IT UNIVERSITY OF COPENHAGEN

Using algorithms to identify climate activism trends on TikTok

Bachelor Research Project
IT University of Copenhagen (Spring 2024)

Author

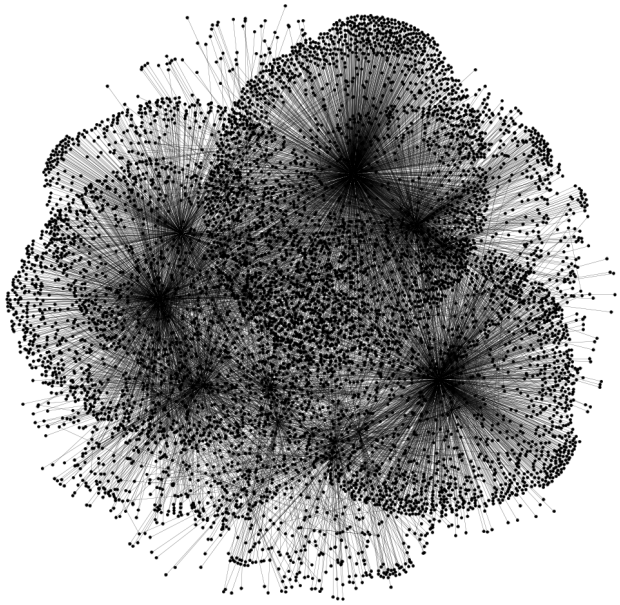
Francesco Caglianone, Student N. 21750

Supervisor

Prof. Luca Maria Aiello, Associate Professor @ ITU Copenhagen
Department of Computer Science

GitHub Repository

TikTok-network-crawler: FrankCaglianone



Representation of the network of TikTok climate activists used in the research.

Abstract

This research project explores the network dynamics of TikTok users engaged in climate activism, employing network science methodologies to map and analyze interactions and influence within this digital community.

Using the TikTok Research API, a network was constructed starting from a set of users known for their climate activism on TikTok. Through the snowball sampling method, implemented using the Breadth-First Search (BFS) technique, the study expanded the network to include users interconnected by following relationships.

The PageRank algorithm was then applied to determine the influence of users within the network, dividing them into quartiles to determine if certain trends and topics exert more influence within the network through hashtag analysis. Additionally, the Louvain algorithm was employed to detect communities within the network, to identify the predominant topics discussed within each community, and to explore how different topics link communities that share similar interests.

The study also implemented network backboning techniques to filter out the less significant interactions and focus on the most important relationships between nodes, providing a clearer view of how certain topics spread within the community.

Findings explain the structure around climate activism within the TikTok network, highlighting the diverse yet interconnected topics shared by various groups and showing how users strategically use popular hashtags to engage with a broader audience.

This research deepens our understanding of TikTok's role as a social media platform in climate activism by identifying the most influential topics and how they vary among different user groups. It provides new insights into how climate activism resonates across this rapidly growing platform, which has yet to be extensively studied.

Keywords

TikTok, Networks, Social Network Analysis (SNA), Climate Activism, Snowball Sampling, Page Ranking, Network Backboning, Community Detection, Trends and Topics, Algorithms.

Table of Contents

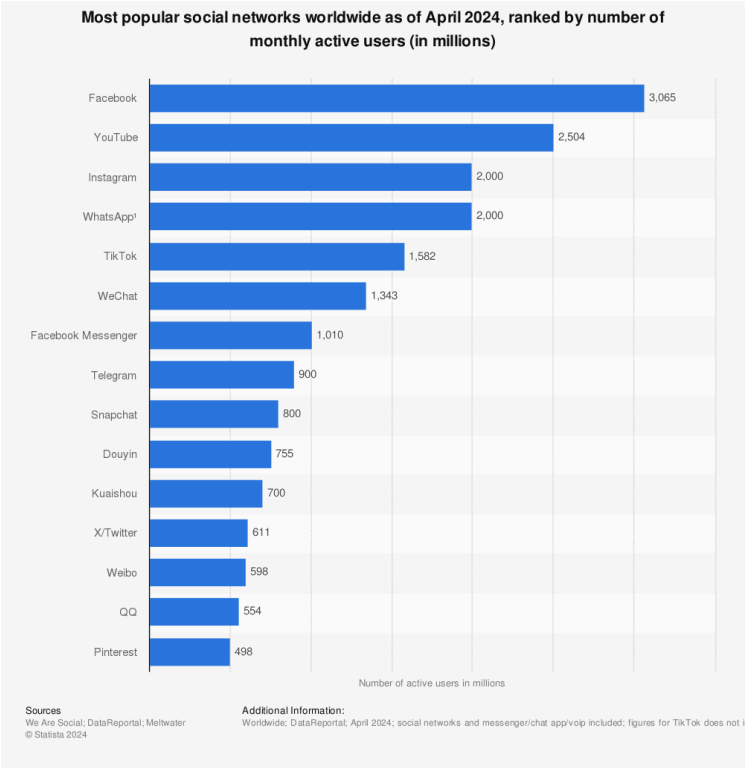
Tools & Technologies

- Python
- TikTok Research API
- Requests Library
- Igraph
- Numpy
- Matplotlib
- Gephi (Yifan Hu Layout)
- Network Backboning Algorithm by Michele Coscia and Frank M. H. Neffke [18]
- ChatGPT (In the making of this paper, the AI tool ChatGPT (OpenAI. (2023) was used to write in more academically inclined and proper English.

Introduction

The rise of social media has transformed social activism, particularly regarding climate change. Today, social activism is shaped mainly by user interactions on social media platforms like X/Twitter and Reddit [1] [2] [3], where activists spread awareness, engage in discourse, and mobilize action. This is why these platforms have been focal points for various studies and research on behavioral change in digital environments, specifically focusing on social media. [3]

Despite the plentitude of research on platforms like X/Twitter, Reddit, and more, [3] TikTok, a newer and rapidly growing platform, has not yet been studied as extensively, even though as of April 2024, it boasts more than double the number of monthly active users compared to X/Twitter. TikTok's unique content delivery mechanisms offers a different dynamic in how information is shared and consumed. The app provides an algorithmically personalized content feed, the so-called For You Page, which has become a fixed part of their user's everyday routines, with some people describing it as addictive. [4] This gap in research leaves a portion of digital climate activism potentially misunderstood or underrepresented in academic studies.



This paper proposes an innovative study designed to explore the trends surrounding climate activism, specifically on TikTok. Starting from a known set of TikTok users actively engaged in climate activism, the algorithm begins by mapping these users' followings, which are fetched via post request to the TikTok Research API. It then iteratively maps the followings of their followings to construct a relevant subgraph of the TikTok network. Various SNA (Social Network Analysis) tools and metrics are then employed, such as PageRank to assess the importance and influence of users within

the network and the Louvain Algorithm to identify communities within it. Finally, the analysis examines the distribution of popular hashtags across different levels of user influence to understand what topics trend among famous influencers compared to less influential users and across various communities to understand what topics and trends are connected with each other.

Focusing on TikTok, this study will contribute new insights into the patterns of digital activism in an evolving social media landscape.

This paper will outline the methodological framework, discuss the data cleaning and analysis processes, and interpret the implications of emergent trends in digital climate activism, trying to answer the following questions:

1. What topics and trends are commonly associated with climate activism?
2. Where are most climate activists located?
3. How can we get more people involved in climate activism?

Literature Overview

Network analysis is fundamental to understanding complex systems, particularly in digital and social media environments. "*A network is a mathematical object consisting of a set of vertices (or nodes) that are connected to each other by edges.*" [5] This simple description corresponds to a wide range of systems in the real world, from technological ones, such as the Internet and the World Wide Web, to infrastructural ones, such as roads and power grids, all the way to biological networks, such as the nervous systems or protein interactions. [5]

Social media ranks among the most extensively analyzed complex systems, evidenced by many research articles, mainly focusing on platforms like X/Twitter and Reddit [3]. A significant contribution to this field is the study by A. Pera, L. M. Aiello, and G. De Francisci Morales, titled "Measuring Behavior Change with Observational Studies: a Review," which advocates for the use of data from a broader array of digital platforms, to enable researchers to examine behavior changes across different contexts and demographics, potentially uncovering unique patterns and trends [3]. The paper also points out that there is a lack of deep theoretical grounding in many of those studies, suggesting a stronger integration of theoretical frameworks. [3]

TikTok launched in 2017, and despite its relatively young age, it has seen remarkable growth in recent years, with a significant rise during the Coronavirus pandemic as a means of escape and relief that users could not find elsewhere. [4] TikTok's popularity is attributed mainly to its algorithmically personalized content feed, the so-called "For You Page," which tailors content feeds based on each user's interests, making the content feel more relatable. [4] A. Schellewald describes this process in his paper titled *"Understanding the Popularity and Affordances of TikTok through User Experiences"* as 'reading for personalization.' This involves users skipping over 'basic' videos, navigating towards 'hits,' and marking them with a 'like' to signal their preferences to the algorithm. Schellewald's paper also highlights user testimonials indicating that people prefer TikTok over other platforms like Instagram and X/Twitter because it provides relatable content from everyday people instead of established influencers and supermodels. [4]

The environmental crisis is one of the main concerns in today's social, political, and media agendas, causing widespread engagement in environmental activism.

In recent years, Social media has become the main platform for environmental and climate activists to spread awareness and mobilize communities, leading to the rise of a phenomenon known as "eco-influencers." These are individuals who use social media platforms like Instagram and X/Twitter to share content focused on sustainability and climate change. [10] [11]

Identifying influential nodes in complex networks is a critical issue in Social Network Analysis (SNA). [6] In literature, common centrality measures such as degree centrality, edge-betweenness, betweenness centrality, closeness centrality, and eigenvector centrality are widely used to estimate an individual's influential power within social networks. [6] [7] However, in my opinion, and in the context of this paper, the most effective measure is an eigenvector centrality, PageRank. PageRank has been extensively utilized and validated in various studies to identify key influencers in social networks. [7] [9] The problem with degree centrality is that it measures an individual's importance merely depending on the number of friends that person has, without also considering the influence of those friends. For example, having connections with highly influential individuals like the Pope, Obama, and Berlusconi is more significant than having many connections with less influential individuals, such as family members. [7] [8] Moreover, other centrality measures, such as edge-betweenness, betweenness centrality, and closeness centrality, are often computationally expensive due to the need for shortest-path calculations, which require $|v|^3$ operations. In contrast, eigenvector centralities like PageRank use random walks instead of shortest paths and employ clever linear algebra techniques, which can reduce the computational complexity of $|v|^2$ operations. [8] This can reveal hidden patterns of what topics regarding climate activism have a higher impact and influence on social media.

Network backboning is the problem of taking a network that is too dense and removing the connections that are likely to be not significant. [8] In social network analysis, which deals with a vast array of connections, it is an essential process to reveal key patterns of interaction and influence, that might otherwise remain hidden. This research used the Network Backboning algorithm from Michele Coscia and Frank M. H. Neffke. Their paper “Network Backboning with Noisy Data” introduces this new algorithm to extract the significant edge backbone from a noisy, complex network. [18] It employs the Noise-Corrected (NC) backbone extraction method, together with a significance threshold, to remove less significant connections, ensuring that the backbone of the network consists only of the most important connections. [8]

Community detection can reveal hidden patterns of collaboration and support among users, as well as understandings of interconnected topics of interest within these communities. This research employs the Louvain algorithm for community detection, which is known for its efficiency in discovering high-modularity communities in large networks. Modularity is a scalar value that measures the density of links inside communities as compared to links between communities, so a higher modularity value means a better separation of communities within the network. [12] Identifying these communities is essential because it allows for examining sub-groups that are closely connected by shared interests or common content.

The first hashtag ever used was *'#barcamp,'* which appeared on Twitter in 2007, [13] after the proposal of Chris Messina, a social technology expert, who suggested using the "#" symbol to group related tweets together and facilitate conversations on the platform. [17]

Defined as "a string of characters preceded by the pound symbol #" [16], the actual etymology of the word comes from:

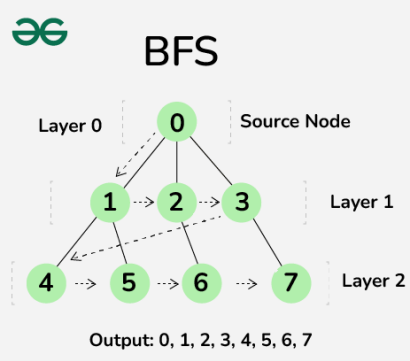
- *Hash*: In the context of computing and technology, "hash" refers to the pound sign "#" character.
- *Tag*: In the context of an information system, it is a metadata tag; it's a keyword or phrase used to categorize content or make it easily searchable.

The use of hashtags allows users to attach metadata to their posts, thus increasing the probability that their posts will be found and followed. [16] Michele Zappavigna [14] refers to this tagging method as 'ambient affiliation,' highlighting that it connects users who share interests in the same topics, even if they have never directly interacted. Zappavigna [15] further observes that hashtags represent a shift from users searching for content to searching for other communities with similar interests. This change highlights how hashtags help create online communities around shared interests, which is exactly what we are trying to discover in this research.

Methodological Framework

Snowball Sampling

To map the network of TikTok users involved in climate activism, the *'user_following_query.py'* script was employed. This script utilized the snowball sampling method implemented with the Breadth-First Search technique to systematically expand the network.



Picture from GeeksforGeeks

Snowball sampling is a non-probability sampling technique used in research to extend the sample by asking initial subjects to identify more participants. The method is, in fact, named "snowball" because, like rolling a snowball down a hill, the sample size grows as the study progresses.

Breadth-First Search (BFS) is a graph traversal method. It begins at a root node and explores all its neighbors at the current depth before moving on to nodes at the next depth level.

In the context of this research, the process started with a predefined set of seed users known for their engagement in climate activism on TikTok.

- unclimatechange
- bbcnews
- climateadam
- dr_gilbz
- dw_planeta
- extinctionrebellionxr
- greenpeace_international
- guardian
- ourchangingclimate
- tedtoks
- juststopoil
- minuteearth
- metoffice
- zbtheclimateoptimist
- greenrupertread
- friends_earth
- parleyfortheoceans
- action4climate
- pique_action
- margreen_s
- nowthiseart

A queue was used to maintain a list of users yet to be explored to manage BFS exploration efficiently. After inserting the seed users in the queue, the algorithm started exploring all their direct connections (users they follow). This first layer of connections formed the first "wave" of network expansion. Subsequently, the connections of each user in this first wave were explored, adding a second wave of users to the network, and so on. This iterative process continued until the desired network size was achieved.

When the script finishes running, whether for successful completion or encountered exception, it generates several .csv files:

- **`parsing_list.csv`**: This file links usernames to integers that indicate the outcome of their post requests:
  1 = Parsed successfully
  2 = Status code 403: User cannot be accessed
  3 = Status code 500: User not existent or found
  4 = Unknown error
- **`queue.csv:`** This file contains a list of usernames that have been fetched but not yet parsed, which helps in resuming the process if interrupted.
- **`saved_jsons.csv`**: Stores the JSON responses received for each user query.
- **`time_stamps.csv`**: Records the timestamps for each response.
- **`network.csv`**: The list of edges, tuples of "source" (username) and "destination" (their following), which represent the network structure.

After reaching the desired network size, the graph was cleaned by removing any connections (edges A -> B) where either user A or user B had not been successfully fetched.

Pagerank Analysis

PageRank is an algorithm developed by Google and named after one of its founders, Larry Page. It is designed to rank websites based on their importance. The algorithm determines a page's importance by counting how many other important pages link to it, which is a circular definition. How can we determine the importance of a page in the first place?

The algorithm starts by assigning a numerical weight to each page. The process involves a "walker" who navigates randomly from one connected page to another, counting the number of visits to each page. The higher the score, the more influential the page.

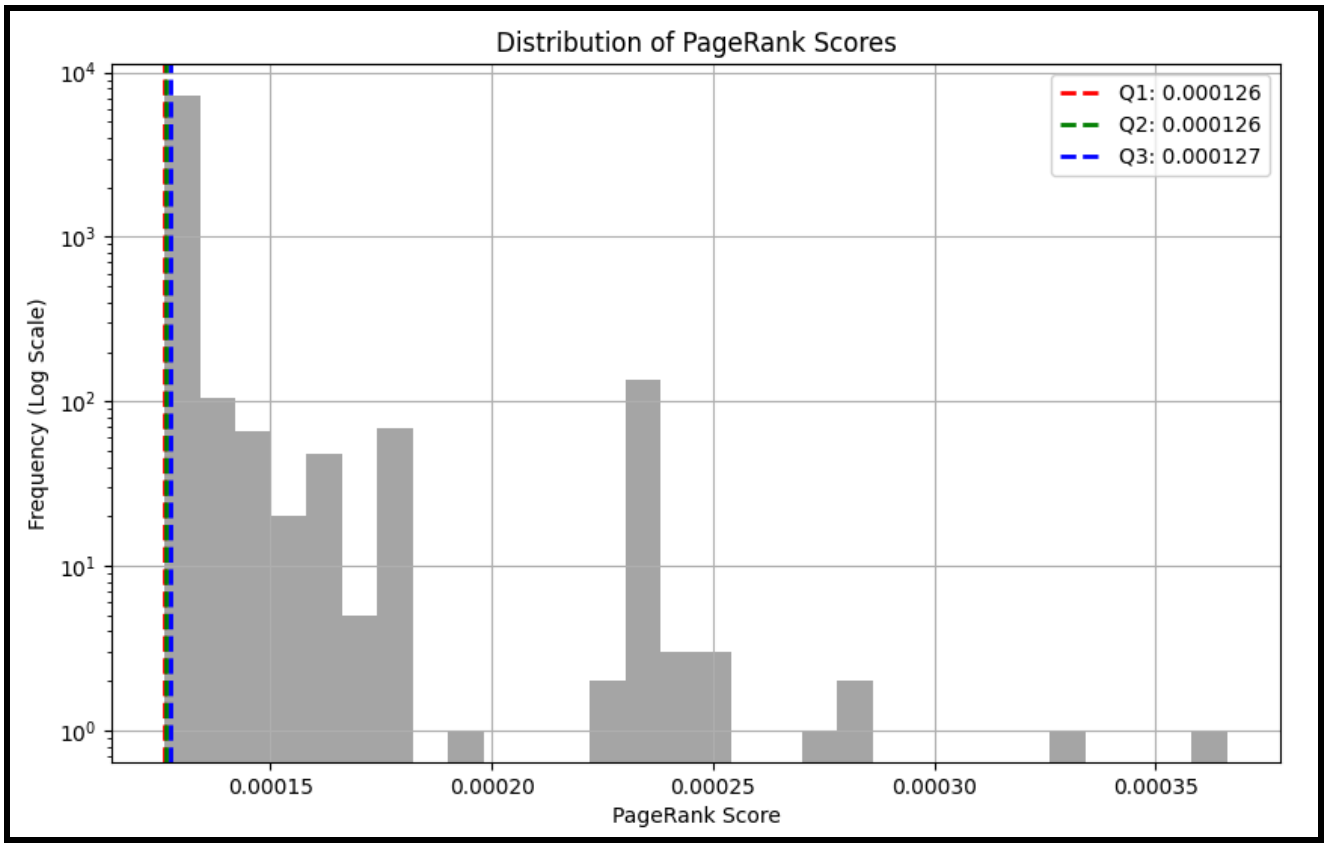
However, if the network of pages is not fully interconnected, the walker will only explore a set of pages, leaving the rest of the network unexplored. To overcome this limitation, Google gave the walker a "teleportation" device. At each step, the walker can randomly teleport to a different page, ensuring no part of the network remains unexplored. [8]

The teleportation parameter is denoted by ε, commonly referred to as the damping factor $(1 - \varepsilon)$ in literature. The typical value for ε is 0.15, which establishes the damping factor $d = 0.85$. This means that 85% of the time, the walker follows direct links between pages, and the remaining 15% of the time, the walker teleports randomly. [8]

This method ensures that, over time, the algorithm can accurately calculate the PageRank of all pages in the network.

In the context of this research, Pagerank was utilized to identify the most influential users within this network, meaning users with the ability to affect the spread of information regarding environmental issues. In this framework, the network is conceptualized as a directed graph, where each node represents a user, and directed edges are the relationships between them. It's important to understand that in this directed setup, an edge from person A to person B means A considers B important, which contributes to B's score.

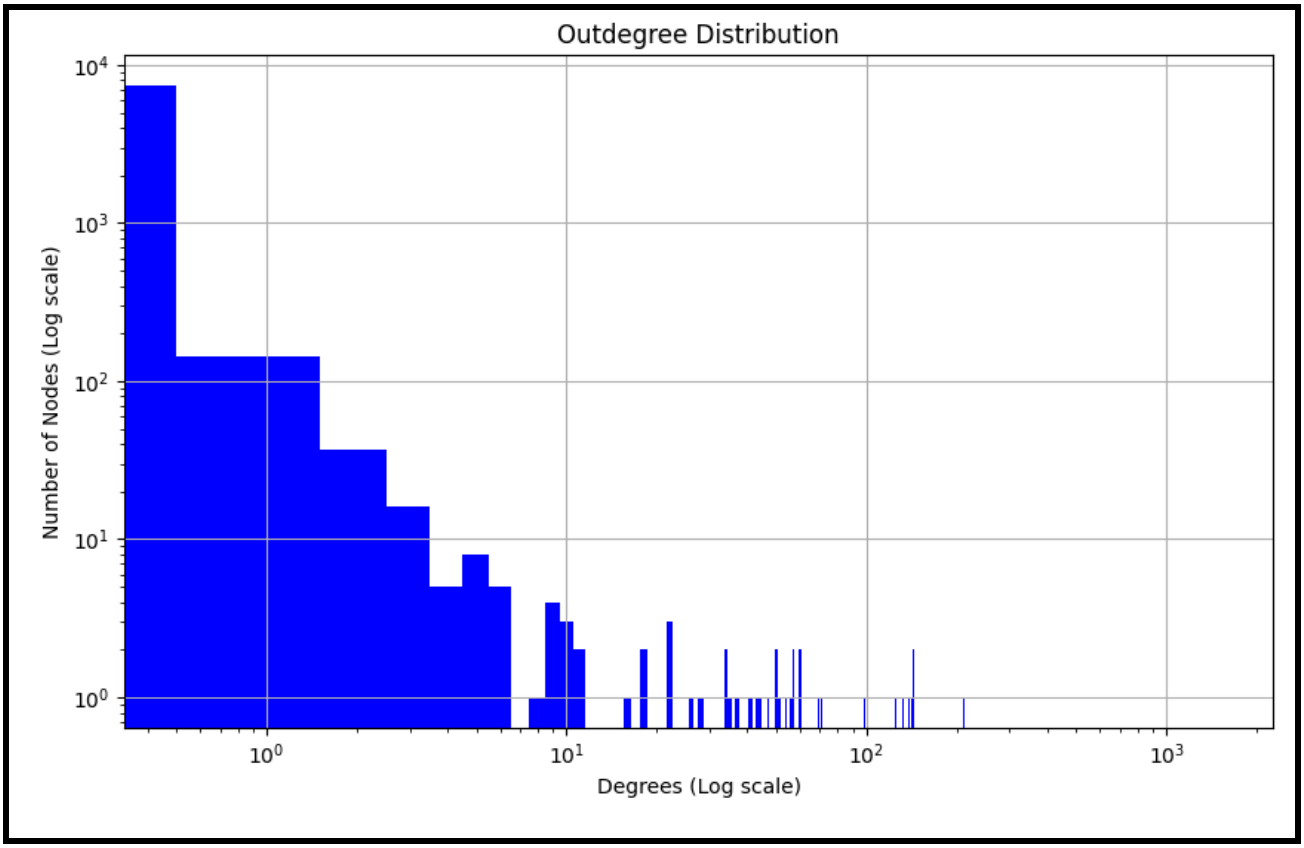
In the *'network_analysis.py'* script, the igraph library was used to compute the PageRank score for each user using a damping factor $d = 0.85$. After calculating all the scores, the users were divided into four quartiles based on their PageRank scores, categorizing them from lowest to highest influence.



The histogram reveals that the first three quartiles (Q1, Q2, and Q3) are very close to each other yet significantly different from the maximum value. This pattern suggests a right-skewed distribution, highlighting that most users have similar, low influence (low PageRank scores). Conversely, a few users have extremely high influence (high PageRank scores). This is shown by the few sparse bins toward the right side of the histogram, showing that while rare, there are users with significantly higher influence or connectivity compared to the rest.

In terms of our network's structure, this means that the majority of users act as peripheral nodes with limited connectivity, while a small number act as central hubs, having disproportionately larger influence or connectivity within the network.

By dividing the data into four subsets, we can later retrieve and analyze the hashtags used by each user within these groups. This will enable us to compare the hashtags and topics between the most influential subgroup (the 4th subgroup) and the least influential (the 1st subgroup). Through this analysis, we aim to determine if certain hashtags and topics trend more and exert more influence within the network.



Outdegree distribution for a better understanding of the PageRank Distribution

Network Backboning

Network backboning is an analytical technique used to filter out the less significant interactions within a complex network, thereby highlighting the most important relationships between nodes. It is an essential process in large social networks like TikTok, where the vast number of connections can hide key patterns of interaction and influence.

Starting from our complete network graph of TikTok users involved in climate activism, the Network Backboning algorithm from Michele Coscia and Frank M. H. Neffke was used to remove less significant connections. Their paper “Network Backboning with Noisy Data” introduces this new algorithm to extract the significant edge backbone from a noisy, complex network. [18]

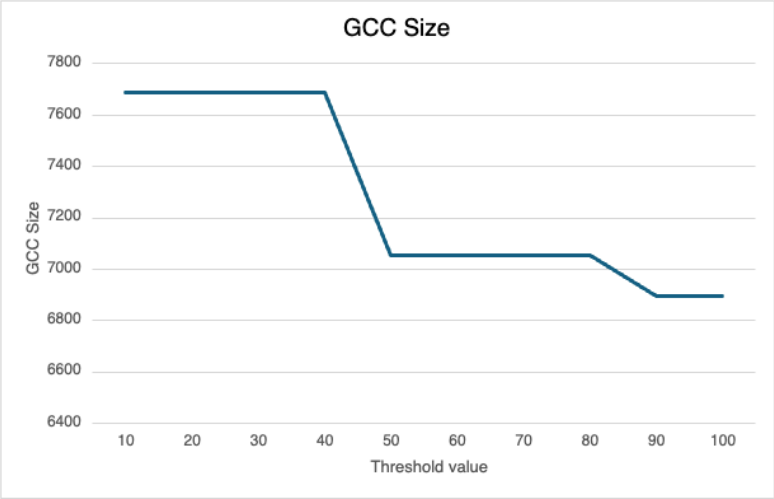
This algorithm employs the Noise-Corrected (NC) backbone extraction method, which isolates significant edges in the network based on their weights. It calculates the expected weight of each connection as if it were formed randomly and compares this to its actual weight, identifying edges that represent meaningful interactions.

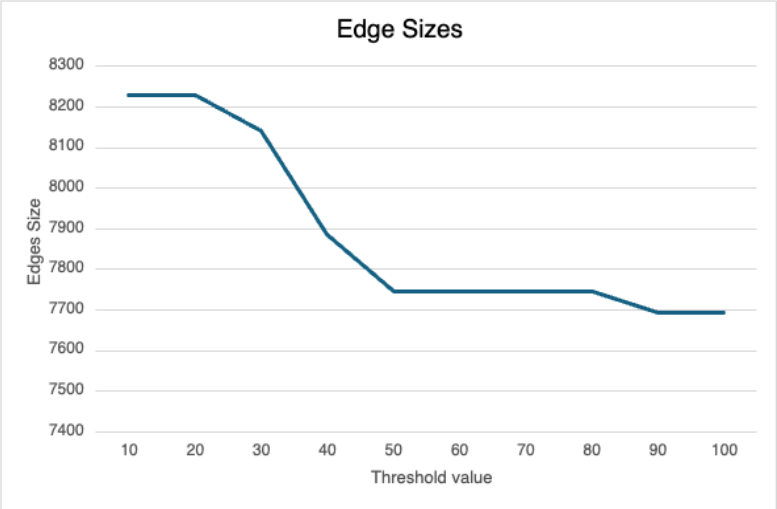
Furthermore, a significance threshold was applied to decide which edges to include in the reduced network. Only edges that met or exceeded this threshold of significance were kept, ensuring that the backbone of the network consisted only of the most important connections. [8]

In this research, various threshold values were tested to identify the one that best preserves the size of the Giant Weakly Connected Component (GWCC), keeping it as close as possible to the size of the original graph while eliminating the maximum number of edges.

The GWCC refers to the largest subgraph in which there is a path between any pair of nodes. Before applying the algorithm, the network was represented as an undirected graph. This decision was based on the understanding that whether a connection is from A to B or B to A, the mutual interests between users A and B are likely similar, thus making the direction of the connection irrelevant for our analysis. So, given that the graph is undirected; initially, the size of the GWCC in our network should equate to the total number of nodes in the graph, demonstrating that all nodes are interconnected.

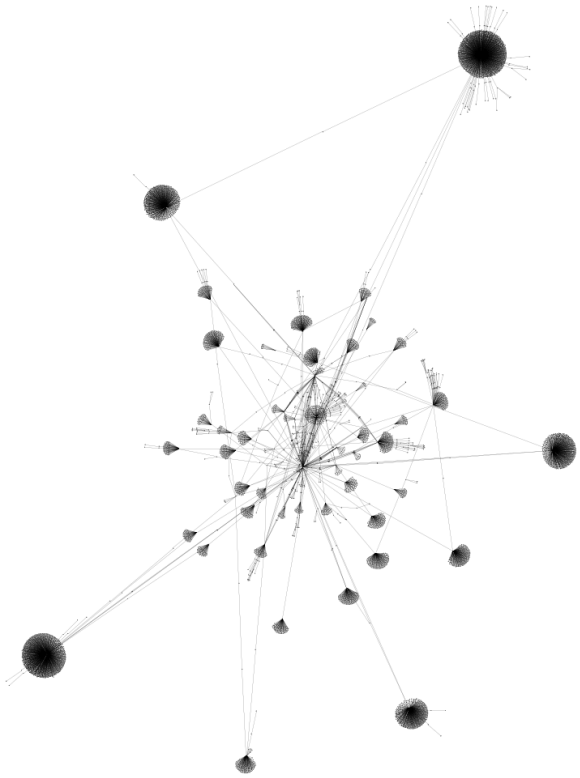
This approach aims to preserve the core structure of the graph while eliminating noise.





These graphs show that the ideal threshold value is 40, which allows us to maintain the graph's structure while eliminating a significant number of edges.

Despite the relatively sparse network of TikTok climate activists, employing network backboning has significantly improved our understanding and visualization of the user connections. It facilitated the analysis for detecting communities and examining their interactions, thereby providing insights into how climate activism is coordinated on the platform.



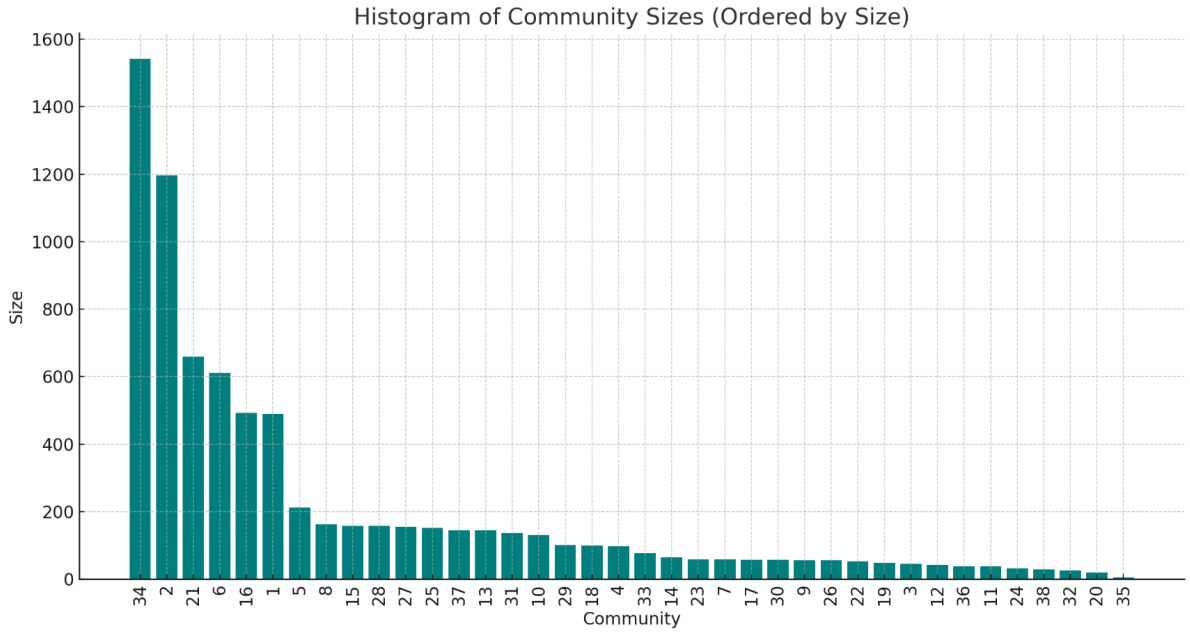
Community Detection with Louvain

Community detection in networks is crucial for understanding the structure and characteristics of various clusters within a large network. In the context of this study, the Louvain algorithm is used to identify communities within the TikTok network of users focused on climate activism.

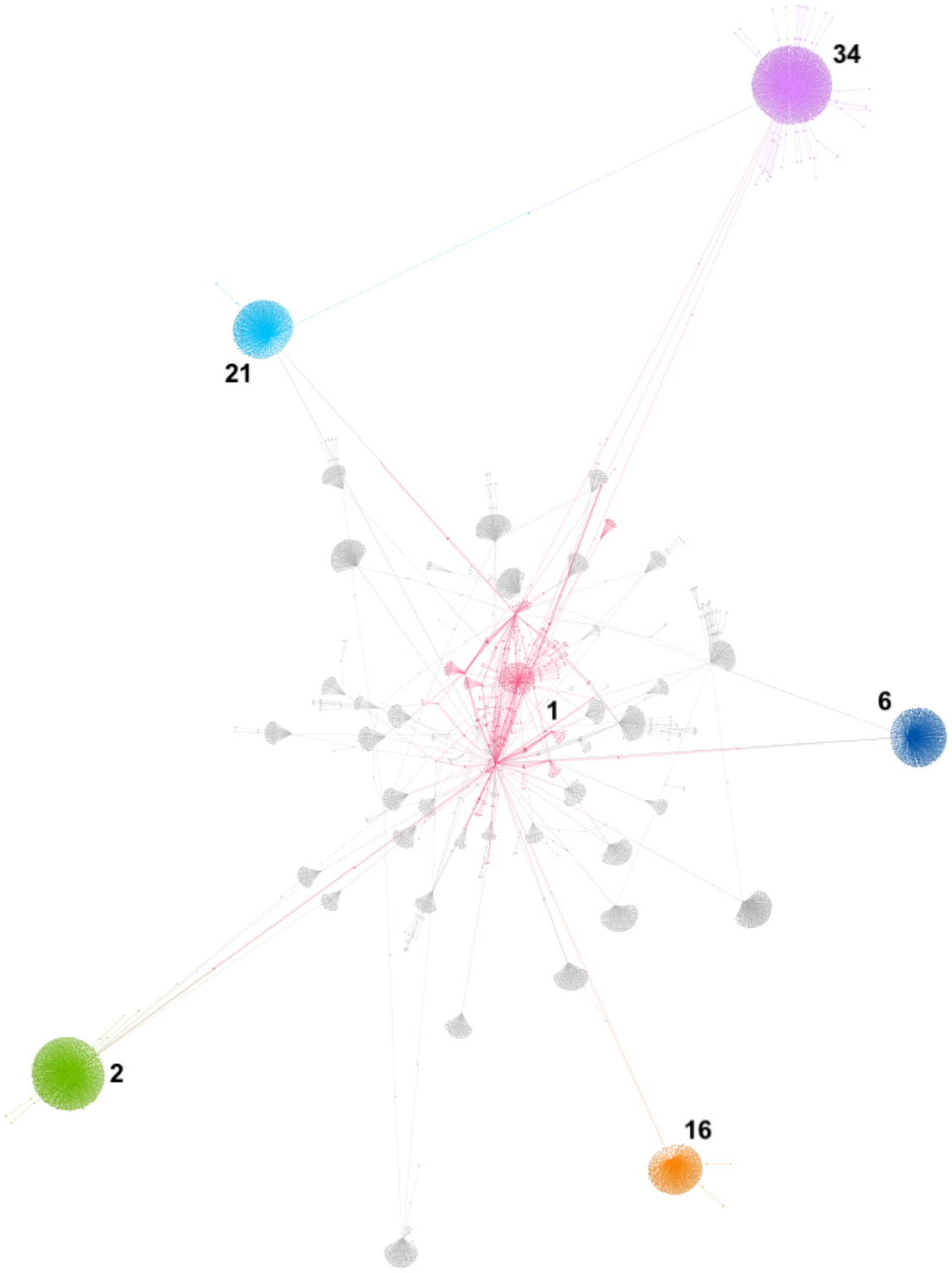
The Louvain algorithm is based on modularity optimization. Modularity is a scalar value that measures the density of links inside communities as compared to links between communities. The higher the modularity, the more defined the community structure within the network. [12]

Initially, each node is assigned to its own community. Then, in each iteration, nodes are moved to neighboring communities to evaluate modularity gains. This means that if moving a node to a neighboring community increases modularity, making that community more densely connected, the move becomes permanent. If not, the node remains in its original community. After all nodes have been evaluated, each community is combined into a single "meta-node," and the process is repeated until no further improvements can be made, resulting in the network being divided into optimal communities. [12]

The communities were detected within the previously established backboned network using Python's igraph library. The algorithm detected a total of 38 communities with a modularity of 0.9. We will concentrate our analysis on the first six communities by largest size.



Having identified the communities, we can analyze the hashtags used by each user within these groups to define their characteristics and determine the predominant climate topics discussed, exploring how different topics link communities that share similar interests.



Hashtag Analysis

Hashtag analysis is crucial for identifying the prevailing topics and trends among TikTok users involved in climate activism. This methodology section is dedicated to collecting and analyzing the hashtags from each user's most recent posts on the network.

For each user within the network, the latest content they posted in a 30-day timeframe was retrieved via the TikTok Query Videos API, including the hashtags used in their posts. To avoid duplicates within the same user, all hashtags were collected in a set for each user.

The first analysis was a frequency analysis to identify the most commonly used hashtags within each quartile. The goal is to explore correlations between the popularity of specific hashtags and the influence levels of users. This included analyzing whether higher-ranked individuals tend to use certain hashtags more frequently or whether some themes are universally popular across all influence levels.

This process involved collecting all the hashtags used by users within each quartile, counting the occurrences, and then ordering them by frequency. Furthermore, the data was categorized into two types: one based on Term Frequency (TF) and the other on Term Frequency-Inverse Document Frequency (TF-IDF). The first category focused solely on hashtag occurrences, while the second adjusted the importance of hashtags that appeared in two or more quartiles, prioritizing unique terms.

Term Frequency (TF) results ordered by the number of occurrences.

Q1	Q2	Q3	Q4
#fyp	#fyp	#fyp	#fyp
#eclipse	#grwm	#stitch	#earthday
#stitch	#ootd	#foryou	#climatechange
#humor	#foryou	#viral	#stitch
#greenscreen	#ad	#earthday	#foryou
#capcut	#fashion	#capcut	#viral
#ad	#makeup	#funny	#capcut
#grwm	#travel	#travel	#eclipse
#viral	#fypシ	#greenscreen	#climateaction

#parati	#foryoupage	#fypシ	#science

#grwm ("Get ready with me"), **#parati** (“for you” in Spanish), **#ootd** (“outfit of the day”)

We can notice how the *#fyp* (For You Page) remains consistently popular across all quarters, indicating its role as a fundamental hashtag for content visibility. Other hashtags such as *#stitch*, *#viral*, and *#capcut* appear multiple times, suggesting trending content types or popular editing tools.

Hashtags such as *#earthday* in Q3 and Q4, along with *#climatechange* and *#climateaction* in Q4, show a trend towards environmental themes among the network's more influential users. This suggests that topics related to climate activism are more popular within the network.

Furthermore, we can see how Q1, Q2, and Q3, which were all clustered together, frequently use hashtags related to fashion, humor, and lifestyle, while Q4 uses hashtags associated with science and climate. This highlights the distinction between users in the first three quartiles, who have similar low influence, and those in Q4, representing a set of users with significantly higher influence in the network.

Term Frequency-Inverse Document Frequency (TF-IDF) results ordered by the number of occurrences.

Q1	Q2	Q3	Q4
#aprendeentiktok	#humour	#trini_tiktoks	#plasticpollution
#cdmx	#díadelatierra	#trinidadandtobago	#weather
#publicidad	#styleinspo	#trinidad	#space
#anime	#pourtoi	#caribbean	#climatejustice
#recuerdos	#cultura	#caribbeantiktok	#biden
#comida	#students	#sgfoodie	#election
#gamerentiktok	#bananabread	#trini	#olympics
#cine	#styletips	#contentcreator	#ukpolitics
#dancer	#bedroommakeover	#trinitiktok	#quiz
#contentcreator	#olympics	#tobago	#sciencetok

#aprendeentiktok (Spanish for “learn in TikTok”), **#cdmx** (Short for Ciudad de México), **#publicidad** (Spanish for “advertising”), **#recuerdos** (Spanish for “memories”), **#comida** (Spanish for “food”), **#gamerentiktok** (Spanish for “gamer in TikTok”), **#cine** (Spanish for “cinema”), **#díadelatierra** (Spanish for “Earth Day”), **#pourtoi** (“for you” in French), **#cultura** (Spanish for “culture”)

Here, we notice that the majority of hashtags in Q1 are in Spanish, with several Spanish hashtags in Q2 and a focus on Trinidad and Tobago and the Caribbean in Q3. This indicates that a large proportion of users in the first three quartiles are from Latin America. In contrast, the hashtags in Q4 are all in English, with some specifically referencing topics like *#ukpolitics* and *#biden*, suggesting these users are from the UK and the US. This shows that content creators from Latin America trend less frequently than their American and British counterparts, as users in Q4 have significantly more influence within the network.

There's a dynamic range of topics, including entertainment (*#humor, #bananabread*), global events (*#dialadetierra, #olympics*), and political campaigns (*#biden, #election, #ukpolitics*). However, there is still a noticeable trend, just like in the TF analysis. Q1, Q2, and Q3 predominantly cover entertainment and lifestyle topics, whereas Q4 focuses on more serious themes such as climate activism (*#plasticpollution, #weather, #climatejustice*), science, and politics, showing that those topics trend more.

The second analysis was also a frequency analysis to identify the most commonly used hashtags within the communities in the network. The goal is to reveal patterns of collaboration and support among users, as well as understandings of interconnected topics of interest within these communities. This involved examining whether the communities mainly use hashtags directly related to climate activism or include topics that may be indirectly associated with climate activism. This process involved collecting all the hashtags used by the users in each community, counting the occurrences, and then ranking them by frequency. Additionally, the data was divided into two categories: TF and TF-IDF.

Term Frequency (TF) results ordered by the number of occurrences.

Comm. 34	Comm. 2	Comm. 21	Comm. 6	Comm. 16	Comm. 1
#fyp	#fyp	#fyp	#fyp	#fyp	#earthday
#stitch	#humor	#earthday	#trini_tiktoks	#grwm	#fyp
#greenscreen	#eclipse	#gaza	#trinidadandtobago	#ootd	#climateaction
#ad	#parati	#climatechange	#viral	#ad	#climatechange
#grwm	#mexico	#tiktok	#stitch	#makeup	#sustainability
#nyc	#cdmx	#travel	#trinidad	#foryou	#stitch
#ootd	#comedia	#foryoupage	#foryou	#fashion	#science

#foryou	#aprendeentiktok	#israel	#fypシ	#fashiontiktok	#climate
#vlog	#capcut	#nature	#caribbean	#duet	#learnontiktok
#capcut	#publicidad	#palestine	#ad	#fypシ	#environment

#grwm ("Get ready with me"), **#ootd** (“outfit of the day”), **#parati** (“for you” in Spanish), **#cdmx** (Short for Ciudad de México), **#aprendeentiktok** (Spanish for “learn in TikTok”), **#publicidad** (Spanish for “advertising”)

We can notice how the #fyp (For You Page) remains consistently popular across all communities, indicating again a general strategy to maximize visibility on the platform. Across different communities, there is a wide variety of hashtags encompassing topics like fashion and lifestyle, climate activism, and politics, as well as location-based tags such as #nyc, #mexico, and #trinidad.

Community 34 focuses on general entertainment, including popular TikTok tags like *#fyp, #stitch,* and *#greenscreen*, with a strong focus on fashion and lifestyle topics such as #ootd, #grwm, and #nyc, and also includes branded promotions (#ad).

Community 2 appears to be located in South America, as indicated by the Spanish hashtag, more precisely in Mexico (*#mexico, #cdmx*), with a focus on humor and entertainment (#humor, #comedia, #publicidad).

Community 21 focuses on more serious content addressing social issues (*#gaza, #israel, #palestine*) and climate activism (*#earthday, #climatechange, #nature*).

Community 6, likely based in Trinidad and Tobago, utilizes local tags (#trini_tiktoks, #trinidadandtobago, #trinidad) and focuses on entertainment-related content (#viral, #stitch), most likely about showcasing their beautiful country.

Community 16 is predominantly focused on fashion and lifestyle (*#ootd, #makeup, #fashion, #fashiontiktok*) and suggests branded promotions (*#ad*).

Community 1 is deeply engaged in environmental activism, using hashtags like *#climateaction, #climate, #enviroment, #sustainability, #climatechange*, along with educational tags (*#science, #learnontiktok, #earthday*).

Term Frequency-Inverse Document Frequency (TF-IDF) results ordered by the number of occurrences

Comm. 34	Comm. 2	Comm. 21	Comm. 6	Comm. 16	Comm. 1
#pointeshoes	#recuerdos	#idulfitri1445h	#trini_tiktoks	#parisianstyle	#inflationreductionact
#eastersunday	#dúo	#mudikceriapenuhmakna	#trinidadandtobago	#risotto	#defendourjurys

#contemporarydance	#tiktokmehizover	#zakat	#caribbeantiktok	#tunisie	#justicefortrudi
#thriftwithme	#videojuegos	#perú	#trini	#youngmom	#notguilty
#workoutroutine	#diadelniño	#DiIndonesiaAja	#trinitiktok	#seamoss	#earthdefenders
#tubipartner	#chisme	#pesonaindonesia	#tobago	#steviewonder	#peacefulprotesters
#dancerlife	#vacaciones	#wonderfulindonesia	#dancehall	#singingchallenge	#wearealltrudi
#danceclass	#japon	#harikartini2024	#pcosweightloss	#parisjetaime	#climatememes
#balletclass	#peliculas	#banggabuatanindonesia	#trinidad🇹🇹	#neoncarnival	#ourocean2024
#nycliving	#mascotas	#banggaberwisatadiindonesia	#jamaicatiktok	#focaccia	#ecowarrior

#recuerdos (Spanish for “memories”), ***#dúo*** (Spanish for “duet”), ***#tiktokmehizover*** (Spanish for “TikTok made me see”), ***#videojuegos*** (Spanish for “videogames”), ***#diadelniño*** (Spanish for “day of the child”), ***#chisme*** (Spanish for “gossip”), ***#vacaciones*** (Spanish for “vacations”), ***#japon*** (Spanish for Japan), ***#peliculas*** (Spanish for “movies”), ***#mascota*** (Spanish for “mascot”), ***#idulfitri1445h*** (Idul Fitri, also known as Eid al-Fitr, marks the end of Ramadan, the year "1445 H" refers to the Islamic year, which corresponds to 2023-2024 in the Gregorian calendar.), ***#mudikceriapenuhmakna*** (Indonesian for “Joyful and Meaningful Homecoming”), ***#zakat*** (is one of the Five Pillars of Islam, it is a religious tax, typically 2.5% of a Muslim's total savings and wealth over a year.), ***#DiIndonesiaAja*** (Indonesian for "Just in Indonesia"), ***#pesonaindonesia*** (Indonesian for “Wonderful Indonesia”), ***#harikartini2024*** (Hari Kartini, or Kartini Day, celebrated on April 21st each year in Indonesia, is a special day that commemorates the birth of Raden Adjeng Kartini, a pioneer in the area of women's rights and education in Indonesia.), ***#banggabuatanindonesia*** (Indonesian for “Proudly Made in Indonesia”), ***#banggaberwisatadiindonesia*** (Indonesian for “Proud to Travel in Indonesia”), ***#risotto*** (Italian food), ***#tunisie*** (French for “Tunisia”), ***#parisjetaime*** (French for “Paris I love you”), ***#focaccia*** (Italian food), ***#dancehall*** (a genre of Jamaican popular music).

Here, we can notice how each community has distinct themes and interests, which are likely influenced by geographic, cultural, or specific activity-based preferences.

Community 34 suggests a focus on dance and related activities featuring hashtags such as #balletclass, #contemporarydance, and #pointeshoes, alongside lifestyle tags (#thriftwithme, #workoutroutine), probably geolocated in the USA (**#***nycliving*).

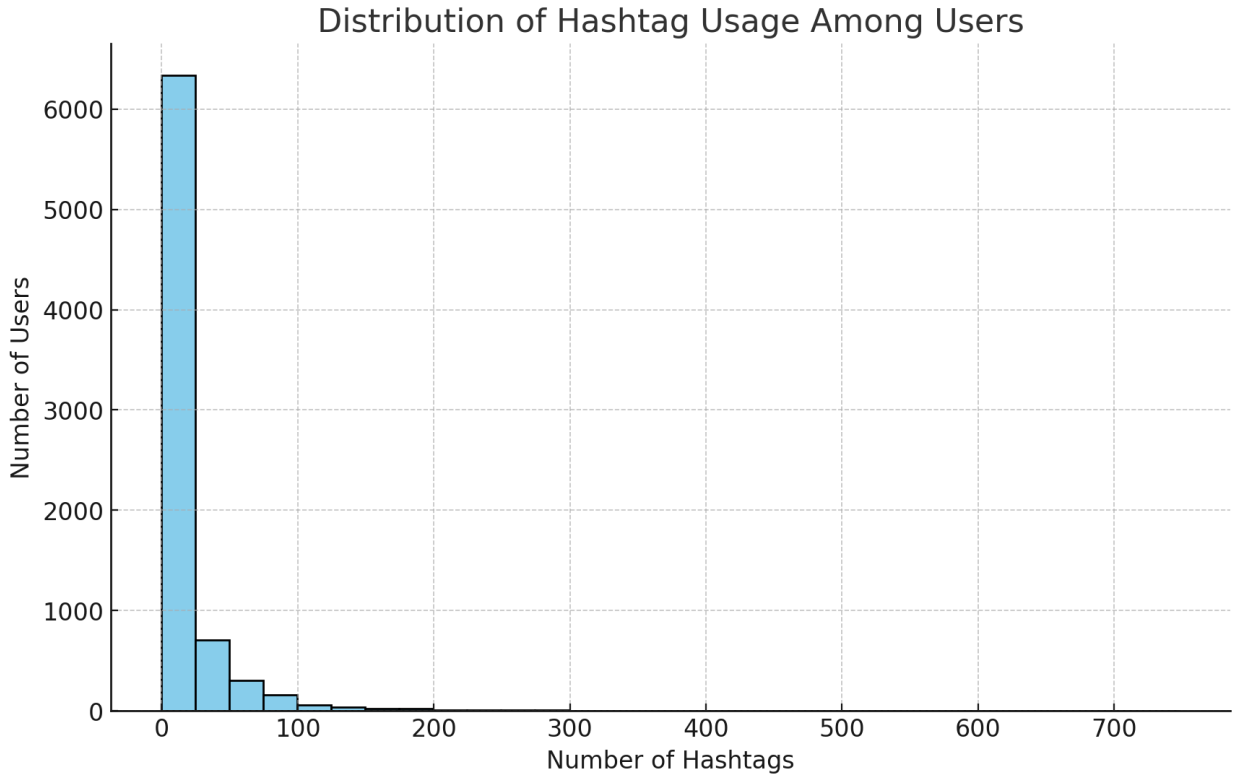
Community 2 still hints at being likely based in a Spanish-speaking country, with its content centered on entertainment, with tags such as #dúo, #videojuegos, #peliculas, #mascotas, and #recuerdos

Community 21 appears to be located in Indonesia, focusing mainly on cultural and religious aspects of Islamic and Indonesian traditions.

Community 6 is mainly located in the Caribbean, particularly Trinidad and Tobago and Jamaica, showcasing the beauty of their countries and cultures.

Community 16 has a strong focus on fashion and lifestyle, with indications of being based in Europe, as suggested by French (#parisianstyle, #parisjetaime) and Italian (#focaccia, #risotto) tags.

Community 1 suggests a focus on political activism and social justice (#inflationreductionact, #defendourjurys, #justicefortrudi, #peacefulprotesters) alongside climates and environmental activism with tags such as #earthdefenders, #climatechange, #ourocean2024, and #ecowarrior.



Conclusions

By identifying the most influential topics and how they vary among different user groups, the study offers insights into how climate activism resonates across the platform. This analysis is crucial for anyone aiming to improve the effectiveness of digital campaigns by aligning their content with the topics that have the highest engagement and influence within the TikTok community. Especially those focused on raising awareness and sparking action on environmental matters.

The research began with a snowball sampling from a selected group of users known for their engagement in climate activism on TikTok. This method effectively expanded the network, establishing a core group centered around climate activism before branching out to include other, more diverse topics.

The research findings suggest that users who are focused on climate activism also tend to be involved in other important social and political issues, like politics, wars, and elections, showing there's a strong connection between different types of activism.

Moreover, we notice that most of the people interested in climate activism are from Europe and the USA, while other communities focused on topics such as fashion, lifestyle, and entertainment are from Asia and South America. Even though they might not be directly involved in climate activism, they could still be engaged in it.

For instance, Community 6 in the Caribbean, showcasing their country's nature, indicates a possible interest in climate activism due to its relevance to their surroundings. Furthermore, communities centered around fashion and lifestyle, such as Communities 34 and 16, could contribute by promoting environmentally friendly brands or developing eco-friendly fashion concepts.

This research also led to side findings such as the common strategy among all users, regardless of their influence or community affiliation, in using hashtags such as *#fyp, #foryoupage, #foryou, and #fypシ*, typically used to maximize visibility on TikTok's algorithmically personalized content feed, the "For You Page." [4]

In summary, this project not only explains the structure around climate activism within the TikTok network, highlighting the diverse yet interconnected topics shared by various groups but also shows how users strategically use popular hashtags to engage with a broader audience.

References

1. HerdaĞdelen, A., Zuo, W., Gard-Murray, A., & Bar-Yam, Y. (2013). An exploration of social identity: The geography and politics of news-sharing communities in twitter. Complexity, 19(2), 10-20.
   An exploration of social identity: The geography and politics of news-sharing communities in twitter

2. Saura, J. R., Herráez, B. R., & Reyes-Menendez, A. (2019). Comparing a traditional approach for financial Brand Communication Analysis with a Big Data Analytics technique. IEEE access, 7, 37100-37108.
   Comparing a traditional approach for financial Brand Communication Analysis with a Big Data Analytics technique

3. Pera A., De Francisci Morales G., and Aiello L. M.. (2023). Measuring Behavior Change with Observational Studies: a Review. 1-34.
   https://arxiv.org/abs/2310.19951

4. Schellewald, A. (2023). Understanding the popularity and affordances of TikTok through user experiences. Sage Journals, 1568-1582.
   Understanding the popularity and affordances of TikTok through user experiences

5. Ghoshal, G., Mangioni, G., Menezes, R., & Poncela-Casanovas, J. (2014). Social system as complex networks. Social Network Analysis and Mining, 4, 1-2.
   Social system as complex networks

6. Malik, H. A. M. (2022). Complex network formation and analysis of online social media systems. Cmes-Comr Model Engg & Sci, 130(3), 1737-1750.
   Complex network formation and analysis of online social media systems

7. Zhang, B., Wang, Y., Jin, Q., & Ma, J. (2015). A pagerank-inspired heuristic scheme for influence maximization in social networks. International Journal of Web Services Research (IJWSR), 12(4), 48-62.
   A pagerank-inspired heuristic scheme for influence maximization in social networks

8. Coscia, M. (2021). The atlas for the aspiring network scientist. arXiv preprint arXiv:2101.00863.
   The atlas for the aspiring network scientist

9. Wang, P., Bo, X., Yang, H., Sun, S., & Li, S. (2015, October). An improved PageRank algorithm for Social Network User's Influence research. In 2015 3rd International Conference on Mechatronics and Industrial Informatics (ICMII 2015) (pp. 542-547). Atlantis Press.
An improved PageRank algorithm for Social Network User's Influence research

10. Sanford, M., Witkowska, M., Gifford, R., & Formanowicz, M. (2023). Emotional framing in online environmental activism: Pairing a Twitter study with an offline experiment. Frontiers in Psychology, 13, 1099331.
Emotional framing in online environmental activism: Pairing a Twitter study with an offline experiment

11. San Cornelio, G., Ardèvol, E., & Martorell, S. (2021). ENVIRONMENTAL INFLUENCERS ON INSTAGRAM: CONNECTIONS AND FRICTIONS BETWEEN ACTIVISM, LIFESTYLES AND CONSUMPTION. AoIR Selected Papers of Internet Research.
ENVIRONMENTAL INFLUENCERS ON INSTAGRAM: CONNECTIONS AND FRICTIONS BETWEEN ACTIVISM, LIFESTYLES AND CONSUMPTION.

12. Meena, S. S. (2023). Community detection using Louvain. Medium.
COMMUNITY DETECTION USING LOUVAIN | by Shyam Sundar Meena | Medium

13. Mahfouz, I. (2020). The Linguistic Characteristics and Functions of Hashtags:# Is it a New Language?. Arab World English Journal (AWEJ) Special Issue on CALL, (6).
The Linguistic Characteristics and Functions of Hashtags:# Is it a New Language?

14. Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. New media & society, 13(5), 788-806.
Ambient affiliation: A linguistic perspective on Twitter

15. Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. Social semiotics, 25(3), 274-291.
Searchable talk: The linguistic functions of hashtags

16. Caleffi, P. M. (2015). The 'hashtag': A new word or a new rule?. SKASE journal of theoretical linguistics, 12(2). The 'hashtag': A new word or a new rule?

17. Wikipedia contributors. Hashtag. In Wikipedia. Hashtag - Wikipedia

18. Coscia, M., & Neffke, F. M. (2017, April). Network backboning with noisy data. In *2017 IEEE 33rd international conference on data engineering (ICDE)* (pp. 425-436). IEEE. Network backboning with noisy data