Exploring Confirmation Bias Via Social Network Analysis

GitHub:

A description of the phenomena and why it matters

I. What is the phenomenon?

In this project, I'm exploring the Confirmation Bias phenomenon by analyzing social network connections between clusters of users. Confirmation bias refers to the human tendency to seek out evidence that confirms their current beliefs and reject evidence that contradicts what they believe.

II. Why should we be interested in this phenomenon?

Confirmation bias gets heavily exploited by various information and propaganda campaigns. Namely, because of the confirmation bias, once a person gets "hooked" up by a certain belief system, that person becomes very reluctant to change that belief system. Therefore, such campaigns don't need to ensure they provide the best evidence possible to retain their followers. Instead, their followers are likely to remain with them indefinitely regardless of the quality of that belief system. For this reason, we can make a conclusion that it is not a lack of evidence that makes people adopt false beliefs, but rather a lack of willingness to explore contradicting evidence.

This bias suggests that increasing the diversity of information will not diversify people's thinking, as no matter what evidence is out there, people will still seek out evidence that favors their current beliefs. In this work, my goal is to demonstrate confirmation bias by visualizing social network connections and showing that people form clusters around opposing belief systems and that people within those clusters are not interconnected with each other.

A brief review of current research trends regarding the phenomenon

*_Comparison study 1_: The Influence of Confirmation Bias on Memory and Source Monitoring

This work conducts a classical **empirical study of the confirmation bias** phenomenon. Namely, the groups of people are presented with two opposing views. Later they are asked to recollect the evidence supporting and opposing their beliefs. As a result, the researchers found that it is much harder for people to recall evidence that opposes their beliefs than the other way around.

**_Comparison study 2_: Social Network Visualization in Epidemiology

In this paper, the authors are using network visualizations to investigate epidemiological phenomena and design possible interventions. Namely, the paper is focused on understanding the structure and function of social networks and their effects on public health. In this work, the authors graph social networks for various people in order to identify **clusters** of people that exhibit a particular behavior. (e.g. **Obesity, Smoking, Happiness clusters, etc.**) For this purpose, they are using 3D graph visualizations.

***_Comparison study 3_: Political Polarization on Twitter

In this work, the authors are examining a similar phenomenon as I do but for a different belief system. Namely, the researchers are exploring **political polarization** on Twitter within the US political system. This work concludes that the network of political retweets exhibits a highly segregated partisan structure, with extremely limited connectivity between left- and right-leaning users. They further use graph visualization tools to demonstrate such polarization. (Based on visualizations, we are even using the same software package for visualizing graphs!)

What information regarding the phenomenon I targeted

In this project my aim is to construct and analyze clusters of people on social networks that constitute a particular belief system. I've chosen Twitter as my target social network and Russo-Ukrainian war as a target belief system.

Namely, in my analysis, I'm taking the following approach:

The first step is to identify cluster centers. By cluster centers I'm referring to Twitter users that belong to a particular belief system and have a large number of followers. I'm performing this step manually. Namely, I selected 9 Russian language sources which are in supported of Russia's war against Ukraine and 6 Russian language, 7 Ukrainian language, and 11 English language sources which are against the war.

The second part is to collect followers for these accounts. Namely, for the cluster centers, I'm collecting their twitter usernames, user ids, and follower counts. And for the cluster followers I'm collecting their usernames and user ids.

The final step is to visualize and analyze the data. In this step, I'm focusing on three primary data attributes:

1. Connection analysis and visualization. This attributes helps us identify connections that exist between the users, where connections are represented using follower relations.

2. Tweet content analysis and visualization. In this part I'm collecting a number of recent tweets posted by the cluster center accounts and analyzing the content of those tweets. This part of the project further splits into two sub parts.

    1. In the first sub part I'm collecting all tokens (words) present in tweets, cleaning them, and visualizing them as a word cloud. The word cloud shows us which words are the most frequently used by the cluster sources and gives us a better understanding of the overall tweet content.

    2. In the second sub part I'm performing entity recognition on each tweet from the cluster center using NLP software. This helps us gain a better understanding of what each tweet is about. Entity recognition algorithms identify entities such as people, places, companies, counties, etc. This data is later used together with the sentient analysis data to create a joined sentiment-entity visualization.

3. The final attribute that I am analyzing in this research is tweet sentiment. Namely, I'm using NLP algorithms to calculate sentiment for each tweet (positive score vs negative score) and using this data in conjunction with entity recognition data from the previous step to visualize and analyze the attitude of a particular cluster towards a particular entity. For insane if a particular tweet contains Russia as its primary entity and the sentiment is 80 % negative it's highly likely the tweet author has a negative attitude towards Russia.

What data was available for the phenomenon and how I used and accessed that data

There are two primary data sources available for my research. The first one is Twitter itself. Twitter provides a free API for developers, which allows users to access the data I needed. The main limitation of Twitter API are rate limits. Twitter places strict limits on the number of calls developers can make and the amount of data that can be fetched with these calls. After the limit has been exhausted, the client has to wait for 15 minutes to resume querying Twitter. This proved to be a significant inconvenience for my research, as, for example, I was only able to query 15000 follower names within 15 minute period, whereas some of the cluster centers have millions of followers. Furthermore, I initially intended to get the number of followers of center cluster followers. But, since that would require us to make a separate API request per follower, such a query would not be complete within a reasonable amount of time.

The second major data source is Kaggle. Kaggle has an amazing dataset on Russo-Ukrainian war called ****"Ukraine Conflict Twitter Dataset". This dataset contains millions of tweets crawled from Twitter daily, which contain references to the war in Ukraine. The main advantage of using this dataset is that we don't have to query Twitter directly and deal with rate limits. However, for the purposes of this project, it's main drawback is the diversity of data sources. Namely, it contains thousands of different data sources and there is no easy way to identify which belief system do each of these sources correspond to.

For the reasons outlined above I decided to use Twitter API for this project. All I needed to do in this case is to request a developer access from Twitter, and within a couple of days I was able to start fetching the data.

I have selected cluster centers manually. They consist primarily of Russian, Ukrainian, and US media sources, as well as several "independent" journalists and government officials. Most of them were identified by me using my personal knowledge, and some were found online based on a simple web search.

Information visualization techniques I used

1. User-Cluster-Group Visualization

Goal: visualize the follower-followed data

In this project, my primary data structure for data processing is Pandas DataFrame. Namely, I'm representing user attributes as columns and users as rows. After pulling the follower data from Twitter, my data has 3 primary dimensions. In particular, follower *username*, *cluster name*, and belief *group id*. My first visualization is using a dynamic 3D scatter plot to visualize these dimensions. 3-D data structures are ideal for visualizing three-dimensional data and make it very easy to identify the special relations between them. I'm further using tooltips to display the same information to the user and make it easier to see.

Tools:

- Python
- Pandas
- Plotly Express

Grammar of Graphics:

- *Coordinate System*: Cartesian
- *Geometric Objects*: Points corresponding to followers
- *Scale*: one to one
- *Aesthetics*:
  - Axis: x (username) ; y (cluster name) ; z (group id)
  - Colors: based on a group id
- *Data*: Twitter cluster center follower list

2. **Graph Visualization**

Goal: understand who constitutes a cluster

The main objective of our project is to analyze connections between users on Twitter. Therefore, our data is topological in nature. Namely, we need a way of representing connections between users. If we were to represent users as nodes and "follows" as edges, we wouldn't need to care about the exact layout in which the users should be arranged. Just the fact that one set of users is connected to another set of users suffices. For this reason, one of the best ways to visualize such data is in a form of a graph.

Also, our data has a direction. Namely, if user A follows user B, it doesn't mean that user B even knows about user A. For this reason, I'm choosing a directed graph data structure to visualize the relationships between followers and cluster centers.

Furthermore, graph visualization allows us to visualize other dimensions of our data, such as the size of a cluster, visualized by making the node size correspond to the number of followers the cluster has and the belief system the cluster belongs to, by making the node colors correspond to cluster groups. To summarize, we use a directed graph to visualize the following dimensions of our data:

- **Users**: Visualized using **graph nodes**.
- **Connection**: which users are connected to which other users. Visualized using **graph edges**.
- **Relationship**: which user is the follower and which user is the cluster center. Visualized using graph **edge directions**.
- **Cluster Size**: how big is each cluster. Visualized using **node sizes**.
- **Belief Systems**: which belief system does the cluster relate to. Visualized using **node color**.

Tools:

- Python
- Tweepy
- Pandas
- NetworkX
- Matplotlib

- Bokeh

Grammar of Graphics:

- *Coordinate System*: None
- *Geometric Objects*: Points correspond to follower users and clusters users
- *Scale*: node sizes are directly proportional to the number of followers of a given user
- *Aesthetics*:
  - Colors: based on identified communities
  - Arrows: represent follow relations between users
- *Data*: Twitter cluster center follower list

3. **Word Cloud Visualization**

Goal: understand what each cluster is talking about

Word Cloud works in the following way: the more a specific word appears in a specific text body, the bigger and bolder it appears in the word cloud. Word Clouds help us get a general sense of which terms are the most important in a given data source. In this project, I'm using the word clouds to understand what each cluster is posting about. Namely, I'm pulling the 100 most recent tweets from each cluster center. After the tweets have been fetched, a number of transformations need to be applied to them in order to make them usable for a word cloud visualization. Namely, we need to transform the data as follows:

Extract the tweet text body and clean it by removing hyperlinks, images, etc.

1. **Tokenize** the tweets. i.e. break each tweet into a list of tokens (words).
2. **Clean** the tokens by removing punctuation and stop words. Removal of the punctuation is the same for all three languages I'm working with. However, the removal of stop words (e.g. the, a, on, to, etc.) is different for all three. For English and Russian I'm using NLTK library. For the Ukrainian, I'm using a manually compiled text file containing a list of stop words.
3. **Combine** all tokens within the same string and **visualize** them in a word cloud format.

Tools:

- Python

- Tweepy
- Pandas
- WordCloud
- NLTK

Grammar of Graphics:

- *Coordinate System*: None
- *Geometric Objects*: Words
- *Scale*: the word size is directly proportional to the word frequency
- *Aesthetics*:
  - Colors: words are color coded for visibility purposes
- *Data*: Textual content from tweets published by different clusters

4. **Sentiment-Entity Visualization**

Goal: understand the attitude of a given cluster towards other entities

The final major part of my project is the creation of sentiment-entity visualization. In this part I'm collecting a subset of most recent tweets published by each given cluster center. Then I'm using NLTK sentiment analysis tool as well as Vader entity recognition tool to calculate positive and negative sentiment scores from each tweet as well as extract entities mentioned from each tweet. After this data is calculated, it's being packaged into Pandas data frame.

Finally, I'm visualizing the data on a 2D scatter plot. Since our sentiment scores are not binary, the sentient data is 2 dimensional, and each tweets has a combination of positive and negative score. Scatter plot with sentiment scores on its axis is an ideal visualization for this data.

After visualizing the sentiments on a scatter plot, I'm augmenting additional information about the tweet to each data point. Namely, I'm augmenting cluster username, cluster followers counts, and the list of entities identified in that given tweet.

Similarly to the graph visualization above, we use the sizes of points on a scatter plot to illustrate the number of followers each cluster has. We further color code the points according to the usernames of clusters these points belong to.

Tools:

- Python
- Tweepy
- Pandas
- Plotly
- NLTK
- Vader

Grammar of Graphics:

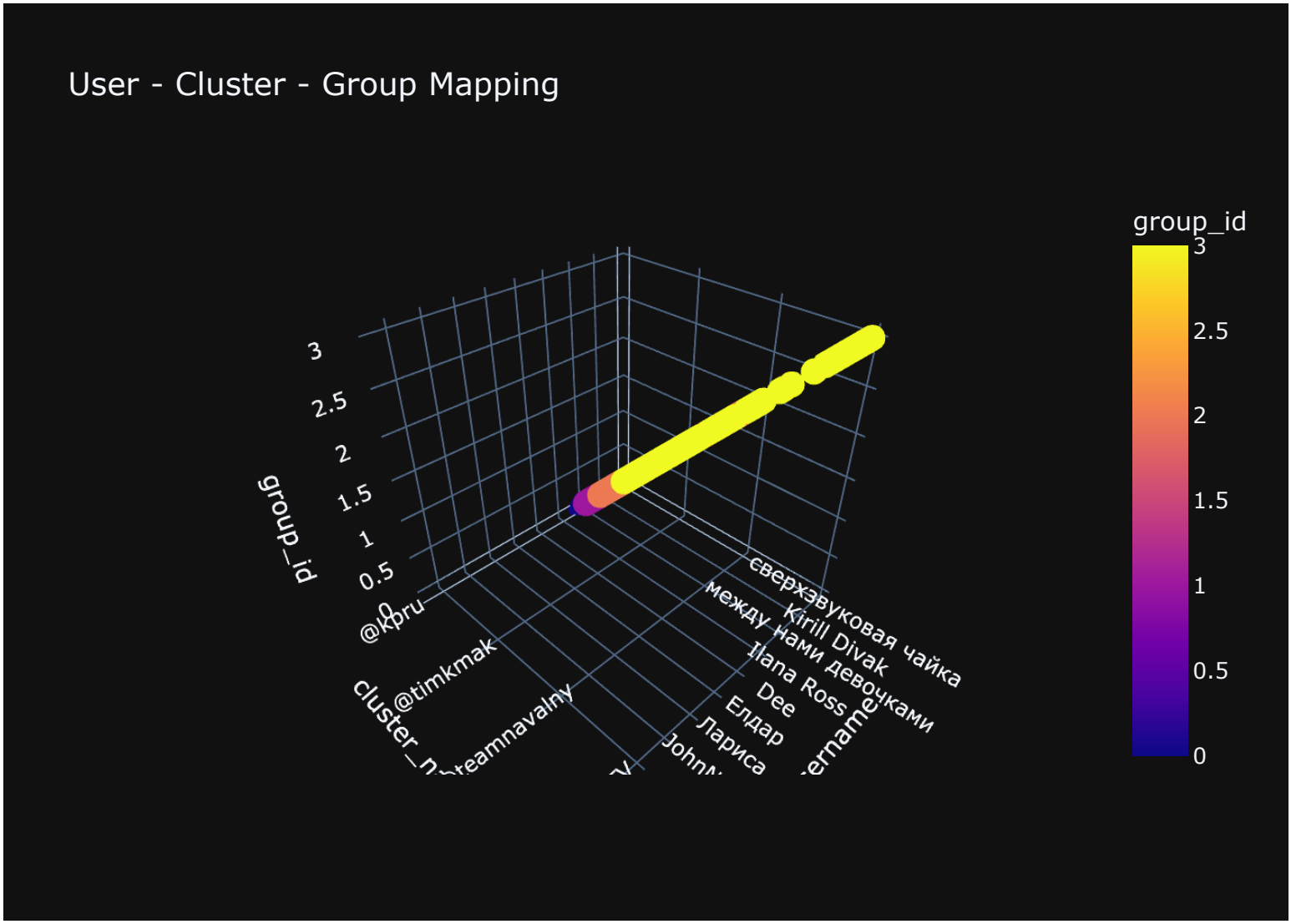
- *Coordinate System*: Cartesian
- *Geometric Objects*: Points correspond to individual tweets
- *Scale*:
  - The point size is proportional to the number of followers that a given user has
  - The magnitude of values on the x-axis and y-axis directly correspond to the sentiment score values.
- *Aesthetics*:
  - Colors: point color corresponds to the cluster name that published a tweet
- *Data*: Textual content from tweets published by different clusters

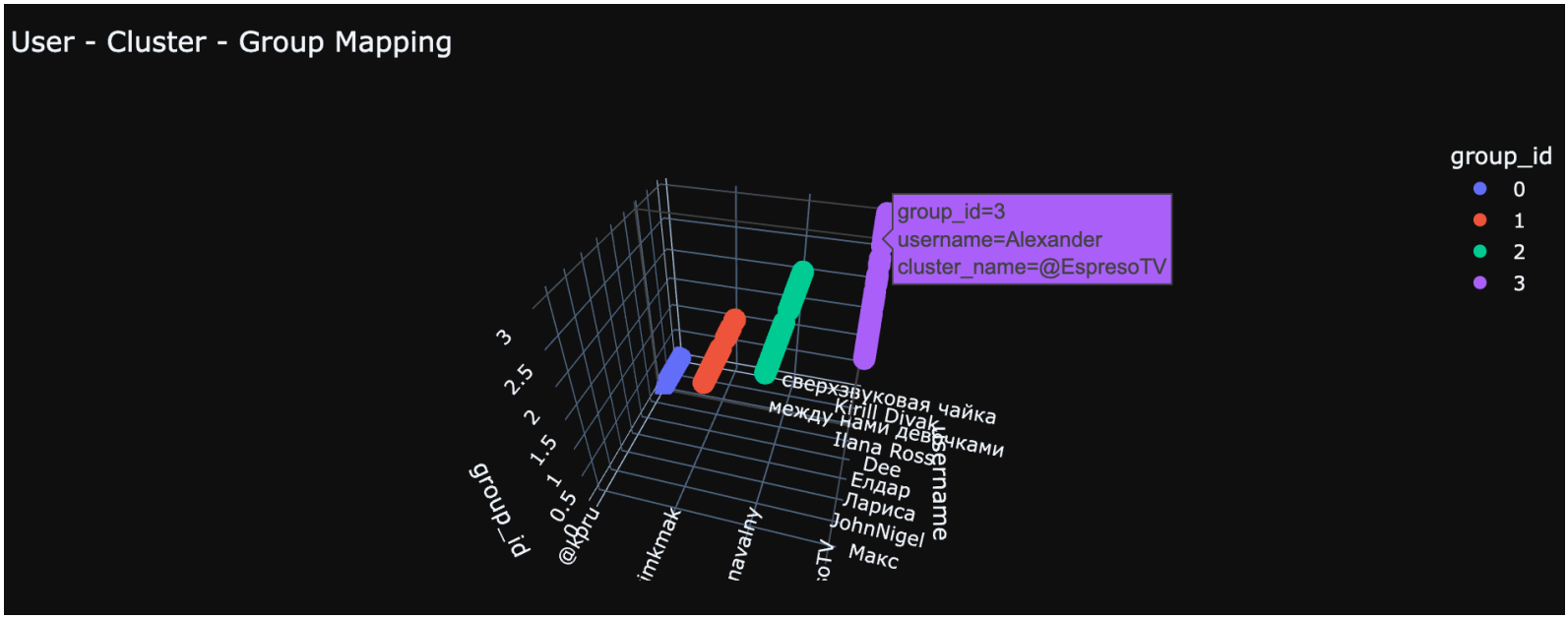
Information Visualization Results & Interpretation

The visualizations contain 2K users per cluster, thus approximately 102 000 users and 100 tweets per cluster, which results in approximately 5 100 tweets.

I. User-Cluster-Group Visualization

This visualization illustrates a mapping of all followers to their respective clusters. This visualization is an interactive 2D scatter plot. Clusters are color-coded with distinct colors. The follower usernames are positioned on the x-axis, the follower cluster names are positioned on the y-axis, and the clusters (group id) is positioned on the z-axis. In addition, there's a tooltip for each point on a scatter plot which displays the group id, username, and the cluster name.

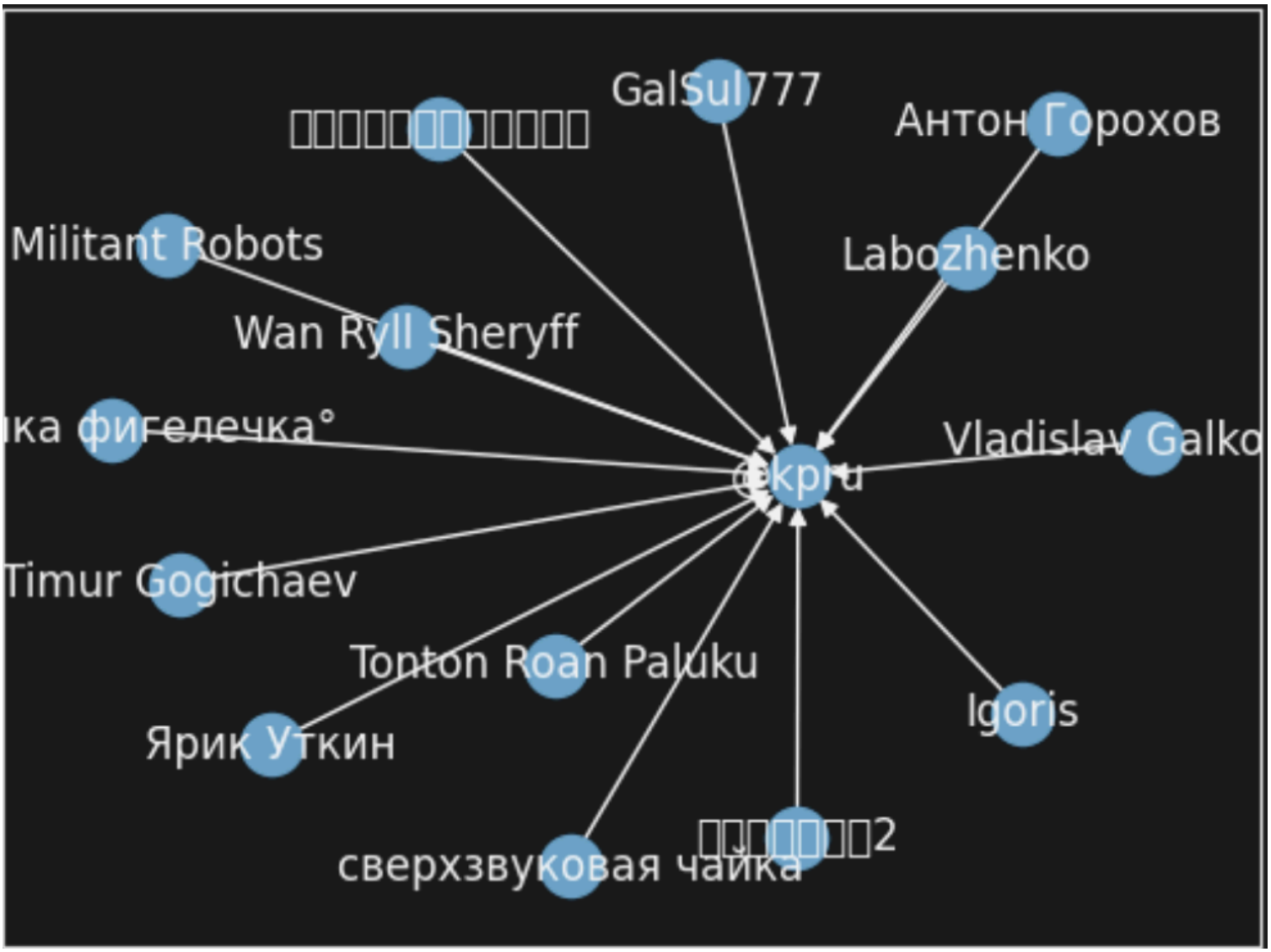

User - Cluster - Group Mapping
group_id
3
2.5
2
1.5
1
0.5
0
group_id
cluster_name
username
@kbru
@timkmak
navalny
сверхзвуковая чайка
Kirill Divak
между нами девочками
Ilana Ross
Dee
Елдар
Лариса


User - Cluster - Group Mapping
group_id
0
1
2
3
group_id=3
username=Alexander
cluster_name=@EspresoTV
group_id
cluster_name
username
@kbru
timkmak
navalny
сверхзвуковая чайка
Kirill Divak
между нами девочками
Ilana Ross
Dee
Елдар
Лариса
JohnNigel
Макс

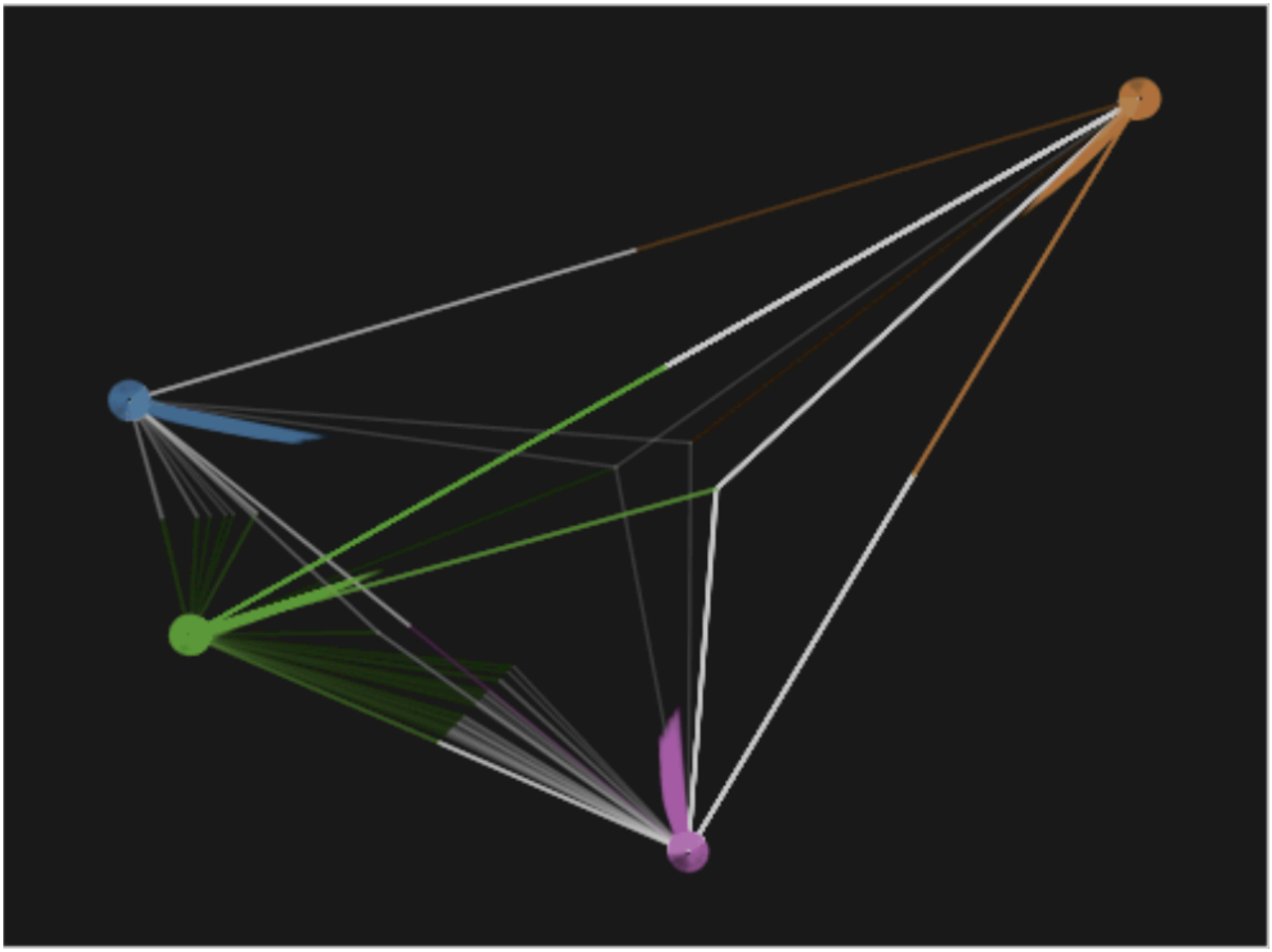
II. **Graph Visualization**

After constructing a graph in NetworkX, we start by visualizing a subgraph of that graph consisting of 15 nodes. The graph represents cluster center (@kpru) in the middle and cluster followers to the sides. The arrows indicate follower-followed relations.

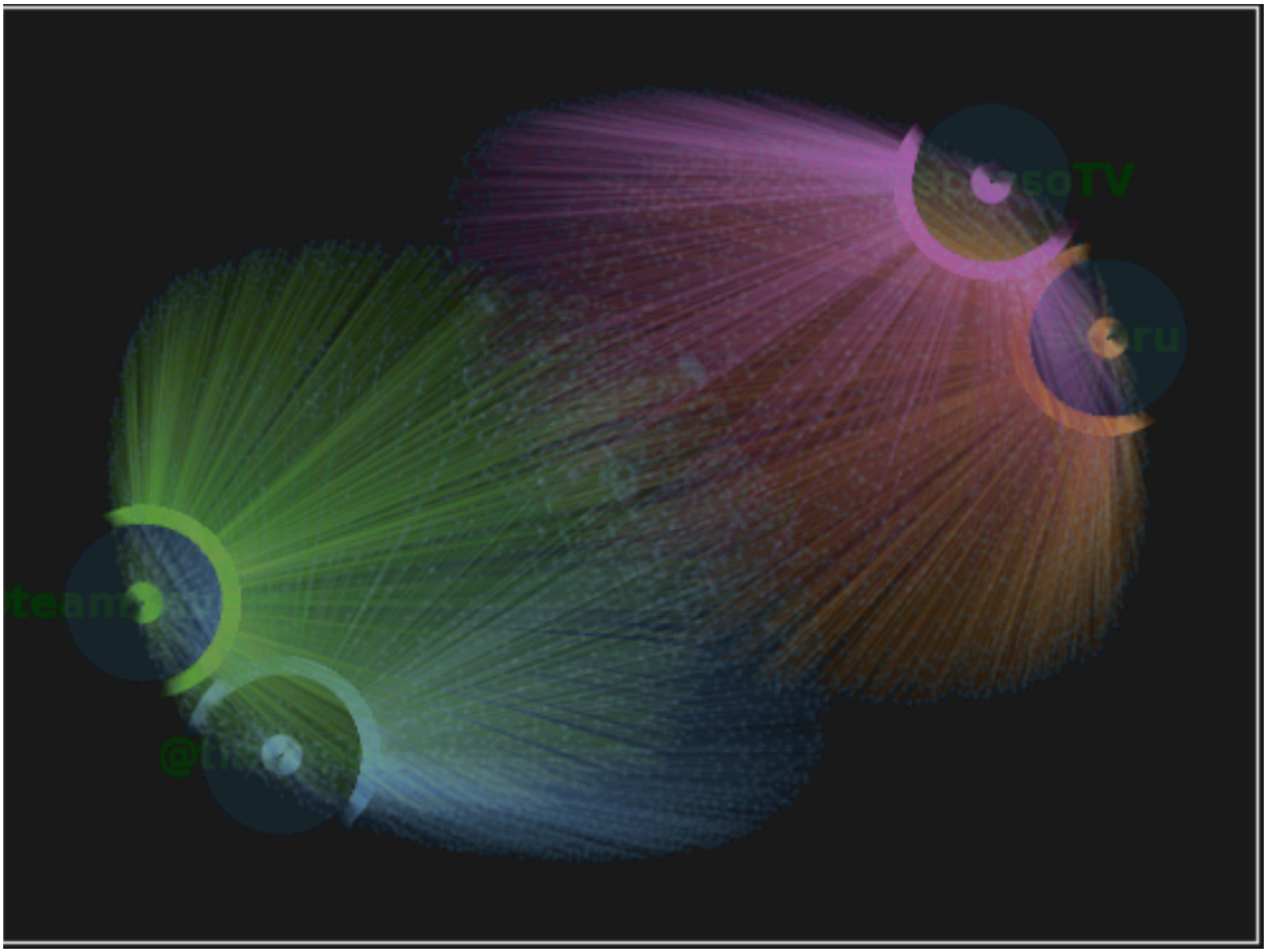
Interpretation: graph visualizations help us understand the relationships between the users and answer the primary question of this research: are people interconnected or do they form mostly disjoined clusters around certain belief systems.



The next graph visualization illustrates a complete graph. The nodes are separated using a community detection algorithm, which correctly identifies 4 different communities in our network. Node sizes are based on the number of followers a given cluster has. The node and edge colors are based on the community that a given node belongs to.

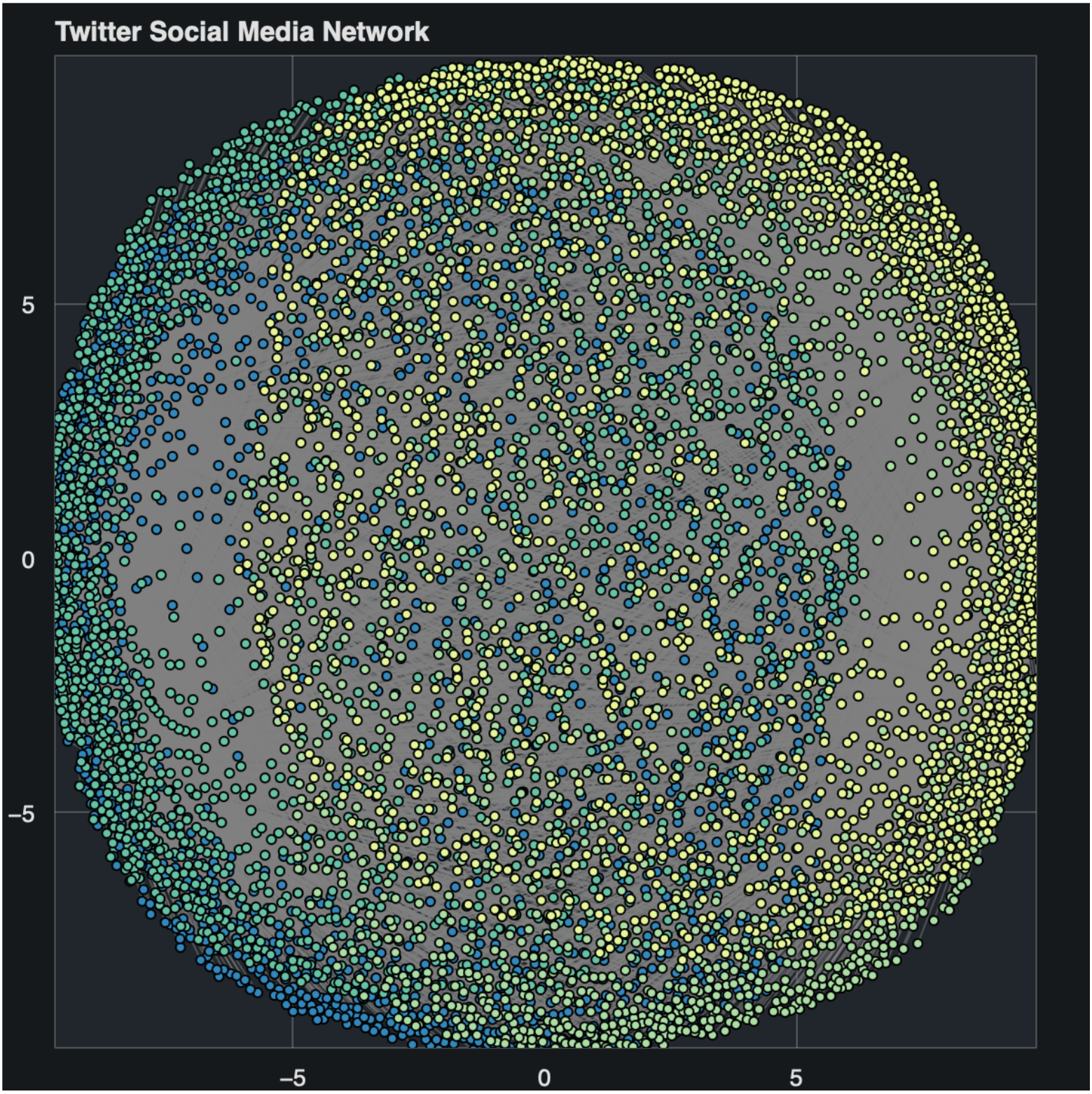


Another graph visualization below is based on the same graph as the one above, however, it's using a different separation coefficient (k = 0.15). This results in nodes

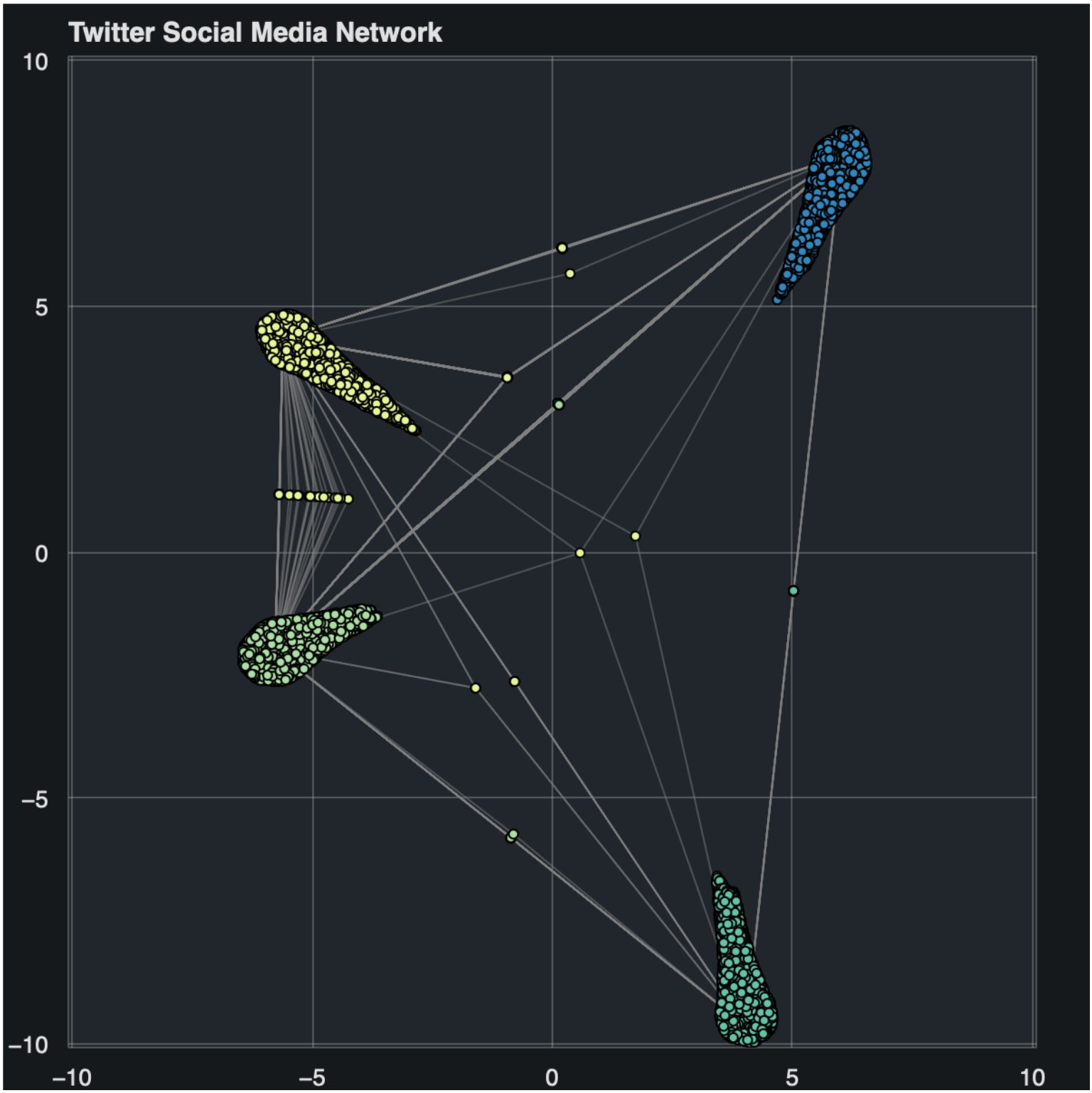


from different clusters being pushed further apart, which allows us to see the distinct clusters of followers much more clearly.

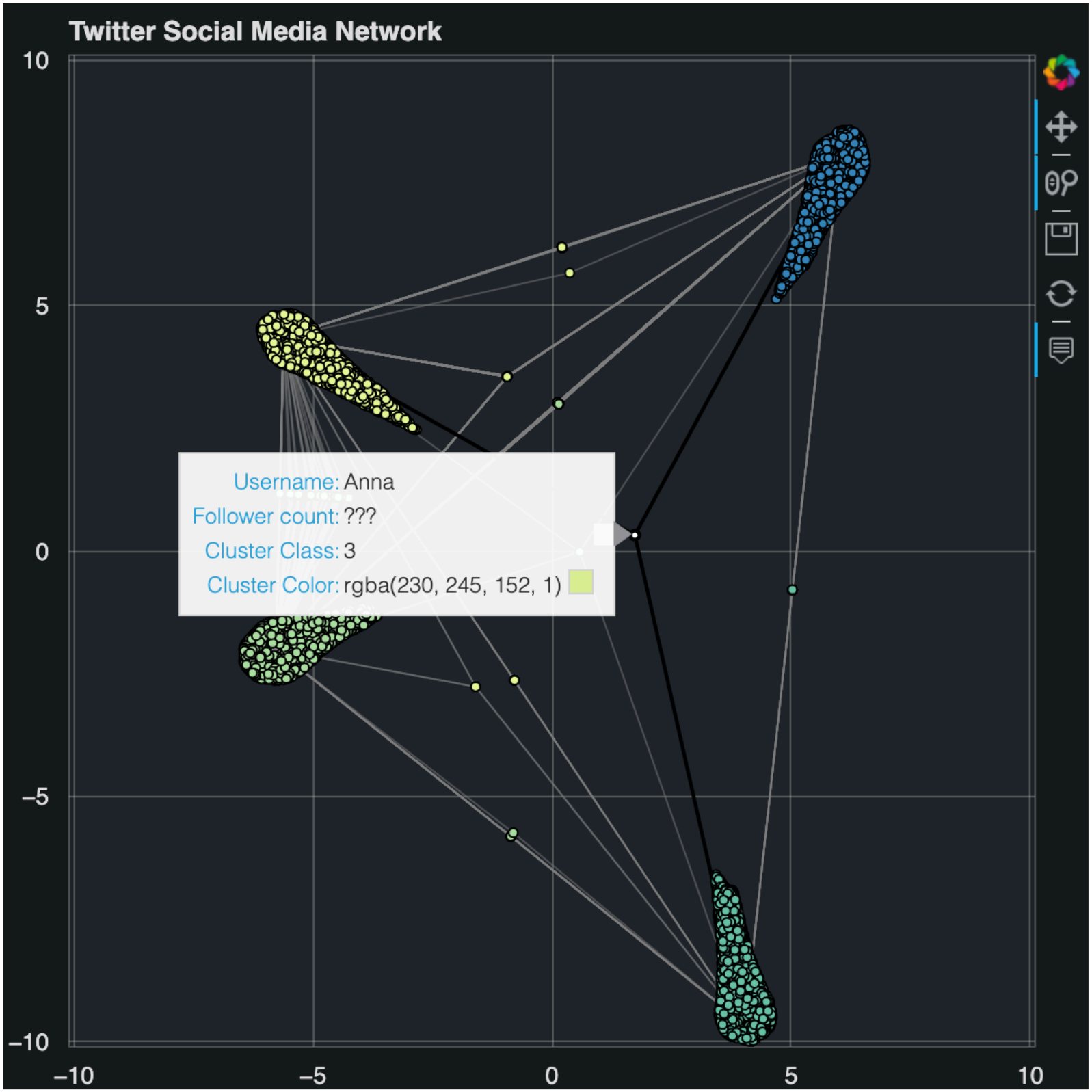
The following visualizations are produces using the NetworkX graph but in combination with bokeh plotting library. It's interactive visualization which visualizes all clusters and



connections of users to those clusters. The network clusters are bigger in size than their followers. The nodes are color-coded based on their affiliation to a given cluster.



Below is the same graph visualization but with a larger separation coefficient (k = 0.02). On this visualization we can clearly see how followers separate into **almost disjoined clusters**. There are only a few users which are connected to multiple clusters at the same time.

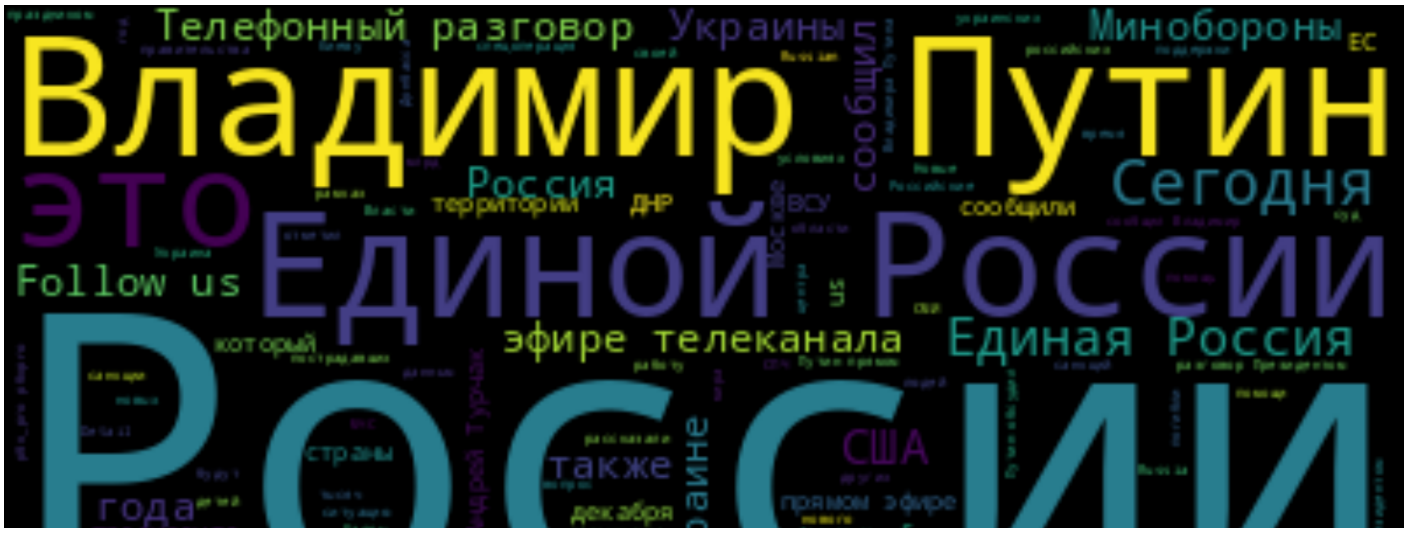


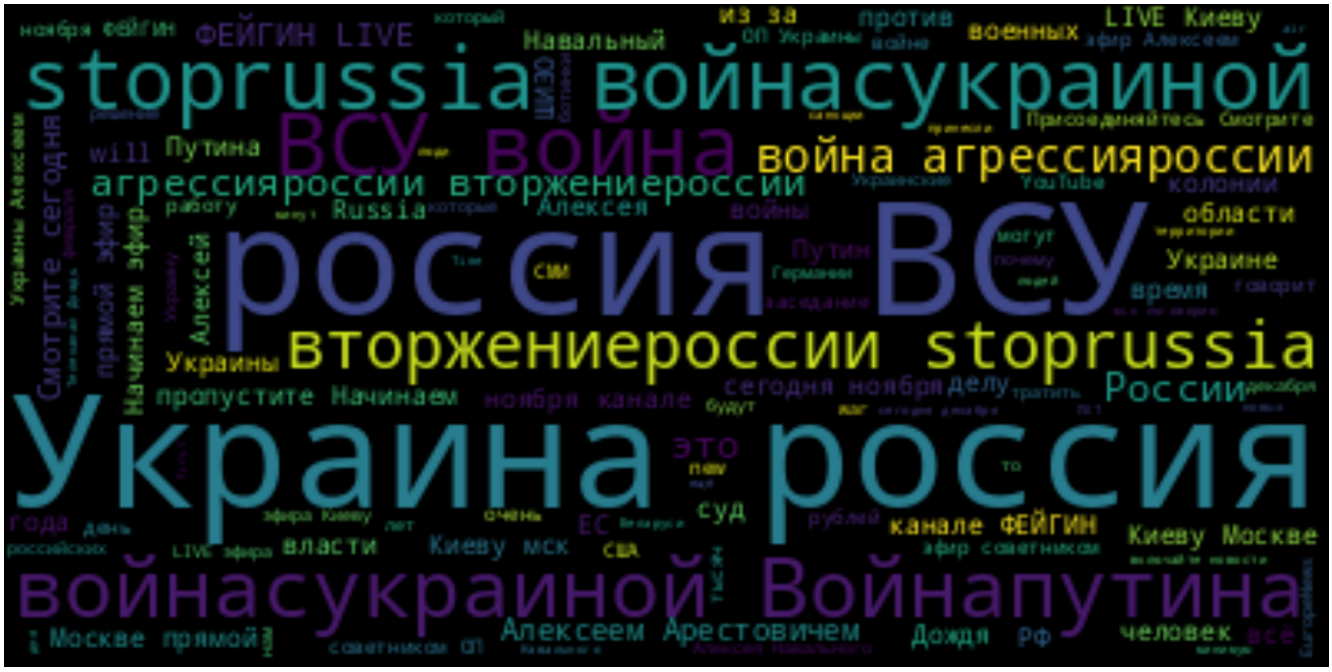
The following visualization illustrates the tooltip feature of this visualization as well as one of the few users which are connected to multiple belief systems. We can clearly see the connections by the dark highlighting of the edges on hover.

III. **Word Cloud Visualization**

The visualizations below are Word Cloud visualizations for different clusters. (Pro-Russia Russian language, Pro-Ukraine Russian, English, and Ukrainian)

Interpretation: these visualizations help us understand what the authors of tweets are discussing. What are some of the most important topics.






Ukraine
Russia
Ukrainian
Russian
war
Putin
attack
said
people
one
Kherson
Kyiv
drone
still
strike
killed
many
back
fighting
end
time
weapon
even
military
missile
Kremlin
Russia s
report
day
today
month
need
story
new
Poland
region
Ukraine s
right
great
CNN
invasion
Governor
first
two
say
know
country
must
Video
city
way
explosion
Moscow
according
civilian
without
official
see
want
Zelensky
troops
Dec
year
China
President
NATO
latest
Russian forces
defense
world
peace
rocket
million

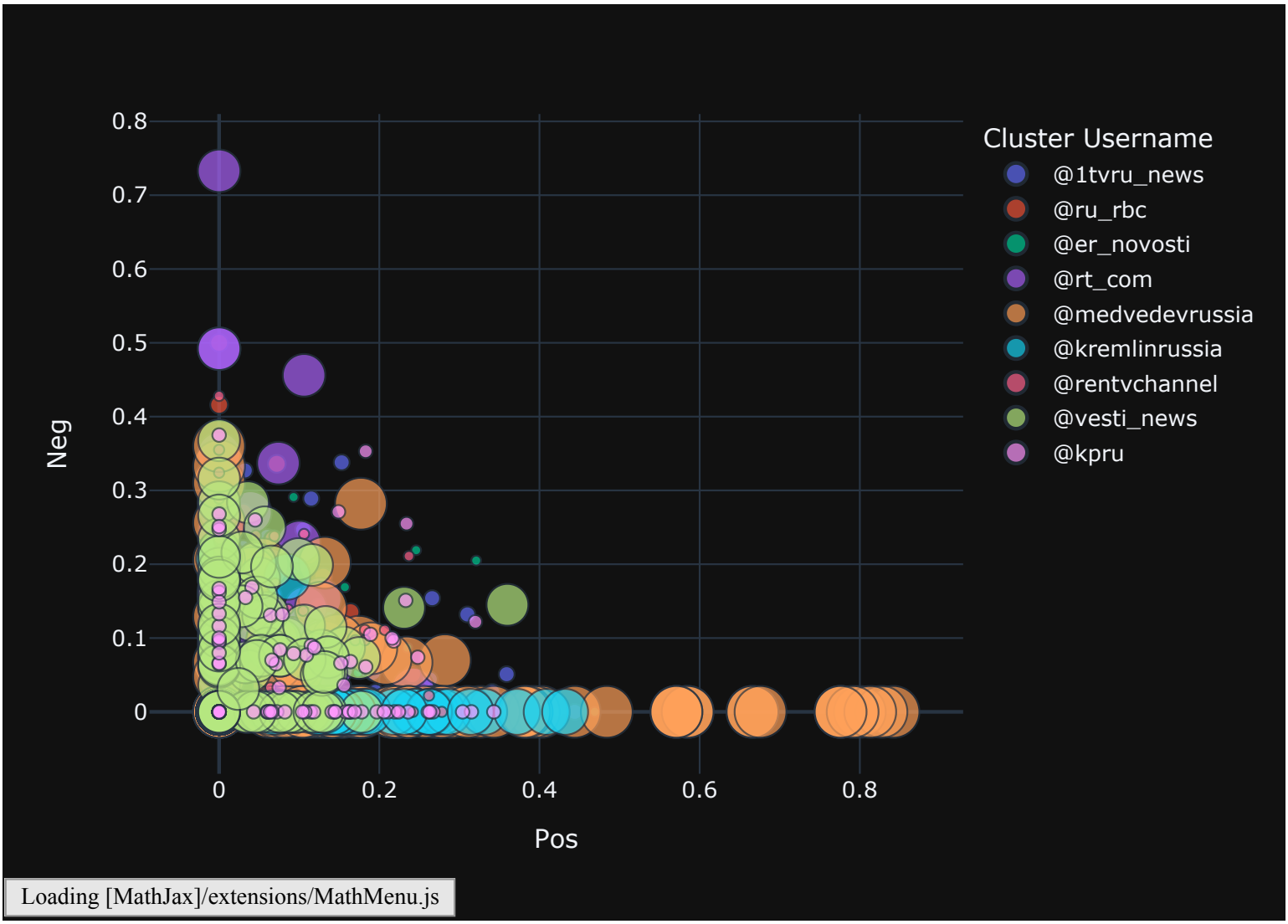

україні
новини тсн
рф
новини
україни
росії
зсу
дивіться youtube
росія
генштаб
україна
заявив
україну
рішення
голова
війни
зеленський
допомогу
києві
грудня
путін
путіна
збили
людей
день
загинули
окупантів
українських
обстріли
повідомив
донеччині
російських
війну
наслідки
упц
мп
сша
суд
безпеки
нато
відео
оборони
трьох
ударів
ракети
окупанти
росіяни
російські
російського
області
військових
тисячі
жінок
атак
дітей
обстрілу
розповів
зокрема
має
ЄС
фото

IV. Sentiment-Entity Visualization

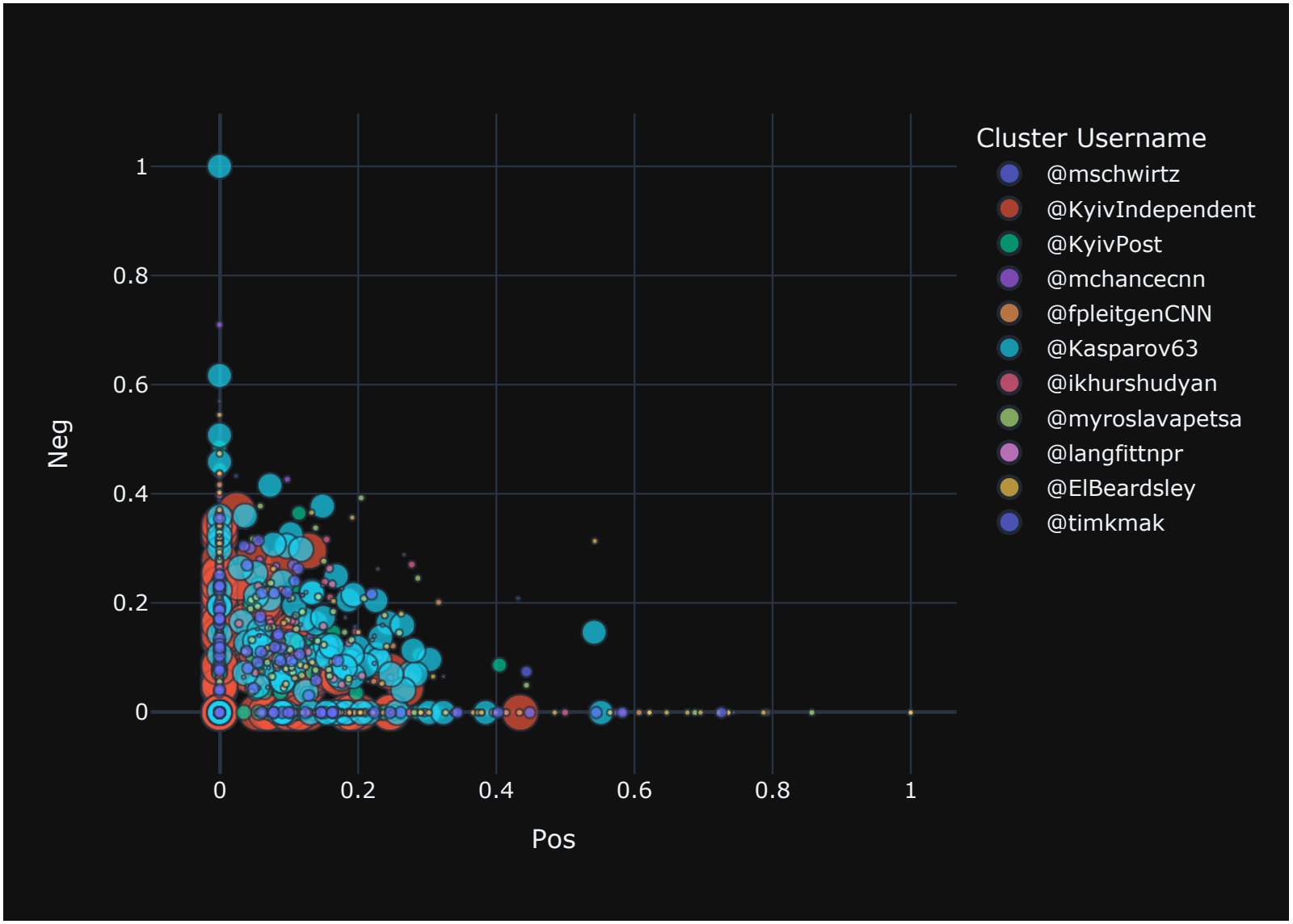
As a final part of my project, below we can see sentiment-entity visualizations for 4 different clusters. These visualizations were formed from approximately 5000 tweets collected from all clusters. Visualizations are presented as interactive scatter plots with negative sentiment on the y axis, positive sentiment on the x axis and tooltip containing cluster name, follower count, sentiments, and the list of entities mentioned in a given tweet.

Interpretation: These visualizations help us understand the attitude of particular clusters to certain entities. By combining the two together we can infer how a particular source feels about a given entity. For example, if the word Russia is an entity in a given tweet and a sentiment is terrible, we can assume the author of the tweet is against Russia.

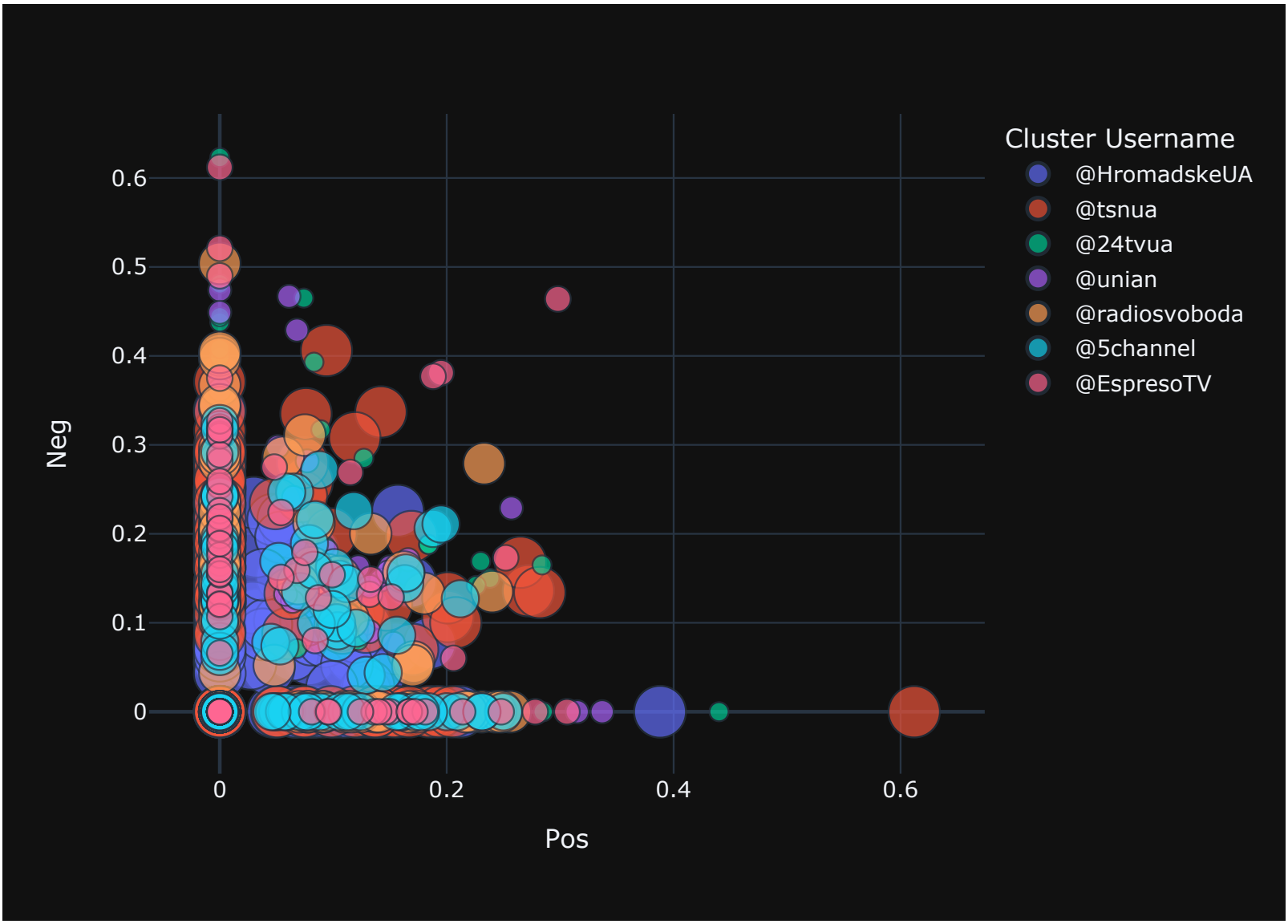
Pro-Russian cluster Russian language.



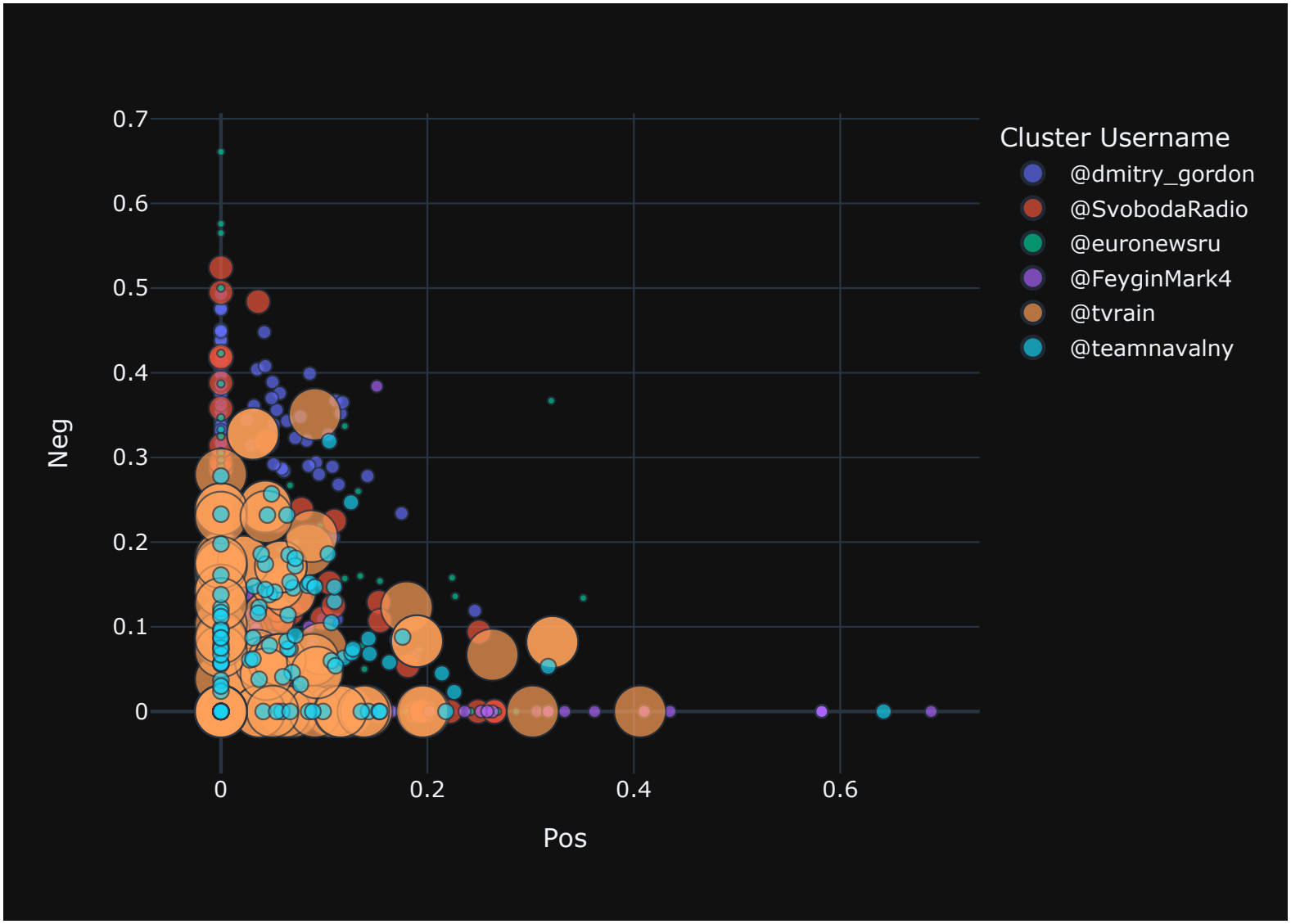
Pro-Ukrainian English language cluster.



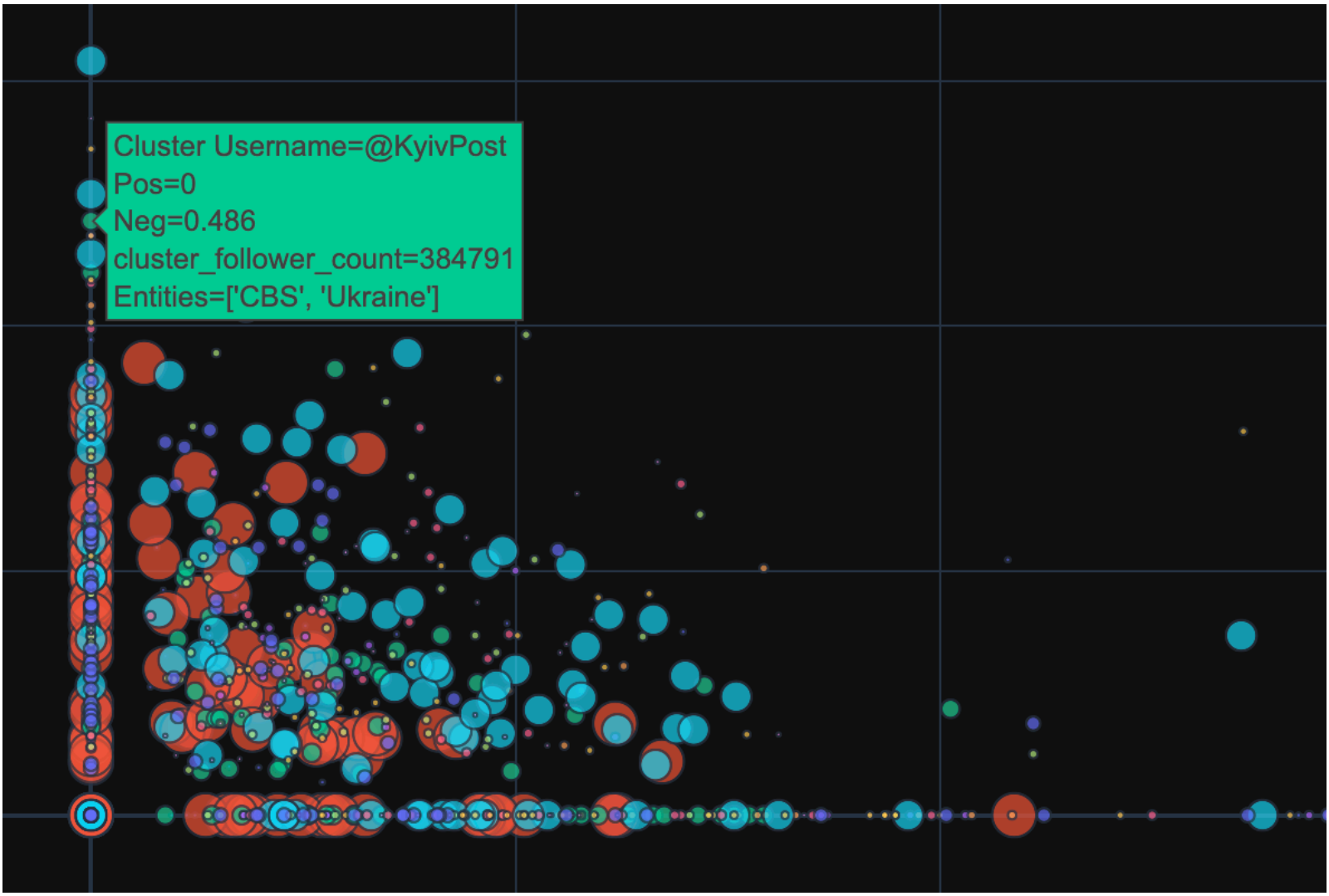
Pro-Ukrainian Ukrainian language cluster.



Pro-Ukrainian Russian language cluster.



Tooltip example.



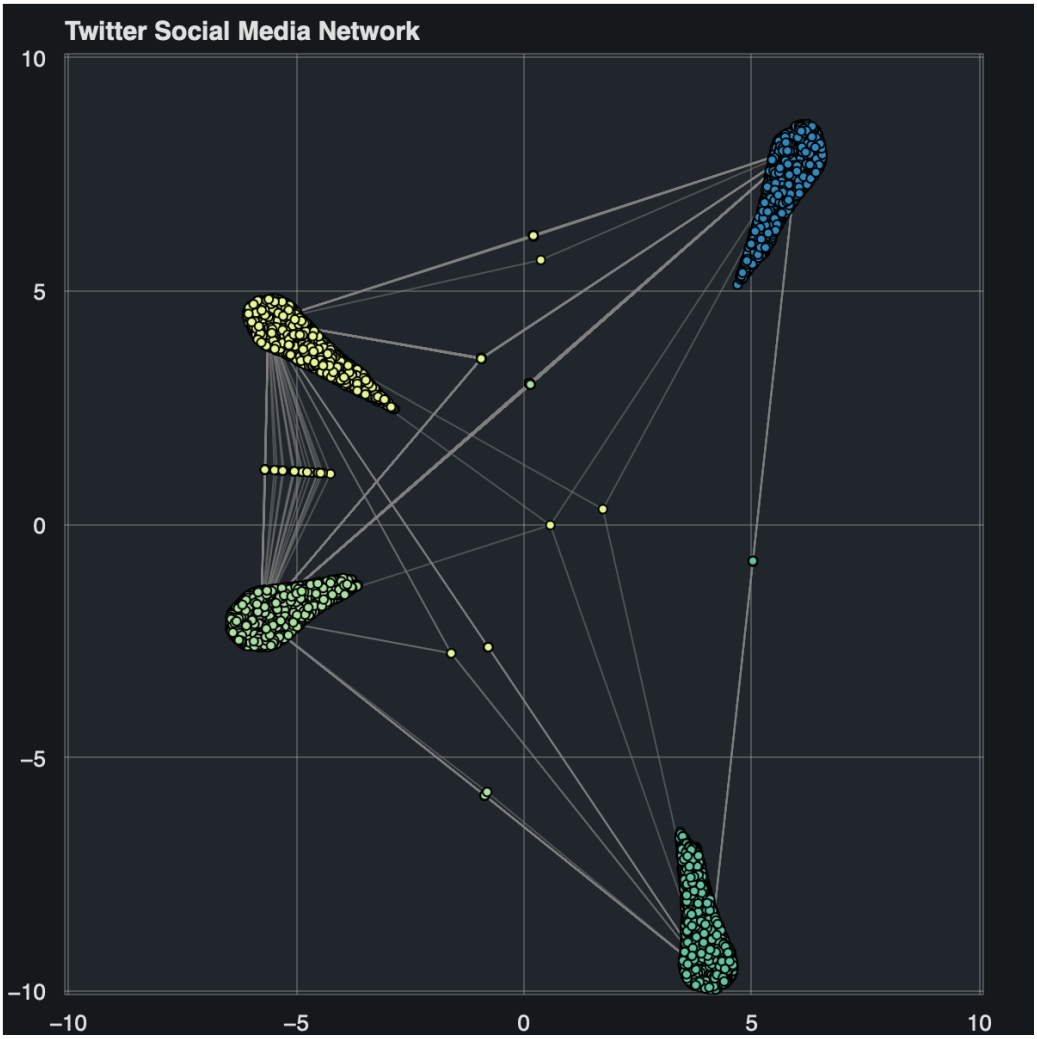
Findings

I. To summarize, in this project I was able to prove that people on social networks form clusters that correspond to a particular belief system. Once users become affiliated with a particular cluster, they stop seeking out alternative information sources.

II. Let us now discuss in greater detail what information visualization tools helped us learn about the phenomenon.

Graph visualizations helped us answer one of the most important questions of this work. Namely, do people form disjoined clusters around certain belief systems or do they exchange ideas in interconnected fashion? And the answer is resounding yes. People do form clusters around opposing belief systems. As we can see from our graph visualizations only several users are connected to multiple belief systems.

Namely, let us consider again the visualization below.



We can see that two of the clusters on the left share a considerable number of connections. The bottom left cluster is called @teamnavalny, which is a Pro-Ukrainian Russian media source. The top left cluster is called @EspresoTV, which is a Ukrainian media source publishing in Ukrainian. Even though the languages are distinct, they share the same belief system. On the other hand, the top right cluster is @kpru, which is a Russian propagandist media source publishing in Russian, and indeed there are very few shared connections with that source from the other two sources. The bottom right source is a pro-Ukrainian English language source. There are also few shared connections to that source, however, the likely cause in this case is the language barrier.

The WordCloud visualization for the pro-Russian cluster shows the words Russia, Putin, president, Ukraine, and the US. This shows huge attention being dedicated to advertising Putin as their great leader with additional discussions about the nazis in Ukraine. The pro-Ukrainian word clouds show that some of the most frequent words are Ukraine, Russia, Ukraine Armed Forces, Putin, aggression, occupiers, etc. This shows that Ukrainian media sources are focused on discussions related to war.

Finally, the sentiment-entity visualization shows that all media sources are biased toward negative news. In addition, we can clearly see negative sentiment from pro-Ukrainian sources towards Russia and the refusal of some of the western countries to provide additional support to Ukraine. On the other hand, pro-Russian sources display a negative sentiment towards Ukraine and the west as a whole.

References

* Peter Frost, Bridgette Casey, Kaydee Griffin, Luis Raymundo, Christopher Farrell & Ryan Carrigan (2015) The Influence of Confirmation Bias on Memory and Source Monitoring, The Journal of General
Psychology, 142:4, 238-252, DOI: 10.1080/00221309.2015.1084987

** Christakis NA, Fowler JH. Social Network Visualization in Epidemiology. Nor Epidemiol. 2009;19(1):5-16. PMID: 22544996; PMCID: PMC3337680.

*** Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2021). Political Polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 89-96.

**** @misc{bwandowando_2022,

title={🇺🇦 Ukraine Conflict Twitter Dataset},
url={https://www.kaggle.com/dsv/4675622},
DOI={10.34740/KAGGLE/DSV/4675622},
publisher={Kaggle},
author={BwandoWando},
year={2022}
}

***** More references specific to the code in references.md file in GitHub repo.