Keiser University - Latin American Campus

Mathematical Models for the Analysis of the Creation and Dissemination of Fake News on TikTok, Facebook, and Twitter in the Hispanic Community

Elisa Lucía Acosta Vásquez

Elisa Francheska Granizo Gadea

Alexamari Hawkins Martínez

Introduction

Fake news is a ubiquitous concept, encompassing misinformation, disinformation, and mal-information (Carson, 2021). It can be defined as false or fabricated or camouflaged information or rumor, either deliberate or accidental, that misleads people (Muigai, 2019). Fake news is not new. However, the phenomenal growth of online social media coupled with the ease of publishing unverifed content, and click-based advertisement revenue, have increased the use of fake news to drive discussions (Samanth 2017). According to Murayama et al. (2021), "As smartphones become widespread, people are increasingly seeking and consuming news from social media rather than from traditional media" (p. 1). However, it also appears to have developed into a hub for fake news, which could have detrimental effects on society.

In recent years, the spread of fake news on social media has grown exponentially, affecting public perceptions of social, political, and economic issues. This phenomenon, amplified by widespread access to platforms like Facebook, Twitter, and TikTok, has raised concerns due to the rapid influence these spaces can have on the opinions and behavior of users.

In the Spanish-speaking context, vulnerability to fake news has been heightened by the ease with which these contents are shared. According to a study conducted by the digital security company Kaspersky (2020), on average, 70% of Latin Americans do not know how to detect or are not sure how to recognize fake news from real news on the Internet.

The danger of disinformation caused by the phenomenon of fake news in mass media, such as social networks, highlights the need for accurate mathematical models to help analyze this problem with a view to mitigating its effects. Therefore, this paper seeks, from in-depth research and literature review, to find an appropriate mathematical model to describe and analyze the phenomenon of fake news in the Hispanic community. And from this model, analyze relevant data that shed light on this misinformation problem.

Problem Statement

In recent years, fake news has spread rapidly across social media, leading to confusion, misinformation, and, in some cases, harmful societal effects. The Hispanic community has been particularly vulnerable to false information on platforms like Facebook, Twitter, and TikTok. This project seeks to explain this phenomenon by employing accurate mathematical models supported by consistent scientific theories.

Research Questions

- Are there accurate mathematical models that describe the phenomenon of fake news spread by social networks in the Hispanic community?
- What are the key features or patterns of fake news that can be effectively identified using mathematical models in social media platforms popular among the Hispanic community?
- How can we apply models based on filtering theory to social phenomena such as fake news in the Hispanic community?

Justification

This study is important because the spread of fake news can influence crucial decisions on social, economic, and political issues. Understanding how fake news is disseminated and how the Hispanic community on social media is impacted will allow for the development of strategies to mitigate its negative effects, promote digital literacy, and strengthen trust in legitimate media outlets.

Objectives

General Objective:

To determine the existence of accurate mathematical models describing the phenomenon of fake news spread by TikTok, Facebook, and Twitter in the Hispanic community.

Specific Objectives:

- To analyze the factors influencing the spread of fake news on TikTok, Facebook, and Twitter within the Hispanic community using mathematical models
- To identify key patterns and features of fake news that can be modeled mathematically, leading to a more accurate representation of this phenomenon within the Hispanic community on social media.
- Analyze specific fake news shared by the Hispanic community in social networks using accurate mathematical models.

Background Information

The phenomenon of fake news has existed for centuries, but the digital age and the rise of social media have greatly increased its reach. Today, users can produce and share information with ease, allowing for the rapid circulation of news—whether true or false. This is of particular concern in environments such as the Hispanic community, where 70% of Hispanic social network users have stated that they do not know how to identify when a news item is false or not (Kaspersky, 2020). This causes a chain of misinformation that expands in social, political, and healthcare contexts.

Hypothesis

Mathematical models based on the filtering theory that describes the phenomenon as a relationship between noise and signal can help determine key elements of the fake news shared by the Hispanic community on Facebook, TikTok, and Twitter.

Methodology

Research Design.

A quantitative methodology will be applied. The research will be exploratory and descriptive, seeking to determine key factors of the phenomenon of fake news on social networking platforms such as Facebook, Twitter, and TikTok within the Hispanic community. The objective is to apply mathematical models based on filtering theory to describe and analyze the phenomenon.

Literature review.

A literature review will be conducted to identify the most relevant theories and mathematical approaches used in previous studies on the spread of fake news. This exploration will focus on:

- Mathematical foundations previously used to describe the phenomenon of disinformation in social networks, such as diffusion models, graph theory, sorting, and filtering algorithms.
- Applications of communications theory, more specifically filtering theory in the context of fake news.

- Previous case studies on the phenomenon of fake news in the Hispanic community and other similar communities, with emphasis on those specifically analyzing behavior within Twitter, Facebook and TikTok.

Data collection.

Specific and relevant examples of fake news propagated through Twitter, Facebook and TikTok will be determined.

Data analysis.

Using the mathematical models chosen beforehand will be applied to the collected data, to determine the interaction of relevant factors in a given news story.

Interpretation of results.

The results obtained from the mathematical models will be interpreted, identifying key patterns and common characteristics of fake news on social platforms. This will provide a quantitative representation of the phenomenon in the Hispanic community.

Scopes or Goals

- Determine which mathematical models and theories are best suited to describe the propagation of fake news on social media platforms.
- Identify key patterns and common characteristics of fake news shared on Facebook, Twitter, and TikTok within the Hispanic community.

- Develop a quantitative case analysis applying mathematical models to specific cases of fake news on social networks to assess key factors influencing the proliferation of the phenomenon.
- Provide an analytical approach to help develop future strategies in the process of mitigating the phenomenon of disinformation in vulnerable communities.
- Provide results that can be used as a reference for further research in the area of fake news propagation in social networks.

Theoretical Framework

Understanding Social Media

Social Network Algorithms: TikTok and its Addictive Algorithm

Although ByteDance, the company that owns TikTok, has given clues about the social network's algorithm, since it is not open-source software, it is difficult to determine with certainty how TikTok's algorithm works. However, multiple investigations have been developed with the aim of determining what is the algorithm of recommendations for this social network, based on content metrics, and user experience. According to Klug, Qin, Evans & Kaufman (2021), unlike Facebook, Instagram, or YouTube and their short-video versions, Instagram Reels and YouTube Shorts, TikTok does not generate video feeds based on content from accounts followed. Rather, the TikTok recommendation algorithm customizes video content for the individual user's 'for you' page based on previous and continuous user engagement with presented video content through video viewing time, liking, commenting, and sharing. In this sense, we can say that the TikTok algorithm deduces the user's preferences based on a series of "signals" that indicate when the user is interested in a certain video. These signals will be described below.

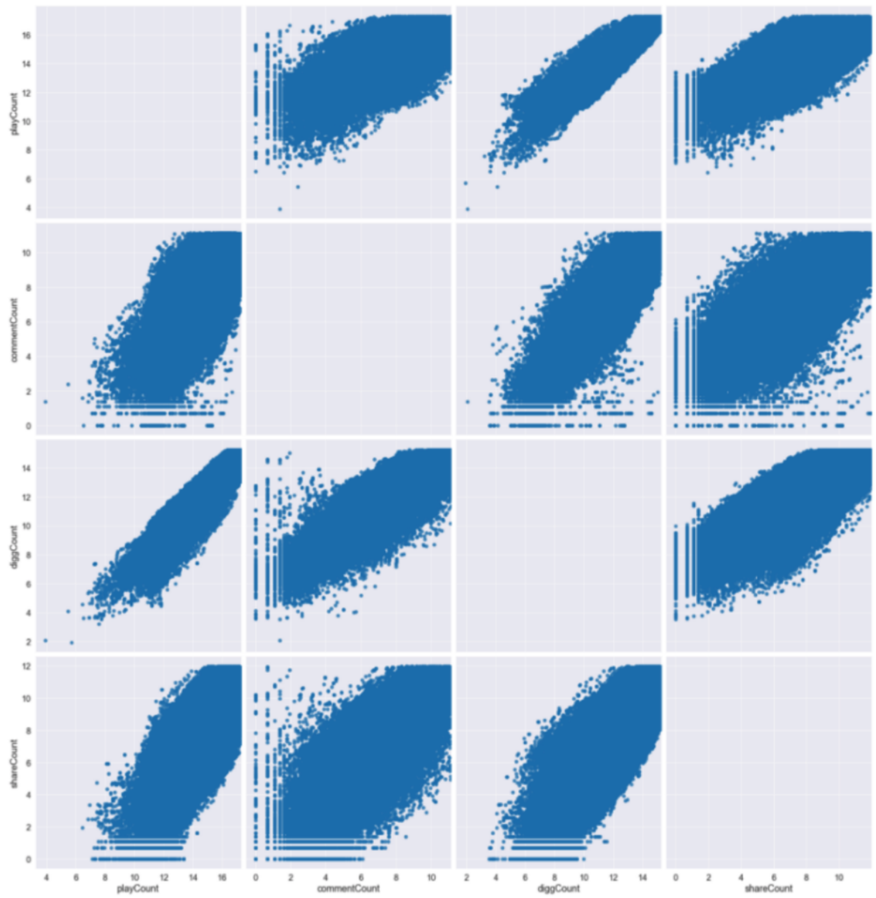


Figure 1: Correlations between number of video plays ('playCount'), number of comments ('commentCount'), number of likes ('diggCount'), and number of shares ('shareCount').

In 2020, to counter accusations of links to Chinese government entities, the company ByteDance explained for the first time in a blog post how the platform's engagement signal system works. According to Matsakis (2020), TikTok relies on a number of signals to identify what kinds of videos users want to see, some weighted more heavily than others. Strong signals include whether you watched a video to the end, whether you shared it, and if you followed the creator who uploaded it afterward. Weak signals are things like the type of device you have, your language preference, and whether you're in the same country as the person who posted a video. TikTok also considers negative feedback on a video, like whether a user tapped "Not Interested," or if they choose to hide content from a certain creator or featuring a specific sound. TikTok says the For

You page algorithm isn’t optimized for any specific metric, but rather is designed to take into account many factors, like whether people like using the app and if they choose to post content themselves.

Social Network Algorithms: **X Designed an Algorithm For You**

X's algorithm personalizes your feed by evaluating your interactions: what you like, share and the accounts you follow. It aims to provide a more engaging experience by filtering out inappropriate content and continually adjusting what you see to keep you interested. Content is divided into three main tabs:

“For You” tab:

- Displays tweets that the algorithm deems interesting, including content from accounts you don't follow.
- It uses a process called “candidate sourcing” to show tweets outside your usual network.
- Prioritize your past interactions, popular tweets and a mix of familiar and new content to keep you engaged.

“Following” tab:

- Organizes tweets in a primarily chronological order from the accounts you follow.
- Slightly adjusts the visibility of tweets with high engagement (likes, retweets, replies).
- This tab is more controlled and only shows content from the accounts you have decided to follow, no external surprises.

Explore tab:

- Highlights what's trending and trending news, even from people you don't follow.

- It's designed to show what's generating a lot of engagement, helping you discover new content and stay informed.
- Organized into sections such as “News”, “Sports” and “Entertainment” to make it easy to find topics of interest.

Social Network Algorithms: Facebook Algorithm Does not go out of Fashion

The Facebook algorithm is a sophisticated system designed to prioritize content based on what’s most relevant to each user. Here’s how it works step-by-step:

1. Inventory: The algorithm starts by gathering all available posts from friends, groups, and pages the user follows. This collection excludes flagged content that violates Facebook’s Community Standards.
2. Signals: Next, the algorithm evaluates "signals" to gauge each post's relevance. Signals include details like:
   - Who posted the content
   - The content type (text, image, video)
   - Past interactions between the user and the poster
   - Even environmental factors like internet speed
3. Prediction: Based on these signals, the algorithm predicts how a user might interact with a post. Key predictions include whether the user might like, comment, share, or watch a video entirely. These predictions aim to identify content that will likely foster engagement, like starting a conversation.
4. Relevance Score: The algorithm then assigns each post a relevance score, determining the order posts appear in a user’s feed. Posts with higher scores appear first, as they are most

likely to capture the user's interest. Finally, recommended content and ads are woven into the feed based on similar user preferences.

Relevance score for Facebook posts

Number of posts: p

Members per group: g

Replications: s

Average number of Facebook friends:

Impact: I

$$I = (p \cdot g) + (s \cdot \bar{a}) \qquad (1)$$

I = (p - g) + (s - a¯) (1)

The above mathematical formula is explained as follows. The product of n and g is obtained, obtaining the number of times a fake news was published in different groups by the number of members of the various groups. Separately by a sum and limiting the direct relationship by multiplication, we find the product of the number of times the fake news was shared and the average number of friends of a Facebook user. In this way, we obtain the total impact, without taking into account the limitations transposed by the algorithm (García & Martínez, 2021).

$$I = ((p \cdot g) + (s \cdot \bar{a})) \cdot \frac{1}{10} \quad (2)$$

$$I = \frac{pg}{10} + \frac{\bar{a}s}{10} \quad (3)$$

Echo chambers: a new form of interaction

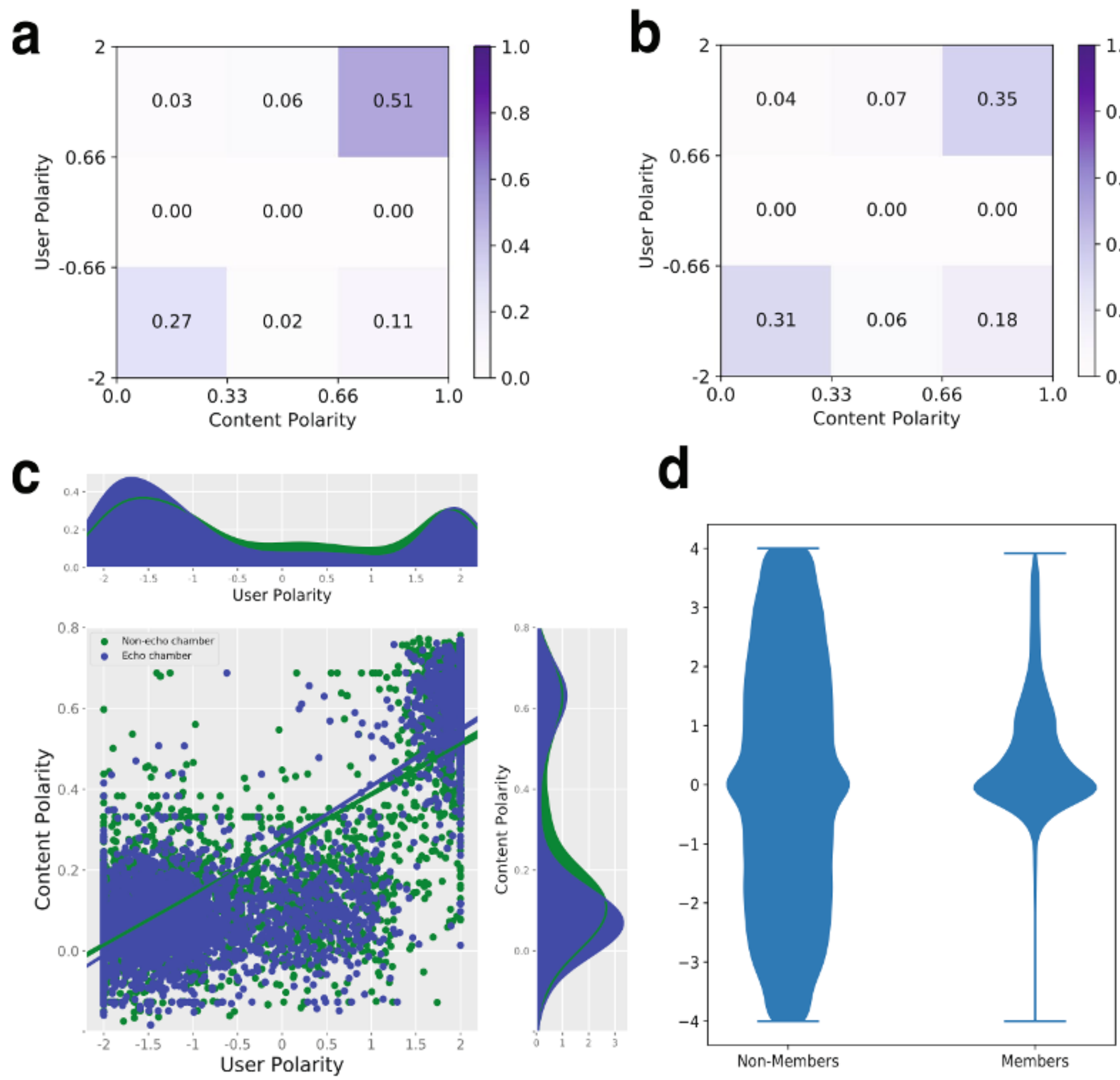
Social media such as Twitter or Facebook allows users with diverse backgrounds (e.g., political viewpoint, race, or gender) to share news, information, or opinions. Despite such a diversity in social media, studies have reported that online users with similar interests tend to be gathered and eventually form a homogeneous cluster, known as *echo chamber*. Selective exposure is a social phenomenon that a user tends to selectively consume informa- tion that he/she would like to believe21, which is considered as a distinct property of echo chambers

User homogenity calculation: *user homogeneity* between two users *i* and *j* in an echo chamber as $\sigma i = \sigma i \sigma j$, where σi is the user polarity of user *i*. The user homogeneity score is close to −4 if two users are in opposite positions, and close to 4 if their positions are similar.

Chart a shows the user homogenity between users in the same echo chamber, and **chart b** shows user homogenity between non-members of the echo chamber.

In both figures, the portions of the users in the **left-bottom and right-top** cells tend to be higher than others, meaning that users participating in rumor spreads tend to show <u>the selective exposure</u> in general. Each cell is calculated by averaging the homogeneity of the content among users with these characteristics.

The sum of the portions of user in the **left-bottom and right-top** cells of echo chamber and non-echo chamber are 0.78 and 0.68, respectively, meaning that <u>*echo chamber members show the relatively stronger selective exposure than non-members.*</u>



Fake News: More than Just Rumors

Understanding the Fake News

What is a fake new?

This is false information that is presented as true, with the intention of misleading or manipulating the audience. This news may be deliberately fabricated to disseminate a particular narrative or may arise as a result of unintentional misinformation.

It is news that is not true or that has been distorted. This phenomenon has existed since humans began to communicate through language. For example, "gossip" is often fake news. Someone may make up a rumor for their own benefit or to hurt another person they are angry with. The difference is that "Fake News" is rumors or lies that are spread through the Internet and social networks, such as Google, Facebook or YouTube.

Online news has become an easy solution to obtaining information without reading the newspaper. Several online news sources provide free access for readers because the cost of building online news is not expensive and difficult if we are comparing it with the newspaper.

Are we in the post-truth era?

Post-truth" refers to a situation in which objective facts and truth have less influence in shaping public opinion than appeals to emotions and personal beliefs. In other words, in the post-truth era, people tend to place more importance on what they feel or believe, rather than considering evidence or verified facts. This phenomenon has been linked to the rise of social networks and disinformation, where false or manipulated information spreads rapidly, generating an environment in which the truth is difficult to distinguish.

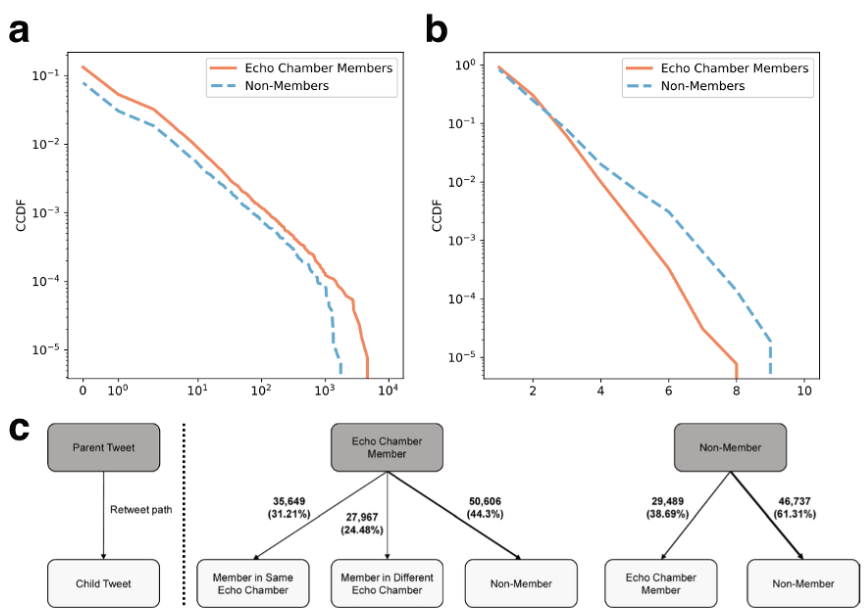
Fakes News Propagation

How does Fake News spread?

Social media has revolutionized how information is disseminated, as it allows users to share content with minimal effort, at a higher speed and larger scale (Koohikamali and Sidorova, 2017). Its very nature implies that users are constantly called upon to validate the content in circulation and engage with others (Boyd et al., 2010). Individuals therefore can have a crucial role in the spreading of fake news.

In fact, individuals consider that news sharing is an appropriate approach to preserve and extend their relationships and networks. It enables them to talk, interact and communicate with their friends using different aspects such as commenting, posting, chatting and liking news stories (Lee and Ma, 2012). These inherent characteristics of social media, allowing high levels of interactivity and diffusion of uncontrolled content, make it especially conducive to the sharing of fake news (Apuke and Omar, 2021).

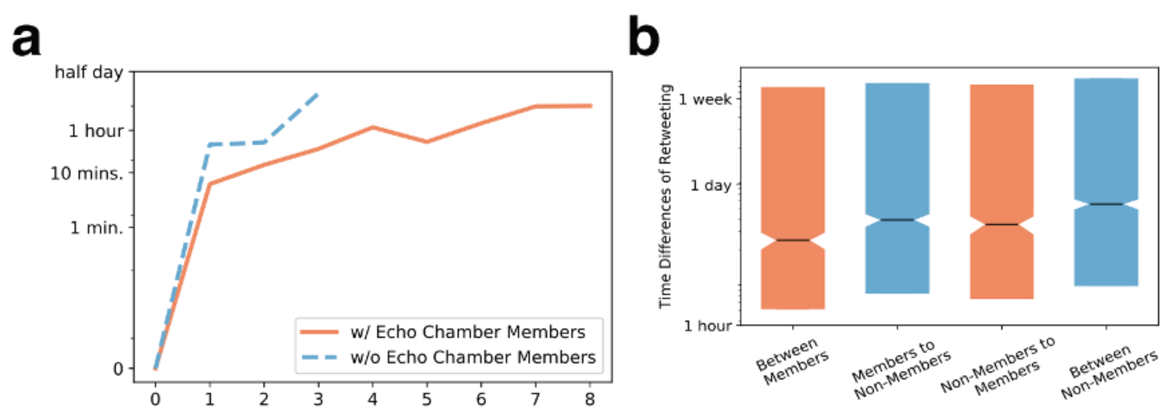
Echo chambers in rumor propagation



Graph a compares the depth of retweets from echo chamber members (orange line) and non-members (dotted blue line). On a **logarithmic scale**, it shows how echo chamber members tend to retweet more deeply **(cascade),** reflecting greater homogeneity in their interactions. **Graph b**, on the other hand, shows the number of retweets generated by both groups.

Most echo chamber members tend to be located at close to the roots of rumor cascades, and are likely to generate more retweets than non- members. e numbers (and portions) of retweets from members/non-members to members/non-members are described in (c). Overall, rumors tend more to be propagated to non-members while the portions of retweet paths from both members and non-members to non-members are different (44% and 61%, respectively). The portion of retweet paths between the members in the same echo chamber is higher than the ones in different echo chambers, *meaning that rumors written by an echo chamber member tend to be more propagated among non-members or members in a same echo chamber.*

Rumor propagation speed



These results imply that echo chamber members contribute to propagate rumors quickly, by not only retweeting the rumor quickly but also eliciting other users' quick responses.

How can we identify a fake new?

In recent years, research in artificial intelligence (AI) has been considered the best solution to find the best results in detecting fake online news.

LSA is more powerful at extracting meaningful concepts within semantic meaning [31]. Through LSA, the text data (TDM) are subjected to singular value decomposition (SVD) to construct a semantic vector space that can be used to represent the concept–word– document association. Latent semantic analysis (LSA) was employed to extract significant sentences from an input document in order to create a summary by identifying a latent or hidden semantic structure. When analyzing collected documents, the LSA method uses singular value decomposition (SVD) to derive concepts from collected documents by modeling them as a term–document matrix (TDM).

Step 1. Dataset collection

Step 2. Data pre-processing and NLP

Step 2.1: Tokenization

The Python Natural Language Toolkit (NLTK) was used for text tokenization. This study employed “unigram” to segment sentences.

Step 2.2: Data cleaning

In this step, we eliminated stop words such as “the” and “or”, non-English words, and meaningless icons to obtain clean data for further analysis.

Step 2.3: Lemmatization

This step involves reducing complex forms of a single word to their most basic form, such as “ate” and “eaten” to “eat”.

Step 2.4: Word Frequency Counting

The authors performed a word frequency count and eliminate words with a frequency of less than 5.

Step 2.5: Constructing a Term–Document Matrix (TDM)

In this step, the authors created a term–document matrix (TDM) using the TF–IDF (term frequency–inverse document frequency) weights shown in Equation (1).

$$TF - IDF = TF(t_i, d_i) \times \log\left(\frac{N}{N(t_i)}\right)$$

where ti is the ith term; di represents the jth document; N is the total number of all documents; N (ti) denotes the number of documents which contain ti features.

Step 3: Latent Semantic Analysis (LSA)

Step 4: Naming the extracted concepts

Step 5: Explaining Results

Step 6: Drawing discussions and conclusions

Fakes News Metrics and Statistics

What metrics can we get from fake news?

X (former Twitter)

1. Impressions: Number of times a tweet appears on someone's timeline, either because it was followed or shared.
2. -Interactions: Total actions taken on a tweet, such as clicks, retweets, replies and likes.
3. Interaction rate : Percentage of interactions divided by the total number of impressions, which indicates how engaging the content is.
4. Retweets: Number of times a tweet has been shared by other users.

5. Likes: Number of times a tweet has received a “like”.
6. Replies: Number of replies a tweet has received.
7. Profile clicks : Number of times someone clicks on the profile from a tweet.
8. Mentions: Frequency with which other users mention the account.

Facebook:

- Reactions: Include “Likes” and other reactions such as “I love it”, “I'm amused”, “I'm amazed”, “I'm saddened”, and “I'm angry”.
- Comments: the total number.
- Shared: Number of shares with users, on other pages and groups.
- Reach: the number of unique users who have seen the post.
- Impressions: The number of times the post has been viewed, regardless of whether it was the same profile.
- Clicks: Clicks made on the post, either to expand the image, see more text, play a video or click on a link.
- Click on links.
- Post Engagement: The sum of all interactions with the post.
- Engagement rate: A percentage that measures the number of engagements in relation to the total reach of the post.
- See more clicks: The number of times users have clicked to see the full text of a long-form post.

- Video views: If it is a video post, it shows how many times it has been played. It can include more specific metrics, such as 3-second plays, 10-second plays, or average audience retention.

Tik Tok:

1. Views: Total number of views.
2. Likes: Number
3. Comments: Number
4. Shares: Number.
5. Followers: Number of users who follow the account.
6. Retention rate: Percentage of people who watch your video until the end.
7. Active Days: Number of days you have published content in a specific period.
8. Referral traffic: Traffic sources that generate views on your videos (search, profile, other videos).
9. Views per profile: Number of times your profile has been visited.
10. Demographic metrics: Information about the age, gender, and location of your audience.

Relevant Statistics

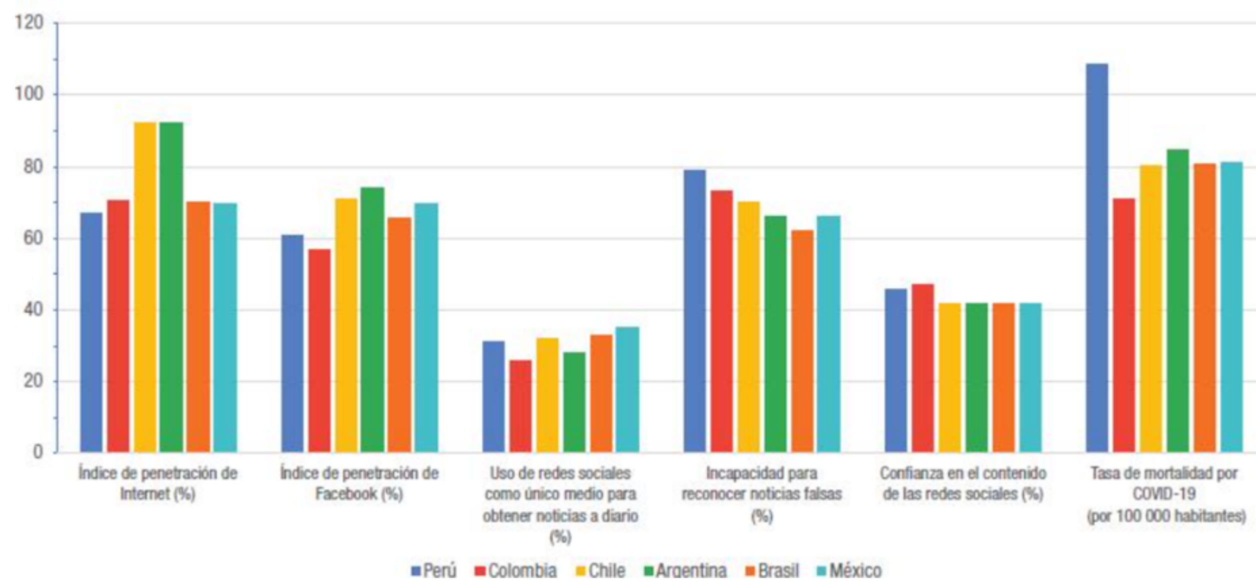
Latin America's post-pandemic social networking landscape.

“82.5% of Latin Americans will have access to social networks in 2020” according to Statista, (Escala, 2022). Allowing the number of social networking profiles in the region that year to be:

- Brazil: 130 million
- Mexico 84.9 million

- Argentina: 29 million
- Colombia: 32.9 million
- Peru: 23.5 million
- Chile: 13.9 million

According to the information collected, Chile and Argentina were the countries with the highest Internet (92.4% and 92.0%, respectively) and Facebook (70.9% and 73.9%, respectively) penetration rates, and also make considerable use of social networks as the only means of obtaining news on a daily basis (32.0% and 28.0%, respectively); Brazil and Colombia showed intermediate behavior according to the Internet penetration rate and had relatively low Facebook penetration rates, compared to the other countries analyzed. Mexico had the highest use of social networks (35%), while Peru (79.0%) and Colombia (73.0%) had the highest values for the index of inability to recognize fake news and the lowest Facebook penetration rates (60.7% and 56.9%, respectively), (Escala, 2022).



Of the countries analyzed, Peru-whose population had the highest percentage of inability to recognize fake news (79.0%) and which had the second highest trust in social network content

(46.0%)-had the highest mortality from COVID-19 (108.7 per 100,000 population), (Escala, 2022).

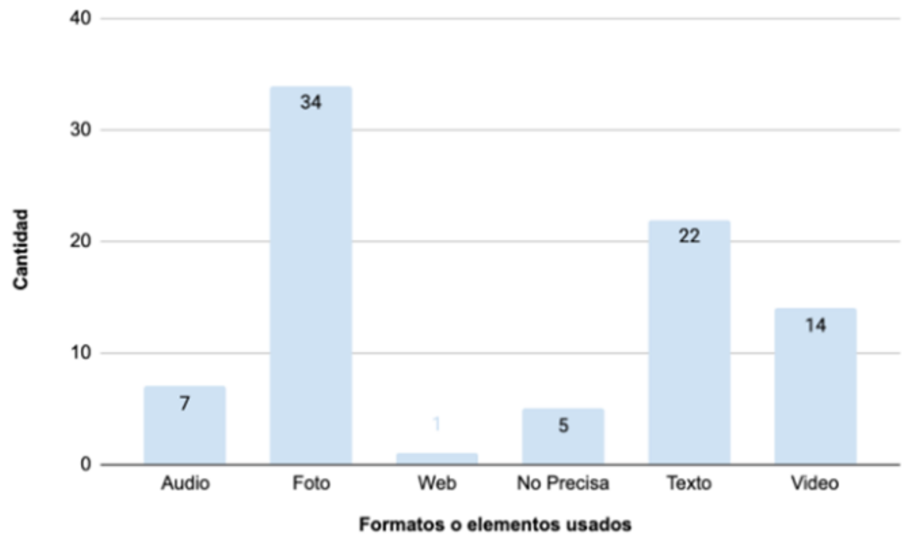
When assessing the population's inability to recognize fake news, a more favorable pattern was observed in Brazil (62.0%), Argentina (66.0%) and Mexico (66.0%), which also had lower mortality rates (between 81 and 85 per 100 000 inhabitants) than Peru. Contradictorily, Colombia (73.0%) and Chile (70.0%) had a high inability to recognize false news and low mortality due to COVID-19 (71 and 80 deaths per 100,000 inhabitants, respectively), (Escala, 2022).

Article on viral disinformation content (Peru-Covid):

Tabla 2. Medios digitales usados para difundir contenido de desinformación

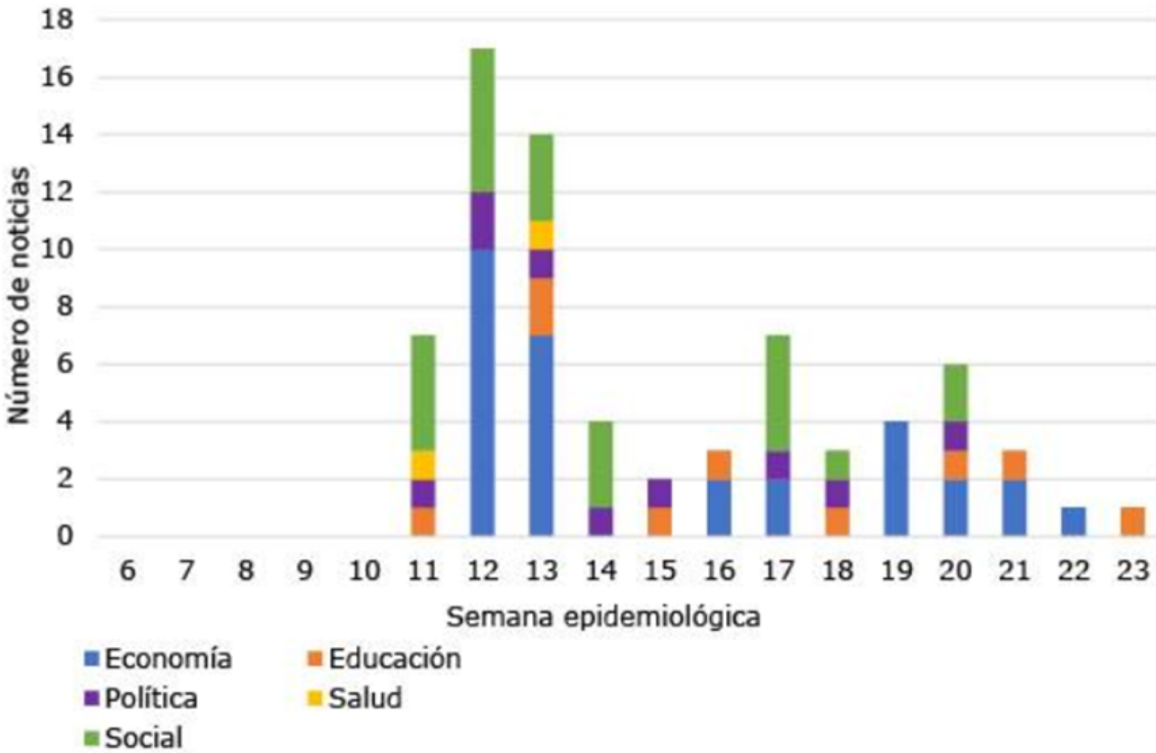
Medios digitales usados	Cantidad
Web	1
Email	1
Facebook	15*
Instagram	2
SMS	2
Redes Sociales**	28
TV	3
Twitter	6
WhatsApp	23
YouTube	2*

Figura 2. Formatos o elementos empleados para los contenidos de desinformación



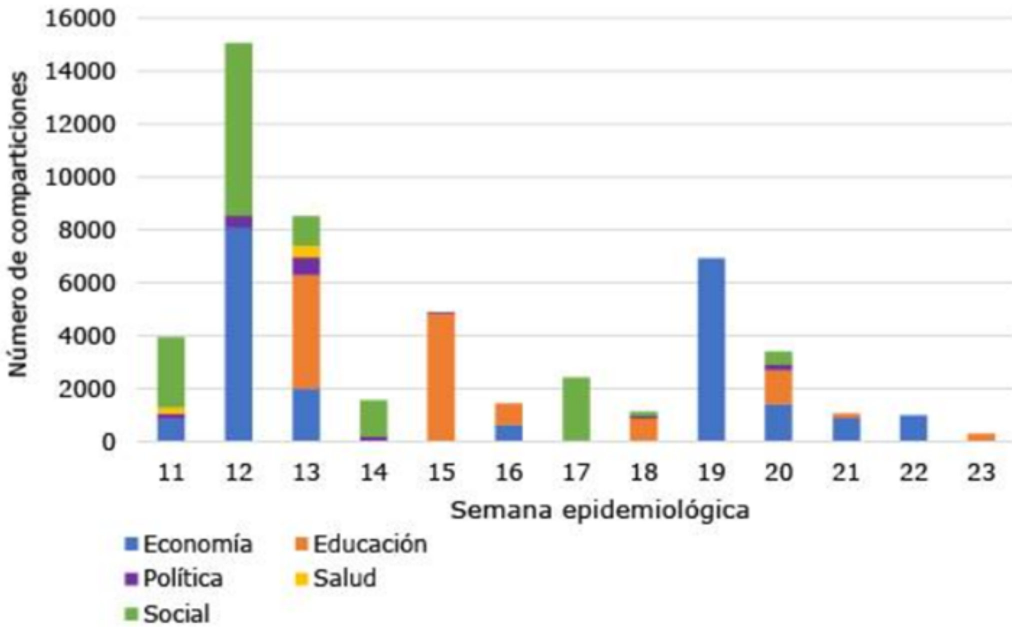
Identification of fake news study

Seventy-two news items were identified as false by Peruvian State institutions on the social networks Facebook and Twitter between February 1 and June 4, 2020. No posts disproving news were found between February 1 and March 8. The first news identified as fake was published on March 9, 2020. The news was mostly about an economic (41.7%) or social (30.6%) topic and, to a lesser extent, education (12.5%), politics (12.5%) and health (2.8%). The type of falsification was impersonation (51.4 %), and indirect and mentioned (25 % and 22.2 %, respectively). Only one false news item could not be categorized according to this classification because it did not allude to any of the State institutions. The distribution of news identified as false by topic and epidemiological week is shown in Figure 1 (Scielo, 2022).



Number of times they were shared

For October 13, 2020, the 72 news items identified as fake were shared 51 768 times, with a median of 369 shares per news item. Economic news was the most shared (21 118 times), followed by social (14 710 times), education (13 422 times), politics (1822 times) and health (696 times). The shares of news identified as false by topic and epidemiological week are shown in Figure 2, (Scielo, 2022).



Number of times news identified as false were shared, by topic and epidemiological week, between February 1 and June 4, 2020, (Scielo, 2022).

2021:

"In Mexico, WhatsApp and Facebook were the social networks with the highest percentage of users. In Brazil, Instagram is positioned as one of the most popular social networks with more than 109 million users in June 2021, a number of users that surpasses any other country in Latin America, (Escala, 2022).

Top most used social networks in the world in 2021-2022

- Facebook with 2,740 million
- Youtube with 2,291 million
- Whatsapp with 2,000 million
- Facebook Messenger with 1.3 billion
- Instagram with 1,221 million

- TikTok with 689 million

- Twitter with 353 million

2022:

Social network users in 2022 by region:

- Latin America: 392.6 million

Top 6 most used social networks in Latin America in 2022

According to Statista, the most popular social networks are:

- YouTube
- Facebook
- WhatsApp
- Facebook Messenger
- Instagram
- Twitter

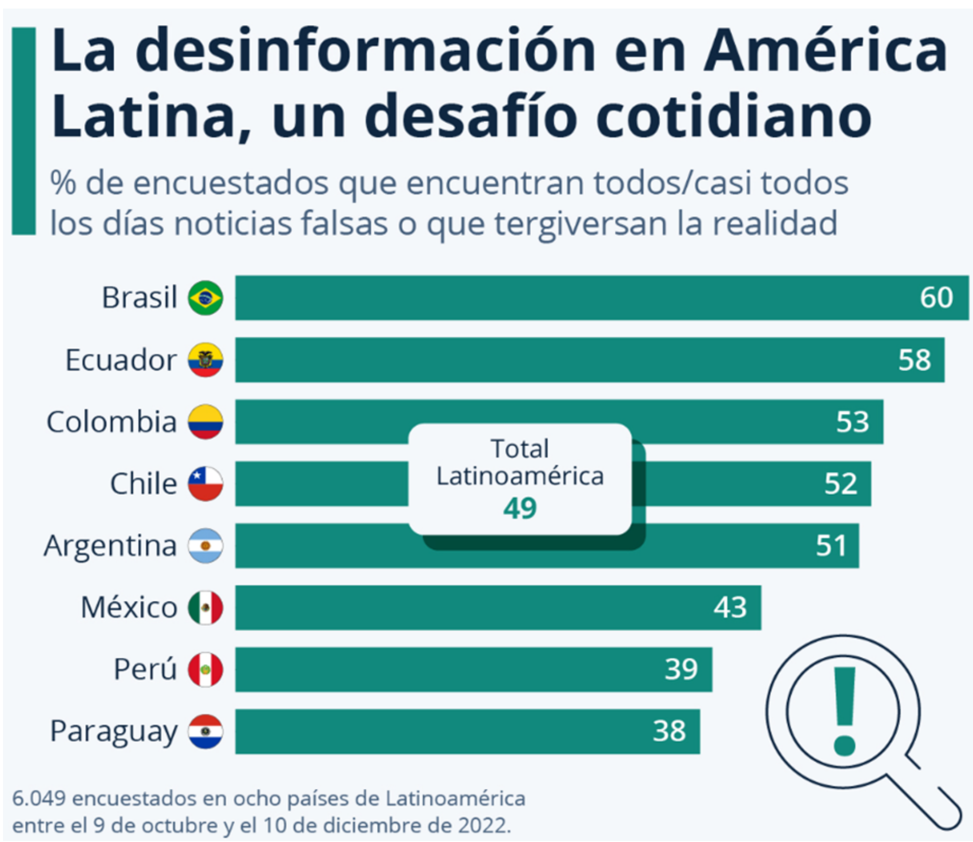
Facebook is currently the most popular social network in Latin America with more than 180 million users. The platform is expected to continue to grow in popularity, with an estimated 225 million users by 2022, (Escala, 2022).

Likewise, the Latin American countries that use social networks the most are:

- Chile with 92.8% (Rank #1)
- Uruguay with 90.2% (Rank #2)
- Argentina with 86.3% (Rank #3)
- Peru with 83.8% (Rank #6)
- Colombia with 81.3% (Rank #8)

Porcentaje de interacciones por país en cada red social en 2022

	Facebook	Instagram	Twitter	YouTube
Argentina	47%	45%	5,9%	2,1%
Brasil	54%	37%	6%	3%
Chile	47%	45%	4,5%	1,5%
Colombia	47%	42%	7,5%	35%
México	51%	37%	7%	5%
Perú	47%	21,5%	3%	3,5%



Analyzing Fake News From a Mathematical Approach

Filtering Out the Noise of Truth

Filtering theory and fake news application

According to Brody & Meier (2021), we can define a fake new as “information that originates from the “sender” of fake news, is transmitted through a communication channel, and is then received, typically, by the general public.” Therefore, to analyze a fake new we can rely on the approach of communication theory, more specifically on the filtering theory.

The Filtering Theory, better known as Wiener Filters or Wiener Theory, was formulated by Norbert Wiener, and proposes certain techniques in the field of signal processing to extract a desired signal in the presence of noise. According to Vaseghi (1996), “Wiener filters play a central role in a wide range of applications such as linear prediction, signal coding, echo cancellation, signal restoration, channel equalization, system identification, etc.”(p. 3).

Conventional applications of filter theory are prediction, filtering, and interpolation (Wiener 1949, Kailath 1974). Recently, however, a new application has emerged for the description and modeling of observed phenomena, known as phenomenology (Brody & Meier, 2021).

In the context of fake news we can apply phenomenology to analyze the behavior of an individual in the face of fake news because people's behavior is guided by the predictions obtained after cleaning the “noisy information” presented to them (Brody & Meier, 2021).

Considering this approach, the problem of the propagation of fake news is no longer seen as a problem of lack of logical discernment, but as a problem of lack of filtering capacity. Applying the filtering theory, the “noise” (the fake news) alters the real information (the signal) in such a way that people may receive a distorted or misleading view of reality.

Filtering theory and evolution of information

By applying filtering theory and Bayes' theorem, we can model how the reception of new information influences decision making. We will use this model later to model fake news.

1. Initial preference

   Given a decision making problem between option A and option B, X is a binary variable, such that X = 1 means that A is better, and X=0 means that B is better. P(X=0)=p and $P(X=1)=1-p$, these probabilities represent the preference initiates;

2. Information evolution

   As new information reaches an individual (e.g., hearing on the radio that A is better than B), the X value remains constant, but our estimate of which is the best option is adjusted. This evolution is described by the flow of information.

$$\xi_t = \sigma X t + B_t.$$

   Here, σ is a constant representing the signal-to-noise ratio; it determines the importance of signal (initial preference) versus noise (new information received). The term Bt is a Brownian motion representing noise or random uncertainty, and is assumed independent of X to represent unbiased noise.

3. Application of Bayes' theorem to calculate the post-noise presence probability.

As we receive more information (changes in ξt), the probability of each option is adjusted. Using Bayes' theorem, we update the probability of X given the observed value of ξt.

$$\mathbb{P}(X = x_i|\xi_t) \;=\; \frac{\mathbb{P}(X = x_i)\rho(\xi_t|X = x_i)}{\sum_j \mathbb{P}(X = x_j)\rho(\xi_t|X = x_j)}.$$

The conditional density function $\rho(\xi t\ |X{=}\ xi)$ is the density of ξt given a value of X, which follows a normal distribution, given that the noise Bt is a Brownian motion:

$$\rho(\xi|X = x_i) = \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{(\xi - \sigma x_i t)^2}{2t}\right)$$

Recalling that (x0, x1) = (0, 1) we thus obtain

$$\mathbb{P}(X = x_i|\xi_t) = \frac{p_i \exp\left(\sigma x_i \xi_t - \frac{1}{2}\sigma^2 x_i^2 t\right)}{p_0 + p_1 \exp\left(\sigma \xi_t - \frac{1}{2}\sigma^2 t\right)},$$

4. Decision Making.

   The posterior probability $P(X{=}xi \mid \xi t)$ is used to make a rational decision at time t, considering the information accumulated up to that point. If the a posteriori probability of an option (e.g., taking the bus) is greater than 0.5, then that option is considered better.

Mathematical model for fake news processing

Given the nature of fake news, when such news spreads, disinformation is superimposed on true information (Brody & Meier, 2021). Applying filtering theory, false information can be understood as “noise” on a true signal. Following this logic Brody & Meier (2021) propose the following model for information processing in the presence of fake news:

$$\eta_t = \sigma X t + B_t + F_t,$$

- ηt: Represents the total process of information perceived by people at a specific time t. That is, it is the combination of true signals, noise, and fake news that affect the perception of reality.
- σX: It is the true signal component about the variable X that the public is trying to discover (the “true value of X”). Here, X represents the underlying information or objective reality, and σ is the “information flow” that measures the rate at which that truth is spreading over time.
- Bt : This is the noise component. It represents rumors, speculation and other forms of unsubstantiated, but unbiased, information. These elements are unintentional and can hinder the understanding of the truth, although on average they do not skew perception in a specific direction.
- Ft : This is the fake news component. It represents information that contains intentional bias (or “fake news”), with the goal of influencing or skewing the perception of the truth. Unlike noise, fake news is not neutral and its expectation is not zero ($E[Ftt]$≠0 E[Ft]=0), so it can actively skew people's perception in a particular direction.

The goal is to help distinguish between the components that affect the perception of truth so that, despite the presence of noise and fake news, a more accurate estimate of X, or the "underlying truth," can be arrived at.

Arrival time of a fake news item, and veracity check

While the above model helps us to estimate the true value of a signal X in the presence of noise (fake news), what happens when an individual does not know whether he/she is facing a fake news or not? The following module proposed by Brody & Meier (2021), helps to determine the truth value of X and the arrival time of the false news, if we are in the presence of one.

If the observation is modeled by a process of the form $Bt + \mu(t\text{-}\tau)1\{t \geq \tau\}$, where Bt is a Brownian motion and $\mu(t\text{-}\tau)1\{t \geq \tau\}$ represents the change in the series structure at time τ. The objective is to detect the time τ at which this change occurs and to estimate the signal X in a noisy environment. For this purpose, the model calculates a conditional probability $ft(u)$, which indicates the probability that the fake news arose within a specific small time interval, from u to u+du, for τ is considered in the context of at least one piece of fake news. The model for fake news is given by Ft=$m(t\text{-}\tau)1\{t \geq \tau\}$, where m:$R \rightarrow R$ is an arbitrary function assumed to be at least differentiable once. The conditional density $ft(u)$ is calculated as:

$$f_t(u) = \frac{f_0(u) \sum_i p_i \, \mathrm{e}^{\int_0^t (\sigma x_i + m'(s-u)\mathbb{1}\{s \geq u\}) \mathrm{d}\eta_s - \frac{1}{2} \int_0^t (\sigma x_i + m'(s-u)\mathbb{1}\{s \geq u\})^2 \mathrm{d}s}}{\int_0^\infty f_0(w) \sum_i p_i \, \mathrm{e}^{\int_0^t (\sigma x_i + m'(s-w)\mathbb{1}\{s \geq w\}) \mathrm{d}\eta_s - \frac{1}{2} \int_0^t (\sigma x_i + m'(s-w)\mathbb{1}\{s \geq w\})^2 \mathrm{d}s} \mathrm{d}w},$$

where m' (u) = dm(u)/du and pi = P(X = xi).

Once you have the probability of when the fake news might have arisen, the model uses that information to try to estimate the true value of X. Note that f0(u) is the a priori density for τ , which reflects the initial view on how the release timing is distributed. Hence, the best estimate for the release time of fake news is given by

$$\int_0^\infty u\bar{f}_t(u)\mathrm{d}u.$$

Using the conditional expectation property, it can be written as:

$$\mathbb{E}[X|\{\eta_s\}_{0\le s\le t}] = \mathbb{E}[\mathbb{E}[X|\{\eta_s\}_{0\le s\le t}, \tau]|\{\eta_s\}_{0\le s\le t}].$$

The strategy is to calculate the internal expectation first. Since the information process $\{ntt\}$ is Markovian, this reduces to $Y(\eta t,\tau)$=$E[X|\eta t,\tau]$, which, according to Bayes' theorem, is given by:

$$Y(\eta_t, \tau) = \frac{\sum_k x_k p_k \exp\left(\sigma x_k \eta_t - \frac{1}{2}\sigma^2 x_k^2 t - \sigma x_k m(t-\tau)\mathbb{1}\{t \ge \tau\}\right)}{\sum_k p_k \exp\left(\sigma x_k \eta_t - \frac{1}{2}\sigma^2 x_k^2 t - \sigma x_k m(t-\tau)\mathbb{1}(t \ge \tau)\right)}.$$

Mathematical Model Application

Our approach consists of using language analysis tools to classify the type of user interactions with a post on the three main social networks, and applying the mathematical model proposed by Brody & Meier (2021) to these data.

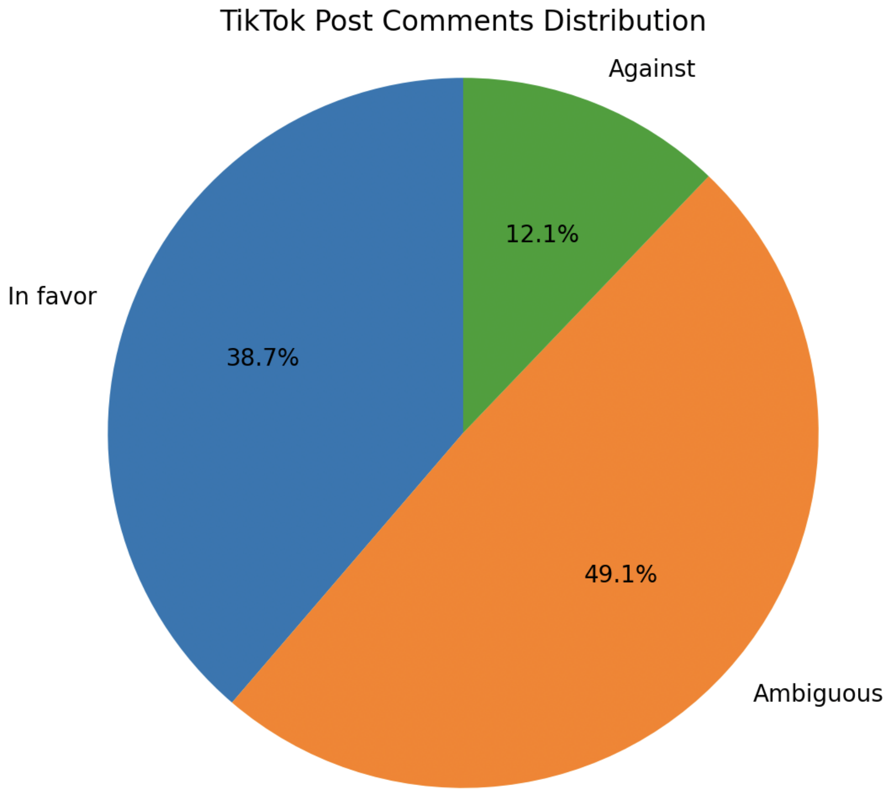
In our case, we focus on a specific fake news story circulating on all three platforms. By downloading and processing comments on the post on each platform, we create a dataset that reflects user reactions. This allows us to apply the sentiment analysis tool VADER (Valence Aware Dictionary and sEntiment Reasoner) to classify interactions into three categories: for, against and ambiguous. These classifications provide insight into how the audience perceives and reacts to the news. It also allows us to determine which part of the news is true signal and which part corresponds to noise.

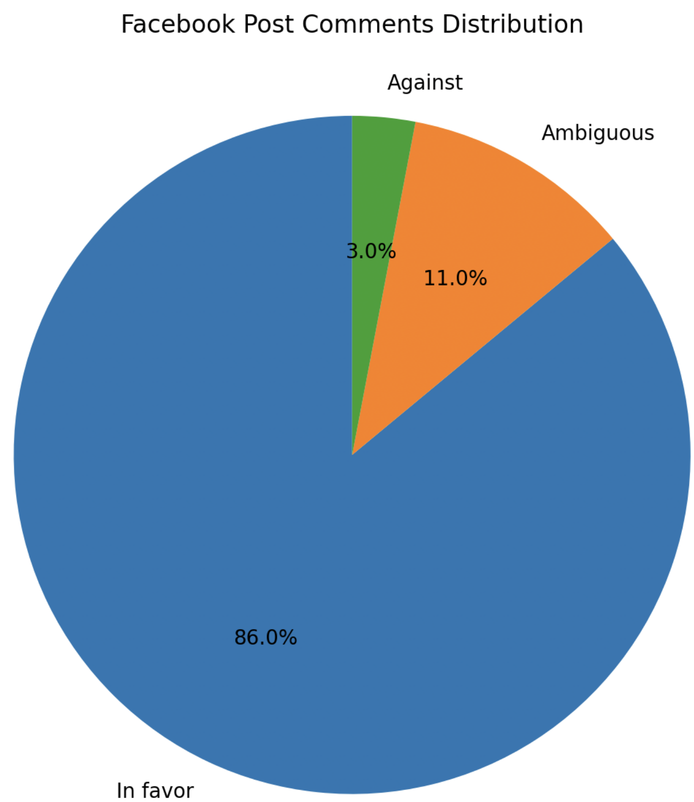
Using the model provided by Brody & Meier (2021), we can relate each category of interaction to the elements of the formula as follows:

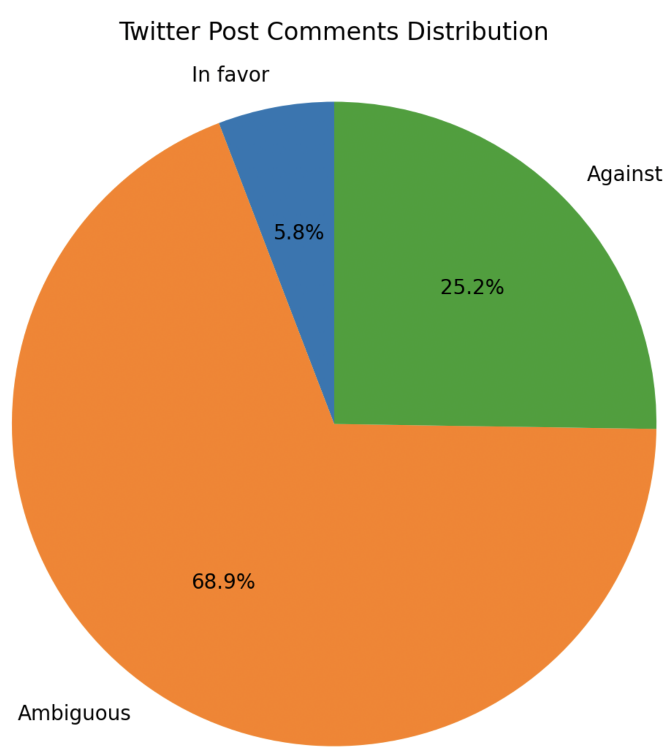
- ηt (Total perceived information process): This is what we want to calculate. It represents how people perceive the news, considering both the truth, the noise, and the fake news.
- σX (True signal): They are those comments that focus on denying the false news. Bt (Noise): It is Brownian motion. It is a random process, modeled as $Bt \sim N(0,b^2t)$ where b is the diffusion coefficient (the noise intensity). Ambiguous comments, which do not provide a value judgment, and which favor the creation of rumors or misleading information.
- Ft (Fake news): Represents false information designed to distort the perception of the truth. They are those comments that are in favor of the fake news and that give it credibility.

Interactions Clasification

As a result of a sample of 200 comments from 3 different TikTok, Facebook, and Twitter posts, related to the same fake news, we obtained the following data in terms of ranking of interactions:


TikTok Post Comments Distribution
Against
12.1%
In favor
38.7%
49.1%
Ambiguous


Facebook Post Comments Distribution
Against
Ambiguous
3.0%
11.0%
86.0%
In favor


Twitter Post Comments Distribution
In favor
Against
5.8%
25.2%
68.9%
Ambiguous

Total Perceived Information Process Calculation

To determine the value of B_t we need to compute the standard deviation for each platform. We applied the binomial standard deviation formula:

$$B_t = \sqrt{p(1-p)}$$

Where *p* is the proportion of ambiguous comments on the platform.

TikTok

Ambiguous comments: *49.1% → p=0.491p = 0.491*

$$B_t = \sqrt{0.491 \times (1-0.491)} = \sqrt{0.491 \times 0.509} = \sqrt{0.249} \approx 0.499$$

Facebook

Ambiguous comments: *11% → p=0.11p = 0.11*

$$B_t = \sqrt{0.11 \times (1-0.11)} = \sqrt{0.11 \times 0.89} = \sqrt{0.0979} \approx 0.313$$

Twitter

Ambiguous comments: *68.9% → p=0.689p = 0.689*

$$B_t = \sqrt{0.689 \times (1-0.689)} = \sqrt{0.689 \times 0.311} = \sqrt{0.214} \approx 0.462$$

Calculate ηt for Each Platform

Next, we use the formula to calculate ηt for each platform. The components for each platform are as follows:

TikTok:

- Comments supporting the fake news: 38.7% → F_t=0.387
- Comments debunking the fake news: 12.1% → σX=0.121
- Ambiguous comments: B_t=0.499

$$\eta_t = 0.121 + 0.499 + 0.387 = 1.007$$

Facebook:

- Comments supporting the fake news: 86% → F_t=0.86
- Comments debunking the fake news: 3% → σX=0.03
- Ambiguous comments: B_t=0.313

$$\eta_t = 0.03 + 0.313 + 0.86 = 1.203$$

Twitter:

- Comments supporting the fake news: 5.8% → Ft=0.058
- Comments debunking the fake news: 25.2% → σX=0.252
- Ambiguous comments: Bt=0.462

$$\eta_t = 0.252 + 0.462 + 0.058 = 0.772$$

Results Analysis

A high value of ηt, indicates that the public's perception of the information is strong and noticeable, i.e. the news has a significant impact on the flow of information.

Facebook users has the highest ηt indicating that users are more heavily influenced by fake news and are more susceptible to its impact. TikTok sits in the middle, with a moderate influence of fake news and some level of ambiguity in comments. This could indicate that fake news has moderate to little impact on the information flow of its users. FinallyTwitter shows the lowest impact of fake news, suggesting that users may take a more critical approach to the information they encounter.

Conclusions

From the analysis of the mathematical models, we can conclude that a determining factor in the propagation of fake news is the lack of capacity to adequately filter the "noisy" information (fake news) that is disseminated in the communication channels.

By applying filtering theory we can have an insight into how individuals adjust their decisions upon receiving new information. This adjustment does not occur in a linear fashion, but is influenced by the user's ability to discriminate between what is useful information (the signal) and what is misleading information (the noise). By applying these models, we can infer that people tend to adjust their beliefs and preferences as they receive information, but this adjustment depends to a large extent on the relationship between the signal (verified information) and the noise (fake news).

The exploratory analysis of the mathematical model indicated that each platform exhibits unique dynamics in the propagation and reception of fake news. And even though the data sample was relatively low considering the enormous flow of information, we can notice a clear pattern in the perception of fake news on these social networks. Facebook's vulnerability highlights the role of algorithms amplifying engagement-driven content, including misinformation, as well as a potentially less skeptical user base. TikTok's moderate impact reflects its nature as a platform with mixed user behavior: although fake news spreads, it is tempered by ambiguity or mixed signals. Twitter's resilience may be due to more frequent fact-checking and critical discourse, along with users' tendency to question viral narratives.

Therefore, to counteract the spread of fake news, it is crucial to develop methods that strengthen filtering capabilities, both at the individual level and through the recommendation algorithms of platforms, allowing users to distinguish more efficiently between legitimate and manipulated content. This mathematical approach provides a solid foundation for future research and applications in the fight against misinformation on social networks.

References

Vista de Desinformación en la era digital: el papel de las redes sociales en la propagación de noticias falsas durante conflictos globales. (2024.). Redilat.org. Retrieved November 15, 2024, from https://latam.redilat.org/index.php/lt/article/view/1865/2404

Proquest.com. Retrieved November 15, 2024, from https://www.proquest.com/scholarly-journals/analyzing-online-fake-news-using-latent-semantic/docview/2829699262/se-2?accountid=35796

Klug, D. (2021). *Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm*. https://dl.acm.org/doi/10.1145/3447535.3462512

Meier, D. C. B. D. (2021). *Mathematical models for fake news*. https://arxiv.org/pdf/1809.00964

Zamora, A. (2021). *Fake News: Medicion de Impacto en Redes Sociales*. https://www.researchgate.net/publication/353451549_FAKE_NEWS_MEDICION_DE_IMPACTO_EN_REDES_SOCIALES_A_TRAVES_DE_CANALES

Statista. (2024). *Noticias falsas y desinformación en el mundo - Datos estadísticos*. https://es.statista.com/temas/10126/fakes-news-y-desinformacion/#topicOverview

Melo, M. F. (2024). *La desinformación en Latinoamérica, un desafío cotidiano*.

https://es.statista.com/grafico/31618/encuestados-que-encuentran-todos-casi-todos-los-dias-informacion-noticias-falsas-o-que-tergiversan-la-realidad/

ACTIVA. (02 de abril de 2023). *Fake news Desinformación en Chile y LatAm*. https://chile.activasite.com/wp-content/uploads/2023/04/222287_Estudio-Desinformacio%CC%81n_0329_V1-2.pdf

Newman, N. (2023). *Resumen ejecutivo y hallazgos clave del informe de 2023*. https://reutersinstitute.politics.ox.ac.uk/es/digital-news-report/2023/dnr-resumen-ejecutivo

Valenzuela, M. C. C. (n.d.). *La Veracidad de la Información Expuesta en las Redes Sociales*. https://dialnet.unirioja.es/descarga/articulo/6148884.pdf

LARAZON. (2023). *Así han sido las "fake news" más sonadas de este 2023*. https://www.larazon.es/actualidad/asi-han-sido-fake-news-mas-sonadas-este-2023_2023121865804f72d7b0c3000196945d.html

Blog, E. (2022). *Redes sociales más usadas en Latinoamérica en 2022 y cómo aprovecharlas en tu empresa*. https://escala.com/redes-sociales-mas-usadas-latam/

Nieves-Cuervo, G. M. (2021). *Infodemia: noticias falsas y tendencias de mortalidad por COVID-19 en seis países de América Latina*. SciELO. https://www.scielosp.org/article/rpsp/2021.v45/e44/

Habana, L. (2022). *Publicación de noticias falsas durante la pandemia de COVID-19 en el Perú*. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132022000100025

Pichihua Vegas, S. P. (2022). *Tipología de los contenidos virales de desinformación durante los primeros meses de emergencia sanitaria por la COVID-19 en el Perú*. https://revistadecomunicacion.com/article/view/2930/2495

Annexes

Source code for Cataloging Comments on a Publication

```
import tkinter as tk
import matplotlib.pyplot as plt
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import chardet
from textblob import TextBlob

# Función para clasificar comentarios

def clasificar_comentarios(comentarios):
    analyzer = SentimentIntensityAnalyzer()
    categorias = {'In favor': 0, 'Ambiguous': 0, 'Against': 0}

    palabras_clave = ['mentira', 'falso', 'invento', 'fake', 'rumor', 'marketing']
```

```
    palabras_buena = ['bendiciones', 'felicidades', 'dios', 'bendiga']

    for comentario in comentarios:
        comentario = comentario.strip().lower()  # Eliminar espacios y poner en
minúsculas

        # Verificar si hay palabras clave
        if any(palabra in comentario for palabra in palabras_clave):
            categoria = 'Against'  # Si contiene palabras clave, lo clasifica como
'Desmintiendo'
        elif any(palabra in comentario for palabra in palabras_buena):
            categoria = 'In favor'
        else:
            # Analizar sentimiento solo si no contiene palabras clave
            sentiment_score = analyzer.polarity_scores(comentario)

            if sentiment_score['compound'] >= 0.05:
                categoria = 'In favor'  # Alegría
            elif sentiment_score['compound'] <= -0.05:
                categoria = 'Against'  # Enojo
            else:
                categoria = 'Ambiguous'  # Sorpresa

        # Incrementar las categorías
        categorias[categoria] += 1

    return categorias

# Función para mostrar el gráfico de pastel
def mostrar_grafico(categorias):
    # Datos para el gráfico de pastel
    labels = list(categorias.keys())
    sizes = list(categorias.values())

    plt.figure(figsize=(6, 6))
    plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)
    plt.axis('equal')  # Para que el gráfico sea un círculo
    plt.title('Twitter Post Comments Distribution')
    plt.show()

# Función para cargar los comentarios desde un archivo CSV
def cargar_txt():
    comentarios = []

    try:
        # Cambia la ruta al archivo .txt
        archivo = 'twitter.txt'  # Aquí defines el nombre y la ruta del archivo txt
        with open(archivo, 'rb') as file:
            raw_data = file.read()
            result = chardet.detect(raw_data)
            encoding = result['encoding']
        with open(archivo, encoding=encoding) as file:
            # Leer cada línea del archivo de texto
            comentarios = file.readlines()

        if not comentarios:
            print("El archivo no contiene comentarios.")
            return

        # Limpiar las líneas para quitar saltos de línea y comillas si es necesario
```

```
        comentarios = [comentario.strip().replace('"', '') for comentario in
comentarios]

        # Clasificar los comentarios
        categorias = clasificar_comentarios(comentarios)

        # Mostrar el gráfico
        mostrar_grafico(categorias)

    except Exception as e:
        print(f"Ocurrió un error al cargar el archivo: {e}")


# Ejecutar la carga y el análisis
cargar_txt()
```