# Oxycodone and Illicit Drug Abuse on Twitter

Patrick Morabito, Christopher Homan

July 2015

**Abstract**

Something or other about Oxycodone and Illicit Drugs.

## 1 Introduction

(INSERT SOMETHING HERE.)

## 2 Oxycodone Tweet Selection Methodology

We start the process for finding relevant tweets begins by filtering tweets by a selection of keywords. The keywords that we selected were based on some simple quick and dirty on-line sleuthing for information relating to what sort of slang phrases that individuals were using relating to the group of pain killing drugs using oxycodone as one of their ingredients. Additionally, the keyword set contained a listing of different corporately issued market names for the drugs that contain oxycodone as an ingredient. We also included some very simple permutations of each word in the keyword list which were also searched for, to try to quickly account for some possible variation in things such as spelling regarding the phrases. During the search process, we would remove hashtag symbols from the tweets so that the phrase contained in the tweet's hashtag could also be searched for by our keywords, without generating a separate list of hashtag specific phrases.

During the initial sorting process as we fine-tuned the list of keywords, there were some phrases that we had initially sorted for that we removed. We removed some keywords because based on some initial inspection of the filtered tweets, we were finding a massive volume of falsely flagged tweets regarding the keywords relative to small, if any, correctly associated uses of the keywords within the tweets. Therefore, for this initial pass of searching for tweets, we removed some phrases from our keywords so that we could more swiftly hone in on some more usable tweets.

All of the tweets that were flagged in our initial search as containing keywords are sorted into separate file containing just these potentially relevant tweets. We then sorted the tweets by their authors, and then by time and date, so that we were able to gain a slightly more holistic picture of the meaning within a user's tweets. We then inspected each tweet by hand to decide if the tweet was properly flagged as being relevant to our search, possibly relevant to our search (but unable to decide either way based on the contents of the tweet), or not being relevant.

From our list of flagged tweets that we have hand sorted to ensure relevance, we then compile a list of the user ID numbers for everyone who tweeted one of the said tweets. Using this list, we then sort through our data set and pull all pull all of the tweets from all of the users on our list; again sorting them by user and date. This effort is to try to build a more complete picture of the tweeting habits, content, and context for each user's narrative.

## 3 Oxycodone Tweet Results

From our collection of tweets from users in the Rochester area, numbering (INSERT NUMBER OF TWEETS HERE) that were from (DATE) to (DATE), we found 1372 tweets that were flagged by our keyword search

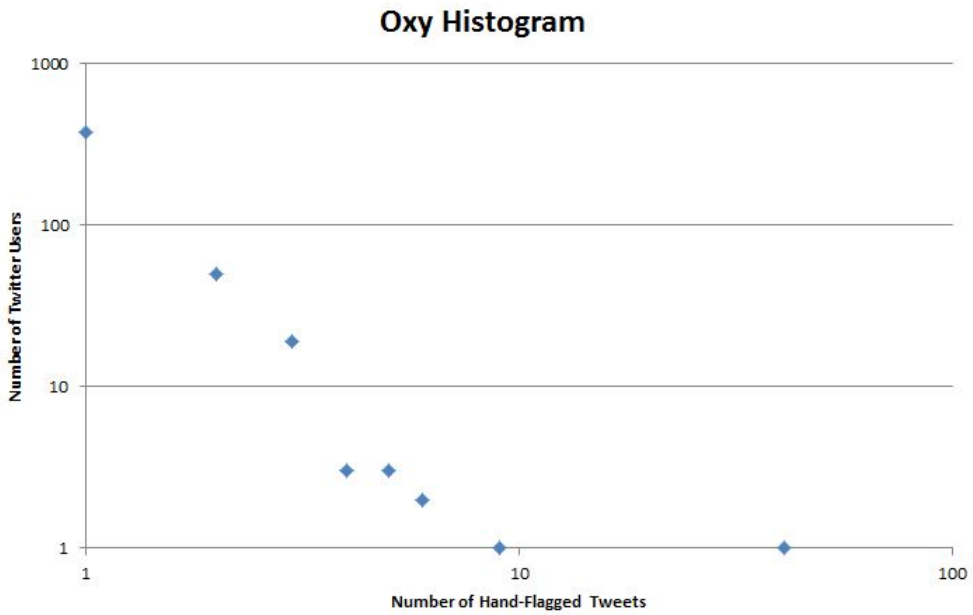


Figure 1: Histogram of positively flagged, hand inspected, oxycodone related tweets.

methodology, of which 621 were found after hand inspection to have been flagged correctly regarding the content of the tweet we were interested in. These 621 tweets were authored by 454 different users. The histogram shown in Figure 1 gives the breakdown for how many users tweeted how many of these tweets.

To give a little bit of the flavor of the kinds of tweets that were flagged, Figure 2 shows the 250 most commonly used words from the 621 tweets.

# 4 Interesting Oxycodone Narratives

After finding users who tweeted regarding oxycodone, we sought to see what sort of narratives the individuals were putting forth overall through their Twitter feed. To do this we first gathered all of the tweets from the aforementioned 454 users who where flagged. (Patrick's Note: Currently I only have the tweets from 196 of the 454 users, because they were the ones that tweeted in the 1 year data file I have.) Next we targeted several users as a starting point, based on the specific content of the oxycodone tweets; preferencing ones we felt would have interesting narratives.

# 5 Oxycodone Users' Interconnectivity Methodology

From our list of users we sought to find any sort of follower and following relations between different users within the list. To do this we check the social graph information for each of these flagged users for any connections to other individuals on the list.

After we would find all of the connections, we then refine our search even more to look for only reciprocal following relations between pairs of users from our list. From these reciprocal pairs users, we sought to see if there were any third parties that formed reciprocal connections with both of the users in the pair, thus forming a triangular relationship between the three individuals.

We then would sort through our body of tweets and pull all of the tweets from each of the users in this new set of users who formed triangles with our initial grouping. We did so using the same methodology from when we gather all the tweets from the flagged users.


Figure 2: Word-Cloud of the 250 most common words from the hand inspected, positively flagged tweets. Words are scaled based on log(n).

## 6 Oxycodone Users' Interconnectivity Results

Using the 07-2015 social graph file, we found that of the 454 users who tweeteded a positively flagged tweet, we had the social graph information from 354 of them. Of the 354, 122 users have some form of interconnectivity among each other. This includes 54 reciprocal connection pairs of users made up of 97 distinct users.(Therefore some users were part of more than one reciprocal pair.) Of these 97 users, we have the social graph information for 84 of them. Of the 54 pairs of reciprocal users, 38 shared reciprocal connections with at least one third user; forming a triangle. From the remaining 16 pairs, 15 were missing one of the two users from the social graph file, meaning building triangles from them is impossible. Only 1 pair from those where we had social graph information for both users formed no triangles. From these 38 pairs who from triangular reciprocal connections with other users, there are a total of 1348 of these triangles that were formed between 1320 unique users. 28 of the 1320 unique users form triangles with two flagged-user pairs. Of the 1320 unique users, zero of them are from the original set of 454 flagged users. Figure 3 shows the histogram for the distribution of the triangles that were formed amongst the pairs of reciprocal users.

These figures varied slightly when we switched to using the 13-2015 graph file. For the newer, denser file, we find graph information for 361 of the 454 users, which 120 of them forming connections amongst each other. There are 55 reciprocal pairs between 99 distinct users, of whom 89 are in the graph file. 43 of the 55 pairs form 1678 reciprocal triangles with 1599 additional individuals. Only one of the reciprocal pairs had both individuals in the graph file, but formed no triangles. Zero triangles had all three members as oxycodone tweeters. Shown below in Figure 4 is the graph of the connections between the 120 aforementioned users. From the 1599 additional people who formed reciprocal triangles with the oxycodone users, 78 of them form more than one reciprocal triangle. Figure 5 is the graph of the connections between the 78 multiple triangle forming users, the 120 oxycodone tweeting users, and 12 other additional oxycodone users who didn't form connections with other oxycodone tweeters, but did form connections with the 78 triangle forming individuals.

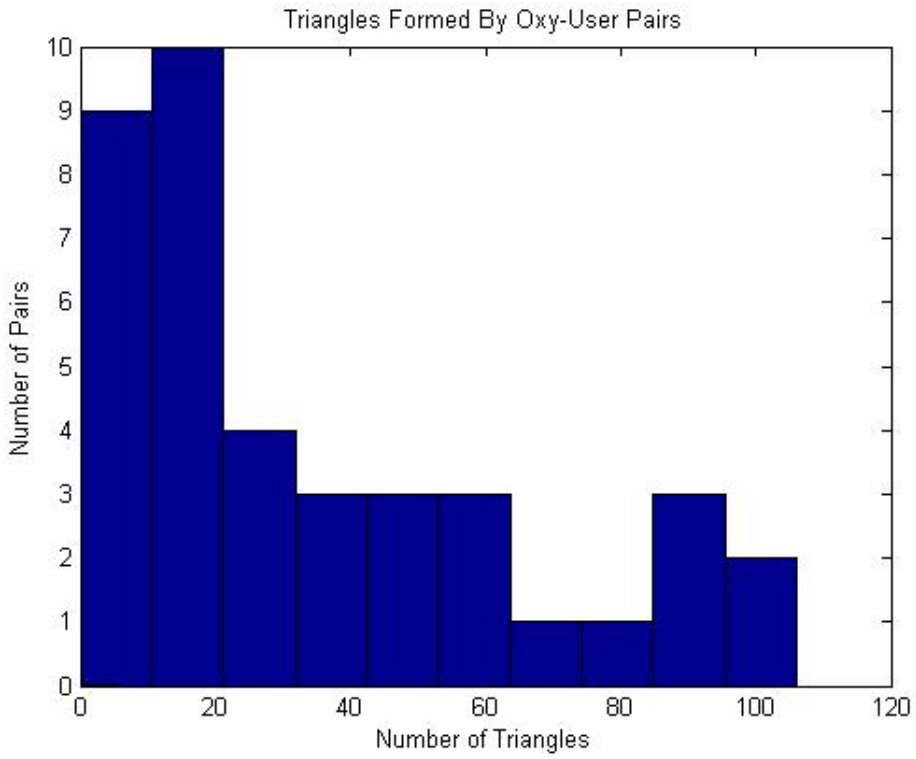


Figure 3: Histogram of reciprocal triangles formed by two oxycodone tweeters and a third individual.

# 7 Oxycodone User Narratives

We sought to find individuals with interesting narratives regarding their lives, and the impact of oxycodone drug use, shown through their tweets. We began this search during the process of inspecting the tweets by hand to see if they were properly flagged during our keyword search. We selected a small number of tweets that we felt had the potential to have an interesting narrative associated with them, and use these users as a starting point for our investigation.

# 8 Illicit Drugs Tweet Selection Methodology

We use the same methodology for searching for tweets relating to illicit drugs using essentially the same methodology that we employed for searching for oxycodone tweets. As before, we began with a set of keywords that we searched our body of tweets for. However, this set of keywords was composed of words and hashtags that (INSERT AN EXPLANATION OF WHERE THE LIST WE GOT FROM HENRY IS FROM). As with our oxycodone keyword search, we also had to removed several keywords after some initial inspection of the kinds of tweets that were being flagged, which were falsely flagging significant numbers of tweets.

After this honing of keywords, we found 20469 tweets that were flagged from only the 1 year tweet file, from 7094 different users. (We would therefore expect 2 3 times as many additional tweets to be found in the much larger data sets at the U-o-R.) After visual inspection of a selection of these tweets, they seemed to show much higher amounts of correctly flagged tweets than the oxycodone keyword search did. This initial work definitely begs for a much more sophisticated language model to sort tweets.

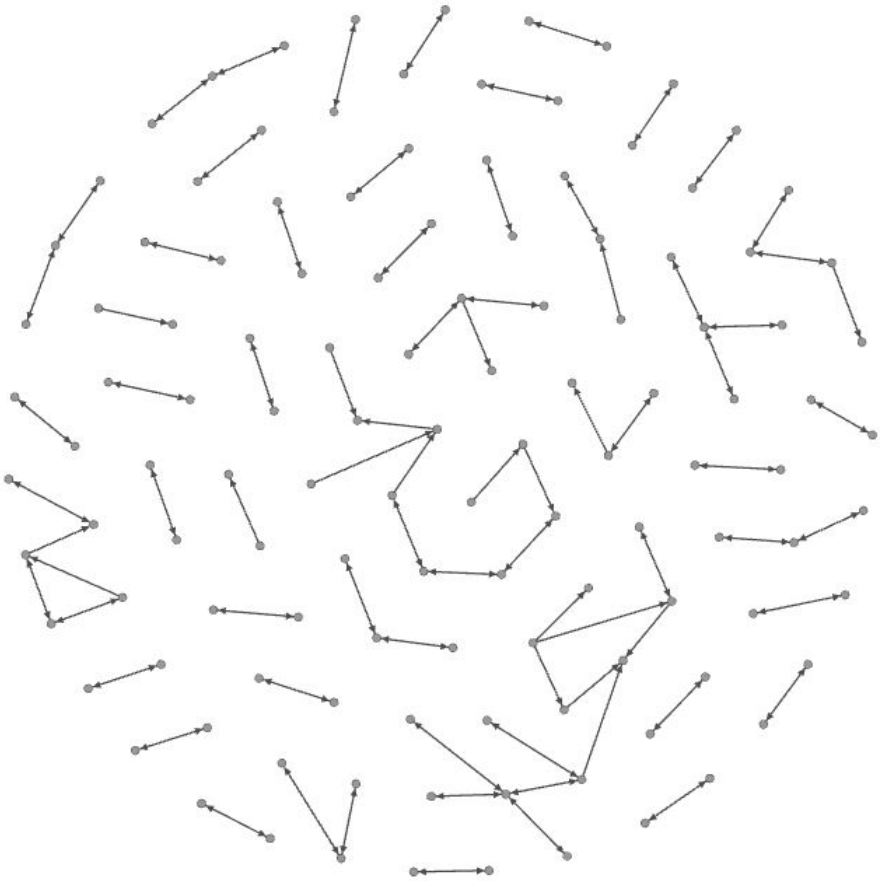

Figure 4: Directed graph of the connections between people tweeting about Oxycodone.

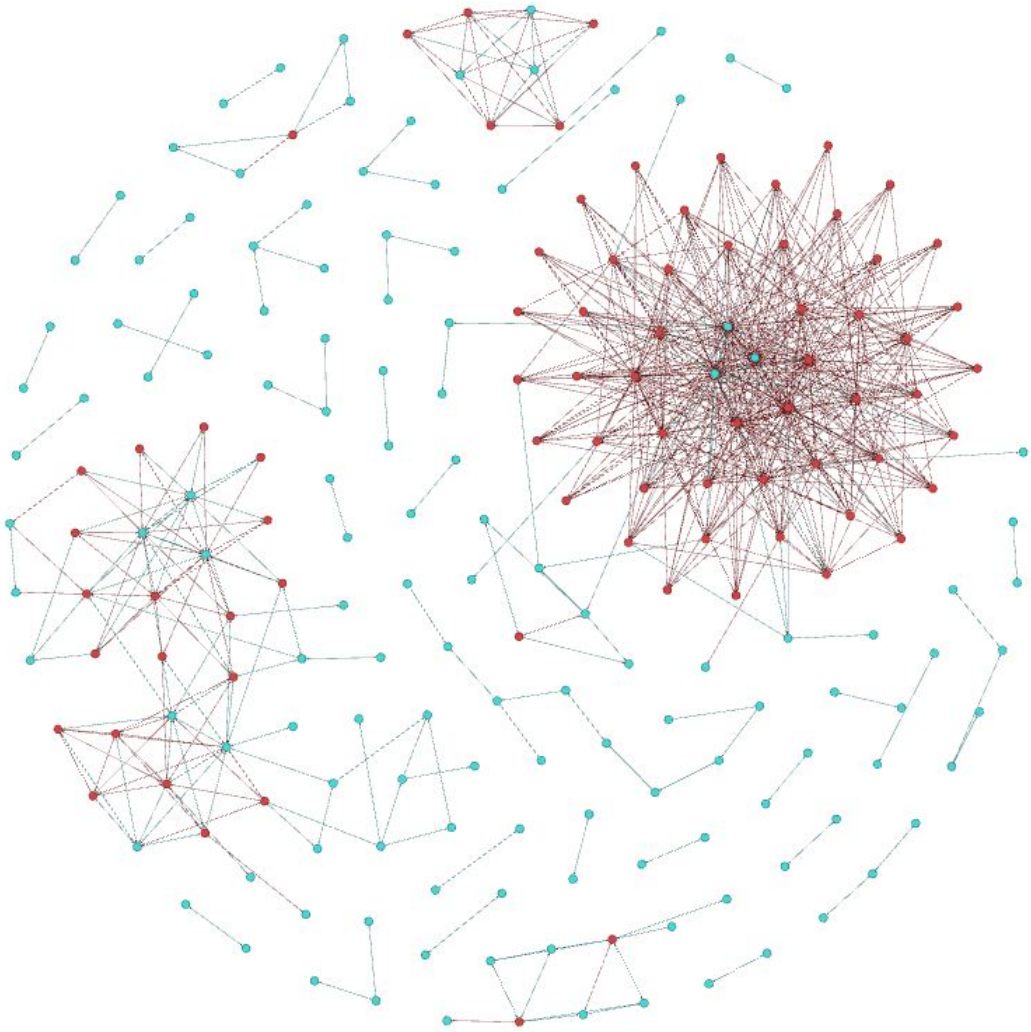

Figure 5: Directed graph of the connections between people tweeting about Oxycodone (teal) and those forming multiple reciprocal triangles with those tweeting about Oxycodone (red).

# 9 Illicit Drug Users' Interconnectivity Results

Due to the shear volume of flagged drug tweets, at this stage in our work we felt it would not be the best use of our resources to allocate time to sorting through all of the tweets by hand, and will set that work aside for a more sophisticated methodology to be used later. That being said, because of the much higher rate of positively flagged tweets, we went ahead and took an initial look at the network structure for those tweeting about illicit drugs.

From the 7094 users who tweeted about illicit drugs, 3919 of them who were in the 13-2015 graph file. 5732 of the 7094 formed some kind of connection amongst each other, including 2700 reciprocal connections between 3927 different individuals, 3337 of whom are in the graph file. From the 2700 reciprocal pairs, 1659 of which formed 47,289 reciprocal triangles with 30,677 additional individuals. Though, only 40 of the 2700 reciprocal pairs had both users in the social graph and formed no reciprocal triangles. From the 30,667 users that formed reciprocal triangles with illicit drug tweeters, 8662 of them formed more than one reciprocal triangle. Additionally, from the 47,289 triangles that were formed, 5852 of them had all three members, 2157 unique ones in total, from the illicit drug tweeters.

Figure 6 shows the connections between the 5732 users in the social graph who tweeted about illicit drugs. Additionally, Figure 7 shows the connections between only the 2157 users who form reciprocal triangles amongst each other.

# 10 Conclusion

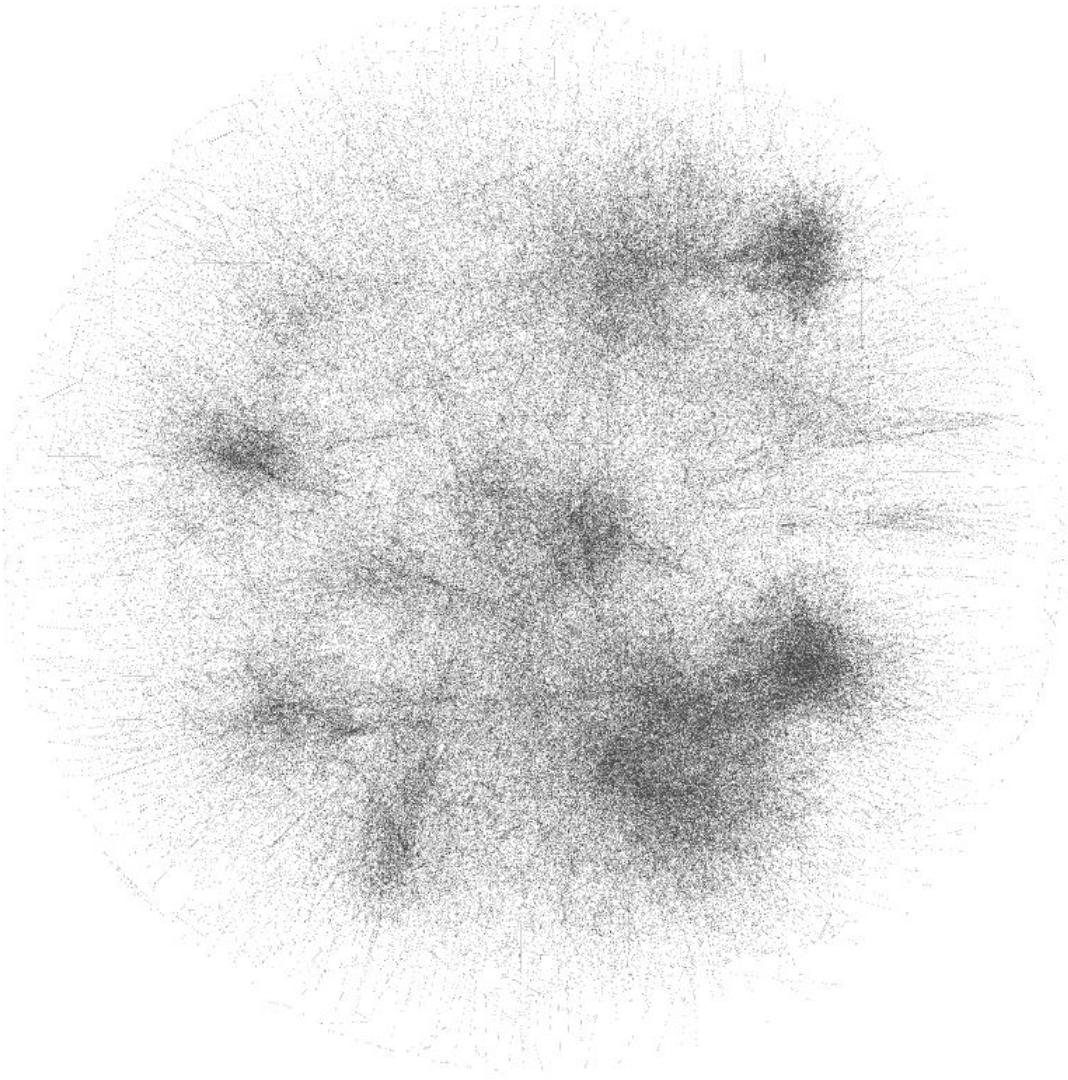

Figure 6: Directed graph of the connections between people tweeting about illicit drugs.

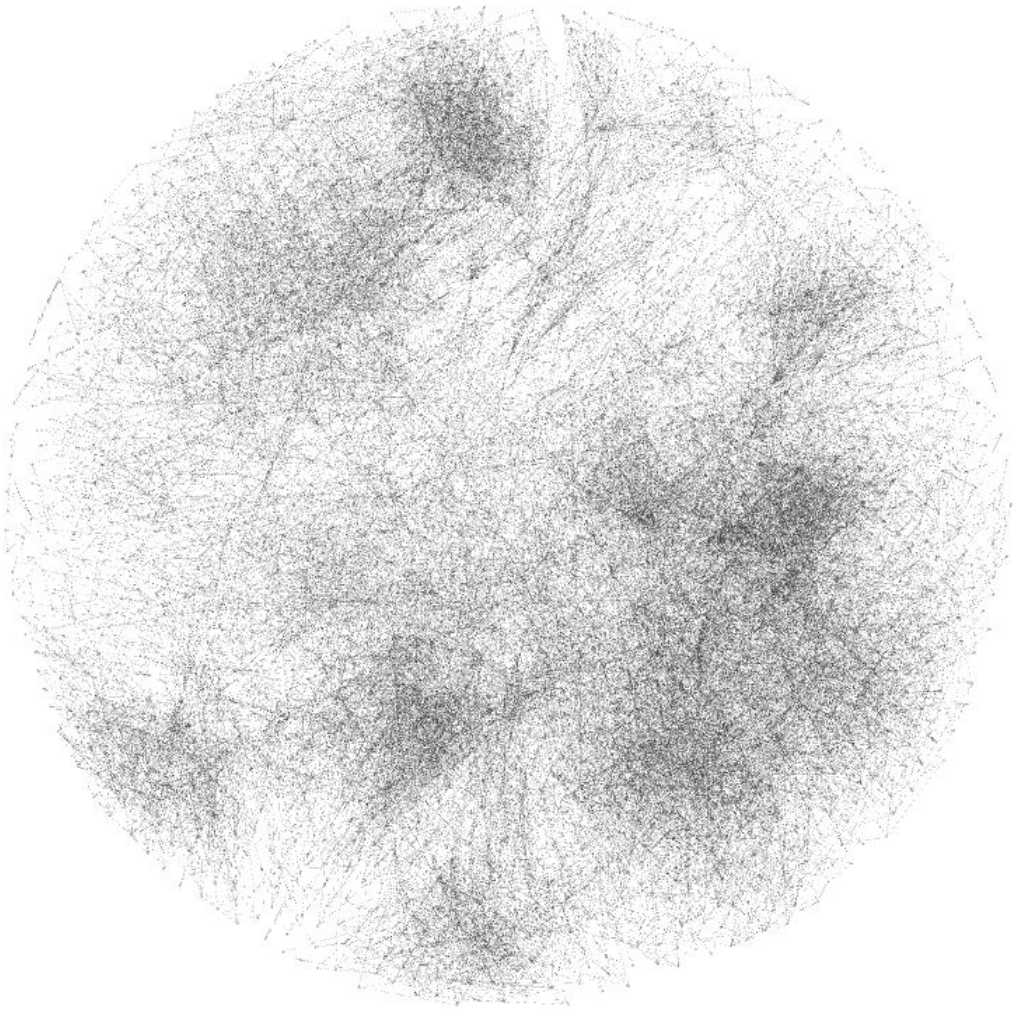

Figure 7: Directed graph of the connections between people tweeting about illicit drugs who form complete reciprocal triangles amongst each other.