

**Climate Conversations on Twitter: Assessing the Usefulness of the Similarity Network**

BY

ALINA RASHID

AMSAL MALIK

ZAIN UL HAQ

FINAL PROJECT – GROUP 4

FOR THE SOCIAL NETWORK ANALYSIS

BSC SOCIAL DEVELOPMENT AND POLICY/BSC COMPUTER SCIENCE

HABIB UNIVERSITY

KARACHI, PAKISTAN

2023

Table of Content

[Abstract 3](#_Toc153144182)

[Introduction 4](#_Toc153144183)

[Literature Review/Related Work 6](#_Toc153144184)

[Data Extraction and Network Construction 7](#_Toc153144185)

[Data Collection: 7](#_Toc153144186)

[Data Processing: 7](#_Toc153144187)

[Sentiment Analysis: 7](#_Toc153144188)

[Methodology for the Analysis using Gephi and R Studio: 8](#_Toc153144189)

[Analysis and Discussion 10](#_Toc153144190)

[Metrics Incorporated: 10](#_Toc153144191)

[Discussion: 11](#_Toc153144192)

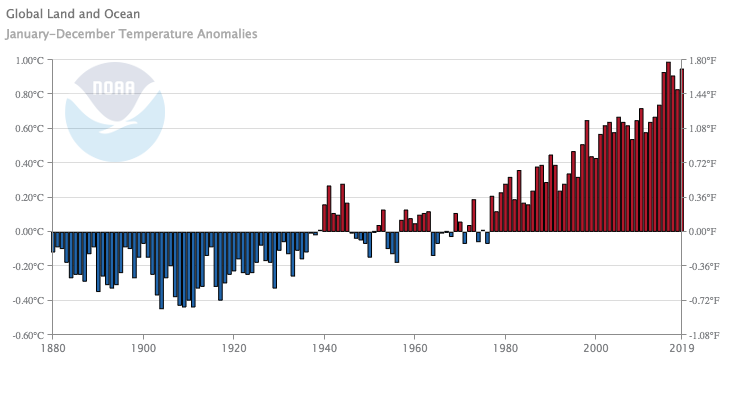
[References 18](#_Toc153144193)

Abstract

Twitter, having billions of downloads, serves to be one of the crucial social media platforms that serves as an opportunity to establish a connection between various actors of society and the masses. This dialogue helps in starting an intellectual revolution which can help in making the world a better place. With Climate Change being the pressing issue in the modern era, which requires immediate steps, the aim is to explore significant Twitter conversions forming linkages based on a similar sentimental score (Threshold=0.05) to ignite actions that can have the utmost important impact. This was done by analyzing the topology of the networks formed using the "Climate Change Twitter Dataset” from Kaggle, where sentimental values are allotted. Tweets from the three most frequent topics, 1) The Impact of Resource Overconsumption 2) Weather Extreme 3) The Seriousness of Gas Emissions were used to form a social network. Contrary to our hypothesis, the results show that Twitter networks using similarity threshold do not exhibit the properties of “small world” phenomena, however, the topology could still be attributed to being significant due to its unmatchable resilience. In fact, the network is significantly different from other Online Social Networks (OSN) like Facebook. However, this could be pivotal for users to locate and organize themselves with users having the same sentimental value towards a topic for a greater impact.

Introduction

Climate change has been one of the pressing issues on the twenty-first issue with a radical increase in greenhouse emissions across the globe. The persisting emissions rates hint at the 1.5 degrees Celsius global warming between 2030 and 2052 (Wisconsin Department of Natural Resources, n.d.). Owing to this massive change, the world has already shown enormous signs of unprecedented natural disasters causing massive social and economic destruction. On one hand, climate change has been at the forefront of issues during contemporary times, while simultaneously the use of social media has escalated and grown exponentially over the past two decades. While there tend to exist various social media platforms to connect individuals across the globe, the most influential tends to be Twitter. Being one of the oldest social media platforms, it allows diplomats, and educationists, and policymakers to start a discourse and engage in a dialogue with the common public. This makes Twitter the hub of change, where ideas are discussed and a revolution is spread. Twitter sentiments were also concluded to have a predictive power to anticipate the returns of major cryptocurrencies (Kraaijeveld & De Smedt, 2020).



**“Fig. 1: Global temperature anomalies from 1880-2019. National Oceanic and Atmospheric Administration (NOAA)” Source: Wisconsin Department of Natural Resources**

Our research aims to explore the power of Twitter social networks to influence the climate change discourse if the users form edges based on similar sentiments attributed to certain topics. This will be vital in finding patterns across communities having similar sentiment scores. Consequently, tightly-knit communities (clustering coefficients and average densities) would be identified who are passionate about the issue, and reveal the information flow amongst them and if this restructuring is efficient for an actual impact. Our hypothesis predicts that the restructuring of the Twitter social network would be vital to organize mass-level impact as well as it would enable the user to collaborate with more individuals, thereby forming a small world network. The paper begins with discussing the related work, methodologies incorporated, and finally analysis of the network of tweets.

Literature Review/Related Work

The study by Mary Sanford *et.al* on “Controversy around climate change reports: a case study of Twitter responses to the 2019 IPCC report on land” tries to gauge the impact of a report on Twitter and the influence of the social media platform in setting up the agenda for the solution of Climate Change. The study analyzed 6000 tweets, related to the Special Report on Climate Change and Land (SRCCL) published by the Intergovernmental Panel on Climate Change (IPCC). After the analysis of the toxicity and sentiment polarity, it was seen that the discourse on the meat and dietary options originated which was merely discussed in the report. The finding was pivotal in depicting how the focus of the masses has shifted from critiquing climate science to now critiquing potential solutions, that too with their rigid stances. All in all, the study captures the role of Twitter in portraying the stance of the public especially on Climate change, and taking action accordingly (Sanford et al, 2021)

Another vital study, “Scientific Networks on Twitter: Analyzing Scientists’ Interactions in the Climate Change Debate” by Stefanie Walter *et.al* stresses how scientists interact with a diverse audience, and how they strategically vary their tone. The study assessed the tweets ranging between October 2017 to March 2018 and analyzed different networks of scientists, journalists, politicians, and others whose profession was unidentified. The results revealed that although the scientific community engages mostly amongst their peers, there still tends to be some sort of connection with journalists, politicians, and the masses. However, it is important to note that their tone toward politicians changes significantly as the stress on the action, indicates that Twitter provides a platform to establish a connection between various stakeholders of society (Walter et al, 2019)

Data Extraction and Network Construction

Data Collection:

The dataset for this study was sourced from a publicly accessible collection on Kaggle, named the "Climate Change Twitter Dataset," which is retrievable at the following URL: <https://www.kaggle.com/datasets/deffro/the-climate-change-twitter-dataset>.

This collection comprises a variety of Twitter posts collated through specific search terms such as "climate change." Despite the dataset's extensive coverage, spanning over a decade of Twitter activity, the scope of this research was intentionally narrowed to encompass only the tweets from the calendar year of 2019. This delimitation was aimed at simplifying the analytical process.

Data Processing:

We started by picking out all the tweets from 2019 to keep our study within just one year. We then cleaned the data, removing any tweets that were missing information or weren't from 2019. Due to technical limitations, the tweet focusing on the following topics were used:

1) The Impact of Resource Overconsumption

2) Weather Extreme

3) The Seriousness of Gas Emissions

To make our visualization easier to handle, we only used tweets from April 22, 2019, which is Earth Day. We chose Earth Day because we expected more people would be talking about climate change then. This helped us have enough relevant tweets to look at closely but not too many to manage.

Sentiment Analysis:

The dataset we used already had sentiment scores for each tweet, which told us how positive or negative each tweet was about climate change. Because these scores were already part of the dataset, we didn't need to use any special tools to find out the sentiments ourselves. Although we don't know exactly how these scores were calculated, they were handy for quickly understanding how people were feeling about climate change in their tweets.

Methodology for the Analysis using Gephi and R Studio:

Using the extracted data from Kaggle called “The Climate Change Twitter Dataset”, and after dividing the data into three CSV files, we needed to prepare our data even further for analysis on Gephi. We now used just three topics that were tweeted about, namely “Impact of Resource overconsumption”, “Weather Extremes” and “Seriousness of gas Emissions”. We now just sampled all the tweets from one day for each individual topic and read it using an R script which further calculated the Euclidean distance between the sentiment scores for each tweet. We built a network based on this, our nodes being the tweets and similar sentiment scores meaning edges between the nodes. We added certain properties from our initial CSV to the network, namely the temperature and the stance for each tweet made. This helped us make a more sound analysis. We generated graphml files, to analyses this data further on Gephi.

We now opened our graphml files on Gephi, now. First, we increased the size of all the nodes, the minimum size being 5.5 and the maximum size being 16 for all the node degrees. Similar degrees matched one another based on the sentiment scores. We now further visualized the stances of each node with a different color. The stances were “Believer”, meaning those who believe climate change exists, “Denier” meaning those who believe climate change does not exist and “neutral” who don’t have much of an opinion on whether or not it exists and use a more central approach. We found that nodes of the same color generally made an edge with one another, the most occurring stance being “Believer”. We figured this out using two layouts for each topic, namely Fruchterman Reingold and Force Atlas. Fruchterman Reingold was useful for us as the data was so carefully put in a circle which made it easier for us to visualize it. Force Atlas, on the other hand, was a bit hard to comprehend, but it simply showed us an accurate representation of how our network is shaped with a snake-like structure. We made the same analysis for temperature after this whereby, we visualized each common numeric temperature value using a different color and found that tweets having the same average temperature temperature were mostly linked with one another. We also found that whether the temperature difference was negative or positive did not make much of a difference. When the amount of difference was the same, the tweets generally made edges with one another. The data for only one day was used for visualization because the data was proving to be difficult on gopher, since large files comprised up to 25 million edges for one topic.

Analysis and Discussion

The analysis of the tweets was shortlisted based on the frequency of the topic/theme highlighted in the tweet. The following three topics were used, 1) The Impact of Resource Overconsumption 2) Weather Extreme 3) The Seriousness of Gas Emissions.

Metrics Incorporated:

**Degree Centrality**: Degree centrality refers to the number of edges a node has which explains the frequency of connections with other nodes. The higher degree indicates the higher number of linkages of the node with its neighbors. This explains that under a certain aforementioned topic, the tweets with a high tendency of similarity with other tweets would have a higher degree of centrality. In other words, other Twitter accounts also talk about a topic in a similar way and associate similar sentimental value as the certain tweet is talking about. This could also help us discover how a certain topic is being perceived across the globe.

**Clustering coefficient:** The clustering coefficient refers to how densely and interconnected are the nodes. This indicates the nature of clusters and communities that exist within the grander social network. With regard to our current scope of the study, it indicates how much of the tweets from the communities share similar information and sentimental value. It also implicitly tells about the flow of information shared within a mega cluster as well as small clusters.

**Density:** The density measure conveys the ratio of the edges that have been formed in the graph and the total potential edges that the graph could have encapsulated. Graph density illustrates whether the nodes in the graph and densely connected or are forming a sparse real-world network. In this context, the density of the graph would illustrate how much of the sentimental saturation would have occurred in the social network about a certain topic of the tweet. In other words, it would hint if the sentiment of the tweet related to a certain topic is mostly similar, or if there is diversity in terms of different perspectives.

**Average Path Length:** Path length measures the shortest number of steps a node takes to connect with all the other existing nodes. Consequently, the average path length inculcates all the nodes' path lengths to generate a cumulative and average number of steps taken. This informs about how disconnected or connected the nodes are in the social network. This metric on three tweet networks generated under our analysis would indicate how widely are nodes connected which translates into the intensity at which the flow of information takes place.

**Diameter:** The diameter of the social network denotes the longest and shortest distance between the two farthest-located nodes in the network. In other words, it depicts the steps required via other neighbors to form an interpersonal link between two distant nodes. Under our research, the metric indicates the connectivity amongst the Twitter community vocalizing for a similar topic, while also hinting towards the extremes of sentiments attached to a certain topic from one entity to another.

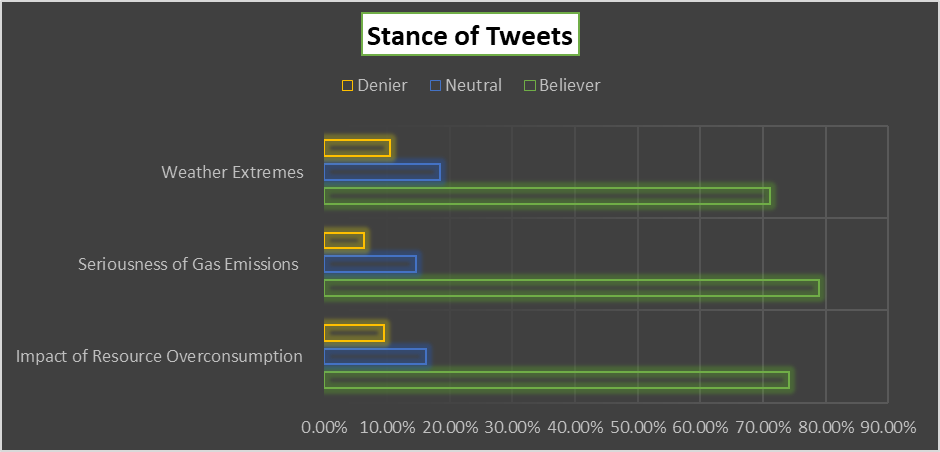
Discussion:

The tweets for one year on the topics of “The Impact of Resource Overconsumption”, “Weather Extremes”, and “The Seriousness of Gas Emissions” were analyzed using the inferential statistics and metrics above. Table 1 summarizes the statistical results which highlights that although different topics, if linked using similar sentiments tend to exhibits similar topology of network. A detail analysis would reveal potential advantages and shortcoming of this formulation of social network.

|  |  |  |  |
| --- | --- | --- | --- |
| **Metrics** | **The Impact of Resource Overconsumption** | **Weather Extremes** | **The Seriousness of Gas Emissions** |
| Total Number of Nodes | 7857 | 26367 | 13197 |
| Total Number of Edges | 2330859 | 25898653 | 6509236 |
| Average Degree | 593 | 1293 | 986 |
| Maximum Degree | 1226 | 3424 | 1542 |
| Minimum Degree | 4 | 13 | 2 |
| Clustering Coefficient | 0.82 | 0.82 | 0.79 |
| Density | 0.076 | 0.081 | 0.074 |
| Average Path Length | 10.2 | 10.3 | 10.2 |
| Diameter | 40 | 39 | 38 |

**Table 1: Descriptive Statistics**

The initial analysis reveals that the “weather extremes” was the most talked topic on the Twitter linked with climate change with over 26,000 tweets in the entire year of 2019. This indicates that the public took a lot of interest in conversing about it, mainly because they were directly impacted by the extreme weather conditions. The high volume of tweets also showcase that people shared the events of climate extremes, like floods, hurricanes, and heat waves, reporting about real-life incidents, engaging in debates, and highlighting possible solutions to the unprecedented catastrophe. It was followed by “The seriousness of Gas Emissions” with approximately 13,000 tweets which indicated that there is a persistent escalation of the awareness of the direct and indirect impact of greenhouse gases like carbon dioxide and methane. With this regard, there has been a growing trend in questioning human practices like industrial process where fossil files are burned. The topic of “The impact of Resource Overconsumption” with approximately 8,000 tweets was relatively less popular as compare to above-mentioned topics. The figure helps us depict that most of the individuals who are concerned with climate change, are other topics are conversed up to 3 times more than resource overconsumption. One of the prime reason for this is that the twitter users are mostly residents of the urban area who are hustling and bustling in their work lives who are not aware about the phenomena’s of deforestation and depletion of natural resources. However, social activists and intellectuals play an imminent role in sparkling debates and make other platform users aware about these underlying issues impacting climate change.

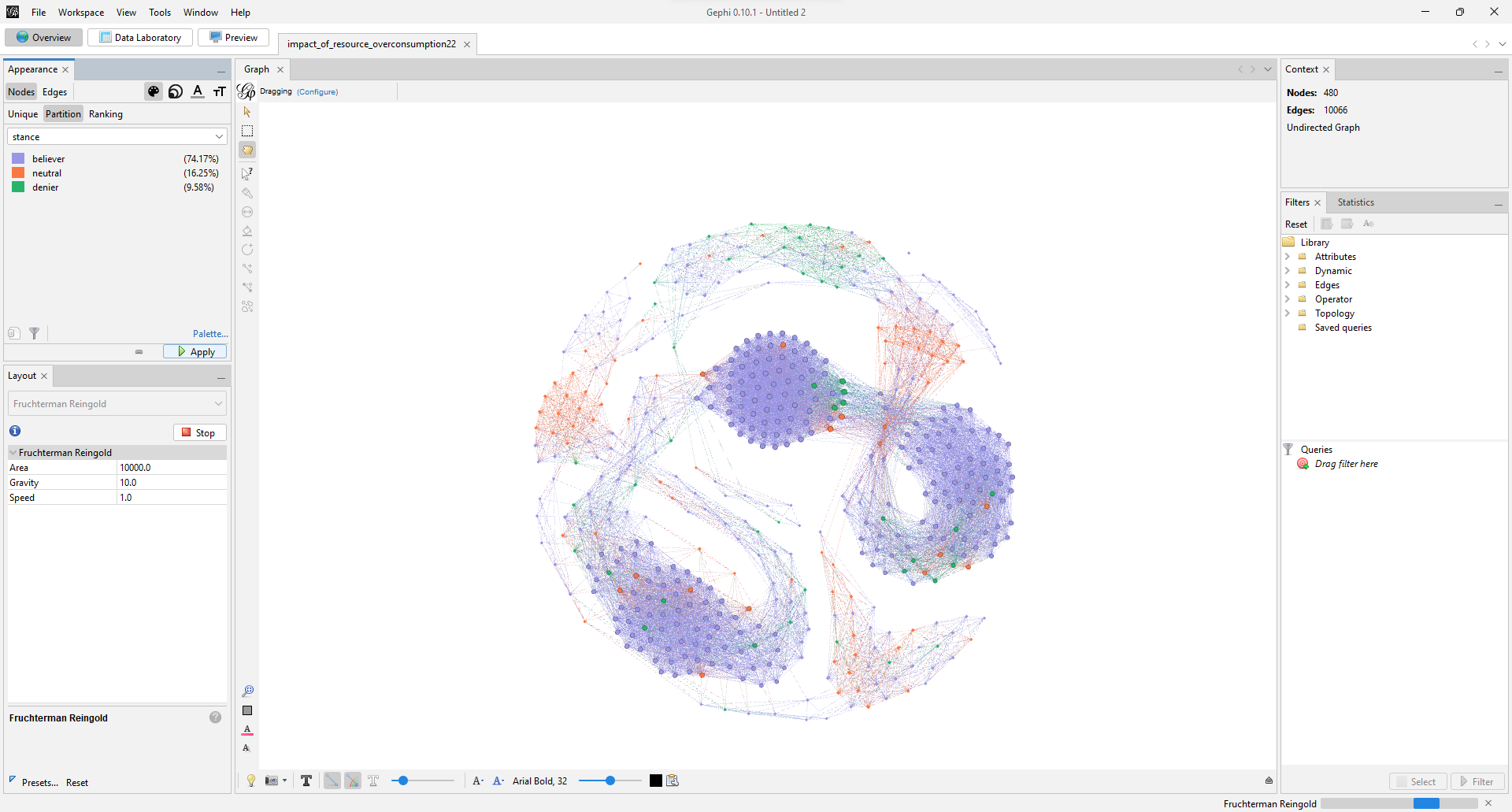


**Fig 2: Bar chart for stance of tweets as per topic**

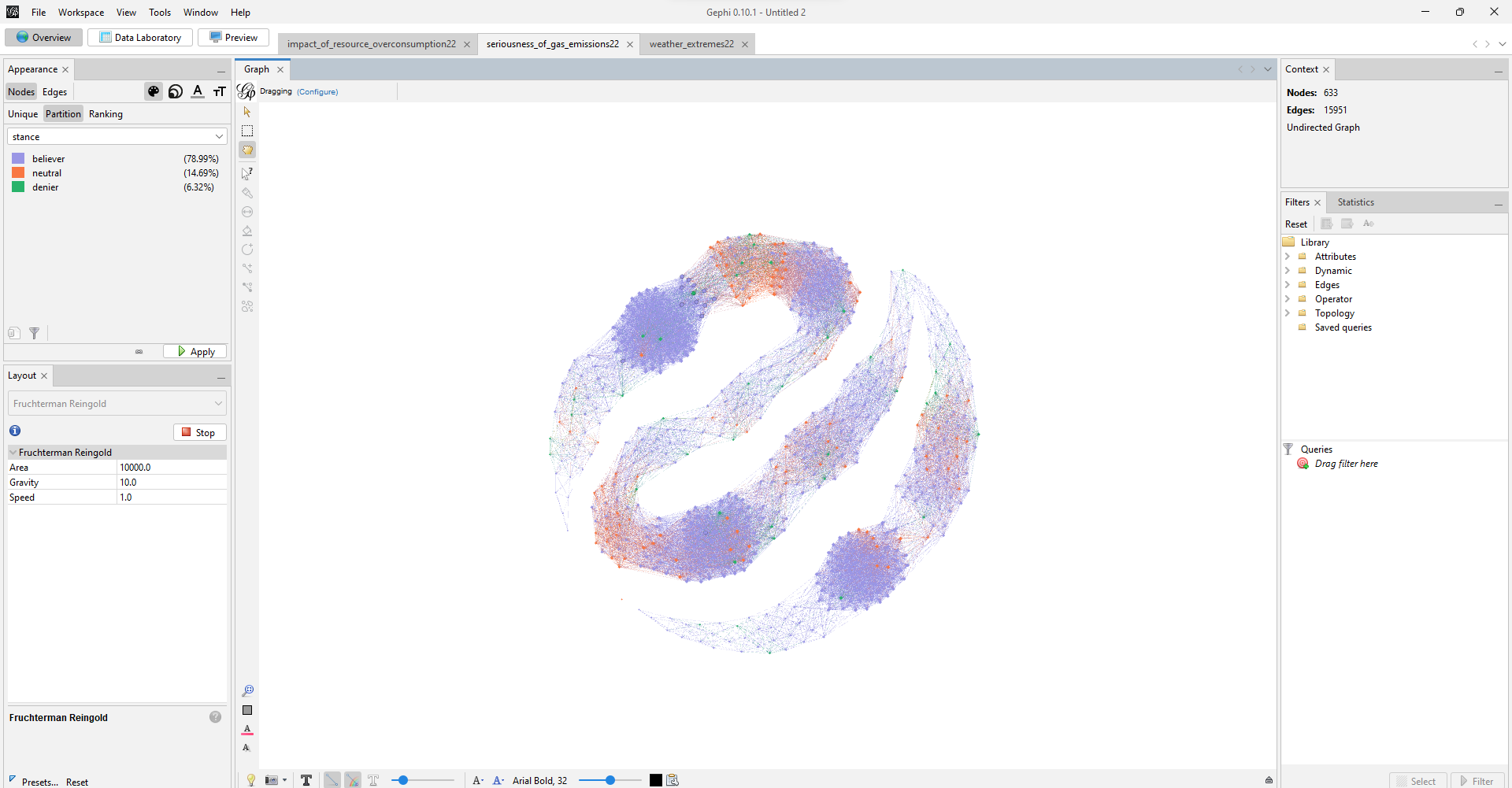
Although the frequency of the tweets vary according to the topic, the analysis reveals that the stance for each topic exhibited on twitter is to large extent same. The believer stance is adopted by roughly 70% of the tweets, while the neutral and denier stance is adopted by approximately 15% and 5% of tweets, respectively. This is vital to understand that the twitter is being used by the users who are firm believer that there is an urgent need of addressing the climate change issue, and connecting them based on a shared sentiment is beneficial/

The discourses ignited at this powerful social media platform, does not only need the number of tweets but it also require persistence and consistent social presence that could lead toward the voices and ideas being brought into action. The unconventional ideas could mean inflicting losses to large conglomerates and organizations that have been doing business in conventional ways. Therefore, if an ever-lasting impact has to be made, the network has to be robust and resilient to avoid network failure in case of any attack on the network. Upon analysis of the formulated network, the average clustering coefficients of all three graphs generated fall towards the higher range (approximately 0.8), meaning all of the tweets with similar sentiments towards the topic could exist in local clusters. This not only helps in the rapid flow of information to mobilize mass communities such as for a protest but also helps its robustness against the failure of individual nodes. In other words, all of the networks formed are resilient from a foreign attack since there are alternative paths to connect with other neighbors and make a collective impact. Although this high communal structure helps make the network resilient and ensures long-term impact, it also undermines the flow of information across different communities in the macro setting.

Visualization of one-day tweets ascertains the potential of different local communities strongly interconnected with each other, having a similar stance of the “believer” (Purple Color) for advocating for “the impact of resource overconsumption”, and consequently climate change. However, if certain high degree nodes are removed from each community, the link between the two communities is completely eradicated. Hence, there exists the dilemma that while local groups are well connected, strong overall collaboration for an efficient impact is missing in this network formation.

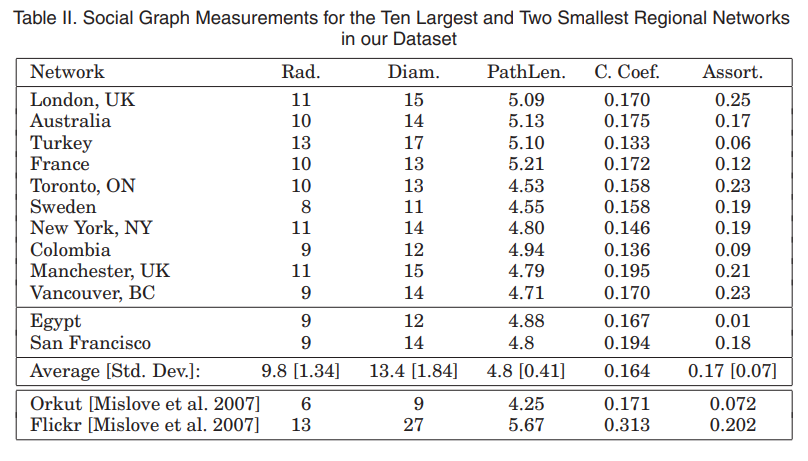
**Fig 3: Visualization of one day tweets under “The Impact of Resource Overconsumption”. Nodes color allocated - Believer (Purple), Neutral (Orange), Denier (Green)**

In fact, a relatively higher average path length of 10.2 across three networks, indicates that on average a tweet has to take 10 steps to connect with another tweet. Although the nodes represent tweets, instead of Twitter accounts, this social network still represents an ineffective flow of information. Hence, the network does not coincide with the “small world “phenomena which could be vital to advocate for climate change. However, this could also be said to be effective to some extent, as different NGOs and other organizations working on climate change might be better off and more efficient when working separately and independently.

**Fig 4: Visualization of one day tweets under “Seriousness of Gas Emissions”. Nodes color allocated - Believer (Purple), Neutral (Orange), Denier (Green)**

However, Fig 4 for a relatively popular topic with almost double the number of tweets reveals that as a certain topic and idea start to gain popularity, the local communities still exist, however, now with more significant collaborations amongst all the advocates. Worthwhile, there is also a significant increase in linkages with deniers, which indicates advocates taking the opportunity to interact and engage in a dialogue with them to increase their support. The veracity of this claim is also reflected in a reduction in diameter for gas emissions topic as compared to resource overconsumption, from 40 to 38.

However, as the network tends to increase with the number of nodes, it starts to move away from the small-world properties. As shown in Table 1, although diameter, density, and clustering coefficient decrease as the number of nodes increases to around 14,000. However, all these properties tend to again decrease as the nodes are increased up to 26,000 tweets.



**Table 2: Source: Beyond Social Graphs: User Interactions in Online Social Networks**

**And their Implications by Christo et al., University of California Santa Barbara**

Along the same lines, it is important to compare our social network of Twitter being drive with that of the other Online Social Networks (OSN). Table 2, shows that a study at the University of California derived the following results from their Facebook social network. The comparison reveals that Facebook strikes a balance between clustering coefficient and average path length, exhibiting the small world properties since it serves a platform where the friend of a friend is also a friend, which also helps to overall diameter (WILSON et al., 2012). While this can be effective in connecting with people, Twitter users prefer to form and engage with their communities to have a common voice. This might be the key characteristic which sets Twitter apart from other OSNs, where individuals are interconnected within their network.

**Conclusion and Future Work**

Climate change is an urgent issue that needs to be addressed to be saved from the unprecedented catastrophes that the future holds. Twitter has been a platform for decades that has been used to advocate for social change and ideas since it provides an efficient way to engage with other users. A high number of tweets having a “believer” stance validated this claim. The network of tweets formed using similar sentiment thresholds for various topics under the realm of Climate Change, reveals that although all these networks have different frequencies of nodes, they could be organized to have similar network topology. If organized based on the metric closely-knit communities, it would be beneficial to instantly share information and to build resilient structures to increase the longevity of the ideas and impact. There also tend to be few collaborations and similar sentimental value amongst members of different communities, enhancing the scope of collaborations and combined resistance. This social network being distinctive from other OSNs helps in understanding the need and efficiency of Twitter network to be used for climate change actions. Hence, the research helps in understanding how Twitter users could be organized by sharing similarities in terms of sentiments they attach, however, it also calls for Twitter (Now X) to enable the change-makers to locate other Twitter users sharing similar sentimental values by using its algorithm.

This research, although being conducted with the intention of producing reliable results, only used the three topics due to time and technological constraints. In fact, the network for entire one-year tweets could not be visualized due to the same limitations. Future work could cater to these shortcomings, as well as locate these tweets geographically to see if a true impact has been achieved due to advocating on the Twitter platform.

References

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for

Forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188.

Sanford, M., Painter, J., Yasseri, T., & Lorimer, J. (2021). Controversy around climate

change reports: a case study of Twitter responses to the 2019 IPCC report on land.

*Climatic change*, *167*(3-4), 59.

Walter, S., Lörcher, I., & Brüggemann, M. (2019). Scientific networks on Twitter: Analyzing

Scientists’ interactions in the climate change debate. Public Understanding of Science, 28(6), 696-712. <https://doi.org/10.1177/0963662519844131>

WILSON, C., SALA, A., P. N. PUTTASWAMY, K., & ZHAO, B. Y. (2012, November).

*Beyond Social Graphs: User Interactions in Online Social Networks and their Implications*. Retrieved December 10, 2023, from <https://people.cs.uchicago.edu/~ravenben/publications/pdf/graphs-tweb12.pdf>

Wisconsin Department of Natural Resources. (n.d.). *The Science of Climate Change | the*

*world is warming | Wisconsin DNR*. Wisconsin DNR. Retrieved December 10, 2023, from https://dnr.wisconsin.gov/climatechange/science