Dynamics of Opinion Formation and Influence in Networked Communities: How Reddit Users Shape Urban Air Pollution Discussion

Yuhou Chen, Department of Sociology

McGill University, Montreal

April 2025

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Arts in Sociology

TABLE OF CONTENTS

I. **Introduction** 8

II. **Literature Review** 10

2.1 Public Understanding of Science and Social Media Knowledge Translation 10
2.2 Opinion Formation and Social Influence 11
2.3 Influence in Networked Communities 14
2.4 Influence Flows with Time 16
2.5 Combining Topic Modeling and Social Network Analysis 18

III. **Analytical Strategy** 20

IV. **Methods** 21

4.1 Data Selection 22
4.2 Text Processing 23
4.3 Network Building 25

V. **Data Analysis and Results** 25

5.1 Static Topic Modeling 26
5.2 Dynamic Topic Modeling 29
5.3 Social Network Analysis 33
5.4 Looking into One Local Network Case 41

VI. **Discussion** 44

6.1 Influence Flows with Time and Engagement 44
6.2 The Temporal Structure of Opinion Formation 46
6.3 Visibility within Reddit 48
6.4 Limitations and Future Research 49

ABSTRACT

Social media has become an important platform for knowledge translation and public opinion formation on urban air pollution. Social influence on social media is a critical and intriguing research agenda within networked communities. Reddit, with its semi-anonymous, community-driven structure, allows influence to emerge organically through user interactions, offering a decentralized network to examine opinion formation beyond traditional models. This study examines how influence and public opinion are shaped on Reddit through discussions about urban air pollution. It demonstrates how different types of users engage in topic evolution and how influence dynamics unfold over time within the network structure. Using a mixed-method approach, this research integrates dynamic topic modeling (BERTopic) with social network analysis (SNA) to analyze over 26,000 Reddit comments from 2020 to 2023 related to urban air pollution. The topic modeling identified 19 topics within four themes and traced their evolution over time. Seven of the most fluctuating topics were selected for SNA to map the structure of user interactions and influence flows. Users are categorized as Early Adopters, Early Majority (based on topic contributions), and Hubs (based on network positions). The findings reveal that influence flows with time and engagement rather than static authority, where occupying a central network position does not necessarily equate to influence. Hubs act as amplifiers rather than opinion leaders, strategically selected by Early Adopters to gain visibility and amplify emerging topics. The temporal structure of opinion formation is shaped at the early stage by Early Adopters, while the Early Majority fills the existing network by echoing established opinions. These findings contribute to sociological theories of social influence, knowledge translation, and public opinion formation, as well as the integration of topic modeling and SNA in decentralized social media.

Keywords: Social Influence; Opinion Formation; Urban Air Pollution; Social Network Analysis; Topic Modeling; Networked Communities

ABRÉGÉ

Les réseaux sociaux sont devenus une plateforme importante pour la traduction des connaissances et la formation de l'opinion publique sur la pollution atmosphérique urbaine. L'influence sociale sur les réseaux sociaux constitue un axe de recherche essentiel et intrigant au sein des communautés en réseau. Reddit, avec sa structure semi-anonyme et orientée par la communauté, permet à l'influence d'émerger de manière organique par le biais des interactions entre utilisateurs, offrant ainsi un réseau décentralisé permettant d'examiner la formation de l'opinion au-delà des modèles traditionnels. Cette étude examine comment l'influence et l'opinion publique sont façonnées sur Reddit à travers les discussions sur la pollution atmosphérique urbaine. Elle montre comment différents types d'utilisateurs participent à l'évolution des sujets et comment les dynamiques d'influence se déploient au fil du temps au sein de la structure du réseau. En utilisant une approche mixte, cette recherche intègre la modélisation dynamique de sujets (BERTopic) et l'analyse des réseaux sociaux pour analyser plus de 26 000 commentaires sur Reddit de 2020 à 2023 concernant la pollution atmosphérique urbaine. La modélisation de sujets a identifié 19 sujets répartis en quatre thèmes et a suivi leur évolution au fil du temps. Sept des sujets les plus fluctuants ont été sélectionnés pour l'analyse des réseaux sociaux afin de cartographier la structure des interactions entre utilisateurs et les flux d'influence. Les utilisateurs sont catégorisés en tant qu'Early Adopters, Early Majority (basé sur la contribution aux sujets) et Hubs (basé sur les positions dans le réseau). Les résultats révèlent que l'influence se déploie au fil du temps et de l'engagement plutôt que par une autorité statique, où occuper une position centrale dans le réseau ne signifie pas nécessairement avoir de l'influence. Les Hubs agissent comme des amplificateurs plutôt que comme des leaders d'opinion, étant stratégiquement sélectionnés par les Early Adopters pour accroître leur visibilité et amplifier les nouveaux sujets. La structure temporelle de la formation de l'opinion est façonnée au stade précoce par les Early Adopters, tandis que les Early Majority remplissent le réseau existant en reprenant les opinions établies. Ces résultats contribuent aux théories sociologiques de l'influence sociale, de la traduction des connaissances et de la formation de l'opinion publique, ainsi qu'à l'intégration de la modélisation de sujets et de l'analyse des réseaux sociaux dans les médias sociaux décentralisés.

ACKNOWLEDGMENTS

Completing this thesis marks the end of a two-year journey at McGill. Nestling on the slope of Mont-Royal has been not only an academic experience but also a sensory one, reminding me to breathe, to pause, and to keep going. The changing seasons—blooming spring, vivid summer, colorful autumn, and quiet winter—have been filled with learning, reflection, and steady effort.

I would first like to express my sincere gratitude to my supervisor, Professor Peter McMahan, for all his guidance and support throughout these two years, especially during the writing of this thesis. Your clarity, patience, and encouragement have been invaluable, not only for this master but also for my life journey. Also, many thanks to my committee member, Professor Jason Carmichael, who helped me shape the direction of this research.

To my parents, thank you for your powerful love and constant belief in me. Your support has been a source of stability, even across distance, and I carry it with me in all that I do. To my friends, thank you for holding me up and carrying me through every moment, making this journey lighter and even more meaningful. To myself, thank you for moving forward one step at a time, for every choice and decision made after a struggle.

So here I stand, with grateful heart, a chapter closed, a brand-new start. This thesis is more than words on pages—A tale of growth, of joy, of age.

LIST OF FIGURES

Figure 1. [Dynamic Topic Modeling Result: Topic Trend Over Time] 30
Figure 2. [Sampling Time Segment] 31
Figure 3. [Global Network Plots: Positions and Community Detection] 33
Figure 4. [Example of Early Adopters and Early Majority in the Network] 35
Figure 5. [Heatmap of Hubs' Comments: Topic × Subreddit] 38
Figure 6. [Special Nodes' Position and Neighbors] 40
Figure 7. [The Formation and Structure of Topic 5 Local Network] 42
Figure 8. [Heatmap of All Comments Topic ×Subreddit] 57
Figure 9. [Hubs' Comments Distribution] 58

LIST OF TABLES

Table 1. [Keywords and Subreddit for Data Collection] 23
Table 2. [Descriptive Results of Static Topic Modeling] 26
Table 3. [Representative Topics for SNA] 31
Table 4. [Average Topic Contribution of Hub, Early Adopters, and Early Majority] 32
Table 5. [Logistic Regression Results] 34
Table 6. [Average Minimal Path Length between Hub, Early Adopters, and Early Majority] 34
Table 7. [Special Nodes' Interactions]..40

I. Introduction

Air pollution is one of the most pressing environmental issues that has an impact on human health and the sustainability of cities. Epidemiological studies show adverse effects of pollutant particles on physical and mental health, with 90% of urban residents exposed to PM2.5 levels exceeding WHO guidelines (Pai et al., 2022) and 3 to 9 million premature deaths caused by air pollution each year globally (Lelieveld et al., 2019). Citizen's scientific knowledge and environmental awareness are essential to effectively assess and communicate the facts and impact of air pollution (Rickenbacker et al., 2019). Understanding how the public processes and conveys scientific information about air pollution has become a vital tool for mitigating damage (Nisbet & Scheufele, 2009). In the digital age, social media has changed the way of information dissemination and consumption and has strongly influenced the forms of knowledge and public opinion (Williams et al., 2015).

Unlike traditional media, which is presumed to operate within a top-down model, social media allows for a networked form of knowledge framing and dissemination among diverse audiences (Comfort & Park, 2018). In the social studies of science and knowledge, this model of participation resonates with work on *knowledge translation*, which suggests that scientific knowledge neither diffuses from experts to the public nor passes directly into social narratives, but is through a process of reinterpretation and adaptation (Latour, 1987; Star & Griesemer, 1989). In the case of air pollution, public discussions on social media serve as a form of boundary work, bridging the gap between expert knowledge and lay interpretations (Stilgoe et al., 2014). Digital platforms are characterized by decentralization and interactivity, promoting public understanding of science, which in turn builds societal attitudes and collective opinions toward air pollution (Karell et al., 2023).

Reddit, a widely used social media platform in North America, provides a useful platform for researching networked influence and opinion formation. Unlike platforms dominated by public figures, Reddit's semi-anonymous and community-driven structure fosters an environment where influence grows organically through engagement and user interactions rather than existing authority and followers (Jang et al., 2024). This decentralized model of influence allows how users shape public opinion of urban air pollution. The influence of users on public discourse is earned not only through their structural position in the network but also from their content and engagement patterns (Huffaker, 2010; Kozitsin, 2023).

This study explores the influence, opinion formation, and networked communities within Reddit discussions on urban air pollution from a micro-dynamic perspective. Rather than assuming that highly connected users are always influential, I evaluate influence as a context-dependent and time-sensitive phenomenon, shaped by structural positioning and engagement over time. To achieve these aims, this research employs dynamic topic modeling and social network analysis to analyze Reddit opinion evolution on urban air pollution and the network positions of different types of users. The study addresses the formation of public discourse on environmental issues in semi-anonymous social media and answers three main research questions: 1) How do the main topics of discussion related to urban air pollution on Reddit emerge and evolve over time? 2) Are there regular discursive or structural user roles associated with the emergence of topics? 3) If so, how do these roles influence community-wide dynamics of topic prevalence?

II. Literature Review

2.1 Public Understanding of Science and Social Media Knowledge Translation

Public understanding of science and knowledge translation provides a framework for analyzing how individuals perceive and engage with scientific knowledge. Knowledge is a key driver of public opinion, acting as a bridge between the scientific community, policymakers, and the general public (Stilgoe et al., 2014). However, scientific knowledge does not "speak for itself" – it must be actively communicated, contextualized, and given meaning within social and political discourses (Stilgoe et al., 2014). Scientific knowledge translation is not a linear process of disseminating information, but a socially embedded and interpretative process (Liyanage et al., 2009). It is influenced by cognitive capacities, societal contexts, values, and the availability of diverse information sources (Bandura, 2008). Environmental issues are intersections of scientific data, political discourse, media representation, and personal experience. Uncertainty surrounding air pollution levels, health consequences, and regulatory responses contributes to diverse perceptions of the issue, often leading to conflicting knowledge understanding and social narratives (Rickenbacker et al., 2019). These challenges emphasize that science communication is not merely a process of "informing" but also producing social meanings, as individuals interpret and integrate information within their own epistemic frameworks (Nisbet & Scheufele, 2009). To understand this process, theories of knowledge translation emphasize that before knowledge is accepted by the public, it must be reinterpreted, adapted, and mediated by various actors (Callon, 1984; Latour, 1987). This translation involves mediators who transform complex scientific concepts into forms that resonate with public experiences (Wæraas & Nielsen, 2016).

Social media emerges as a crucial platform for knowledge translation and science communication (Liang et al., 2014; O'Connor, 2021). Unlike traditional media, which largely relies on one-way communication from experts to the public, social media supports dynamic, interactive engagement where users can discuss (Comfort & Park, 2018). During this process, people reinterpret scientific information and exchange opinions, and then an echo chamber and polarization may occur, thereby contributing to the formation and structure of public narratives (Jacobson et al., 2016). This interactive model aligns with knowledge translation as a collaborative, social process. The concept of *boundary objects* further explains how user's content on social media acts as a bridge between expert knowledge and public discourse. Boundary objects are artifacts or ideas that inhabit multiple social worlds and facilitate communication across group boundaries (Star & Griesemer, 1989). These boundary objects allow diverse audiences to engage with scientific knowledge from different perspectives, shaping how public narratives emerge and evolve (Star & Griesemer, 1989). Urban air pollution, as an issue involving science, policy, and public health, is worth studying from the perspective of knowledge translation mechanisms to understand how the public understands and frames this problem.

2.2 Opinion Formation and Social Influence

The process of opinion formation is a complex and multifaceted phenomenon, representing an interplay of individual cognition, interpersonal interactions, and broader social structure (Gerard & Orive, 1987; Mueller & Tan, 2018). It is not simply a matter of individuals adopting or rejecting information; rather, opinions are formed, reinforced, or contested through continuous exposure to ideas, social comparison, and group dynamics (Gerard & Orive, 1987).

Research on opinion diffusion suggests that ideas spread through both direct social interactions and indirect exposure to dominant narratives, shaping perceptions and behavioral responses (Christakis & Fowler, 2013). Building on this understanding, the classical two-step flow model (Katz et al., 1955) proposed that mass media influences "opinion leaders," who then disseminate information to broader audiences. This model was later expanded to multi-step diffusion frameworks, incorporating network dynamics, weak ties, and structural influence (Granovetter, 1973).

Social influence theories further elaborate on this process, emphasizing that opinion formation is not static but an evolving interactional phenomenon, continuously shaped by social validation and strategic adaptation (Kozitsin, 2023; Mueller & Tan, 2018; Rohde et al., 2023; Watts & Dodds, 2007). Festinger's (1954) concept of social comparison, highlights how individuals adjust their beliefs and attitudes based on group norms and perceived similarities. Deutsch and Gerard (1955) first distinguished between informational influence and normative influence, the former where individuals conform to group opinions because they perceive them as accurate, the latter where conformity arises from the desire for social acceptance. Kelman (1958) developed social influence theory and divided social influence into three categories: compliance, identification, and internalization. These theories demonstrate that influence is a multi-layered and dynamic process, shaped by both individual motivations and social contexts.

Building on these, research has focused on the networked environments where opinion formation is highly dynamic, where the patterns of connection and interaction determine how information flows and what topics dominate (e.g. Kitsak et al., 2010; Kozitsin, 2023; Rohde et al., 2023; Watts & Dodds, 2007). Opinion formation and diffusion have been studied in networked environments, where interaction explain who gets exposed to information and how ideas spread (Glass & Glass, 2021; Katona et al., 2011; Z. Yang et al., 2023). Social Network Analysis (SNA) provides a methodological framework for examining the relational structures that facilitate or constrain opinion exchange, conceptualizing social influence as a function of network ties, information pathways, and structural positioning (Liang et al., 2014). According to this perspective, a social network consists of nodes (individuals, groups, organizations) and edges (connections between them), with influence emerging not solely from individual attributes but from patterns of connectivity, centrality, and clustering (Mueller & Tan, 2018). Gerard and Orive (1987) elaborated on opinion formation by emphasizing its dynamic and reciprocal nature. They argued that individuals not only receive influence from their social environment but also actively contribute to shaping the opinions of others. This conceptualization moves beyond classical media influence models by emphasizing that opinion formation is not passive reception but a dynamic and reciprocal process of co-construction (Boyd, 2010). Marsden and Friedkin (1993) further emphasize the role of relational structures within networks in facilitating information exchange and shaping opinions. Furthermore, research has challenged the rigid classification of opinion leaders, arguing that influence should be understood as a fluid phenomenon rather than a static role assigned to a specific class of individuals (Watts & Dodds, 2007). This aligns with the approach of this study, which examines influence as a dynamic and time-sensitive process shaped by the structural positions and interaction behaviors of different types of users, rather than a permanent characteristic of specific users. Thus, by integrating classical social influence theories with contemporary network perspectives, this study seeks to explore opinion formation in decentralized network communities.

2.3 Influence in Networked Communities

Social influence represents a critical mechanism through which attitudes, opinions, and behaviors emerge and spread (Cialdini & Goldstein, 2004). It explains why individuals align with prevailing norms, adopt certain ideas, and modify their stances in response to group dynamics (Brass, 2022). In networked spaces, influence is often understood through patterns of interaction, communication, and structural positioning, where individuals are more or less influential depending on their embeddedness within the network (Marsden & Friedkin, 1993). Network-based models often assume that resources like information flow faster through shorter paths and that central actors are more likely to adopt or quickly receive information and transmit it to others (Marsden & Friedkin, 1993). Rather than being a static trait of particular individuals, social influence is better conceptualized as a dynamic process, fluctuating based on a user's engagement, relational ties, and temporal position in discourse.

Social influence manifests in digital networks through processes such as information cascades, opinion shifts, and discourse amplification (Christakis & Fowler, 2013). Users influence each other not only through direct interactions but also through indirect exposure to dominant narratives, shaping their perceptions and behaviors over time (Gerard & Orive, 1987). In this regard, influence can emerge through both central and peripheral actors, as users on such platforms accumulate social capital and influence through structural ties (Katona et al., 2011). Research on social contagion and collective attention highlights that influence often occurs through networked exposure rather than intentional persuasion (Granovetter, 1973). Thus, influence is not solely tied to fixed social status or network centrality but is mediated by patterns of interaction, timing, and engagement strategies.

Unlike traditional influencers who rely on established authority, social media is a web-based service in which users can establish public or semi-public profiles and connect with other relevant people to form a list of relationships (Watts & Dodds, 2007), including following, boosting, liking, replying, etc. However, studies often assume influence as a function of network centrality rather than as an evolving property that shifts over time. They treat “influencers” as a predefined category, assuming that certain users inherently possess a greater ability to shape discussions, analyze their followers, friendship (Milani et al., 2020; Zou et al., 2021), and interaction (Ganley & Lampe, 2009; Stovel & Shaw, 2012; Yang et al., 2023). These approaches have been particularly dominant in studies of Twitter and Facebook, where influence is visible through followers and algorithmic amplification. However, Reddit’s voting system and subreddit-based communities allow influence to emerge more organically. This creates a decentralized system where any user, regardless of prior engagement or social capital, can influence discourse if their contributions resonate within specific conversational contexts.

This study distinguishes between influence as content formation versus influence as the structural center in information diffusion. The influence of opinion formation may not correlate directly with network centrality, which means these two forms of influence may not necessarily align (Gulati, 1995). In this case, users who shape discourse content may not be those who occupy central structural positions. Instead, central nodes serve as structurally important points in a network, capable of accelerating information spread without necessarily contributing original insights or arguments (Kitsak et al., 2010). For example, an account widely tagged in debates may function as an amplifier rather than an originator of ideas (Bruns & Burgess, 2020), which means their presence in the network structure is crucial for the flow of information rather than its content (González-Bailón et al., 2013). Studies have shown that holding a central position does not necessarily equate to high credibility or persuasive power (Kitsak et al., 2010). Therefore, this study introduces the concept of In-Degree Hub as a distinct category of users who hold central position within the network. By distinguishing In-Degree Hub from other forms of influence, this study aims to explore the various roles that different types of users play in shaping discussions on urban air pollution.

2.4 Influence Flows with Time

Building on the distinction between influence in opinion formation and influence as structural importance, this research focuses on how influence flows over time. Rather than categorizing users as "influencers" or "non-influencers," influence can be understood as an emergent and fluctuating property within networked interactions (Huffaker, 2010; Kozitsin, 2023). This perspective challenges the assumption that influence is a static trait of individuals and instead positions it as a fluid characteristic that varies depending on contexts. Users do not hold fixed levels of influence; instead, their ability to shape discourse may vary based on their engagement at specific moments, the topics they engage with, and the evolving structure of the network over time. Reddit's emphasis on user interaction rather than user identity, where users are not explicitly marked by follower counts or institutional legitimacy, can still gain prominence through participation (Jang et al., 2024; Parsa et al., 2022).

To observe influence in a time dimension, this study introduces the concepts of Early Adopters and Early Majority. This study does not assume that Early Adopters and Early Majority are inherently "influencers." Rather, it conceptualizes influence as a time-sensitive and network-dependent process, where users gain prominence based on when and how they engage in discussions. Early Adopters are individuals who engage with new topics before they gain mainstream attention, initiating discussions and introducing key arguments (Rogers et al., 2008; Sziklai & Lengyel, 2022). This concept originates from Everett Rogers' Diffusion of Innovations Theory (1962), which categorizes individuals in society into five groups based on their adoption timing. Early adopters tend to be well-connected within specific subgroups but are not necessarily central within the broader network (Valente, 2010; Sziklai & Lengyel, 2022). These users contribute significantly to shaping the initial framing of a topic, acting as the first wave of participants before broader adoption occurs (Kadushin, 2012). Their importance lies in their ability to bring attention to emerging topics rather than their overall network centrality. Early Majority engages when a topic is already gaining traction. They are less likely to take risks but play a crucial role in accelerating diffusion (Lynn et al., 2017). Empirical research suggests that the Early majority tend to adopt innovations through a mix of social influence and personal evaluation, making them an essential bridge between niche adoption and mainstream acceptance (Rong & Mei, 2013; Seebauer, 2015). By distinguishing these two roles by opinion formation time, this study explores how influence changes with time within the network.

2.5 Combining Topic Modeling and Social Network Analysis

Several studies have combined topic modeling and SNA in examining public discourse on social media. Topic modeling has proven to be a valuable method for opinion measurement in social media (Sobkowicz et al., 2012), which can extract underlying topics within large-scale data and examine how specific narratives are communicated and collective opinion is formed (Huang, 2019). Jang et al., (2024) applied static topic modeling and dynamic topic modeling based on latent Dirichlet analysis (LDA) to compare the discourse about stalking on Reddit before and after the COVID-19 pandemic. Similarly, Dahal et al., (2019) used LDA-based topic modeling to compare climate change discussions between different countries and over time on Twitter. Parsa et al., (2022) used Non-negative Matrix Factorization topic modeling to analyze climate change discussions on Reddit, and the study also traced embedded URLs to identify influential information sources. However, they primarily focused on macro-level opinion trends without fully addressing the mechanisms through which influence propagates within the network.

SNA provides a framework for examining how influence operates through relational ties rather than individual attributes (Kadushin, 2012). It allows researchers to map opinion formation and communication interaction by measuring network centrality, clustering structures, and bridging positions (Bedi & Sharma, 2016). Many studies have attempted to identify influencers on social media by SNA, measurements including in-degree centrality (Rohde et al., 2023), eigenvector centrality (Alazazi, 2023), closeness (L. Yang et al., 2018), interest and exclusivity of users for a specific topic (Riquelme et al., 2019). Some studies have integrated topic modeling and SNA to examine how discourse develops over time within networked communities. Rohde et al. (2023) explored the role of community structures and opinion leaders in facilitating discussions on e-cigarettes within Reddit. Using degree centrality metrics in SNA, combined with non-parametric ANOVA and negative binomial regression, the study identified opinion leaders and examined their influence on network composition and participation. Similarly, Alazazi (2023) analyzed vaccine-related discussions on Reddit using LDA topic modeling and SNA to identify both local and global opinion leaders. The study found that global opinion leaders foster broad participation, while local opinion leaders deepen intra-community exchanges. Jang et al. (2024) examined changes in discussion topics on Reddit before and during the pandemic, emphasizing the emergence of new pandemic-related topics such as victim support and outbreak tracking. Through SNA, the study demonstrated that the pandemic intensified interactions within like-minded groups, revealing increased polarization in online networks. Cau et al. (2024) studied echo chambers in sociopolitical discussions on Reddit, using LDA to analyze user-generated content and SNA to map interaction patterns. The findings emphasize the evolution of echo chambers over time, influenced by both user interactions and the contentious nature of the topics discussed. Additionally, Lee et al., (2016) combined LDA with an empirical model to examine relationship formation in location-based social networks. Their analysis offered perspectives on how topic modeling can be combined with social network analysis to explore opinion formation mechanisms.

As argued, Reddit's visibility is not determined by followers and authority, which allows a more organic growth of the influence dynamics. This decentralized model creates an environment in which influence fluctuates over time and is shaped by both content engagement and network positioning. While previous research has separately analyzed thematic content and network structures, the relationship between the evolution of opinions and the structural dynamics of influence remains insufficiently explored. Few studies have explicitly combined topic modeling and social network analysis to examine how different types of users emerge and shape public discourse in semi-anonymous contexts. Therefore, this study combines dynamic topic modeling with social network analysis to trace the evolution of discussions on urban air pollution, identify key roles at different phases of engagement, and analyze their influence and network positions over time.

III. Analytical Strategy

This study explores the micro-dynamics of influence and opinion formation in networked communities on urban air pollution discussions. How public opinion emerges and evolves and who leads the influence mechanisms in semi-anonymous social media remain insufficiently examined. To observe these dynamics, the analysis proceeds in three distinct stages. First, I employ topic modeling to map the main topics discussed on urban air pollution on Reddit and their evolving trends over time. I then identify different types of users who engage actively at different points during the shift of the topic. Finally, I use social network analysis to analyze these users' network positions and discursive roles in the process of topic transition. Rather than assuming the existence of "influencer" or "opinion leader", this study conceptualizes influence as a continuum of user engagement and network positioning. Influence is not a fixed attribute but a fluid characteristic that shifts based on user participation, discussion context, and network structure.

I begin with static topic modeling using BERTopic (Grootendorst, 2022) to identify the main themes discussed, showing the range of issues addressed in discussions and establishing a baseline understanding of discourse content. Next, adding timestamps to the dynamic topic modeling, I analyze the trends of these topics over time to capture the evolution and emergence of topics. I trace critical transition points (i.e., changepoints and peaks) and identify users who are notably active during critical phases—specifically, those who contribute to the topic significantly before the topic emerges and between the changepoint and peak periods. By analyzing user activity during these key phases, the study distinguishes between Early Adopters—users who engage significantly before a topic reaches mainstream visibility—and Early Majority—users who become active between the changepoint and peak period. Then, SNA is employed to examine whether users with greater topical contributions occupy more central or bridging positions within the network. I assess centrality measures (e.g., Eigenvector, Betweenness, Closeness, and clustering coefficients, and PageRank) to determine whether a user's influence, as measured by their topical contributions, is associated with central network positions. I compare the measures of Early Adopters and Early Majority to examine whether they have different positions and roles in the network. Finally, I take a local network as a case to look into the dynamics of the network structure when different users play a role in the discussion. I observe how specific users contribute to shaping opinions and how their interactions influence network structure. This explains previously relatively abstract findings and statements in a symbolic way.

IV. Methods

4.1 Data Selection

The data collection window is from January 1, 2020, to December 31, 2023, reflecting both the pressing public concerns about air pollution (e.g., wildfires and pandemic) and the shifting political landscapes and regulatory changes that frame these discussions. The COVID-19 pandemic, starting in 2020 led to lockdowns that reduced industrial activities and transportation, triggering discussions about the relationship between human activity and air quality. Wildfires in California, Oregon, and Canada between 2020 and 2023 caused severe air quality crisis. Significant policy shifts, such as the United States' re-entry into the Paris Agreement in 2021 and the passage of the Inflation Reduction Act in 2022 gained attention to air pollution control, environmental regulations, and climate mitigation strategies. International initiatives such as COP26 and new WHO air quality guidelines in 2021 and the IPCC Report in 2023 set updated scientific and policy benchmarks.

To collect targeting data, I identified ten of the most relevant subreddits with active discussions of air pollution—*fuckcars*, *science*, *environment*, *conspiracy*, *futurology*, *collapse*, *politics*, *technology*, *urbanhell*, and *electricvehicles*. The decision to extract data from the top ten most relevant subreddits, which I identified qualitatively using keyword frequency analysis, further refines the focus to communities where urban air pollution is actively discussed. Using a list of keywords, implemented as regular expressions[1] to include related terms and variations (Table 1), I filter threads that match these keywords more than once. Following this, I selected authors who had commented at least five times across the identified threads as active participants. This step established a cohort of contributors whose activity was central to the discourse

[1] A regular expression (shortened as regex or regexp), sometimes referred to as rational expression, is a sequence of characters that specifies a match pattern in text. Regular expressions are used to describe regular languages by mathematical notation. See https://en.wikipedia.org/wiki/Regular_expression.

on air pollution within these subreddits. I only focus on active users who posted at least five comments within the sampled subreddits. A total of 28375 comments replying to posts by these authors across Reddit during the same period were collected. Finally, the dataset recorded and counted all reply relationships as (reply_author, parent_author) pairs across the collected comments.

Table 1. [Keywords and Subreddit for Data Collection]

Keyword Regular Expressions	**Subreddits**
\bair\W+?pollut\w*?\b	fuckcars (1356)
\bpollut\w*?\W*?air\b	science (797)
\b(urban|city)\W+?air\W+?quality\b	environment (781)
\bsustainable\W+?cit\w*?\b	conspiracy (637)
\bsmog\b	Futurology (627)
\b(respiratory|lung)\W+?health\b	collapse (605)
\b(vehicle|car|automobile)\W+?emmissions?\b	politics (505)
\bair\W+?quality\W+?index\b	technology (494)
\baqi\b	UrbanHell (442)
\bpollut\w*?\W+?monitor\w*?\b	electricvehicles (304)
\bclean\W+?air\W+?initiatives?\b	
\bgreen\W+?initiative?\b	

4.2 Text Processing

LDA is the most commonly used topic modeling algorithm in social science research, but it tends to neglect co-occurrence in social media text mining (Jaradat & Matskin, 2019). The advent of the Transformer model in 2017 has changed the field of NLP, with neural models based on word embeddings starting to evolve and many new algorithms being developed to overcome some of the limitations. Word, sentence, and document embeddings often lead to better results in a variety of NLP tasks because they capture context, meaning, and semantic relationships between words and text documents (Grootendorst, 2022). A comparative study shows that BERTopic performs better than LDA, Top2Vec and NMF in the analysis of discussions in short social media texts such as community forums (Egger & Yu, 2022). Therefore, this research uses the Python implementation of BERTopic (Grootendorst, 2022), which is a pre-trained transformer-based language model and employs a clustering technique for topic extraction.

The cleaned data contains 26,981 comments. I replaced all instances of the string 'nan' with np.nan, and entries with missing values in either the 'body' or 'date' columns were removed. All text was converted to lowercase. URLs and social media usernames were replaced with the placeholders 'URL' and 'SCREEN_NAME'. Special characters, except for spaces, underscores, and hyphens, were stripped from the text. Additionally, the phrase "air pollution" was removed to prevent overrepresentation of the primary research keyword. Texts containing fewer than ten words were excluded. For the extraction of topic-specific keywords, I used the build-in CountVectorize function. This model extracted unigram and bigram tokens while filtering out stopwords using a custom list that combined NLTK's English stopwords and additional domain-specific terms such as `'http'`, `'https'`, `'amp'`, and `'com'`. Then developed document embeddings using the SentenceBERT (SBERT) framework, which represents textual data in a high-dimensional semantic space. This model is based on a transformer architecture and generates dense embeddings that capture semantic similarities between sentences. These embeddings provided the foundation for downstream dimensionality reduction and clustering tasks. BERTopic applies a class-based TF-IDF (term frequency-inverse document frequency) procedure for developing cluster-word representations. The version of TF-IDF is used to extract the most meaningful words from each identified cluster.

4.3 Network Building

To construct network relations from user interactions, I identified comment replies within discussion threads. Each edge in the network consists of a user who initiates a reply (author) and a user who receives the reply (parent), forming a directed network of discussions, where nodes represent individual users, and directed edges represent these reply interactions. To be clearer, Redditor B commenting on Redditor A's post signifies a directed relationship between B → A, but not the other way around (i.e., A → B). Thus, a node's degree centrality for the current study was further distinguished by its in-degree (responses to a Redditor's submission) and out-degree (a Redditor commenting on a submission). In this case, Redditor A's in-degree centrality would be 1 and Redditor B's would be 0.

This study uses the following network metrics to quantify user position and interaction within the network. Eigenvector Centrality (EC) measures a user's connectivity to other high-influence nodes, indicating whether they occupy a core position; Betweenness Centrality (BC), which is based on the number of network paths that pass through a user, reflects the user's role as an intermediary in information dissemination; Closeness Centrality (CC) reveals the average distance between a user and all others, thus indicating information propagation efficiency, lower scores indicating a more central and important position in the network; Clustering Coefficient (Ccoft) assesses the density of connections within a user's local neighborhood. Out-degree Centrality (outDC) quantifies the frequency with which a user initiates interactions; PageRank evaluates user importance from a random-walk perspective, emphasizing reachability within the information flow.

V. Data Analysis and Results

5.1 Static Topic Modeling

This study first employed static topic modeling to identify the dominant topics discussed in the corpus. BERTopics uses a language model to generate embeddings of the input text, reducing the data dimension through UMAP (Uniform Manifold Approximation and Projection) and clustering the embeddings through HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) which has the feature of considering noisy subjects as outliers. The BERTopic model produced 19 topics based on the corpus of Reddit comments collected from 2020 to 2023. Each comment was assigned a probability distribution across these 19 topics. After qualitatively interpreting the topics based on dominant terms and representative comments, I identified three topics—Sky haze (Topic 11), Plastics and men's health (Topic 10), and Plant-based diets (Topic 15)—as "less relevant" due to their marginal or tangential relationship to urban air pollution. They were excluded from further analysis. The remaining 16 topics were further synthesized into four categories: Area's Smog, Health Impacts, Causes of Urban Air Pollution, and Proposed Solutions.

Table 2. [Descriptive Results of Static Topic Modeling]

Category	***Topic***	***Title***	***Description***	***Top3 Subreddits***
Area's Smog	3	China's Smog and Coal	China's air pollution, referencing smog and coal usage as primary concerns.	Futurology(15.2%), UrbanHell(13.1%), conspiracy(11.9%)
	6	California's Smog	California's poor air quality, linking regional factors to broader environmental challenges.	politics(18.6%), electricvehicles(13.9%), technology(13.4%)
	14	Los Angeles Air Quality	Los Angeles's bad air quality, emphasizing region-specific impacts.	UrbanHell(28.4%), politics(14.1%), fuckcars(9.3%)
	16	India's Air Pollution	India's pollution issues, focusing on urban and industrial sources.	UrbanHell(36.9%), collapse(15.6%), Futurology(9.9%)
Health Impacts	1	COVID and Conspiracies	COVID-19–related lung issues, infection, and mortality, including some conspiracy discussion.	conspiracy(25.6%), science(16.9%), collapse(10.1%)
	9	Marijuana and Lung Cancer	The use of marijuana, tobacco, and cannabis, highlighting potential lung health risks.	science(34.6%), conspiracy(11.1%), Futurology(9.2%)
	12	Masks and COVID	Mask usage and COVID prevention, offering partial links to air pollution health considerations.	conspiracy(42.2%), science(10.7%), politics(9.3%)
	17	Brain and Mental Health	How air pollution may affect brain function (e.g., dementia) and mental health (e.g., depression).	science(41.5%), collapse(16.7%), conspiracy(8.1%)
Causes	0	City Traffic	How car usage contributes to urban pollution, placing blame on drivers.	fuckcars(37.2%), ukpolitics(19.9%), Futurology(5.5%)

	5	Wildfires and AQI	How wildfires influence the Air Quality Index (AQI) and air quality metrics.	collapse(42.0%), ScienceBasedParenting(7.2%), science(7.1%)
	8	Aerosols and Climate	Aerosol outputs and broader climate change implications.	collapse(31.7%), Futurology(13.5%), conspiracy(12.2%)
	13	Gas Stoves and Indoor Air	Gas stove usage and its potential to harm indoor air quality.	environment(19.1%), ukpolitics(14.3%), science(12.1%)
	18	Greenhouse Gas Emissions	Greenhouse gases (CO_2, methane, NO_x) and their role in urban pollution.	Futurology(15.2%), conspiracy(12.6%), science(12.1%)
Proposed Solutions	2	Electric Cars	Electric vehicles reduce greenhouse gas emissions and improve overall efficiency.	electricvehicles(32.2%), fuckcars(14.5%), Futurology(14.0%)
	4	Nuclear Power	Nuclear energy as a potential large-scale solution to air pollution challenges.	Futurology(28.6%), environment(12.9%), technology(10.5%)
	7	EPA Acts and Regulations	U.S. federal Environmental Protection Agency policies under different administrations (e.g., Trump, Obama).	politics(49.4%), environment(13.7%), Futurology(4.3%)
Less Relevant	11	Sky Haze	Smog, fog, and clouds in the sky (too broad).	UrbanHell(31.7%), conspiracy(20.4%), fuckcars(6.6%)
	10	Plastics and Men's Health	Plastics in the air and men's health concerns.	science(27.8%), collapse(13.6%), Futurology(10.1%)
	15	Plant-Based Diets	Diet choices with minimal direct linkage to urban air pollution.	environment(37.0%), science(15.2%), Futurology(12.4%)

The first "Area's Smog" category highlights four regions: China (Topic 3), Los Angeles (Topic 14), California (Topic 6), and India (Topic 16). These topics emphasize the comparative framing between local and global pollution crises, predominantly appearing in UrbanHell and Futurology, suggesting localized environmental critique and drawback of technology.

The second category, "Health Impacts," reflects public concerns about the consequences of air pollution on physical and mental health, especially during COVID-19. Topics include COVID and conspiracies (Topic 1), Masks and COVID (Topic 12), Marijuana and lung cancer (Topic 9), and Brain and mental health issues (Topic 17). These discussions link air quality to broader health and pandemic-related anxieties, framing urban air pollution as a public health crisis. Conspiracy is the most dominant subreddit here, driving COVID-related discussions (42.2% in Masks and COVID, 25.6% in COVID and Conspiracies). These discussions often involve skepticism toward health recommendations and government regulations during the pandemic. Science is another dominant subreddit, particularly in Brain and Mental Health (41.5%) and Marijuana and Lung Cancer (34.6%). Unlike conspiracy, discussions in science focus more on scientific evidence linking air pollution to cognitive decline and respiratory issues.

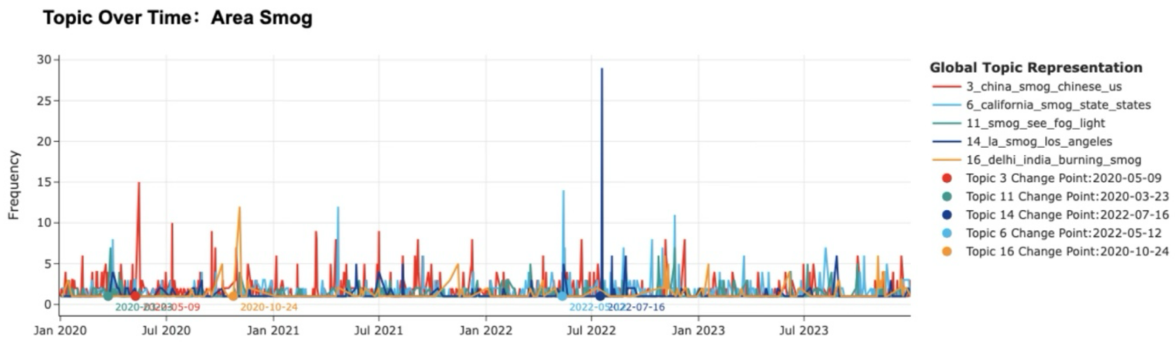
The "Causes" category addresses the various human and environmental factors contributing to air pollution, including City traffic (Topic 0), Wildfires and AQI (Topic 5), Aerosols and climate (Topic 8), Gas stoves and indoor air (Topic 13), and Greenhouse gas emissions (Topic 18), ranging from everyday behaviors to broader industrial and ecological processes. Subreddits are concentrated in fuckcars and collapse, revealing strong ideological and evidence-based debates. Fuckcars is the dominant subreddit for traffic-related pollution (37.2% in City Traffic, 9.3% in Los Angeles Air Quality). This subreddit is highly critical of car-centric urban design, reflecting a growing movement against automobile dependency in climate-conscious discussions. Collapse is highly represented in Aerosols and Climate (31.7%) and Wildfires and AQI (42.0%), suggesting that many users frame air pollution as part of an irreversible climate crisis rather than a solvable policy issue.

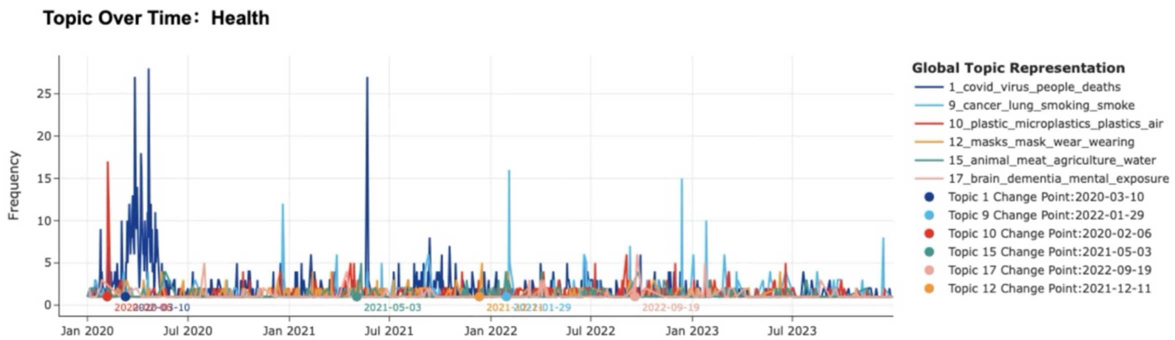
The "Proposed Solutions" is about strategies and mitigations to urban air pollution. This includes Electric cars (Topic 2), Nuclear power (Topic 4), and EPA Acts and regulations (Topic 7). These topics indicate that users are not only engaged in diagnosing problems but also debating actionable interventions and policy solutions. Environment is a key space for pragmatic policy discussions (13.7% in EPA Acts, 12.9% in Nuclear Power, 19.1% in Gas Stoves and Indoor Air), showing that certain Reddit communities actively advocate for environmental policy changes.

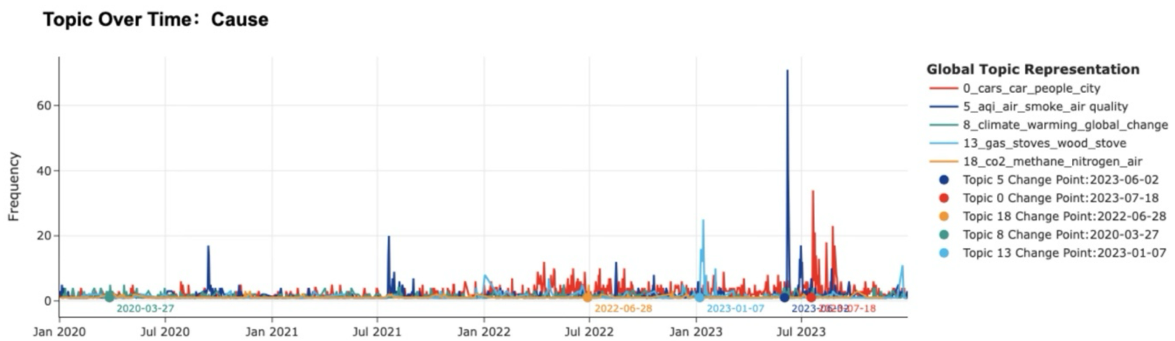
5.2 Dynamic Topic Modeling

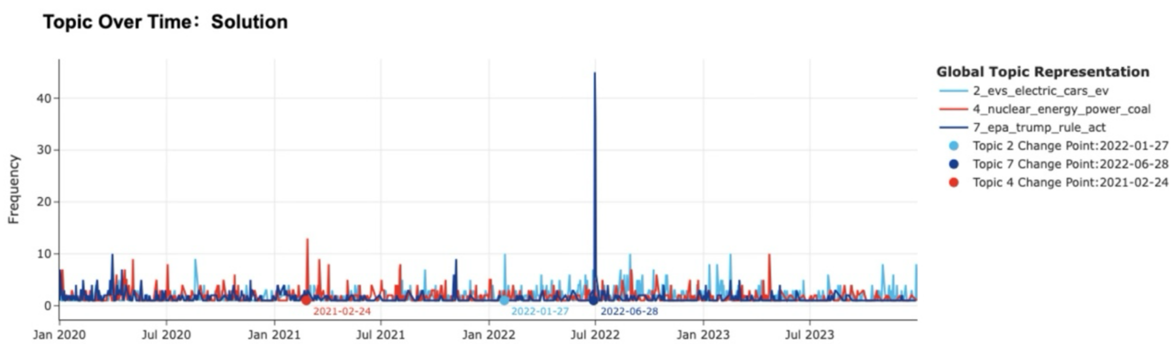
To examine how discussions of urban air pollution evolve over time, the dataset was divided into daily intervals, with each comment assigned a specific timestamp corresponding to the day it was posted. I included "day" timestamps to the model and plotted a time series of post frequencies for each topic. From Figure 1, I observed the trends of each topic and selected the most obvious fluctuated topics to conduct further research on opinion formation.

Figure 1. [Dynamic Topic Modeling Result: Topic Trend Over Time]









Topic 0 (City Traffic), Topic 1 (COVID and Conspiracies), Topic 5 (Wildfires and AQI), Topic 7 (EPA Acts and Regulations), Topic 9 (Marijuana and Lung Cancer), Topic 13 (Gas Stoves and Indoor Air), and Topic 14 (Los Angeles Air Quality) exhibit dramatic fluctuations in their prominence in the discourse, while the other topics remained relatively stable throughout the observed period. I therefore focus on the fluctuation periods of these seven topics in the following analysis. In order to interrogate the process by which topics come to prominence, I identified two points in time for each highly fluctuating topic (see Figure 2): a changepoint (the lowest point immediately before a rapid rise in topic engagement) and a peak (the highest frequency point in the time series). The first segment t_1 covers the month (ensuring that there are enough users who comment on the topic during the shortest possible time) leading up to the changepoint, and the second segment t_2 spans from the changepoint to the peak. The Sampling Time in Table 3 is $t_1 + t_2$.

Figure 2. [Sampling Time Segment]

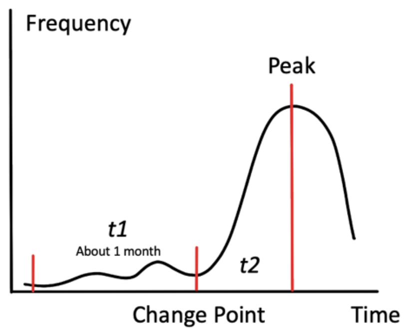


Table 3. [Representative Topics for SNA]

Category	Topic	Title	All	Sampling Time	Comment
Health Impacts	1	COVID and Conspiracies	1858	2020-03-04 to 2020-03-27	137
	9	Marijuana and Lung Cancer	686	2022-11-12 to 2022-12-13	62
Causes	0	City Traffic	2690	2023-06-18 to 2023-07-21	167
	5	Wildfires and AQI	1227	2023-05-03 to 2023-06-07	189
	13	Gas Stoves and Indoor Air	398	2022-12-08 to 2023-01-13	193
Solutions	7	EPA Acts and Regulations	985	2022-05-29 to 2022-06-30	100
Area's Smog	14	Los Angeles Air Quality	398	2022-06-16 to 2022-07-19	298

User's contribution to topics is measured by the sum of each of their comments' topic distribution for each topic calculated by BERTopic. For example, if user A has two comments during t_1, the first comment topic distribution is topic 1 (0.2), topic 2 (0.7), and the second comment is topic 2 (0.1), topic 3 (0.5). User A's contribution to topic 1 during t_1 is 0.2, and topic 2 during t_1 is 0.7+0.1=0.8. This number can be loosely interpreted as the effective number of comments a user makes on a specific topic. Two types of users with substantial contributions in t_1 and t_2 were subsequently identified by summing up the contribution of each topic at the user level. *Early Adopters* refer to users who significantly contribute to a topic one month before the changepoint (the lowest point before a surge in engagement). *Early Majority* are users who contribute significantly from the changepoint to the peak period, meaning they are active when a topic is gaining momentum. While Early Adopters and Early Majority users are identified based on the content of their comments, a third category of the user was identified based on a more traditional network-based measure of influence: *In-degree Hubs* are those who possess the top 2% in-degree centrality, regardless of the topics they discuss. Table 4 shows the number and average contribution of three types of users identified in the seven topics during the sampling time.

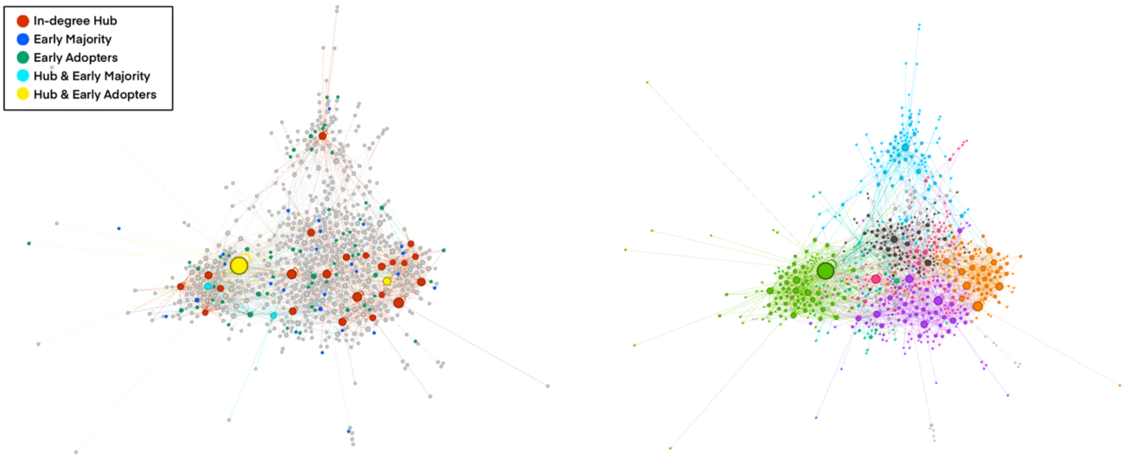
Table 4. [Average Topic Contribution of Hub, Early Adopters, and Early Majority]

	Count	Mean Contribution_t1	Mean Contribution_t2
Early Adopters	64	0.846	0.031
Early Majority	110	0.024	0.937
In-degree Hub	26	0.007	0.005
Others	1353	0.039	0.021
All	1571	0.229	0.249

5.3 Social Network Analysis

To analyze how users' network positions evolve during the distinct time windows associated with the seven rapidly fluctuating topics, I conducted separate social network analyses for each sampling time (see Table 3) and a global network that gathers seven topics together (Figure 3). All comments posted during the sampling time were taken to construct a directed graph in which edges connect authors to the parents of their comments. Within each directed local network, Eigenvector Centrality, Betweenness Centrality, Clustering Coefficient, Closeness Centrality, In-degree Centrality, and PageRank are calculated to analyze how certain users occupy central or bridging positions.

Figure 3. [Global Network Plots: Positions and Community Detection]



5.3.1 Early Adopters VS Early Majority

I conducted a logistic regression analysis based on various network centrality measures and comment attributes. The models assess whether users with higher centrality, clustering, or posting behaviors are more likely to be Early Adopters and Early Majority. The logistic regression models are as follows:

$$logit(t(P(1) = \beta_0 + \beta_1 EC + \beta_2 Ccoft + \beta_3 CC + \beta_4 outDC + \beta_5 PageRank + \beta_6 AvgPostLength$$

Table 5. [Logistic Regression Results]

	Predictor	Estimate	Std. Error	z-value	p-value
Early Adopters					
	(Intercept)	-4.0042	0.2454	-16.316	< 0.001 ***
	Eigenvector Centrality	0.8219	0.3417	2.405	0.0162 *
	Clustering Coefficient	-0.3976	0.3335	-1.192	0.2332
	Closeness Centrality	-0.6461	0.3029	-2.133	0.0329 *
	Out-degree Centrality	-0.1658	0.4575	-0.362	0.7171
	PageRank	-1.2324	0.3135	-3.931	< 0.001 ***
	Avg Post Length	-0.3081	0.2358	-1.307	0.1913
Early Majority					
	(Intercept)	-3.7420	0.2274	-16.457	< 0.001 ***
	Eigenvector Centrality	0.4811	0.3488	1.379	0.1678
	Clustering Coefficient	0.2480	0.1349	1.838	0.0660 .
	Closeness Centrality	-0.5158	0.2437	-2.117	0.0343 *
	Out-degree Centrality	-0.9583	0.4872	-1.967	0.0492 *
	PageRank	-1.4011	0.2769	-5.061	< 0.001 ***
	Avg Post Length	-0.3650	0.1930	-1.892	0.0585 .

The average minimal path length between In-degree Hub, Early Adopters, and Early Majority are calculated. The shortest path between two nodes represents the minimum number of steps required for information to travel between them. Hubs have a longer average path to Early Adopters and Early Majority, which may indicate that Hubs serve more as passive information targets rather than active disseminators within the discussion.

Table 6. [Average Minimal Path Length between Hub, Early Adopters, and Early Majority]

Direction	Average Minimal Path Length
Early Adopters → Hub	2.89
Hub → Early Adopters	3.02
Early Majority → Hub	3.07
Hub → Early Majority	3.14

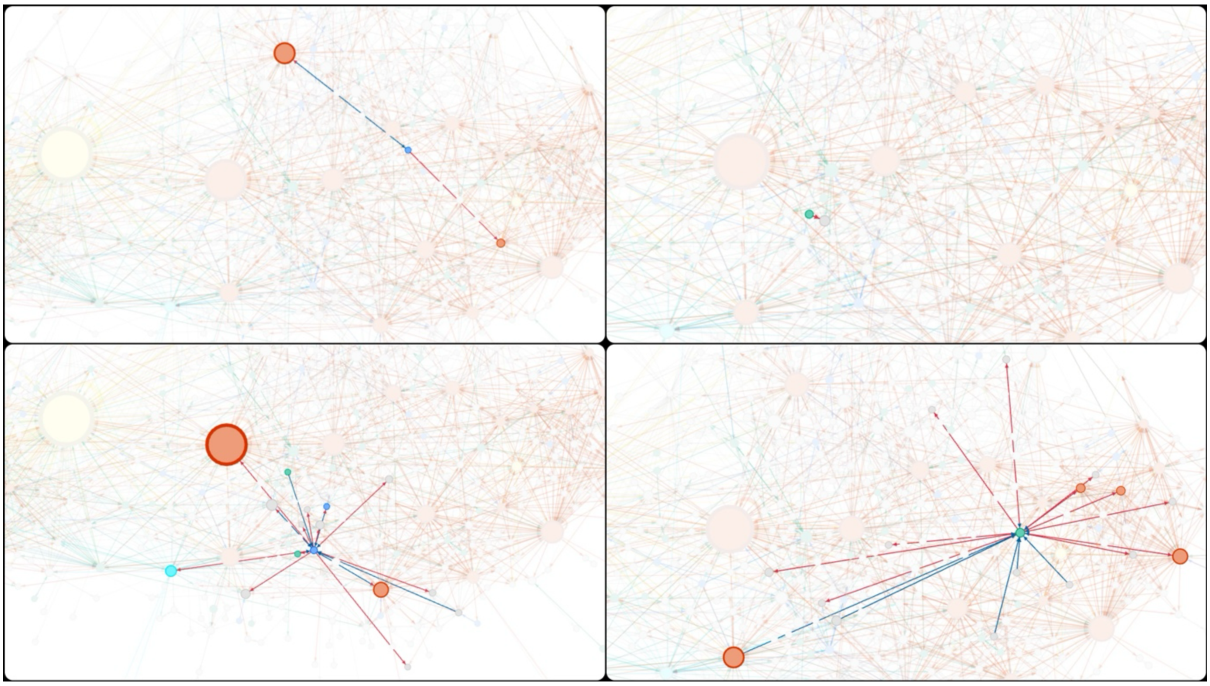
5.3.1 Early Adopters as Strategic Opinion Shapers

Early Adopters emerge as proactive and strategic actors in the initial stages of a discussion's evolution. While Early Adopters and Early Majority users look structurally similar in many ways, Early Adopters are distinguished by significantly higher Eigenvector Centrality (β = 0.8219, p = 0.0162). This suggests that even though they do not receive a large amount of direct attention themselves (as discussed below) they are well-connected to other highly influential (in network structure) nodes. The average shortest path from Early Adopters to Hubs is the shortest among all tested interactions, averaging 2.89 steps. This finding suggests that Early Adopters are highly efficient in identifying and engaging with key figures in the discussion space. However, the reverse path from Hubs to Early Adopters is longer, averaging 3.02 steps, indicating that Hubs do not frequently reciprocate interactions with Early Adopters.

Negatively associated PageRank ($\beta = -1.2324$, $p < 0.001$) also shows that although they actively seek out these high-visibility users but receive little reciprocal interaction, it indicates that these interactions do not translate into sustained influence or widespread recognition. While their connection to influential users may indicate a degree of visibility, that position does not induce other users to engage directly with Early Adopters. Therefore, Early Adopters may serve as specialized initiators; they leave the seed content in the comment section of Hubs to introduce discussions before they gain momentum. Their participation in these high-traffic spaces allows them to amplify their opinion, thus, using Hubs' comment sections as a strategic platform for increasing their own visibility.

The 'back-seat' role played by Early Adopters does not consolidate long-term influence. Their closeness centrality is significantly negative ($\beta = -0.6461$, $p < 0.05$), implying that they are not positioned at the shortest average distance to all other users. Furthermore, their out-degree centrality is not significant, reinforcing the idea that they are not widely engaging others in direct conversations but are targeting the Hubs to reply.

Figure 4. [Example of Early Adopters and Early Majority in the Network]



Early Adopters (Green) Early Majority (Blue)

5.3.2 Early Majority as Echo Reinforcers

Rather than interacting directly with Early Adopters, Early Majority users appear to echo discussions initiated by them in separate spaces. Their significantly lower PageRank scores ($p < 0.001$) indicate that they do not occupy prominent positions within the information flow. Additionally, their negative out-degree centrality coefficient ($\beta = -0.9583$, $p = 0.0492$) shows that they are less likely to reply to others' comments but actively repeat the opinion themselves. The Clustering Coefficient is not significant in Early Adopters, while the Early Majority present a positive trend ($\beta = 0.2480$, $p = 0.0660$), suggesting that close interaction between local groups can promote the rapid diffusion of information when a topic breaks out. The Average Post Length is not significant in the Early Adopters but has marginal significance in the Early

Majority (β = -0.3650, p = 0.0585), which may indicate that concise and easy dissemination of information is more advantageous than lengthy information in the outbreak stage.

The mean minimal path length further shows the differences. The average shortest path from Early Majority to Hubs is longer than that of Early Adopters, at 3.07 steps, suggesting that it takes them more time to reach central figures in the network. Moreover, the path from Hubs to Early Majority is the longest among all tested interactions, at 3.14 steps. These results show that Early Majority users are not actively establishing connections with central nodes but are actively engaging in discussions that have already gained traction. Therefore, Early Majority users may not contribute to the construction of the network but instead, participate in redundant discussions that reinforce existing narratives by repeating the echo.

5.3.3 Hub as an Amplifier

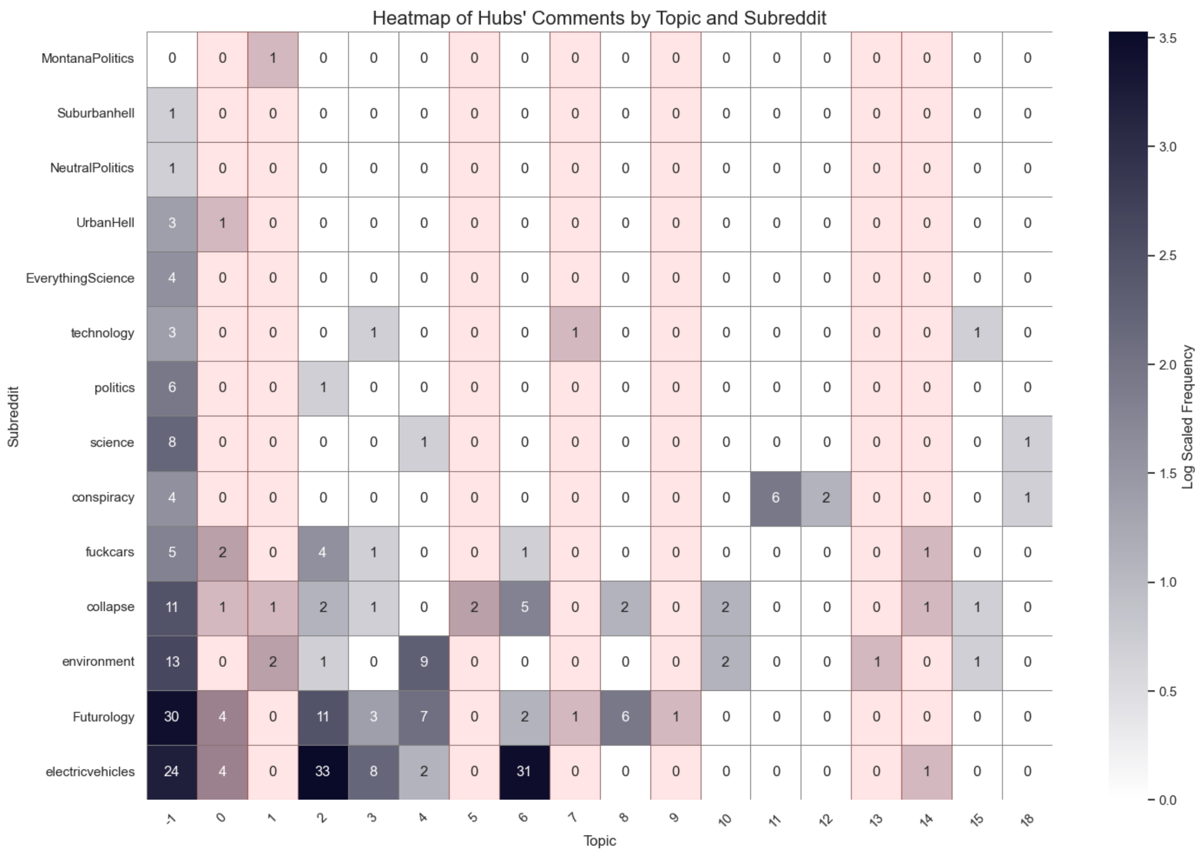
In-Degree Hubs (Red nodes in Figure 3) are defined as those possessing the top 2% of in-degree centrality in each local network, meaning they receive a high volume of replies from other users. Previous research has established that high in-degree centrality does not necessarily equate to influence over content (Kitsak et al., 2010; González-Bailón et al., 2013). In this study, I found that Hubs are amplifiers for Early Adopters rather than active opinion leaders. Hub may serve as aggregation points within a network, enhancing visibility but not necessarily contributing original insights.

Hub's average contributions are very low during both t1 (0.007) and t2 (0.005), indicating that their prominence within the network is only structural rather than content-based.

Information generally travels faster from Early Adopters to Hubs (2.89 steps) than from Hubs to Early Adopters (3.02 steps), suggesting that Hubs are more likely to aggregate information introduced by Early Adopters. Combined with the relatively high Eigenvector Centrality and low PageRank of Early Adopters, Hubs' role in this setting is passive. They almost never engaged in the topics themselves, but were selected by Early Adopters, serving as an amplifier for their opinions. So what are they talking about if they are not discussing the topic exactly?

In Figure 6, the red columns highlight the topics with noticeable fluctuation analyzed in detail above. Hubs' comments are mainly on Topic –1 and other topics that are more stable over time. Topic –1 means the comments were not assigned to any specific topic by BERTopic. Such results suggest two possibilities. First, the comments may be related to air pollution, but the discussion is generalized thus, it cannot be categorized under a single specific topic. Second, the comments may be addressing broader issues where air pollution is only mentioned, although keywords were detected, air pollution is not the topic of comments. In terms of distributing topics with stable trends, on the one hand, Hubs are persistently engaging in discussions on air pollution, thus remaining at the center of the global network, which allows them to be identified by Early Adopters who seek to amplify their ideas through established network nodes. On the other hand, their comments did not trigger a noticeable increase in discussions, which implies that they may not influence these discussions.

Figure 5. [Heatmap of Hubs' Comments: Topic × Subreddit]



5.3.4 Special Nodes

In the global network, I found four highly visible users as exceptional cases: *dumnezero* and *Speculawyer* were identified as both Hub and Early Adopters, *silence7* and *DeaditeMessiah* were identified as both Hub and Early Majority.

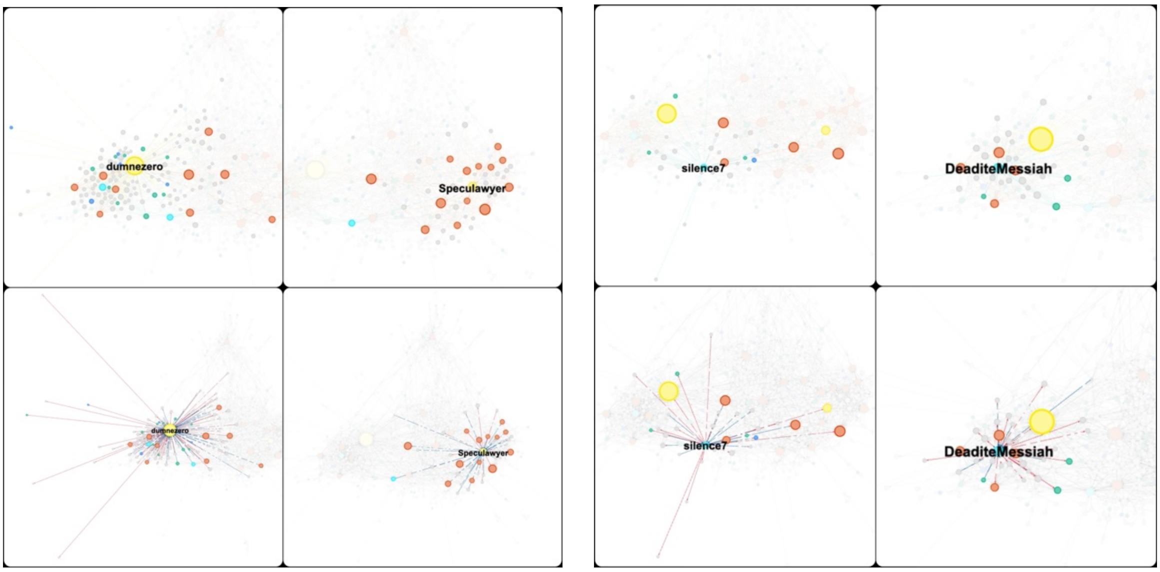
Being both Hub and Early Adopters suggests that they were not only central figures in the network but also among the earliest to engage in emerging discussions. *Speculawyer* demonstrated a typical Early Adopter by strategically engaging with Hubs and stands out among Early Adopters with strong reciprocal engagement from Hubs. Almost all users that *Speculawyer* replied to are Hubs, and the high in-degree from Hubs suggests that these Hubs noticed *Speculawyer*'s engagement. This mutual interaction with Hubs separates *Speculawyer* from other Early Adopters who engage with Hubs but often do not receive replies. Additionally, Speculawyer's engagement is concentrated on two major topics (Topic 2 Electric car and Topic 6 California's Smog), which are not too narrow and popular. *Speculawyer* might highly influence opinion formation and topic discussion on these two topics.

Dumnezero, is the most visible node in the network, with broad engagement across diverse users, and the interaction range is very wide. This engagement pattern is further reflected in *dumnezero*'s topic participation, where it actively contributed to seven different topics (Topics 8, 5, 0, 1, 2, 10, 3)—far exceeding the participation breadth of the other users in the table. This indicates that *dumnezero* was a key participant in the overall urban air pollution discourse, contributing across multiple subtopics rather than specializing in one or two discussions. *Dumnezero* is a highly engaged user in the whole network and takes on a community-wide amplification role, not only raising up topics but also transitioning discussions from early stage to broader adoption within the community.

Table 7. [Special Nodes' Interactions]

	Hub and **Early Adopters**				**Hub** and **Early Majority**			
	Speculawyer		*dumnezero*		*silence7*		*DeaditeMessiah*	
Direction	**Out**	**In**	**Out**	**In**	**Out**	**In**	**Out**	**In**
Hub	11	12	7	8	7	5	3	3
Early Adopters	1	1	6	5	3	1	1	1
Early Majority	0	1	9	11	4	2	3	3
Most Engaged Topics (count >30)	2, 6	2, 6	8, 5, 0, 1, 2, 10, 3	8, 5, 0, 1, 2, 10, 3	4, 2	4, 2	8, 1	8

Figure 6. [Special Nodes' Position and Neighbors]



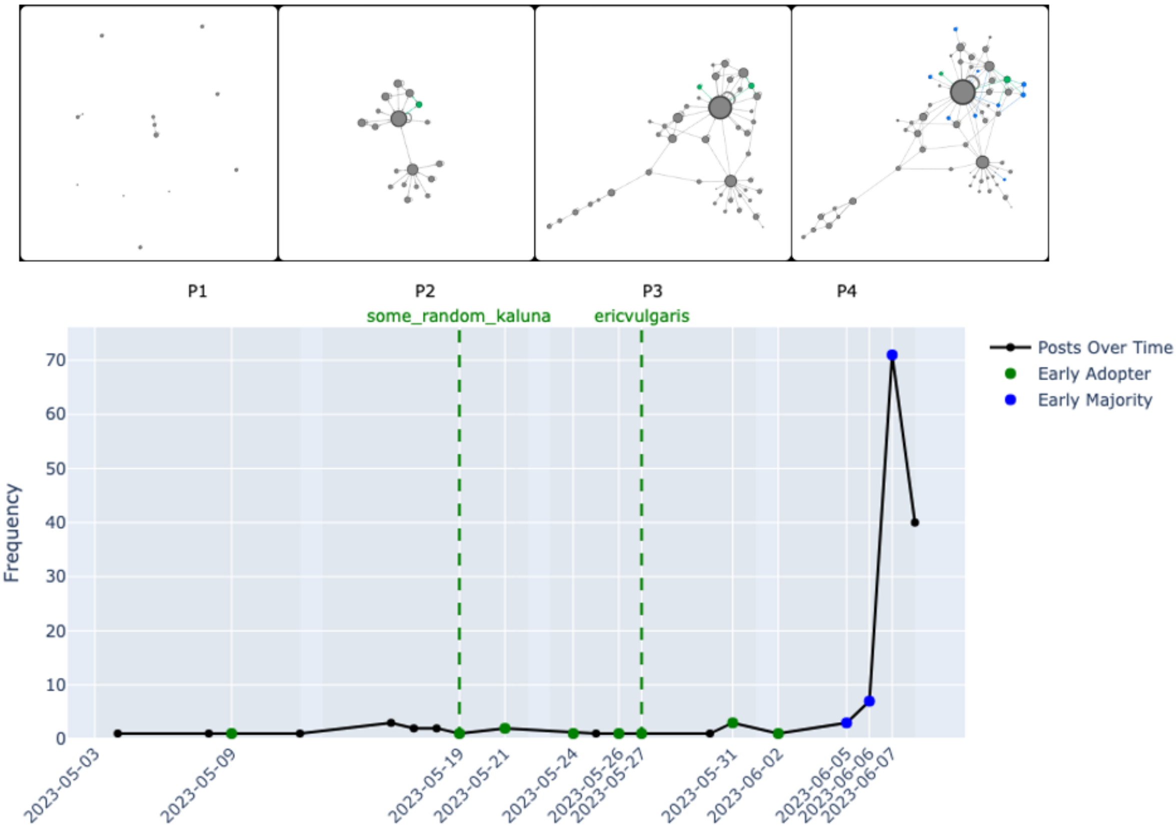
Silence7 and *DeaditeMessiah* did not appear particularly distinct overall in either the network structure visualization or interaction data. However, they stood out within the Early Majority in a phase of rapid expansion where they may have a highly homogeneous discussion. Their difference compared to other Early Majority is that they had reciprocal interaction with other users (especially Hubs like *dumnezero*) rather than being peripherical in the broader discussion. This finding suggests that during the peak phase of topic dissemination on social media, the interaction partners may play a more significant role in determining visibility and influence than the substantive content. If a user is identified both as a Hub and Early Adopter/ Early Majority, which means that it is both prominent in structure and content, I would argue it has an influence on the formation of a topic.

5.4 Looking into One Local Network Case

To look closer at the dynamic of opinion formation and influence transition, I take the local network of Topic 5 (Wildfires and AQI) as a case. I divided the sampling time May 3, 2023, to June 7, 2023 into four time periods (P1–P4) and extracted all comments that have more than

0.05 contribution to Topic 5 within the sampling time. Nodes were assigned to every period based on their comment published and their types (Early Adopters, Early Majority, or Other). Then, a dynamic directed network visualization was constructed in Gephi with four time phases.

Figure 7. [The Formation and Structure of Topic 5 Local Network]



In the first phase (P1), the network was a scattered and disconnected plot. There was no clear Hub or clustering community, indicating that the topic had not yet coalesced into a structured debate. A visible connected group occurred in the second phase (P2) with the entry of an identified Early Adopters, *some_random_kaluna*, into the core discussion area. This Early Adopter replied to a high in-degree centrality user, which is *dumnezero*—a highly active and central user in the whole network. This aligned with the previous regression finding that Early Adopters tend to find an in-degree Hub to reply. The presence of an Early Adopter engaging with a high-degree node suggests that Topic 5 was beginning to gain traction, likely benefiting from the visibility provided by well-connected individuals.

By the third phase (P3), the network showed a clear structure, and another Early Adopter, *ericvulgaris*, entered the network. This Early Adopter also replied to *dumnezero*, which is more visible as a central Hub progressively. While most of the identified Early Adopters did not enter the core network structure, there are several possibilities. One case is that their content might also be seen and adopted by other users like the Early Majority, but since they interacted with other Hubs that were not in this local network, they remained invisible here. Or they just post something highly contributed to the topic themselves without getting noticed by others. In the final phase (P4), the overall structure of the network did not change, but with the entry of multiple Early Majority users that were positioned in the marginal place of the network. The Early Majority did not show strong connections with central nodes, and they expanded the scope of the central network but did not influence the structure. As I discussed before, they were less influential in shaping the discourse and more likely to passively engage with pre-existing discussions. Therefore, combined with the logistic regression results and previous analysis, Early Adopters define the layout and structure of discussions, and Early Majority fill out the existing pattern.

It's obvious that although the Hubs are in the central position, they were not the Early Adopters or Early Majority, which means their content generated in this time period did not contribute much to the topic5. All the comments that the two Hubs in this local network published in the observed time, they did not have a comment that contributed to the topic5 more than 0.1. I ran a logistic regression for Hub and types of contribution and none of the contributions were significantly correlated. This confirms that these in-degree Hubs were not themselves shaping the discussion but rather acting as platforms that provide a space for visibility where others, especially Early Adopters used as amplifiers.

VI. Discussion

6.1 Influence Flows with Time and Engagement

Previous influence frameworks have often conceptualized influence as an inherent characteristic of individuals, such as opinion leaders or high-centrality users (Katz & Lazarsfeld, 1955; Bourdieu, 1986; Freeman, 1979). However, the findings of this study suggest that the discussion of urban air pollution on Reddit is not conducted by so-called "opinion leaders" or by high-centrality users initiated or expanded. Users with high contributions to the topic are often not significant in themselves, but because they first targeted to connect with high-in-degree Hubs. The traditional two-step flow model of communication (Katz & Lazarsfeld, 1955) suggests that mass media influences "opinion leaders," who then disseminate information to a broader audience. Hubs in this research play a similar role in disseminating information, but their impact is indirect and passive. Researchers have measured the influencer by degree centrality or other centralities, and the assumption is that users with numerous connections—especially those with high in-degree centrality—serve as key opinion leaders and agenda setters. However, the findings of this study challenge this assumption by showing that In-Degree Hubs, while structurally central, do not contribute to setting and leading discussions. Instead, they function primarily as amplifiers or platforms selected by Early Adopters, serving as passive aggregators of responses and engagement rather than as leaders of discourse. This phenomenon can often be seen in social media, where messages from people asking for help or comments trying to drive traffic to their personal accounts appear in popular comment sections. The findings of this research underscore the importance of such ‘second-order’ network-structural effects and show those effects may be overlooked if network analysis is undertaken without considering the content of online discussions.

Early Adopters who initiate a topic have a more temporal and interaction-based influence. The study found that the comments through which they introduce topics were not random but selective responses to high in-degree Hubs. Early Adopters often replied to posts or comments made by Hubs, taking advantage of Hubs’ visibility to bring new information to the fore. Even if the Hubs did not respond back, the Early Adopters’ contributions were seen by the many others watching those high-visibility spots. These findings align with and extend the Diffusion of Innovations Theory (Rogers, 1962; Sziklai & Lengyel, 2022), which posits that Early Adopters are key drivers of innovation dissemination. However, classic diffusion models often assume that Early Adopters are a fixed category of socially prestigious or well-informed individuals who are naturally predisposed to innovation adoption. My evidence extends this by showing that early influence on Reddit is not a fixed trait of socially prestigious individuals, but a situational role any user can step into through timely engagement. In our case, Early Adopters are defined by what they do rather than who they are. Early Adopters do not necessarily have high in-degree or persistent visibility to be seen. This finding shows the environment in Reddits and similar digital communities that provide new pathways for influence and also supports a flow of influence in which different users exert influence at different times depending on their interactions.

Previous studies on social contagion (Christakis & Fowler, 2013) and network diffusion models (Kadushin, 2012) suggest that opinion spread is shaped by both structural positioning and exposure frequency. This study emphasizes that structural position or exposure alone is not sufficient for influence, placing in the right structural position at the right time is more important for gaining influence. This finding is in the context of semi-anonymous platforms, where users do not rely on traditional markers of authority such as institutional affiliation or follower count. Instead, influence is constructed through engagement in high-visibility spaces and interaction with key network nodes.

6.2 The Temporal Structure of Opinion Formation

This study finds that the opinion formation of urban air pollution on Reddit follows a structural sequence. First, an initiation phase occurs when Early Adopters introduce a new opinion or perspective. These early comments are often strategically placed in high-visibility space (for instance, as early comments on a popular post), bring new opinions to the table, and form a structural discussion network. Before a growth or peak phase, the structure is relatively stable. Then, the opinion gains traction, and many users join the network of the topic. During the peak period, the discussion thread experiences a surge in comments, and the conversation reaches its widest audience. However, the overall structural pattern has not changed. The findings also resonate with information cascade theory, which finds that opinions spread rapidly through networks when early participants successfully trigger a chain reaction of adoption (Bikhchandani et al., 1992). However, the results of this study indicate that this cascading effect is not infinite in forming the network structure but instead reaches a stable point before the topic peaks.

Early Majority, at the global network level, are structurally embedded within densely clustered subcommunities, meaning that they are highly interconnected with other users who share similar views and interests. Early Majority's involvement reinforces the original topic rather than branching into new directions. These contributions intensify the volume of discussion, but the content remains echoes of Early Adopters. This finding aligns with prior research on network homophily (McPherson, Smith-Lovin, & Cook, 2001), which suggests that individuals with shared interests are more likely to form tightly-knit clusters, leading to the reinforcement of pre-existing narratives. At the local network level, the Early Majority appears to be more peripheral, with a longer distance from central nodes. While they actively engage with the topic, they do not interact as frequently with high-degree Hubs as the Early Adopters do. Additionally, the Early Majority sustains discourse within existing communities rather than expanding the networks. The structure suggests that the Early Majority tends to repeat content within their tightly knit subgroups. In social media discourse, echo chambers emerge when users interact predominantly within their own ideological or interest-based communities, leading to selective exposure and reinforcement of pre-existing beliefs (Sunstein, 2001). Their reinforcement of existing narratives may contribute to the formation of echo chambers.

As the discussion becomes saturated, a decline phase sets in. The conversation might persist among a few enthusiasts, but for most participants, attention moves elsewhere. This stage resembles the "late majority" or laggard phase in diffusion models, where momentum fades (Rogers, 1962). The temporal dynamics are consistent with general models of information spread and opinion formation in previous research that ideas often rise quickly when triggered, then plateau and diminish rather than diffusing at a steady rate (Watts & Dodds, 2007). This sequenced network structure shows that online opinion change is not instantaneous but unfolds in patterned stages shaped by how different users participate over time.

6.3 Visibility within Reddit

Visibility determines which posts and comments attract community attention and subsequent engagement. Visibility on Reddit is heavily influenced by the voting mechanisms, sorting algorithms, and community structures. Reddit's default sorting algorithm strongly favors early upvotes. The first few upvotes on a post or comment are disproportionately influential in boosting its visibility. Thus, timing and immediate reaction are important for content visibility. This dynamic creates a feedback loop: early visibility begets more engagement (upvotes, replies), which in turn sustains visibility. This design can promote rapid surges of attention, but it also means that the visibility (and thus influence) is typically transient. Overall, visibility must be continually earned and is easily lost.

The platform's community-driven ethos further shapes how posts and users gain or lose attention. On Twitter, Facebook, or Instagram, users have followers and personal visibility over the long term, while Reddit emphasizes content over personal profiles. What truly drives visibility is how a user's content contribution resonates within a specific subreddit at a specific time. At the same time, without a permanent following, they must continually re-establish their visibility through new contributions. I observed that some Early Adopters maintained visibility by repeatedly initiating timely discussions across different threads. However, even these active users were subject to the rhythms of the platform: when discussion of a topic died down, their influence naturally faded until the next opportunity arose. In terms of losing attention, negative feedback is also a key factor. Reddit's algorithms penalize posts with a lot of downvotes relative to upvotes, which drop the visibility of controversial or unpopular opinions. In this case, polarization and information cocoons are easily formed, where users who hold minority or dissenting views are systematically silenced, reinforcing the dominance of prevailing narratives. Exposure to opposing views on social media can, paradoxically, reinforce existing biases, further entrenching users within their ideological bubbles (Bail et al., 2018).

6.4 Limitations and Future Research

There are some limitations in the methodology and data. First, the social network edge data did not include timestamps, resulting in the inability to observe the network structure in continuous time. The local network time phases are assigned by selected comments published time but not every reply, which may have a bias in the network formation structure. Second, due to the large volume of data, I included only threads with more than five comments to focus on substantive discussions. While this filtering improved manageability, it caused inconsistency between the datasets used for topic modeling and for social network analysis. Some posts and comments, especially from smaller threads, were excluded from the network dataset. Consequently, certain topics identified in the content analysis might not be fully represented in the network analysis.

The global network plot shows that there are several nodes that may play the role of brokers. An important next step is to explore the role of brokers or bridging users who connect otherwise disconnected communities. Future work could identify users who act as bridges, for example, by posting content, sharing links, or actively participating in multiple subreddits, to examine how ideas travel between subreddits. In network theory, such individuals are described as spanning structural holes, the gaps between unconnected groups, and serving a brokerage function that facilitates new information flow (Burt, 2004). Empirical evidence in social media contexts shows that a small fraction of these brokers can drive a disproportionate share of information transfer between communities on one large network (Ganley & Lampe, 2009; Stovel & Shaw, 2012; Z. Yang et al., 2023). Another promising direction is to incorporate the temporal dimension into the analysis of Reddit discourse. Given that our network data lacked timestamps, future studies could collect interaction data with time for each edge to look into more nuanced changed in small time segment or conduct longitudinal network analysis.

REFERENCE

Alazazi, M. (2023). *Themes and Participants' Role in Online Health Discussion: Evidence From Reddit* [Ph.D.]. https://www.proquest.com/docview/2842764431/abstract/A4FA87AC6A39493APQ/1

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

BANDURA, A. (2008). SOCIAL COGNITIVE THEORY OF MASS COMMUNICATION. In *Media Effects* (3rd ed.). Routledge.

Bedi, P., & Sharma, C. (2016). Community detection in social networks. *WIREs Data Mining and Knowledge Discovery*, *6*(3), 115–135. https://doi.org/10.1002/widm.1178

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026. https://doi.org/10.1086/261849

Boyd, D. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *A Networked Self*. Routledge.

Brass, D. J. (2022). New Developments in Social Network Analysis. *Annual Review of Organizational Psychology and Organizational Behavior*, *9*(Volume 9, 2022), 225–246. https://doi.org/10.1146/annurev-orgpsych-012420-090628

Bruns, A., & Burgess, J. (2020). #Ausvotes: How twitter covered the 2010 Australian federal election. *Communication, Politics & Culture*, *44*(2), 37–56. https://doi.org/10.3316/ielapa.627330171744964

Burt, R. S. (1976). Positions in Networks*. *Social Forces*, *55*(1), 93–122. https://doi.org/10.1093/sf/55.1.93

Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review*, *32*(1_suppl), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Cau, E., Morini, V., & Rossetti, G. (2024). Trends and topics: Characterizing echo chambers' topological stability and in-group attitudes. *PLOS Complex Systems*, *1*(2), e0000008. https://doi.org/10.1371/journal.pcsy.0000008

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, *32*(4), 556–577. https://doi.org/10.1002/sim.5408

Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(Volume 55, 2004), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Comfort, S. E., & Park, Y. E. (2018). On the Field of Environmental Communication: A Systematic Review of the Peer-Reviewed Literature. *Environmental Communication*, *12*(7), 862–875. https://doi.org/10.1080/17524032.2018.1514315

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, *9*(1), 24. https://doi.org/10.1007/s13278-019-0568-8

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Ganley, D., & Lampe, C. (2009). The ties that bind: Social network principles in online communities. *Decision Support Systems*, *47*(3), 266–274. https://doi.org/10.1016/j.dss.2009.02.013

Gerard, H. B., & Orive, R. (1987). The Dynamics of Opinion Formation. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 20, pp. 171–202). Academic Press. https://doi.org/10.1016/S0065-2601(08)60414-1

Glass, C. A., & Glass, D. H. (2021). Social Influence of Competing Groups and Leaders in Opinion Dynamics. *Computational Economics*, *58*(3), 799–823. https://doi.org/10.1007/s10614-020-10049-7

González-Bailón, S., Borge-Holthoefer, J., & Moreno, Y. (2013). Broadcasters and Hidden Influentials in Online Protest Diffusion. *American Behavioral Scientist*, *57*(7), 943–965. https://doi.org/10.1177/0002764213479371

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (No. arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Gulati, R. (1995). Social Structure and Alliance Formation Patterns: A Longitudinal Analysis. *Administrative Science Quarterly*, *40*(4), 619–652. https://doi.org/10.2307/2393756

Huang, R. (2019). Network fields, cultural identities and labor rights communities: Big data analytics with topic model and community detection. *Chinese Journal of Sociology*, *5*(1), 3–28. https://doi.org/10.1177/2057150X18820500

Huffaker, D. (2010). Dimensions of Leadership and Social Influence in Online Communities. *Human Communication Research*, *36*(4), 593–617. https://doi.org/10.1111/j.1468-2958.2010.01390.x

Jacobson, S., Myung, E., & Johnson, S. L. (2016). Open media or echo chamber: The use of links in audience discussions on the Facebook Pages of partisan news organizations. *Information, Communication & Society*, *19*(7), 875–891. https://doi.org/10.1080/1369118X.2015.1064461

Jang, S. H., Kim, D., Zhu, Y., & Kim, C. (2024). Stalking Discourse on Reddit: A Comparative Analysis of Pre- and Post-COVID-19 Pandemic Using Big Data. *Cyberpsychology, Behavior, and Social Networking*, *27*(8), 539–549. https://doi.org/10.1089/cyber.2023.0397

Jaradat, S., & Matskin, M. (2019). On Dynamic Topic Models for Mining Social Media. In N. Agarwal, N. Dokoohaki, & S. Tokdemir (Eds.), *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining* (pp. 209–230). Springer International Publishing. https://doi.org/10.1007/978-3-319-94105-9_8

Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford University Press.

Karell, D., Linke, A., Holland, E., & Hendrickson, E. (2023). “Born for a Storm”: Hard-Right Social Media and Civil Unrest. *American Sociological Review*, *88*(2), 322–349. https://doi.org/10.1177/00031224231156190

Katona, Z., Zubcsek, P. P., & Sarvary, M. (2011). Network Effects and Personal Influences: The Diffusion of an Online Social Network. *Journal of Marketing Research*, *48*(3), 425–443. https://doi.org/10.1509/jmkr.48.3.425

Katz, E., Lazarsfeld, P. F., & Roper, E. (2017). *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Routledge. https://doi.org/10.4324/9781315126234

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of Conflict Resolution*, *2*(1), 51–60. https://doi.org/10.1177/002200275800200106

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, *6*(11), 888–893. https://doi.org/10.1038/nphys1746

Kozitsin, I. V. (2023). Opinion dynamics of online social network users: A micro-level analysis. *The Journal of Mathematical Sociology*, *47*(1), 1–41. https://doi.org/10.1080/0022250X.2021.1956917

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Lee, G. M., Qiu, L., & Whinston, A. B. (2016). A Friend Like Me: Modeling Network Formation in a Location-Based Social Network. *Journal of Management Information Systems*, *33*(4), 1008–1033. https://doi.org/10.1080/07421222.2016.1267523

Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., & Ramanathan, V. (2019). Effects of fossil fuel and total anthropogenic emission removal on public health and climate. *Proceedings of the National Academy of Sciences*, *116*(15), 7192–7197. https://doi.org/10.1073/pnas.1819989116

Liang, X., Su, L. Y.-F., Yeo, S. K., Scheufele, D. A., Brossard, D., Xenos, M., Nealey, P., & Corley, E. A. (2014). Building Buzz: (Scientists) Communicating Science in New Media Environments. *Journalism & Mass Communication Quarterly*, *91*(4), 772–791. https://doi.org/10.1177/1077699014550092

Liyanage, C., Elhag, T., Ballal, T., & Li, Q. (2009). Knowledge communication and translation – a knowledge transfer model. *Journal of Knowledge Management*, *13*(3), 118–131. https://doi.org/10.1108/13673270910962914

Lynn, T., Muzellec, L., Caemmerer, B., & Turley, D. (2017). Social network sites: Early adopters' personality and influence. *Journal of Product & Brand Management*, *26*(1), 42–51. https://doi.org/10.1108/JPBM-10-2015-1025

Marsden, P. V., & Friedkin, N. E. (1993). Network Studies of Social Influence. *Sociological Methods & Research*, *22*(1), 127–151. https://doi.org/10.1177/0049124193022001006

Mueller, S. T., & Tan, Y.-Y. S. (2018). Cognitive perspectives on opinion dynamics: The role of knowledge in consensus formation, opinion divergence, and group polarization. *Journal of Computational Social Science*, *1*(1), 15–48. https://doi.org/10.1007/s42001-017-0004-7

Nisbet, M. C., & Scheufele, D. A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany*, *96*(10), 1767–1778. https://doi.org/10.3732/ajb.0900041

O'Connor, A. (2021). Media and translation. In E. Bielsa, *The Routledge Handbook of Translation and Media* (1st ed., pp. 13–27). Routledge. https://doi.org/10.4324/9781003221678-3

Pai, S. J., Carter, T. S., Heald, C. L., & Kroll, J. H. (2022). Updated World Health Organization Air Quality Guidelines Highlight the Importance of Non-anthropogenic PM2.5. *Environmental Science & Technology Letters*, *9*(6), 501–506. https://doi.org/10.1021/acs.estlett.2c00203

Parsa, M. S., Shi, H., Xu, Y., Yim, A., Yin, Y., & Golab, L. (2022). Analyzing Climate Change Discussions on Reddit. *2022 International Conference on Computational*

Science and Computational Intelligence (CSCI), 826–832. https://doi.org/10.1109/CSCI58124.2022.00150

Rickenbacker, H., Brown, F., & Bilec, M. (2019). Creating environmental consciousness in underserved communities: Implementation and outcomes of community-based environmental justice and air pollution research. *Sustainable Cities and Society*, *47*, 101473. https://doi.org/10.1016/j.scs.2019.101473

Riquelme, F., Gonzalez-Cantergiani, P., Hans, D., Villarroel, R., & Munoz, R. (2019). Identifying Opinion Leaders on Social Networks Through Milestones Definition. *IEEE Access*, *7*, 75670–75677. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2922155

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2008). Diffusion of Innovations. In *An Integrated Approach to Communication Theory and Research* (2nd ed.). Routledge.

Rohde, J. A., Liu, J., & Rees, V. W. (2023). Community and Opinion Leadership Effects on Vaping Discourse: A Network Analysis of Online Reddit Threads. *Journal of Health Communication*, *28*(8), 487–497. https://doi.org/10.1080/10810730.2023.2225447

Rong, X., & Mei, Q. (2013). Diffusion of innovations revisited: From social network to innovation network. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, 499–508. https://doi.org/10.1145/2505515.2505587

Rowley, T. J. (1997). Moving Beyond Dyadic Ties: A Network Theory of Stakeholder Influences. *Academy of Management Review*. https://doi.org/10.5465/amr.1997.9711022107

Seebauer, S. (2015). Why early adopters engage in interpersonal diffusion of technological innovations: An empirical study on electric bicycles and electric scooters. *Transportation Research Part A: Policy and Practice*, *78*, 146–160. https://doi.org/10.1016/j.tra.2015.04.017

Sobkowicz, P., Kaschesky, M., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, *29*(4), 470–479. https://doi.org/10.1016/j.giq.2012.06.005

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001

Stilgoe, J., Lock, S. J., & Wilsdon, J. (2014). Why should we promote public engagement with science? *Public Understanding of Science*, *23*(1), 4–15. https://doi.org/10.1177/0963662513518154

Stovel, K., & Shaw, L. (2012). Brokerage. *Annual Review of Sociology*, *38*(Volume 38, 2012), 139–158. https://doi.org/10.1146/annurev-soc-081309-150054

Sziklai, B. R., & Lengyel, B. (2022). Finding early adopters of innovation in social networks. *Social Network Analysis and Mining*, *13*(1), 4. https://doi.org/10.1007/s13278-022-01012-5

Valente, T. W. (2010). *Social Networks and Health: Models, Methods, and Applications*. Oxford University Press.

Vu, H. T., Do, H. V., Seo, H., & Liu, Y. (2020). Who Leads the Conversation on Climate Change?: A Study of a Global Network of NGOs on Twitter. *Environmental Communication*, *14*(4), 450–464. https://doi.org/10.1080/17524032.2019.1687099

Wæraas, A., & Nielsen, J. A. (2016). Translation Theory 'Translated': Three Perspectives on Translation in Organizational Research. *International Journal of Management Reviews*, *18*(3), 236–270. https://doi.org/10.1111/ijmr.12092

Watts, D. J., & Dodds, P. S. (2007). Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, *34*(4), 441–458. https://doi.org/10.1086/518527

Williams, H. T. P., McMurray, J. R., Kurz, T., & Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, *32*, 126–138. https://doi.org/10.1016/j.gloenvcha.2015.03.006

Yang, L., Qiao, Y., Liu, Z., Ma, J., & Li, X. (2018). Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm. *Soft Computing*, *22*(2), 453–464. https://doi.org/10.1007/s00500-016-2335-3

Yang, Z., Yang, C., Lu, C., Wang, F., & Zhou, W. (2023). Diffusion between groups: The influence of social brokers on content adoption in social networks. *European Journal of Marketing*, *57*(4), 1039–1067. https://doi.org/10.1108/EJM-11-2020-0811

APPENDIX

Figure 8. [Heatmap of All Comments Topic ×Subreddit]

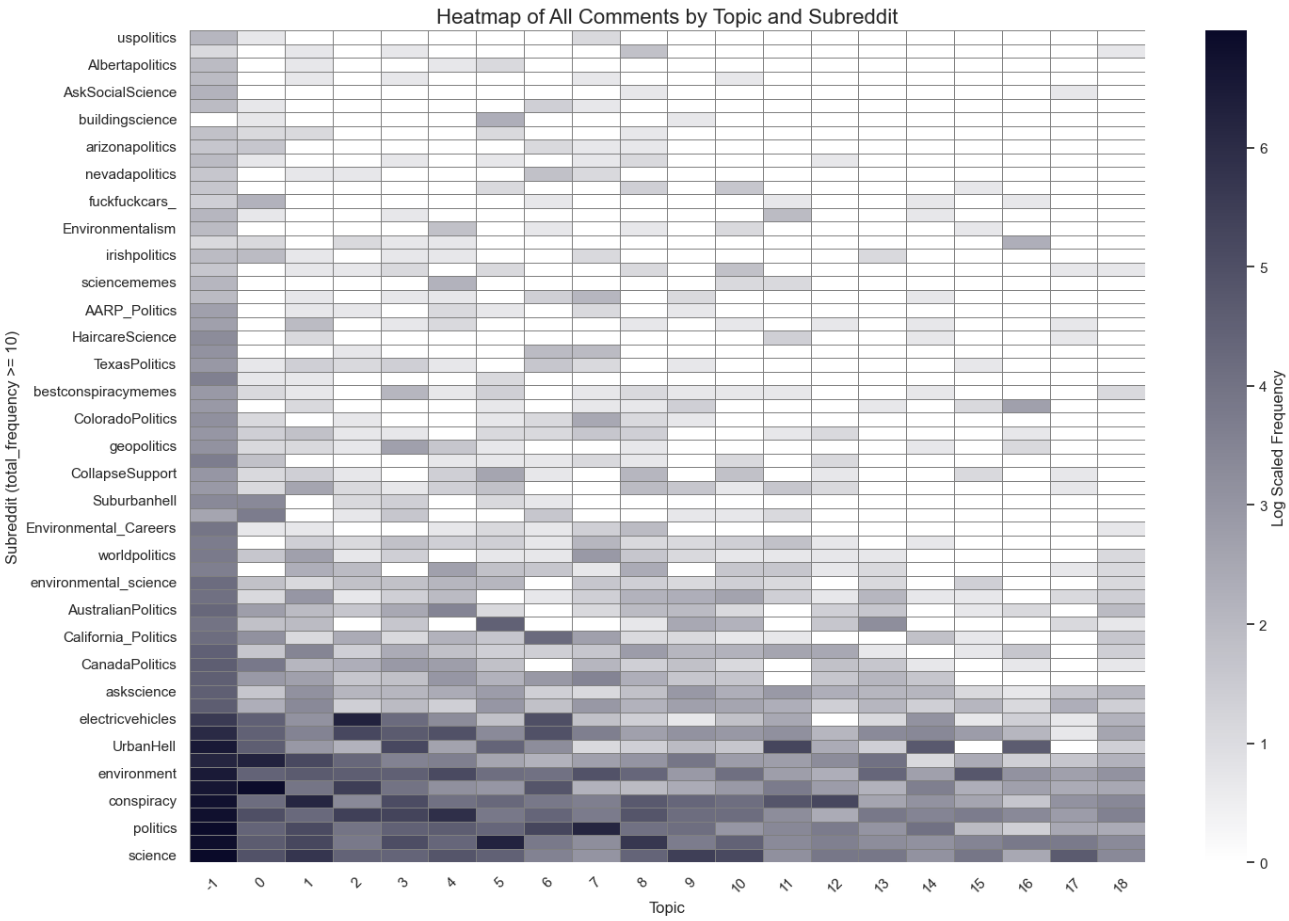


Figure 9. [Hubs' Comments Distribution]