounts with a high number of tweets and a small number of followers.

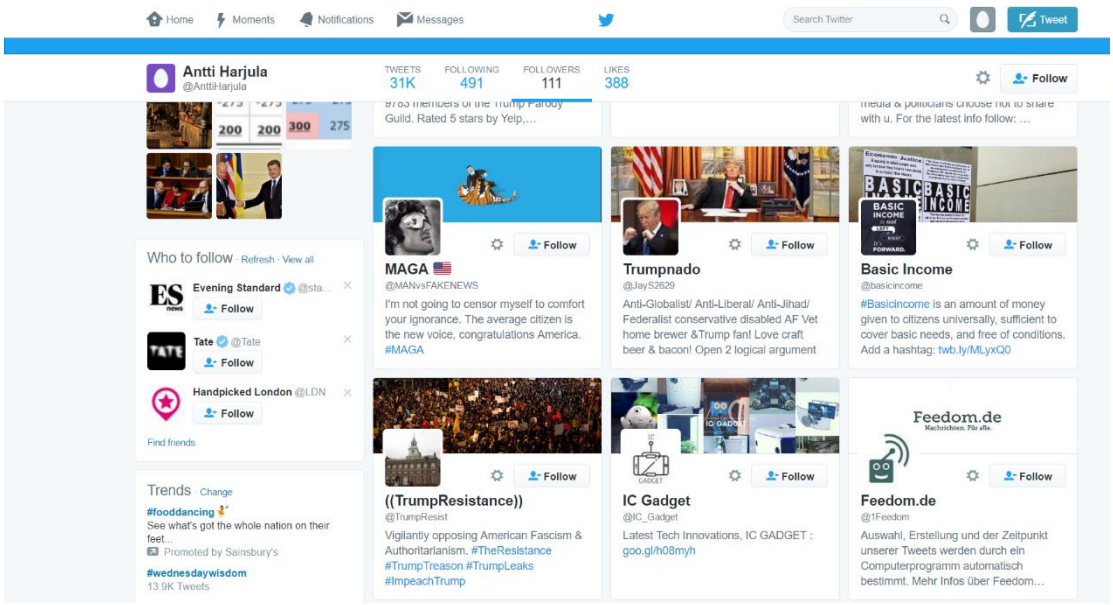
I then exported this as an .csv file that I then put into Gephi, now this is where things got very interesting, because this is where I could see the relationship of the network, what I did was take the data from NodeXL and put it into Gephi where I could see a highly interconnected graph, one that showed me a close community of twitter users interacting with each other.



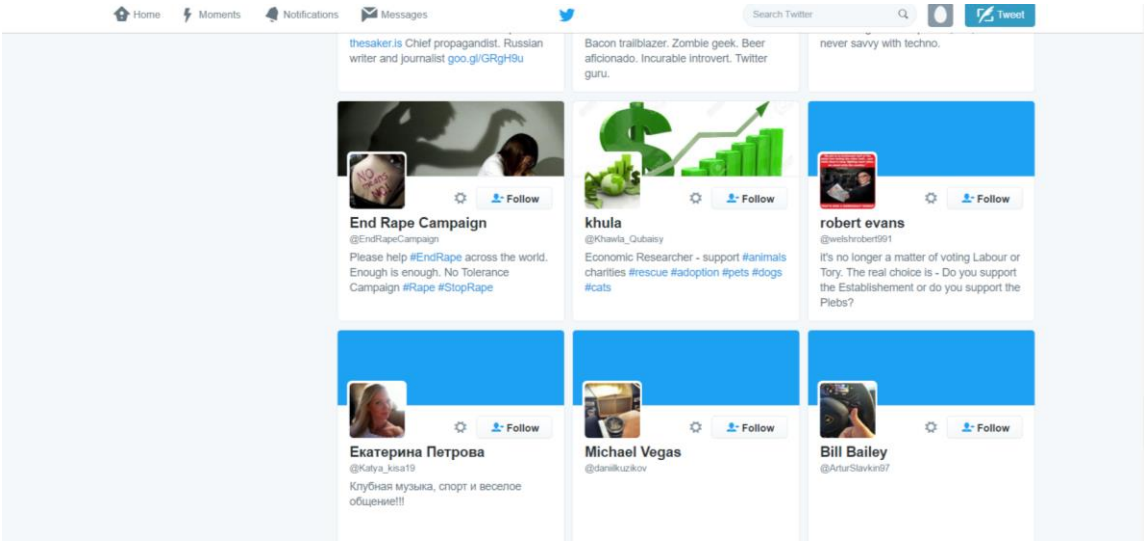
Next came looking at the bots I had found, what I looked for in a bot was a lot of tweets, more than 1000, the next thing I looked for was a mix of liberal and conservative follows then I looked for was a small number of followers, less than 500. From this I figured I had myself a good bot from this. From here I gathered a sample of 5 bots and to make sure that they were bots I put them into a 'bot or not?' checker, http://truthy.indiana.edu/botornot/, from the slides.

From the random sample, I discovered I had four real accounts, although I was unsure of this given that bot or not is not a perfect measure of whether a twitter account is a bot. I then took another random sample of possible pro-trump bots and put them against the bot checker, which gave me an average of 60% possibility of these accounts being bots.

In terms of checking bots, I looked at a number of parameters, for example, you can see below an account that follows an 'anti-globalist, anti-liberal, anti-jihad' trumpnado trump fan, right next to an anti-trump resistance account.



What this led me to was an interesting number of bot accounts, these included an odd mix of pro-trump and pro-liberal retweets, as you can see below there is a bot account that posts in English and follows accounts that speak in English, but I have found a single, purely Russian account that this bot follows amongst all the English following. The account is called AnttiHarjula, however, the botornot web tool told me there was only a 36% chance of this being a bot.



What this did for me was to have solid evidence that I had found a single Russian pro-trump bot. But to make certain of this I wanted to see if I could find the ip address and from there the postcode of the bot.

But there is a problem, firstly, I cannot find the ip address of a twitter, not without getting the account to click on a link, something I could not make the twitter account fall for, despite my attempts to capture their ip address. Secondarily, I was unable to scrape the metadata from the tweets, therefore I could not verify the origin of the tweeter.

Conclusion:

My expected conclusion was to find a number of pro-trump bots with origins in Russia, I found this difficult to achieve because of the sheer number of trump followers and pro-trump tweeters this late in the inaugural process of election. Another issue that perhaps hindered me was that trump has now won, therefore it is possible that the trump bots have now moved onto other subject areas. My actual conclusion was that I found one account that has possible Russian origins, but I was unable to verify its origin.