Investigating Intersubjective Realities From Novel NLP and Chaos Theory Approach

Camila Carreon

Santa Fe Preparatory School

Santa Fe, NM

April 2025

Contents

1 Executive Summary

When does information become important, and how do sentiments gain traction and turn into beliefs, collective and revisionist histories, principles, and ideologies? As wars of human rights and contrasting beliefs and values rage across the Earth, such questions are incredibly important. While the scope of my research is incapable of ending these global conflicts, I start at a smaller scale, analyzing social media data and discourse during the Covid-19 Pandemic taking language as the currency of information to analyze stability and structure. Specifically, I use Natural Language Processing and techniques borrowed from the mathematical field of Chaos Theory to explore what, as historian and scholar Yuval Noah Harari defines as "intersubjective realities," or a "shared, mutual understanding between individuals."[Har24] Ultimately, these realities are the hotbed of shared beliefs, stories, and, at an enlarged scale, ideology, and thus important. Furthermore, this research acts as a form of counterterrorism, looking for patterns in language and structure within these intersubjective realities to assess their potential for becoming influential and dangerous (becoming conspiracy theories).

2 Statement of the Problem

The goals of my research are twofold, first to investigate an issue that is often left out of the quantitative limelight: understanding narrative development and structure, and second addressing rising concerns of misinformed rumors and conspiracy theories on social media platforms like X (Twitter) and Reddit. A time period particularly ripe for such exploration is the Covid-19 pandemic, which in an immense time of uncertainty, as Francesco Farinelli of the Radicalisation Awareness Network (RAN) reports, "Conspiratorial narratives flourish[ed] in such a context and extremist groups exploited the spread of the coronavirus to disseminate fake news and to incite violence."[Far21] We define conspiratorial and uncharged yet popular narratives around the Covid-19 pandemic to be intersubjective realities, which become real once many people believe in it or begin to act in compliance with these narratives, or as the quotation from the RAN suggests, extremist violence in anti-government, anti-establishment, anti-lockdown and anti-restriction protests or AGAAVE—Anti-Government, Anti-Authority Violent Extremism. By identifying the linguistic and structural patterns of conspiratorial narratives we may be able to predict when and how they spiral into harmful ideologies. In other words, this work is not just about studying misinformation—it's about recognizing when belief manipulation turns into a societal threat. The Counter-Terrorism Committee Executive Directorate (CTED) in a 2020 survey reports that, "69% of respondents stated that countering terrorism has become more challenging as a result of the pandemic"[Dir21]. Thus to address such concerns, my work aligns with modern counterterrorism efforts, which in the US as shown here "promote US National security by developing coordinated strategies and approaches to defeat terrorism abroad and secure the counterterrorism cooperation of international partners." My work will be one such strategy and approach.

3 Introduction

The COVID-19 pandemic was not just a biological crisis but also an informational epidemic, where rumors and conflicting narratives shaped public perception. These narratives are part of what scholar Yuval Noah Harari calls intersubjective realities—shared beliefs that exist only in the human mind but are given power through collective belief, or recognition by at least 2 or more people. Yet the

propagation of such narratives reached unprecedented heights, when millions stuck at home resorted to screens and leveraged digital technologies and platforms like social media to stay connected to the rest of the world while quarantined. For example, the amount of Facebook users went up to about 1.9 billion worldwide by the end of 2020, marking an 8.7% increase over 2019[Wil]. And other social media platforms saw similar surges in popularity. Thus to understand the nature of pandemic-related discourse I turn to social media data (Twitter or X), and specifically use the COVID-19 Rumors dataset [Che+21], focusing on how misinformation and competing narratives form and evolve over time. Using Natural Language Processing (NLP), we track how rumors behave within the framework of chaos theory. Unlike static ideological structures, digital discourse is inherently nonlinear, meaning that small fluctuations—such as a single viral tweet—can unpredictably alter the trajectory of public perception. Online discourse, especially on platforms like Twitter, exhibits complex, nonlinear dynamics, making it an ideal system to analyze through chaos theory. The chaotic spread of misinformation, the convergence and divergence of narratives, and their unpredictable shifts mirror the properties of dynamical systems, where small changes in initial conditions can lead to vastly different outcomes. By mapping pandemic-related sentiments in a phase space, we analyze how belief-driven discussions shift, stabilize, or fragment into new patterns—much like turbulent flows in physical systems. These patterns of belief formation are a manifestation of intersubjective realities, where shared ideas constructed through collective engagement shape the social fabric of digital communication. To quantify these shifts, we construct recurrence networks and use symbolic dynamics to assess how belief transitions (support, deny, neutral) emerge, stabilize, or spiral into chaos. Measuring these transitions through entropy helps us understand the unpredictability of discourse and the moments when narratives undergo sudden, irreversible transformations. This study is grounded in the Three V's framework (Lukić), which defines the chaotic nature of online information networks: These properties make digital discourse highly chaotic and sensitive to initial conditions, meaning that small perturbations can lead to significant, unpredictable shifts in dominant narratives. By bridging chaos theory with computational discourse analysis, this research provides a novel framework for understanding the spread of misinformation, uncovering the underlying patterns that govern belief formation in the digital age. Through time series modeling and network-based methods, we map the evolution of online narratives, shedding light on the turbulent

Table 1: Three V's

V	Description
Volume	The vast scale of pandemic-related tweets, reflecting the sheer magnitude of discourse.
Velocity	The rapid rate at which narratives emerge, spread, and transform in real time.
Variety	The diversity of content and sentiment, encompassing conflicting perspectives, emotions, and misinformation.

flow of information and its impact on public perception during the pandemic.

4 Background

4.1 Intersubjective Realities

The swift entrance of Gutenberg's Printing Press in 1440 galvanized the rest of the Afro-Eurasian world into an unprecedented era where information and knowledge became public currency. Information, as scholars often term it, was democratized. Quickly the masses embraced Gutenberg's 42-line Bibles and even flocked to other printed works. One such text that took the medieval world by storm in the 15th century was Malleus Maleficarum, or Hammer of Witches, written by Heinrich Kramer in 1485, and a witch-hunting guide that warned witches were part of a satanic-led campaign to destroy humanity. The books popularity reflected in its immediate impact, as the narrative sparked outrage and violent responses for many in fear of their safety. Preaching to enraptured Alpine peasants, the book extends from Exodus 22:18 that, "You shall not permit a sorceress to live."[Enc]

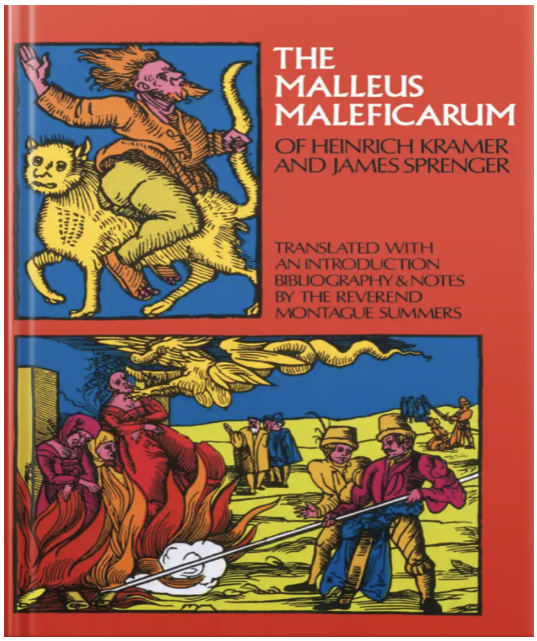

Figure 1: Malleus Maleficarum

The books dual biblical nature and call to legal action, turned credence to violence and mass hysteria. It is estimated that between 40,000 to 60,000 were killed after being tried and accused of witchcraft as a result. Yuval Noah Harari in his book Nexus: A Brief History of Information Networks from the Stone Age to AI considers the European witch craze an intersubjective reality, writing, "But witches became an intersubjective reality. Like money, witches were made real by exchanging information about witches."[Har24] Harari borrows from the philosophical lexicon of Edmund Husserl (1859 - 1938) who describes intersubjectivity as "the interchange of thoughts and feelings, both conscious and unconscious, between two persons or 'subjects,' as facilitated by empathy." [**knuthwebsite**]Husserl introduces intersubjectivity as part of a conceptual hierarchy of realities (objective, subjective and intersubjective) shown below:

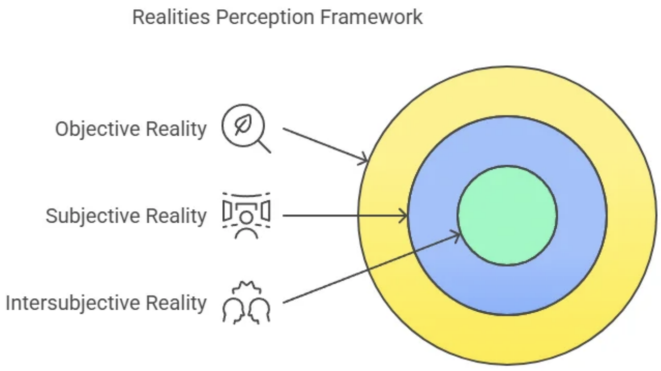

Figure 2: Realities Perception Framework

Harari injects this definition into a modern analysis of society, where he views intersubjectivity as shared fictional realities and social constructs upon which society depends like money, nations, and human rights, which only become real when enough people believe in them and have considerable influence. The European Witch Craze of the late 15th century becomes an intersubjective reality, when mass belief and hysteria overtakes much of Europe and results in extensive corporal loss.

Harari's thesis, as professor of law at Northeastern University Beth Simone Noveck puts it "is profound, albeit obvious: technology is inherently political, shaped by those who wield it, and often reinforces existing power structures."[Nov] In other words, the influence of intersubjective realities is also determined by who and how it is propagated, which as Harari illustrates have become increasingly less human. From the social media algorithms that largely dictate the flow of information in social media platforms and more recently AI, which "increasingly determines what we read about, think about, and talk about"[Nov], studying intersubjective realities in the digital age or age of information introduces a new layer into the research-algorithmic radicalization or the idea that as the Observer Research Foundation expands "Algorithms usually promote emotionally provocative or controversial material by focusing on metrics such as likes and shares, creating feedback loops that amplify polarising narratives."[Awa] For this project this means the structures of discourse and emergence of narratives we discover will be more pronounced than ever, and thus riper for analysis.

4.2 Chaos Theory as a Novel Way to Understand at Linguistic Intersubjective Realities in the time of Social Media

Chaos theory, often associated with physical systems, is a mathematical framework used to describe complex, dynamic systems that are highly sensitive to initial conditions. This sensitivity posits that small changes in the starting conditions of a system can lead to vastly different outcomes over time. The classic paradigm of these sensitivities is the culturally famous "Butterfly Effect" which imagines how the flap of a butterflies wings can cause a hurricane on the other side of the world. In the context of social media and online discourse, this concept is particularly relevant, as seemingly minor posts or viral tweets can trigger widespread shifts in public perception and belief systems.

In the digital age, platforms like Twitter, Facebook, and Reddit serve as a breeding ground for the formation of intersubjective realities—shared beliefs and narratives created through collective human engagement. As Yuval Noah Harari suggests, these intersubjective realities only become "real" when enough people collectively believe in them, regardless of their factual accuracy. The chaotic nature of online discourse, where rumors, misinformation, and competing narratives circulate rapidly, mirrors the principles of chaos theory.

The propagation of misinformation on social media is an inherently nonlinear process. A single viral post or tweet can escalate rapidly, becoming a global phenomenon, while other topics or narratives, despite similar initial traction, may dissipate into obscurity. Additionally, this process is buttressed by the advent of social media algorithms. Twitter's algorithm works by first "learn[ing] about users based on their clicks, likes, and responses. Then, it takes this information and turns it into outputs. In this situation, that information helps create the main 'For You' feed on the Twitter platform." [Tea]On March 31, 2023, Twitter became the first social media platform to release it's engagement formula which is shown in the figure below:

Type of engagement	Weight
Probability the user will like the tweet	0.5
Probability the user will retweet the tweet	1.0
Probability the user replies to the tweet	13.5
Probability the user opens the tweet author profile and likes or replies to a tweet	12.0
Probability (for a video tweet) that the user will watch at least half of the video	0.005
Probability the user replies to the tweet and this reply is engaged by the tweet author	75.0
Probability the user will click into the conversation of this tweet and reply or like a tweet	11.0
Probability the user will click into the conversation and stay there for at least 2 minutes	10.0
Probability the user will react negatively (requesting "show less often" on the tweet or author, block or mute the tweet author)	-74.0
Probability the user will click report tweet	-369.0

Figure 3: Twitter Engagement Formula

Chaotic systems, while they exhibit seemingly random and unpredictable behavior, are governed by deterministic laws. The butterfly and the hurricane both exist in a world governed by the same physical laws, and similarly the social networks on social media platforms are governed by the determinstic rules like the engagement formula above, but still characterized by unpredictability and randomness given it's sheer variety, velocity and volume as per Lukic et al.'s 3 V's Framework as shown in Table 1. This unpredictability in the evolution of online discussions can be modeled using chaotic systems, where the smallest changes can lead to divergent or turbulent outcomes.

Furthermore, by applying chaos theory to linguistic data, we aim to uncover patterns in the way rumors and misinformation spread through online platforms. Similar to the behavior of dynamical systems, these patterns of discourse exhibit critical points where narratives transition from stability to instability, where once solid beliefs splinter into competing factions or where a seemingly benign topic may spiral into a full-blown conspiracy theory. Using techniques from chaos theory, such as recurrence analysis and symbolic dynamics, this study explores how shifts in linguistic patterns and sentiment lead to the formation, consolidation, or collapse of intersubjective realities within the

digital realm.

In summary, chaos theory provides a valuable lens through which to understand the unpredictable and highly sensitive nature of belief systems on social media. It allows us to explore how small fluctuations in discourse—such as the introduction of a new rumor or a change in sentiment—can lead to large-scale shifts in the collective psyche, highlighting the fragile and volatile nature of digital ideologies.

4.3 Natural Language Processing and State of the Field

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and human language. NLP involves designing algorithms and models that enable machines to understand, interpret, and generate human language in a meaningful way. The complexity of human language, with its nuances, slang, and context-dependent meanings, makes NLP a challenging area of AI. However, NLP techniques have made significant advancements in recent years, allowing computers to process and analyze vast amounts of textual data effectively.

In the context of my project, NLP provides the tools necessary to handle the large-scale textual data generated on social media platforms during the COVID-19 pandemic. Social media platforms like Twitter contain millions of posts, tweets, and comments that express varying beliefs, opinions, and narratives about the pandemic. NLP techniques allow for the extraction of relevant insights from this vast and unstructured data, enabling a deeper understanding of the complex dynamics of belief formation and the spread of misinformation.

The primary objective of this research is to analyze the evolution of beliefs, rumors, and competing narratives surrounding the COVID-19 pandemic. Given the massive volume of data, as suggested in the 3 V's framework, involved and the unstructured nature of textual content on social media, manual analysis would be impractical. Thus, by using NLP, we can automatically process, categorize, and analyze textual data to extract meaningful insights that would otherwise be difficult to uncover.

For example, with the help of NLP techniques, we can detect topics that are being discussed, track how those topics evolve over time, and observe how beliefs and ideologies form, stabilize, or fragment in response to external events. The ability to process large datasets with NLP makes it

possible to study the dynamics of public discourse at scale, providing insights into the spread of misinformation, the formation of intersubjective realities, and the relationship between language and belief systems.

4.3.1 State of the Field

Collectively these abilities of NLP, have brought it substantial popularity in research on conspiratorial narratives or other types of discourse. For instance Albladi et. al in their paper *Detection of Conspiracy vs. Critical Narratives and Their Elements using NLP*, use the BERT (Bidirectional Encoder Representations from Transformers) NLP model developed by Google, and the RoBERTa model which proposed in RoBERTa: A Robustly Optimized BERT Pretraining Approach by Yinhan Liu, Myle Ott to identify conspiracy theories[Ais24]. On the more analytical side of this kind of research, Haupt et.al in their paper, *Detecting nuance in conspiracy discourse: Advancing methods in infodemiology and communication science with machine learning and qualitative content coding*, use NLP and qualitative content coding to explore the 5G conspiracy popular during the Covid-19 Pandemic[**Haupt**]. While this is just a small snapshot into the breadth of scholarship on NLP in Thus I was given a wealth of avenues to use NLP in studying Twitter discourse and choose the following techniques and process.

4.3.2 Topic Modeling and Latent Dirichlet Allocation (LDA)

Topic modeling is an NLP technique used to discover the hidden thematic structure in a large collection of texts. The goal of topic modeling is to automatically identify the underlying topics that are present in a corpus of documents, without prior knowledge of what those topics might be. This is crucial for understanding the wide array of subjects being discussed on social media platforms, especially in a context as complex as the COVID-19 pandemic. Below is a graphic explaining the process of Topic Modeling:

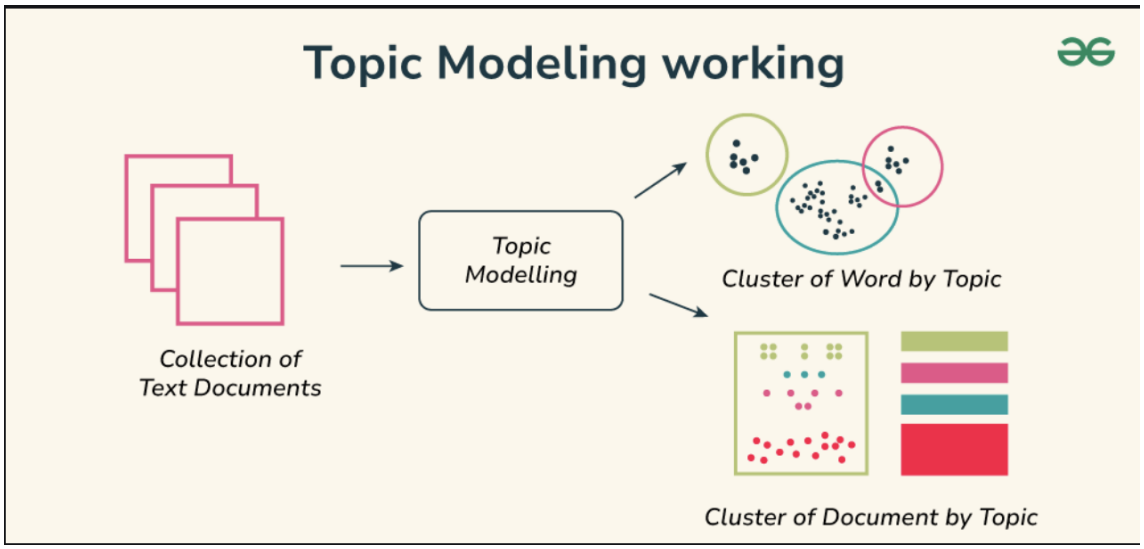

Figure 4: Topic Modeling

One of the most popular algorithms for topic modeling is Latent Dirichlet Allocation (LDA). LDA is a probabilistic model that assumes each document (in this case, a tweet or social media post) is a mixture of topics, and each topic is represented by a distribution over words. Furthermore, LDA generates topics by tracking frequency of co-occurrence (how likely is a word to appear with another word) and individual word frequency, using Gibbs equation to assign topics to words. Gibbs equation is shown below,

$$P(z_i = t \mid z^{-i}, w) = \frac{n_{m,t}^{-i} + \alpha}{\sum_{t'=1}^{T}(n_{m,t'}^{-i} + \alpha)} \times \frac{n_{t,w_i}^{-i} + \beta}{\sum_{v'=1}^{V}(n_{t,v'}^{-i} + \beta)}$$

Figure 5: Gibbs Equation

Where the first ratio is the probability of topic t in some corpus of text d, and the second ratio is the probability of the word w belonging to topic t. The model uses these distributions to identify

latent (hidden) topics in the corpus based on the co-occurrence of words. LDA assumes that there is a fixed number of topics in the collection and seeks to uncover the mixture of topics that best explains the observed word distributions in the data. [Jac]

In my research, LDA plays a critical role in identifying the various themes and narratives being discussed throughout the pandemic. By applying LDA to the COVID-19 Rumors dataset, I can extract a set of topics that represent key areas of discourse, such as "Government Response," "Health Measures," "Vaccines," and "Conspiracy Theories." Understanding these topics allows for a more structured analysis of how public perception evolves over time and how misinformation or competing narratives form and spread.

4.3.3 Preprocessing Techniques

Preprocessing is a critical step in any NLP pipeline. The raw text data collected from social media platforms is often noisy, inconsistent, and unstructured. Therefore, preprocessing techniques are applied to clean and prepare the text for further analysis. In the context of this project, the following preprocessing techniques were employed:

Tokenization: This process splits the text into individual words or smaller units (tokens), which are the basic building blocks for any subsequent analysis. For example, the sentence "Masks save lives" would be tokenized into ["Masks", "save", "lives"].

Lowercasing: To avoid distinguishing between words due to case sensitivity, all text was converted to lowercase. This ensures that words like "Mask" and "mask" are treated as the same word.

Removing Stopwords: Stopwords are common words (such as "the", "and", "is") that carry little meaning on their own and are often removed to reduce noise in the dataset. Removing stopwords helps to focus on the more meaningful words in the text.

Lemmatization: Lemmatization reduces words to their base or root form. For example, "running" is lemmatized to "run". This ensures that variations of a word are treated as the same entity, improving the consistency of the analysis.

Removing Non-Alphanumeric Characters: Social media posts often include special characters like punctuation, hashtags, and URLs. These were removed unless they were relevant to the topic modeling process (e.g., hashtags might indicate the topic of a tweet).

Stemming (if necessary): Although not applied in this project, stemming can also be used to reduce words to their stem form (e.g., "running" becomes "run"). However, lemmatization is generally preferred for its ability to handle words more intelligently.

The cleaned and preprocessed text data is then ready for analysis using NLP models, such as topic modeling or clustering algorithms. Effective preprocessing ensures that the data is structured in a way that allows for accurate insights into the dynamics of online discourse during the pandemic.

5 Methodology

5.1 Preprocessing and Data Collection

5.1.1 Data

Dataset Selection: The primary data source used in this study is the COVID-19 Rumors dataset [Che+21], which contains a rich collection of social media posts, primarily from Twitter, that discuss rumors and narratives related to the pandemic. The dataset includes both original posts and responses, accompanied by engagement metrics (likes, retweets, replies), making it ideal for analyzing social media discourse. There were 2,705 identified tweets and 34,847 retweets/comments associated with the posts. Additionally as shown in the images below the rumor dataset, twitter replies and posts had individual datasets and different accompanying metadata :

Twitter ID	release date	comment	time	replies	retweets	likes	timestamp	stance
1002962201143611	Sun Mar 29 2020	@nypost @JoshMan	Sun Mar 29 03:54:46	0	1	1	11:10.3	comment
1002962201143611	Sun Mar 29 2020	@nypost remember	Sun Mar 29 01:44:43	0	2	2	11:10.3	comment
1002962201143611	Sun Mar 29 2020	@nypost Hats off	Sun Mar 29 02:10:24	0	1	1	11:10.3	comment
1002962201143611	Sun Mar 29 2020	@nypost That's the	Sun Mar 29 01:43:38	0	1	1	11:10.3	comment
1002962201143611	Sun Mar 29 2020	@nypost lol, where a	Sun Mar 29 04:30:06	0	1	1	11:10.3	comment

Figure 6: Twitter Comments

label	content
F	If you can hold your breath without coughing, discomfort, stiffness, or tightness, your lungs do not suffer from fibrosis and therefore you have no COVID-19 infection
F	A homemade hand sanitizer made with Tito's Vodka can be used to fight the new coronavirus
F	Gargling with salt water or Vinegar 'eliminate' the COVID-19 coronavirus from the throat of an infected person's system
U	Patients should avoid taking ibuprofen to relieve pain and fever associated with COVID-19 infections
F	Chinese officials are seeking approval to start the mass killing of 20,000 people in order to stop the spread of new coronavirus

sentiment	reply numbers	retweet numbers	likes numbers
3	2	61	95
3	3	41	67
3	6	39	73
3	0	0	0
4	0	0	0
3	10	256	454

Figure 7: Twitter Posts

The column of metadata most important to me was the stance column. Stance as Cheng et.al define is, "The attitude of the author or editor of the rumor source. We follow classical rumor stance classification and define four classes of stance: support, deny, comment, and query. The stances are labeled and cross-validated manually by going through the context of each website."[Che+21] Other important pieces of metadata to my study are veracity, or labeling the posts/rumors as true(T) or false(F) or unverified(U), which were all manually labeled and cross-validated by referencing authoritative websites like Snopes , Politifact and Boomlive, and the reply, retweet and likes numbers for posts and comments. In their discussion of their dataset, Cheng et al. write,

> "We envision the downstream applications or usage cases of this dataset to include but not restricted to (i) the identification, prediction, classification of rumor, misinformation, disinformation, and fake news; (ii) the study of rumor spread trend and rumor/misinformation/disinformation/fake news combating and/or control; (iii) social network and complex network-related studies in terms of information flow and transition; and (iv) the natural language processingrelated studies of rumor sentiment and semantic."

[Che+21]

The scope of my research speaks to (ii) and (iii), as I am studying the development of rumors and narratives on twitter as a form of risk analysis.

Data Cleaning: In order to ensure the integrity of the analysis, several preprocessing steps were conducted:

Numeric Conversion: Engagement metrics, such as the number of likes, retweets, and replies, were converted into numeric values for easy analysis.

Text Cleaning: The text of each post was standardized by removing stopwords, correcting typographical errors, and eliminating irrelevant content such as URLs, special characters, and repeated symbols.

Topic Modeling: LDA was used to extract topics and classify each tweet into 10 different topics. Additionally I visualized topics with pyLDAvis. Below are the images of my python code for topic modeling using the Gensim library and the visualization outputted for Topic 0:

```
# Train LDA model
lda_model = gensim.models.LdaModel(corpus=corpus, id2word=id2word, num_topics=10,
    random_state=100)

# Extract topics
def get_lda_topics(model, num_topics, top_n_words):
    word_dict = {}
   for i in range(num_topics):
      word_dict[f"Topic # {i+1:02d}"] = [word[0] for word in model.show_topic(i, topn
    =top_n_words)]
  return pd.DataFrame(word_dict)

topics_df = get_lda_topics(lda_model, 10, 10)
print(topics_df)
# Function to assign the most likely topic to each tweet
def get_dominant_topic(text, id2word, lda_model):
  bow = id2word.doc2bow(text)  # Convert text to bag-of-words
  topic_probs = lda_model.get_document_topics(bow)  # Get topic probabilities
  return max(topic_probs, key=lambda x: x[1])[0] if topic_probs else 'Unknown'  #
    Return highest probability topic

# Assign topics to each tweet
tweets_df["topic"] = tweets_df["tokens"].apply(lambda x: get_dominant_topic(x,
    id2word, lda_model))
```

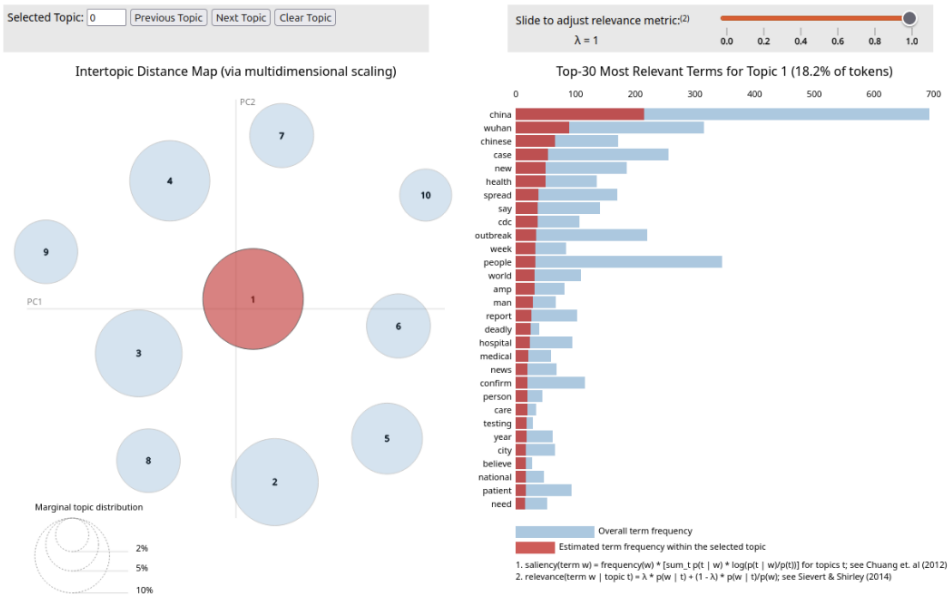

Figure 8: LDA Visualization for Topic 0

Topic Assignment: Topics for each post were assigned using TopicGPT from Pham et.al's *TopicGPT: A Prompt-based Topic Modeling Framework*. Using the generate topic lvl1 function I was able to generate high-level and generalizable topics. This allowed me to analyze how specific rumors and narratives evolved within certain thematic domains, such as government response, health measures, or public perceptions of the pandemic.[Pha+23]

Handling Missing Data: Missing or incomplete engagement data was handled by replacing NaN (Not a Number) values with zero, assuming no engagement occurred. This approach is standard for ensuring that incomplete posts don't disrupt the continuity of engagement analysis.

5.2 Natural Language Processing

Text Vectorization: To facilitate the computational analysis of text, I utilized NLP techniques such as tokenization, lemmatization, and vectorization. Each tweet was converted into a vector of word embeddings, using a pretrained word2vec model that encodes semantic relationships between words.

Stance Detection: The stance of each post was classified into one of three categories: Support, Deny, or Neutral. This classification was accomplished through a supervised machine learning model trained on labeled data. The model uses a variety of linguistic features, including sentiment polarity, word frequency, and topic-specific keywords, to assign a stance to each post.

5.3 Engagement Analysis over Time

Timestamp Processing: To analyze how discussions and engagement evolved over time, the timestamp of each tweet was extracted and standardized. Invalid timestamps were removed, and valid entries were aggregated by day and by topic, allowing for a temporal analysis of engagement metrics (likes, retweets, replies) for each topic.

Engagement Aggregation: Engagement metrics were aggregated at both the post and topic levels, considering interactions across the entire dataset. The number of retweets, replies, and likes for each post was summed daily, providing a time series of engagement for each topic. This step is critical for tracking shifts in public interest and engagement with different narratives over the course of the pandemic.

Stance Ratio Calculation: To examine the spread of conflicting beliefs, I calculated the ratio of support to denial stances for each topic. This ratio was smoothed using a 5-point rolling window to reduce the effect of outliers and to capture long-term trends in belief evolution. This metric gives insights into how narratives shift between support for and opposition to particular topics over time.

5.4 Recurrence Network Analysis

Distance Calculation: To construct a recurrence network, pairwise distances between smoothed stance ratios were computed using the pdist function from the scipy library. These distances were based on the Euclidean distance metric, which measures the similarity between two time points based on their stance ratios.

Recurrence Matrix Construction: A binary recurrence matrix was generated by applying a 10th percentile threshold to the distance values. This matrix indicates the recurrence of similar stance ratios, i.e., whether the discourse at a given time point is similar to that at another time point. This recurrence matrix serves as the foundation for network analysis.

Network Graph Creation: The recurrence matrix was converted into a network graph, where nodes represent time points, and edges represent the similarity (recurrence) between them. This network allows for the visualization of how discourse evolves over time and which time points are most similar to each other. Below is an image of the Network produced for topic 3:

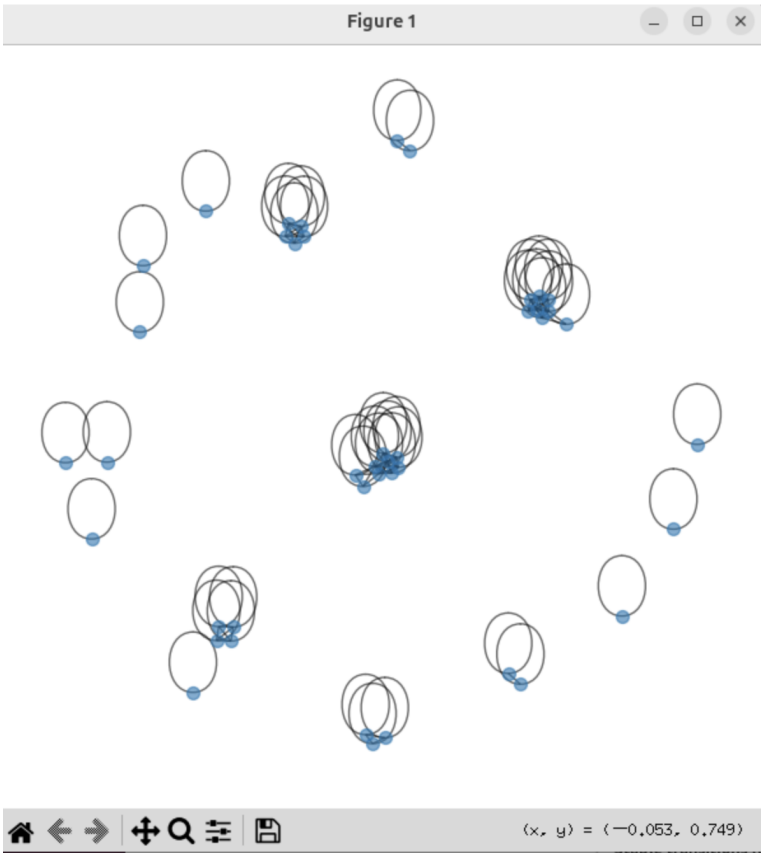

Figure 9: Network produced for Topic 3

Network Metrics Calculation: Several network metrics were calculated to assess the structure of the recurrence network. Specifically, the average clustering coefficient was computed to evaluate how interconnected neighboring nodes (time points) are. A higher clustering coefficient indicates a more tightly-knit network, which suggests that certain beliefs or narratives are more likely to spread within closed groups, fostering echo chambers.

5.5 Symbolic Dynamic Analysis

Discretization of Stance Data: To analyze stance transitions over time, the smoothed stance ratio was discretized into three symbolic categories: Deny, Neutral, and Support. This discretization was achieved by dividing the stance ratio data into quantiles, which allowed us to represent continuous stance changes as discrete states.

Transition Matrix Construction: The transitions between the symbolic states were encoded in a transition matrix, where each element represents the probability of transitioning from one state to another (e.g., from Deny to Support). The transition matrix was normalized to sum to 1 across each row, ensuring that the probabilities are valid. Below is an image of the transition matrix produced for topic 3:

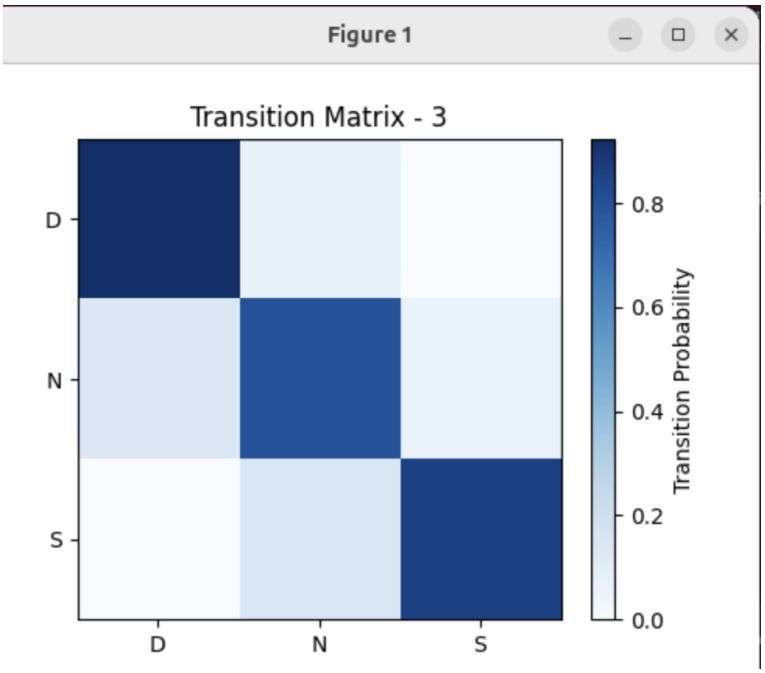

Figure 10: Transition Matrix for Topic 3

Entropy Calculation: To quantify the unpredictability of discourse transitions, I computed the entropy of the transition matrix. Entropy is a measure of randomness or disorder, and higher entropy values indicate more erratic and unpredictable shifts between stances over time. This measure

provides insight into how stable or unstable certain narratives are as they evolve. Below is an image of the calculations for entropy and clustering coefficient of transitions for topic 6:

```
Building recurrence network...
Average Clustering Coefficient: 0.7622
Performing symbolic dynamics analysis...
Transition Matrix:
          D         N         S
D  0.800000  0.133333  0.066667
N  0.217391  0.652174  0.130435
S  0.037037  0.185185  0.777778
Symbolic Dynamics Entropy: 3.0783
```

Figure 11: Calculations on Entropy and Clustering Coefficient for Topic 6

5.6 Social Network Analysis

5.6.1 Network

Network Construction For each of the ten topics, a directed network is constructed using pythons NetworkX library where nodes represent tweets (posts and comments), edges represent interactions between tweets, such as comments and retweets. Additionally, weights on edges correspond to engagement levels (sum of likes, replies, and retweets). Finally posts are assigned a node attribute (label) indicating their veracity, with values mapped as:

1 (True)

0 (Uncertain)

-1 (False)

5.6.2 Network Analysis

Furthermore, for each topic analysis is done using NetworkX's built in functions to calculate betweenness centrality, closeness centrality and degree centrality. The top 5 nodes (labeled with Twitter ID and respective measure of centrality) with the highest of each respective measure of centrality are shown. A description of each form of centrality analysis is featured in the table below and the result for all three forms of centrality analysis for topic 3 is pictured below:

Table 2: Three V's

Centrality Analysis	Description
Degree Centrality	Measures the number of direct connections a node has.
Closeness Centrality	Assesses how close a node is to all other nodes in the network.
Betweenness Centrality	Evaluates the extent to which a node lies on shortest paths between other nodes.

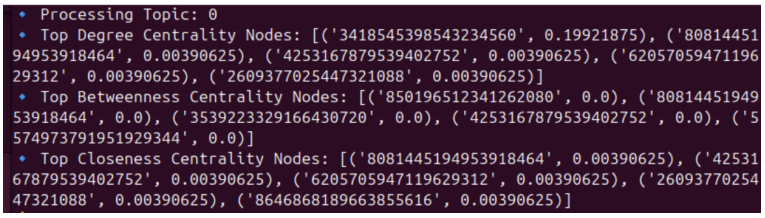

Figure 12: Calculations on Centrality for Topic 0

5.7 Veracity Analysis

Finally I associated each topic with a veracity score shown in the equation below:

$$V = \frac{T}{T + F} \tag{1}$$

Where V is the veracity score and T is the number of posts related to a topic with veracity labels of True (T) and F is the number of posts related to a topic with veracity labels of False (F).

6 Discussion

6.1 Conclusion

My analysis of conspiracy theory discussions on Twitter, leveraging recurrence networks and symbolic dynamics, reveals distinct structural and dynamical patterns in the spread of misinformation. I observed that topics with stronger local connectivity, such as Topic 8 (clustering coefficient = 0.7500) and Topic 6 (clustering coefficient = 0.7622), likely represent tightly-knit echo chambers

where individuals engage primarily within a closed community. These echo chambers may perpetuate misinformation by reinforcing existing beliefs. On the other hand, topics with more diffuse interactions, such as Topic 2 (clustering coefficient = 0.5474) and Topic 4 (clustering coefficient = 0.5714), may foster more open yet still biased discussions, facilitating the spread of conspiracy theories to a broader audience. The symbolic dynamics analysis, highlighting stance transitions, shows that denial (D) and support (S) are highly stable in most topics, contributing to the persistence of conspiratorial narratives. These states reinforce each other, while neutral (N) states exhibit more variability, suggesting that neutral stances play an intermediary role in discourse evolution. Topics with high entropy, like Topic 8 (entropy = 3.1357) and Topic 6 (entropy = 3.0783), show greater unpredictability and dynamic shifts in conversation, potentially fostering more fluid, yet still misinformed, discussions. Furthermore, the integration of veracity scores indicates a correlation between the structural properties of the network and the truthfulness of the discourse. Topics like Topic 3, which show the lowest veracity (0.1069) and high clustering, suggest that misinformation thrives in more rigid, tightly connected communities. Conversely, topics with higher veracity scores, such as Topic 5 (0.2238) and Topic 4 (0.2086), exhibit more openness in their discourse, indicating a greater potential for corrective discourse or less virulent misinformation spread. Additionally, in looking at centrality measures of the network of posts and comments, weighted by retweets and likes, I also looked at veracity. I calculated mean veracity for the top degree centrality, betweenness centrality and closeness centrality nodes. Topic 3 seems to show a neutral trend for degree centrality (mean veracity = 0.0), but betweenness and closeness centrality are slightly more negative, indicating that while many of the most connected nodes might not be spreading misinformation, those controlling the information flow or with closer connections are more likely to be involved in less reliable content.Topic 0 and Topic 9 show that higher degree centrality doesn't necessarily mean more reliable content, with negative mean veracity. Topic 6, on the other hand, shows higher mean veracity for closeness centrality nodes, which suggests that for this topic, nodes that are more central in the network tend to be more truthful.

Table 3: Labels of Topics

Topic	Manual Label
0	Public Health and Data
1	China's Response to Covid-19
2	Government and Official Response
3	Health Measures and Masks
4	Citizens, Infections, and Spread
5	Confirmed Cases and Deaths
6	Testing and Positive Cases
7	China's Fight Against the Outbreak
8	Medical Professionals and Time
9	Hygiene and Cure Efforts

6.2 Qualitative Contextualization

Contextualizing my conclusions within the manual labels assigned to each topic, particular attention should be given to topics 3 (Health Measures and Masks), 4 (Citizens, Infections, and Spread), 5 (Confirmed Cases and Deaths), and 6 (Testing and Positive Cases). My research suggests that Topic 3, with its low veracity and high clustering, is especially vulnerable to misinformation propagation. This aligns with the fact that mask-wearing became a highly polarizing issue during the Covid-19 pandemic, fueling divided beliefs. Conversely, topics related to pandemic statistics, such as Topics 4, 5, and 6, exhibited higher veracity, entropy, and more open structural dynamics, suggesting that discussions in these areas were more fluid and corrective, as factual claims and evidence are harder to dispute. Together, these findings highlight the complex dynamics of misinformation and its spread, demonstrating the chaotic nature of belief systems in the digital age. This research emphasizes the importance of understanding and addressing the forces shaping public discourse, with potential implications for improving communication strategies and combating misinformation in future crises

7 Achievements

My greatest achievement was learning how to perform NLP in python as well as be able to produce a great number of visualizations that succinctly and effectively presented my analysis.

8 Acknowledgments

I'd like to thank my sponsoring teachers Ms.Jocelyne Comstock for providing the space for me to work and help from Dr. Mark Galassi on this project.

9 Data Availability

The Covid-19 Rumors Dataset I used in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/MickeysClubhouse/COVII 19-rumor-dataset.

Works Cited

[Che+21] Mingxi Cheng et al. "A COVID-19 Rumor Dataset". In: *Frontiers in Psychology* 12 (2021), p. 1566.

[Dir21] Counter Terrorism Committee Executive Directorate. "Update on the impact of the COVID-19 pandemic on terrorism, counter-terrorism and countering violent extremism". In: United Nations Security Council, 2021.

[Far21] Francesco Farinelli. "Conspiracy theories and right-wing extremism – Insights and recommendations for P/CVE". In: *EU publications* (2021).

[Pha+23] Chau Minh Pham et al. "TopicGPT: A Prompt-based Topic Modeling Framework". In: *arXiv* (2023). eprint: `2311.01449` (cs.CL).

[Ais24] Albladi Cheryl D. Seals Aish Albladi. "Detection of Conspiracy vs. Critical Narratives and Their Elements using NLP". In: *Notebook for the Lab at CLEF 2024* (2024).

[Har24] Yuval Noah Harari. *Nexus: Nexus: A Brief History of Information Networks from the Stone Age to AI*. Penguin Random House, 2024. ISBN: 9783328603757.

[Awa] Soumya Awasthi. *From clicks to chaos: How social media algorithms amplify extremism.* URL: `https://www.orfonline.org/expert-speak/from-clicks-to-chaos-how-social-media-algorithms-amplify-extremism`.

[Enc] The Editors of Encyclopaedia Britannica. *Malleus maleficarum work by Kraemer and Sprenger*. URL: https://www.britannica.com/topic/Malleus-maleficarum.

[Jac] Eda Kavlakoglu Jacob Murel Ph.D. *What is Latent Dirichlet allocation ?* URL: https://www.ibm.com/think/topics/latent-dirichlet-allocation.

[Nov] Beth Simone Noveck. *The Dark Side of Progress: Harari's Grim AI Predictions in Nexus*. URL: https://rebootdemocracy.ai/blog/nexus.

[Tea] The QuickFrame Team. *How Does the Twitter (X) Algorithm Work in 2025*. URL: https://quickframe.com/blog/the-twitter-algorithm/#:~:text=Twitter's%20algorithm%20learns%20about%20users,feed%20on%20the%20Twitter%20platform..

[Wil] Debra Aho Williamson. *Global Facebook Users 2020: The Pandemic Brought Back Momentum in Lagging Regions and Led to Even Higher Growth in Others*. URL: https://www.emarketer.com/content/global-facebook-users-2020.