Ukraine Conflict Twitter Analysis Using NetworkX

1st Yande Satwika Yasa
Fakultas Informatika
Telkom University
Bandung, Indonesia
satwikayande@student.telkomuniversity.ac.id

2nd Seagata Ade Pratama Barus
Fakultas Informatika
Telkom University
Bandung, Indonesia
seagata@student.telkomuniversity.ac.id

Abstract—This report details the analysis of a Twitter or X dataset concerning conversations on X surrounding the Russo-Ukrainian conflict in 2022. Utilizing data sourced from Kaggle, which was originally collected from twitter, this study focuses on the network structure of user interactions, specifically replies, to determine the most influential people. The dataset is preprocessed to extract user interactions (source and target), forming a directed graph network using NetworkX, to find the most influential people, and then do community detection using Louvain and K-means Clustering. The analysis reveals a sparsely connected network, suggesting interactions primarily occur within smaller, potentially isolated groups or that users do not engage broadly across the entire conversation space; the community detected from the dataset incur that the majority of the tweets are condemning Russia, supporting Ukraine, and discussing while giving the updates regarding the war, the participants are from across the globe using various language to react and giving info/opinion regarding the war. The main limitations of the analysis being restricted to a single day's data due to computational constraints.

Keywords—NetworkX, Social Network Analysis, Page Rank, Degree Centrality, Community Detection

I. Introduction

The Russo-Ukrainian conflict, Commencing in 2022, generated significant global discussion across various digital domains. X or formerly known as Twitter, emerged as a significant platform for sharing information, expressing opinion, and discussing surrounding these unfolding events.

This mid-term project endeavors to analyze the network structure of conversations on X pertaining to the conflict in Ukraine, to dish out the top 10 influential individuals using each centrality metrics. Specifically, the focus lies upon the reply network, where users directly engage with one another's tweets. Through constructing and subsequent analysis of this network, we seek to pinpoint the most influential figures on the network, employing Page Rank and Degree Centrality as instructed by our dear professor.

Further this study does community detection to get a better grasp on the conflict's segmented opinion. The method employed for this purpose is the Louvain algorithm and K-means clustering for a machine learning approach. The Louvain is used for a user centric community detection, and the K-means used for tweet based community detection.

II. Methodology

A. Data Acquisition and Preprocessing

The dataset utilized within this project is hailed from Kaggle [1], having been meticulously gathered through the Twitter API during the conflict in 2022. Owing to computational power limitations, the analysis is limited to data corresponding to a single day '19 August 2022'.

The dataset is filled with the very essence of tweets itself i.e. consisting of the tweets, source of the tweet, reply to who, time, day, etc. Though all the dataset can be used for various things, this project focused on building the network and dissecting its essences.

The raw dataset underwent a preprocessing procedure that seems sufficient and necessary. Rows exhibiting null values in the `in_reply_to_screen_name` column were excluded, as these entries signify tweets that do not constitute direct replies within the observed temporal scope or contextual framework of the dataset.

In the realm of the Louvain algorithm, solely the `username` (indicating the source of the reply) and `in_reply_to_screen_name` (indicating the target of the reply) columns were preserved for its use. These columns were subsequently designated as "Source" and "Target", respectively, to accurately represent the directed edges within the network graph structure.

As for the machine learning approach, KNN Method here is deemed to be fitting and useful for the job, hereby it is being used by us to be deployed for community detection. The dataset first underwent a simple null removal in the columns of 'text','username', and 'in_reply_to_screen_name'. The 'text' column underwent a similar treatment, to remove the unnecessary word like link and extra space. After such a meticulous process, the last play is to vectorize the column 'text' for KNN.

The refined data was subsequently stored in a Comma-Separated Values (CSV) file format, named `edges.csv` to be used in Python as a Data Frame (df).

B. Network Construction

A directed graph (DiGraph) was instantiated from the preprocessed data employing the `networkx` library within the Python programming language, executed in a Google Colaboratory (Colab) environment. The `pandas` library facilitated the ingestion of the `edges.csv` file and converted it into a more manageable Data Frame, that can be easily dissected by python. Each distinct username corresponds to a node within the graph, and a directed edge is established from node A (Source) to node B (Target) contingent upon user A having replied to user B.

C. Network Analysis Metrics

1) Network Structure Metrics: The Basic network properties were calculated for context:

a) Network Size: Showing the number of nodes (users) and edges (replies).

b) Network Density: Used to understand the overall network connectivity sparsity.

2) Influence Metrics (Centrality Analysis): The core analysis focused on identifying influential users using:

a) Degree Centrality: Calculated for each node. In a directed graph, includes:

- *In-degree:* Number of replies received by a user, indicating direct popularity or attention received (acting as the user B).
- *Out-degree:* Number of replies sent by a user, indicating activity level (acting as the user A).

b) PageRank: Calculated using algorithms within NetworkX. PageRank assesses a node's importance based not just on the number of incoming links but also on the importance of the nodes linking to it.

c) Closeness Centrality: a reciprocal of the sum of the shortest paths from a node to all others; measures the average distance between a node and all other nodes. In other words it means someone that is the center of many connections, that can spread information the fastest.

d) Betweenness Centrality: the number of shortest paths from all nodes to all others that pass through that node; measures the frequency of occurrence of a node on the shortest paths between network nodes, in a simpler term the people that most oftenly act as the bridge of connection for shortest path.

3) Social Contagion (Contagion): a method of transmission that does not rely on a direct intent to influence; in normal language it means the method or process of an influence (can be another thing) to pass between people via social links

a) Information Diffusion: process by which a piece of information (knowledge) is spread and reaches individuals through interactions.

D. Network Visualization (Gephi)

Visualization was executed using Gephi after exporting the graph data. Layout algorithms (Yifan Hu Proportional, ForceAtlas 2) were used for spatial organization. Crucially, node attributes were mapped to visualize influence:

1) Node Size: Scaled proportionally to PageRank scores to highlight globally influential nodes. Optionally, size could also represent In-Degree Centrality for direct comparison.

2) Node Color: Determined based on community structures to show the context in which influence operates (just for clarity).

E. Community Detection

Louvain is a heuristic algorithm[6], that try to divide the network to maximize the modularity

1) Cliques: In graph theory and network analysis, a clique is rigorously defined as a "subset of vertices in which every pair of vertices are adjacent to each other" [2].

2) Modularity: Modularity is a scalar metric that "quantifies the density of links within communities compared to links between communities" [5]. Typically modularity spreads between -1 and 1, the higher the modularity the more robust and well-defined the community structure is.

3) Louvain Method: The Louvain algorithm by Blondel et al. (2008) is very simple and elegant. The algorithm optimises a quality function such as modularity or CPM in two elementary phases: (1) local moving of nodes; and (2) aggregation of the network. [3], the 2 steps can be translated as (1)Modularity Optimization and (2)Community Aggregation.

4) K-Means Clustering: K-Means Clustering is an Unsupervised Machine Learning algorithm which groups unlabeled dataset into different clusters. It is used to organize data into groups based on their similarity.

III. Results

A. Network Size

The resultant directed graph comprises:

- Nodes: 10,874
- Edges: 10,282

These numbers indicate that 10,874 unique X users participated in the reply interactions captured within the date of August 19, 2022; involving 10,282 observed replies to be the connections among them (edge).

Given the formula to calculate the maximum number of possible edges in directed graphs, as such the maximum possible number of edges is 118,233,002.

$$\text{Max possible edges (directed)} = N \times (N-1) \qquad (1)$$

B. Network Density

The calculated density of the network is approximately 8.70 x 10-5 (0.000087). This exceedingly low-density value, closely approximating zero, signifies a network characterized by sparse connectivity.

$$\text{Density Calculation (Directed)} = \frac{E}{N \times (N-1)} \qquad (2)$$

C. Network Structure (from Visualization)

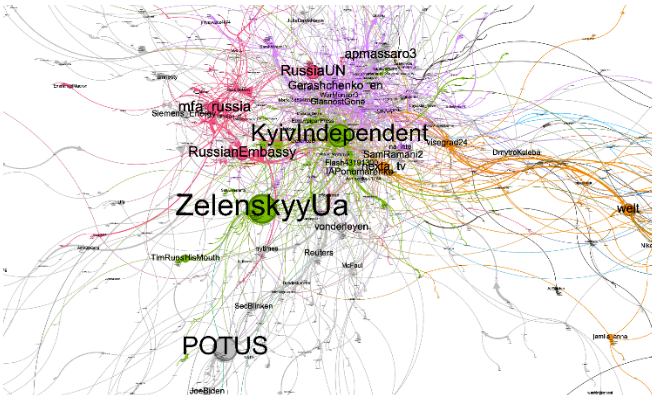

Fig. 1. Gephi Visualization.

It is pertinent to note that although the visual output—the network graph generated via Gephi—the graphical representation is the visualization of the Louvain algorithm, it can also be used as the direct graphic of in-degree centrality.

D. Influence

The key findings concerning community structures identified using function stemmed from NetworkX library, therefore are presented and interpreted herein.

```
[('ZelenskyyUa', 0.007909500597811092),
 ('KyivIndependent', 0.007541616849075693),
 ('POTUS', 0.005978110916950244),
 ('AndrzejGrochow7', 0.005518256231030995),
 ('igorlachenkov', 0.004874459670744045),
 ('SaintMoonKyiv', 0.004874459670744045),
 ('huidneprodym', 0.004782488733560195),
 ('apmassaro3', 0.004690517796376345),
 ('NickYashika', 0.004414604984824795),
 ('nexta_tv', 0.004322634047640945 6)]
```

Fig. 2. 10 Most "Connected" People. (Degree Centrality)

```
[('ZelenskyyUa', 0.007909500597811092),
 ('KyivIndependent', 0.007541616849075693),
 ('POTUS', 0.005978110916950244),
 ('igorlachenkov', 0.004874459670744045),
 ('apmassaro3', 0.004690517796376345),
 ('Gerashchenko_en', 0.004230663110457096),
 ('nexta_tv', 0.004230663110457096),
 ('MAStrackZi', 0.004230663110457096),
 ('mfa_russia', 0.004138692173273246),
 ('SWagenknecht', 0.004046721236089396)]
```

Fig. 3. 10 Most "Connected To" People. (In-degree Centrality)

```
[('AndrzejGrochow7', 0.005426285293847145),
 ('SaintMoonKyiv', 0.004874459670744045),
 ('huidneprodym', 0.004506575922008645),
 ('NickYashika', 0.004138692173273246),
 ('yoga_and_more', 0.004138692173273246),
 ('schayno_karas', 0.004046721236089396),
 ('sukhan1927', 0.003954750298905546),
 ('DaveHorn_CA', 0.003770808424537846 3),
 ('Katerin19848594', 0.003770808424537846 3),
 ('yoksig', 0.003402924675802446 4)]
```

Fig. 4. 10 Most "Connected From" People. (Out-degree Centrality)

```
[('nexta_tv', 0.006169678235467667),
 ('watch_union', 0.004078763865473425),
 ('sternenko', 0.003312421150397376),
 ('pdnetwork', 0.003233737364102343 3),
 ('ZelenskyyUa', 0.002647551283994360 7),
 ('MarkRid89403375', 0.002580927276719425),
 ('POTUS', 0.002216841743303506 6),
 ('KyivIndependent', 0.002082182886638123 7),
 ('igorlachenkov', 0.001942453264944658 7),
 ('KathaSchulze', 0.001638147657510027 5)]
```

Fig. 5. 10 Most "Influential" People. (PageRank)

```
[('ZelenskyyUa', 0.007540706245737238),
 ('KyivIndependent', 0.006975106790623911),
 ('POTUS', 0.005979483617505226 5),
 ('igorlachenkov', 0.00491095607438843),
 ('apmassaro3', 0.004394317278093063),
 ('MAStrackZi', 0.004271721564557029),
 ('mfa_russia', 0.004154922338658631),
 ('Gerashchenko_en', 0.004124420820951613),
 ('SWagenknecht', 0.004036661914834912),
 ('nexta_tv', 0.003858329155083768)]
```

Fig. 6. 10 Most "prominent" people (Closeness Centrality)

```
[('SarahAshtonLV', 3.679852619110992e-06),
 ('schayno_karas', 2.76623403781447e-06),
 ('Pan_M_165', 2.4405459324448765e-06),
 ('AndrzejGrochow7', 2.0302635139922715e-06),
 ('8113Kgreen', 1.7257239868934307e-06),
 ('heyserbernar', 1.5946028016147632e-06),
 ('flunkertungle9', 1.1758609518538572e-06),
 ('tipofthespear42', 9.64375169146329e-07),
 ('AnikaDeMadrid', 9.474563065297267e-07),
 ('huidneprodym', 8.459431308301131e-07)]
```

Fig. 7. 10 Most "well connected" people (Betweenness Centrality)

The Gephi visualization, particularly when scaling node size by PageRank, revealed several nodes with significantly higher influence scores compared to the vast majority (though it is in-degree). These high-PageRank nodes often resided within, or bridged between, the discernible community clusters, and acted as an important node though all of them are not present in the betweenness centrality; showing that they are influential users but not the most crucial one in regarding of connecting many cluster within the network. Similarly, analysis of In-Degree Centrality identified users who received many direct replies, acting as immediate focal points for engagement (popular), and the largest name within the graphic. While there was likely overlap between high PageRank and high In-Degree nodes, differences would highlight users who gain influence through connections from other important users (PageRank) versus those who simply attract a high volume of direct responses (In-Degree).

Setting aside the immediate importance of the people listed by PageRank, in other metrics most of the people listed are different and serve different purposes. The overall sparse structure and community clustering provide the backdrop against which these influential users operate. More will be discussed in discussion.

E. Contagion

In this contagion analysis we dissect the process of how the information spread from the most active user (out-degree) in 19 August 2022, to find out how many infected users would it reach, given he/she is the most active user, and the user is "AndrzejGrochow7" using information diffusion.

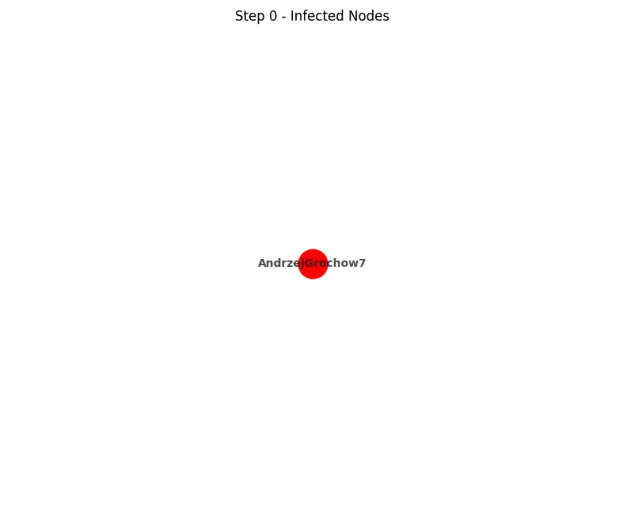

Fig. 8. Step 0 of Information Diffusion from “AndrzejGrochow7”

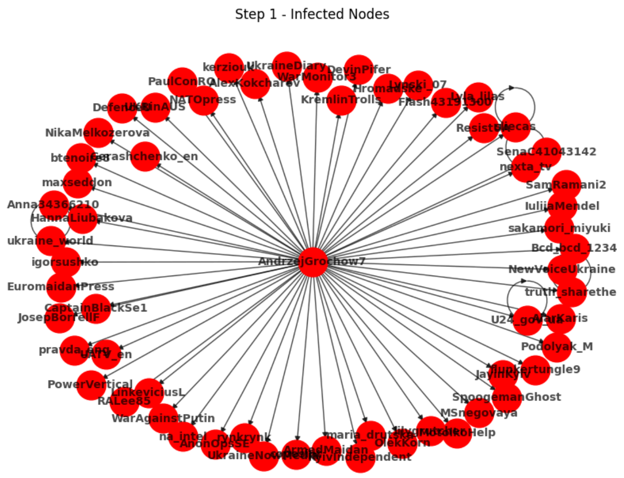

Fig. 9. Step 1 of Information Diffusion from “AndrzejGrochow7”

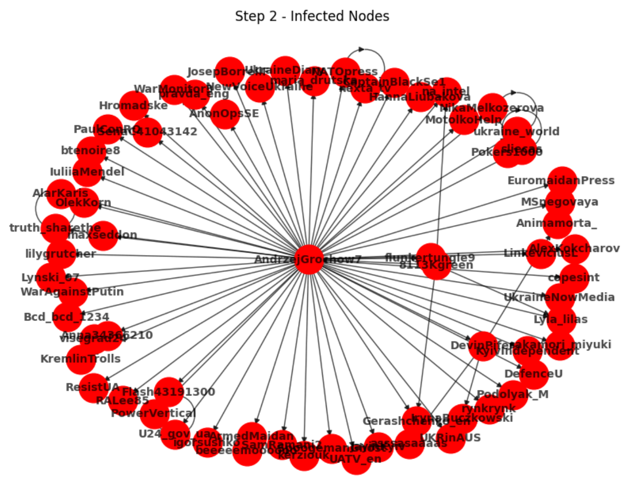

Fig. 10. Step 2 of Information Diffusion from “AndrzejGrochow7”

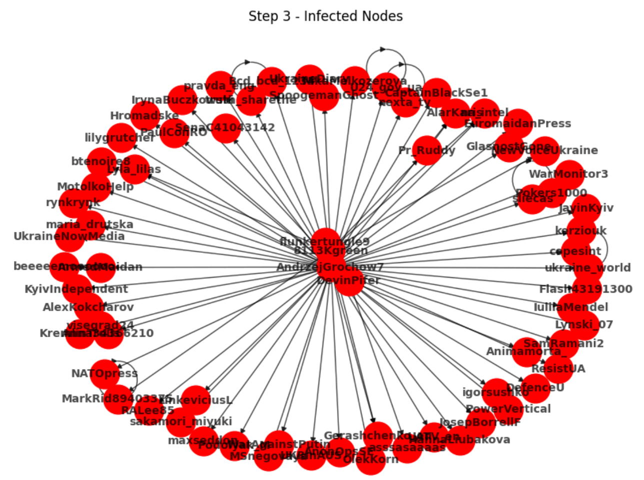

Fig. 11. Step 3 of Information Diffusion from “AndrzejGrochow7”

Within 3 steps it reaches the maximum of nodes it is connected with, and that is 70 users, even though he is the most active user, further will be discussed in the next section.

F. Cliques

Stated in [8] by Wasserman and Faust that cliques need a mutual connection between 2 nodes(directional dichotomous relations), making it not feasible in the term of directed graph, and given the dataset itself is about tweet, prithee making cliques deemed unneeded and not feasible to be done without preprocessing the dataset even further.

G. Louvain Method

Louvain algorithm modularity for directed graph, is calculated by using the following formula [4]:

$$\text{Directed Modularity } Q_d = \frac{1}{m}\sum_{i,j}\left[A_{ij} - \frac{d_i^{in} d_j^{out}}{m}\right]\delta(C_i, C_j) \quad (3)$$

Yielding the results of **2717 communities** with the modularity of the Louvain of **0.9272**; but, running the algorithms few more times will yield in different yet similar result of 2726, 2718, 2721, 2716, 2717 (**in average of 2719**); more shall be explored in discussion.

Although the sheer number of communities might be disheartening to be analyzed; but, by picking the top 5 communities and analyzing them, the verity reveals itself. Verily the top 5 community can be described as follows:

1. Pro-Ukraine, Western Official & Analyst Sphere (Node: 434)
2. International Pro-Ukraine News & OSINT Focus (Node: 415)
3. German-Speaking User & News (Node: 320)
4. Pro-Russian Official & State Media (Node: 287)
5. German-Speaking Alternative/Critical (Node: 274)

More findings shall be laid upon in the discussion.

H. K-Means Clustering

The result of the K-Means bore a resemblance with the Louvain, yet there was a misty difference between the results. The method used by K-Means is classifying the tweet resembling a Sentiment Classification to create a community, thus by analyzing the tweet more data can be used to do a demarcation of the community for a better induction than Louvain. From the total of 121834 tweets, it can be clustered into this 5 Sentiment:

1. Mixed International Reactions & Calls to Action - 35454 Nodes & 42308 Edges (Cluster 2).
2. English-Language Condemnation, Geopolitical Analysis & Support - 23279 Nodes & 27328 (Cluster 4).
3. German-Language Discussion on Russian Aggression & Western Response - 6883 Nodes & 8569 (Cluster 1).
4. Hashtag Activism - "Russia is a Terrorist State" - 4579 Nodes & 5729 Edges (Cluster 3).
5. Personal Reflections and National Identity (Ukrainian Language) - 3197 Nodes & 3370 Edges (Cluster 0).

to show it abruptly the sample of the results as follows:
Cluster 0 :

1. @electrovi4 ну хоч не через тиждень в травму пішов
2. @getmanv86 мало розказую про це, але буду більше
3. @yurchenkost ну коли він зроблений тобою -то краще б ніхто не бачив 😏

Cluster 1 :

1. @jarlkalle @js281135 @welt was gibt's bei #putin denn zu leugnen? er hat es voll auf den #westen abgesehen und meint fast, er sei der wiedergeborene peter der große. #populisten und #nationalisten im westen groß machen, desinformationen erzeugen, in die wahlen vor allem in den #usa eingreifen . . .
2. @holgerkopp eine "befragung" von kunden im kreis halle/saale ist glücklicherweise nicht repräsentativ. die mehrheit der bevölkerung sieht es völlig anders.#standwithukraine
3. 3. ... wer traute schon #scholz, #merkel, #schröder eine neuauflage des hitler-stalin-pakts zu? entscheidend: #deutschland 🇩🇪, mit kaum funktionsfähiger armee, wäre kein auch nur halbwegs ebenbürtiger partner von #russland 🇷🇺 gewesen, sondern hätte unter dessen knute gestanden. 3/6

Cluster 2 :

1. @crazygolub та тема транспорту мені просто близька, бо працюю в ній майже 9 років, тому реальність знаю. допомагав же перші місяці нарішать фури для гуманітарки, бачив що є непорозуміння як то працює, тому й спитав) за уйобків і бариг х.з. за шо ви.
2. @smayagizyasasin evet güzel gözlüm ben seni çok sevdim lütfen bu sessizlik bitsin artık yağız bebek konuşup koşup oynasın lütfen desteklerinizi esirgemeyin #askimemnu #fenerbahceaustriawien @sedat_peker @erenerdemnet @hazalkaya110 #gelsinhayatbildigibi #championsleague
3. @tarasmi @yanasuporovska @cocacola @cocacola supports russians and genocide of ukrainians #russiaisaterrorisstate

Cluster 3 :

1. @rowan_m72 @nafofella #russiaisanazistate. #russiaisaterroriststate #standwithukraine
2. #boycottrussia #standwithukraine #russiaisaterroriststate
3. @visegrad24 #russiaisaterroriststate

Cluster 4 :

1. @gibwuzhere @fullsac2 @aubrlogic @ilias20095187 @freezyfoe @bnonews so could china. i'd bet on #china's nukes to work, not so much for #russia's.
2. @sajidjavid as one #glovepuppet aka @borisjohnson the #bounder is sacked from office, it is replaced with another glovepuppet! the question the uk should be answering, who are the real #puppetmasters? #partygate #costofliving #borisliedpeopledied #covid #brexitreality #brexit #russia #china
3. @russianembassy @mfa_russia @russiaun @fcdogovuk @10downingstreet @bbcworld @skynews @telegraphworld @financialtimes @guardianworld @iiss_org there is only one way #russia can go to is as per below picture #russiaisaterroriststate ://./4ok9sqr8bm

Detailed analysis shall be explored in the discussion.

IV. Discussion

The analysis successfully identified potentially influential users within the Twitter reply network concerning the Ukraine conflict on August 19, 2022, using various Centrality methods. The context for this influence is a large but structurally fragmented conversational domain, indicated by the very low network density (≈ 0.000087) or only 0.0087%, from the maximum possible number of edges in directed graphics. This sparsity suggests that influence might be localized or operate primarily within distinct community clusters, rather than flowing freely across the entire network; that is also the main reason the contagion process is not really effective despite using the most active user, because of the disparity in the network meaning using closeness centrality might be the better option to analyze it.

A. Centrality

Users identified with high In-Degree Centrality represent immediate centers of attention, receiving a large volume of direct replies (as the To target). These might be accounts posting highly engaging or controversial content that prompts immediate reaction or rather the person that is directly act as the center of the conflicts, in this regard it is the president of Ukraine himself, where he is the most influential people in 3 methods (PageRank, Degree Centrality, In-Degree Centrality). Conversely, users with high PageRank scores possess a more global or recursive form of influence; they are replied to not just frequently, but often by other relatively influential users (big names replying to big names). These could represent established authorities, news sources, or figures whose opinions hold significant weight within the network like being said before. Comparing users high in one metric versus the other can differentiate between mere popularity (high volume of replies, potentially superficial) and deeper network influence, and alongside that there is the Out-Degree Centrality which represent the users who give most attention, showing how active these users are in this topics, but hold not much meaning by themselves(as shown in the contagion).

The fragmented nature of the network, visually substantiated by the community clustering in Gephi, as depicted conceptually in Figures 1, implies that the reach of even highly central users might be somewhat constrained by these community boundaries. Influence might propagate rapidly within a cluster but struggle to cross into other clusters, contributing to the echo chamber effect often observed in online discourse, or rather in this case the internet is way too big of a place that makes each interaction might not be connected into others despite talking about the same topic.

And the possibility of how many dead nodes included in this dataset is not really accounted for since it's not the main objective, and we are not the one who gather the dataset, and it might be a big number given how vast and clustered this topics are, not accounting for the different language that people used in tweeting, making this analysis can be continued in the next project or rather by other people.

Each metrics in the centrality contributes to different use case and can be used for various case, but to be frank in this case using PageRank might be the best method given the scope is to find the 10 most influential users, although the other centrality methods can still be used to find the most important user to start and information diffusion from(closeness centrality) or even trying to make a hit list based on it, sky's the limit.

A principal constraint remains the single-day data scope. User influence can fluctuate significantly over time or die over time. Longitudinal analysis tracking changes in Degree Centrality and PageRank would provide a much richer understanding of sustained influence versus transient popularity. Memory limitations also restricted the dataset size.

B. Community Detection

Expanding further into community detection, the results are passable yet dissatisfying as the number of communities are way too sparse for only 10874 nodes. It can be said that 1 community only has 3-4 people in them, making it a sparse community, and knowing the very low network density number, further proves that every community is so small. By analyzing the Louvain method as an example of 2725 communities, only 1906 of them consisted of at least 2 nodes, and 819 only consisted of one; exploring even further only 463 have at least 3 nodes.

As stated in the methodology, Louvain aim to maximize the modularity by dividing the network into communities, but to maximize modularity is a NP problem, meaning it will take astronomical amount of time to calculate and not effective as stated in [7], and thus making it not feasible to do in short amount of time, both algorithms use greedy approach to tackle the issue making it only locally optimal choice even further making why the result of the algorithm is different as the nodes being added in random and making the greedy approach a bit different in each run.

Putting the number aside, analyzing the nodes in top 5 community it can be inferred or rather inducted that the top 5 communities agenda can be broadly classified into:

1. Pro-Ukraine, Western Official & Analyst Sphere
2. International Pro-Ukraine News & OSINT Focus
3. German-Speaking User & News
4. Pro-Russian Official & State Media
5. German-Speaking Alternative/Critical

but that is only inducted from the username and can be misleading, since a name can't tell the whole story, but some are named as the literal standing of the said user in the current conflict. Since the username can and some users with a Pro-ukraine name can be active in other communities and cases as such can happen in other communities, making the analysis deemed not conclusive.

To answer such a problem the most common method in NLP or Data Science that is Sentiment Classification using Machine Learning can be a better solution for the problem. Rather than mapping the user into a network, community detection can be demarcated using the sentiment of their tweet.

By using K-means we can cluster all the text into a few clusters, in this case such a number is 5. Even though this is a better approach in our opinion, a small number of clusters can be proven way too broad to capture all the opinion and matter, such a case can lead japanese tweets to fall into the deutsch community. Despite the drawback, the cluster can be analyzed easily because the content that is being clustered is based on 'text'.

Base on the result of those clustering, we can do a better educated induction about the community, where:

1. Cluster 0: Personal Reflections and National Identity (Ukrainian Language).

   Themed with Ukrainian talking about their nationalism and daily life a few months after the invasion started. To be said generally, it is filled with tweets of Ukrainian people, and also Russian to a certain degree. To say abruptly it is a pro-Ukraine cluster.

2. Cluster 1: German-Language Discussion on Russian Aggression & Western Response.

   This cluster is themed with tweets in Deutsch, though some Japanese & Korean tweets also lumped into this cluster. The general consensus of this cluster is pro-Ukraine and anti Putin/Russia.

3. Cluster 2: Mixed International Reactions & Calls to Action.

   As the biggest cluster, this cluster is filled with English, Italian, and many more languages from across the globe, talking about the conflict. Most tweets are reacting to the conflict, like condemning Russia and supporting Ukraine. Though some tweets are pro-Ukraine and some are out of topic.

making this cluster, is less cohesive than the other one, and can only be described as an international reaction.

4. Cluster 3: Hashtag Activism - "Russia is a Terrorist State".

   This Cluster is pretty self explanatory, it is only filled with hashtags, some of them is #russiaisaterroriststate and replying or tagging other accounts

5. Cluster 4: English-Language Condemnation, Geopolitical Analysis & Support.

   Almost similar to cluster 2, but mainly in english. This cluster is filled with analysis, condemnation, support, and updates regarding the war. Even though it is a broad cluster being the 2nd biggest, it is far more cohesive than cluster 2. This cluster can be labeled as the people that are concerned regarding the situation by giving analysis and updates regarding the war.

To dissect further information from these findings, required another field called Sentiment Analysis, and here we are only doing a surface level analysis from the data of 19 August 2022.

V. Conclusion

This project focused on identifying influential users within a Twitter reply network related to the Ukraine conflict using various centrality methods for analysis. The analysis operated on a network constructed from single-day data, revealing many participants (10,874 nodes/users) but very sparse overall connectivity (10,282 edges/connection). By calculating and visualizing centrality scores, the study successfully highlighted nodes representing users with significant direct engagement (high In-Degree) and/or broader network influence (high PageRank) and many more. These influential actors operate within a fragmented landscape characterized by distinct community clusters. Although constrained by the temporal scope of the project being only for a single day, the findings demonstrate the utility of centrality analysis in pinpointing key voices within large-scale online conversations and suggest that influence in this context is often localized within specific network communities, or around a main actor in this regard the president of Ukraine himself.

Community detection in this project found a diverse kind of communities that dwell in these conflicts, some are pro-Ukraine; some are condemning Russia; and some are discussing the war and giving updates. The 2 approaches of community detection lead to a different result, where Louvain helps demarcate the user into community (user centric), and K-means help demarcate the tweets into community (tweet centric). Resulting in these conclusions:

1. The discourse is significantly fragmented. The community is often separated by language and national identity leading into a different conversational sphere.
2. Inside the sphere mentioned above, the conversation within each sphere is varied, filled with personal analysis and narrative. It can be biased into an echo chamber, helped by hashtags to further reinforce the dominant viewpoint within each group.

References

[1] Bwandowando, "Ukraine-Russian Crisis Twitter Dataset (1-2 M rows)," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows/data. [Accessed: Mar. 16, 2025]

[2] D. Palsetia, Md. M. A. Patwary, W. Hendrix, A. Agrawal, and A. Choudhary, "Clique guided community detection," 2021 IEEE International Conference on Big Data (Big Data), vol. 4, pp. 500–509, Oct. 2014, doi: 10.1109/bigdata.2014.7004267.

[3] V. A. Traag, L. Waltman, and N. J. Van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Scientific Reports, vol. 9, no. 1, Mar. 2019, doi: 10.1038/s41598-019-41695-z.

[4] N. Dugué and A. Perez, "Directed Louvain : maximizing modularity in directed networks," Nov. 20, 2015. https://hal.archives-ouvertes.fr/hal-01231784

[5] N. R. Smith, P. N. Zivich, L. M. Frerichs, J. Moody, and A. E. Aiello, "A guide for choosing community detection Algorithms in social network Studies: The question Alignment Approach," American Journal of Preventive Medicine, vol. 59, no. 4, pp. 597–605, Sep. 2020, doi: 10.1016/j.amepre.2020.04.015.

[6] S. Sahu, "Heuristic-based Dynamic Leiden algorithm for efficient tracking of communities on evolving graphs," arXiv (Cornell University), Oct. 2024, doi: 10.48550/arxiv.2410.15451.

[7] U. Brandes et al., "Maximizing Modularity is hard," arXiv (Cornell University), Jan. 2006, doi: 10.48550/arxiv.physics/0608255.

[8] S. Wasserman and K. Faust, Social network analysis. 1994. doi: 10.1017/cbo9780511815478.