# Disinformation Contagion: Integrating Data-Driven Insights with Theoretical Model

Nitin Agarwal[1,2]

[1] COSMOS Research Center, University of Arkansas at Little Rock, USA
[2] International Computer Science Institute, University of California, Berkeley, USA
nxagarwal@ualr.edu

**Abstract.** The rapid spread of disinformation on social media platforms poses significant threats to public discourse and democratic institutions. This study introduces a novel fractal-fractional epidemiological model, SEDAZR, that partitions users into Susceptible, Exposed, Disinformed, Anti-disinformed, Skeptic, and Recovered compartments. Leveraging real-world datasets from Twitter, Telegram, and TikTok, the model captures memory-dependent dynamics and nonlinear user transitions observed in online environments. We derive the basic reproduction number $\mathcal{R}_0$ and conduct sensitivity analysis using Latin Hypercube Sampling and Partial Rank Correlation to identify key parameters influencing disinformation spread. Theoretical validation is achieved through existence, uniqueness, and Ulam-Hyers stability analyses, confirming the model's robustness under perturbations. Numerical simulations demonstrate the influence of memory effects and transmission rates on user behavior, while model fitting shows strong alignment with platform-specific data. This integrated framework offers practical insights for designing adaptive mitigation strategies and informs future extensions into multi-platform, demographically aware disinformation control systems.

**Keywords:** Epidemiological model · Mathematical modeling · disinformation · social media.

## 1 Introduction

The rapid diffusion of disinformation on digital platforms has emerged as a critical threat to democratic processes and public trust, exacerbated by the sophisticated interplay of technology, psychology, and geopolitics [1,2]. Platforms such as TikTok and Telegram, with their unique algorithmic and structural features, have become fertile ground for false narratives [3]. TikTok's recommendation algorithm, designed to prioritize high-engagement content, often amplifies sensational or emotionally charged posts, allowing conspiracy theories and misleading claims to achieve viral status [4]. Meanwhile, Telegram's encrypted channels and private groups facilitate the unchecked spread of disinformation within echo chambers, protected from public scrutiny [5]. These dynamics are not merely technological quirks but systemic vulnerabilities exploited by malicious actors to distort public discourse. The consequences are stark and multifaceted [6]. In the realm of public health, baseless vaccine conspiracies such as claims linking COVID-19 vaccines to microchips or infertility contributed to hesitancy, with studies estimating that disinformation may have reduced global vaccination rates by up to 15%, prolonging the pandemic [7]. Politically, fabricated election fraud narratives, like the "Stop the Steal" campaign in the 2020 U.S. elections, have eroded trust in democratic institutions, incited violence, and polarized societies [8]. The reach of such disinformation is staggering: a 2023 MIT study found that falsehoods spread six times faster than factual content on social media, aided by automated bots and AI-driven deepfakes that mimic legitimate sources [9].

Anti-disinformation campaigns, recognizing the urgency of this crisis, have adopted multi-pronged strategies. Fact-checking initiatives, such as those led by the International Fact-Checking Network (IFCN), collaborate with platforms to label or remove false content, though their efforts are often outpaced by the sheer volume of disinformation [10]. Educational programs, like MediaWise for youth, aim to bolster digital literacy, empowering users to critically evaluate sources [11]. Grassroots movements, including #ThinkBeforeYouShare, harness civic engagement to counteract false narratives organically [12]. Tech companies, under regulatory pressure, have introduced policies to remove harmful content and enhance transparency in political advertisements [13].

Gaining insight into the dynamics between disinformation and counter-disinformation efforts demands more than conventional analytical methods. This research adopts a novel hybrid framework that combines epidemiological modeling with fractal-fractional differential equations, effectively capturing both the spread patterns and memory-dependent behaviors found in digital environments. We introduce the SEDAZR model, which categorizes users into six states: Susceptible, Exposed, Disinformed, Anti-disinformed, Skeptic, and Recovered. This structure parallels infectious disease models but is tailored to reflect user interactions specific to digital platforms. Exploring these dynamics is essential for formulating targeted interventions that reduce the impact of misleading content and strengthen democratic resilience. The objective is to uncover significant trends in how information, both deceptive and corrective, spreads and engages audiences, thereby revealing the underlying mobilization traits of each campaign. Through this, the study aims to enhance our understanding of digital influence strategies and support the safeguarding of democratic systems in the digital age.

This paper answers the following **research questions: RQ1:** What are the key mechanisms and parameters that drive or mitigate the virality of disinformation in online ecosystems? **RQ2:** How can we develop a mathematically rigorous and platform-adaptable model to simulate the spread and suppression of disinformation on social media? **RQ3:** What role do memory and temporal patterns play in the spread of digital disinformation? **RQ4:** How can we use the proposed model to dynamically evaluate which is spreading faster - disinformation or anti-disinformation?

## 2 Literature Review

The rapid evolution of disinformation ecosystems has spurred a diverse body of research aiming to understand how false narratives propagate, who spreads them, and how they can be countered effectively. This section outlines the current landscape across three key domains: machine learning approaches to disinformation detection, epidemiological models for information diffusion, and emerging fractal-fractional modeling techniques capturing the complexity of online engagement.

### 2.1 Machine Learning Approaches to disinformation Detection

A substantial volume of work has leveraged machine learning (ML) to identify and classify disinformation in real time. Techniques such as natural language processing (NLP), sentiment analysis, and network analysis have been central to these efforts. Pre-trained language models like BERT, RoBERTa, and GPT have shown success in detecting disinformation across platforms including Twitter, Facebook, and TikTok [14]. These models can identify linguistic cues, flag suspicious content, and even detect coordinated inauthentic behavior. However, many ML models focus primarily on static classification and often lack the ability to capture temporal dependencies or feedback mechanisms within online information ecosystems. Studies such as [1,?] have revealed how disinformation super-spreaders exploit platform algorithms to amplify false narratives, suggesting that detection alone is insufficient to mitigate systemic spread.

### 2.2 Epidemiological Models for Information Spread

Epidemiological modeling has increasingly been applied to digital communication networks to study the diffusion of disinformation, drawing analogies between infectious disease transmission and content spread. Variants of classical models—such as SIR (Susceptible-Infected-Recovered), SEIR (Susceptible-Exposed-Infected-Recovered), and SEIQR (Susceptible-Exposed-Infected-Quarantined- Recovered)—have been adapted to account for the stages of user exposure and belief in false narratives [15]. These models offer a conceptual framework for understanding the tipping points and thresholds at which disinformation becomes viral. More recent models include compartments for skepticism or counter-engagement, acknowledging the role of fact-checkers and critical users. For instance, [16] explored quarantine-style strategies for isolating disinformed users or slowing content spread, echoing methods used in public health.

Despite their utility, traditional epidemiological models often oversimplify user behavior and fail to incorporate cognitive biases, memory effects, or the influence of recommendation algorithms. In the context of anti-disinformation strategies, interventions such as fact-checking, user education, and platform moderation can be modeled analogously to vaccination or treatment policies, but these remain underexplored in conventional compartmental frameworks.

### 2.3 Fractal-Fractional Models for disinformation Dynamics

To address the limitations of classical approaches, researchers have turned to fractal-fractional differential equations, which offer a robust mathematical toolkit for capturing complex, memory-driven, and non-local dynamics in social systems [17] and [18]. Fractal-fractional models extend standard differential equations by incorporating historical states and irregular temporal patterns a key characteristics of disinformation campaigns, where influence can persist long after initial exposure. These models are particularly suited to social media, where algorithmic amplification, user memory, and echo chambers create persistent feedback loops. Incorporating fractal-fractional operators into epidemiological models allows for the simulation of both short-term virality and long-term narrative persistence. For instance, memory effects can reflect repeated exposure to the same narrative across platforms, while non-local diffusion captures the sudden jumps of content across unrelated user communities. However, existing applications of fractal-fractional modeling have focused largely on disease outbreaks and financial contagions, with limited adaptation to the sociotechnical dynamics of disinformation.

## 3 Methodology

This section provides an overview of the data collection and methodology used in this paper.

### 3.1 Fractal-Fractional Model Formulation

In this study, we introduce a fractal-fractional model to analyze the spread of disinformation and anti-disinformation within an online network. The model considers six distinct compartments: Susceptible ($S(t)$), representing users who can be influenced by disinformation; Exposed ($E(t)$), referring to individuals who have encountered misleading content; Disinformed ($D(t)$), who actively propagate false information; Anti-disinformed spreaders ($A(t)$), who counteract disinformation; Skeptic ($Z(t)$), representing users who question information before sharing it; and Recovered ($R(t)$), individuals who have become resistant to misleading content. Using a system of fractal-fractional ordinary differential equations (FFODEs), the proposed SEDAZR model captures the complex interactions between these groups. We present a transfer diagram in Fig. 1 and parameters in Table 1 to illustrate the flow of users between different compartments.

The system of ordinary differential equations (ODEs) is given as follows:

$$\begin{cases} {}_{0}^{FFP}D_t^{\alpha,\beta}S(t) = \mathit{\Pi} + \sigma R(t) - \dfrac{\beta_1(A+D)}{N(t)}S(t) + \xi Z(t) - \eta S(t), \\ {}_{0}^{FFP}D_t^{\alpha,\beta}E(t) = \dfrac{\beta_1(A+D)}{N(t)}S(t) - (\psi + \lambda + p + \eta)E(t), \\ {}_{0}^{FFP}D_t^{\alpha,\beta}A(t) = \psi E(t) + \kappa_d D(t) - (\gamma_a + \kappa_a + \eta)A(t), \\ {}_{0}^{FFP}D_t^{\alpha,\beta}D(t) = \kappa_a A(t) + \lambda E(t) + \theta Z(t) - (\gamma_d + \kappa_d + \eta)D(t), \\ {}_{0}^{FFP}D_t^{\alpha,\beta}Z(t) = pE(t) - (\theta + m + \xi + \eta)Z(t), \\ {}_{0}^{FFP}D_t^{\alpha,\beta}R(t) = \gamma_a A(t) + \gamma_d D(t) - (\eta + \sigma)R(t), \end{cases} \tag{1}$$

where ${}_{0}^{FFP}D_t^{\alpha,\beta}(.)$ is the fractal-fractional derivative with the fractional order $0 < \alpha \leq 1$ and fractal dimension $0 < \beta \leq 1$ in the Caputo sense with power law type kernel, and the variables are assumed to be non-negative with appropriate initial conditions.

Incorporating fractal-fractional operators enhances traditional models by considering memory effects, spatial heterogeneity, and anomalous diffusion, providing a more accurate representation of SEDAZR dynamics. This approach is crucial for devising effective control strategies.

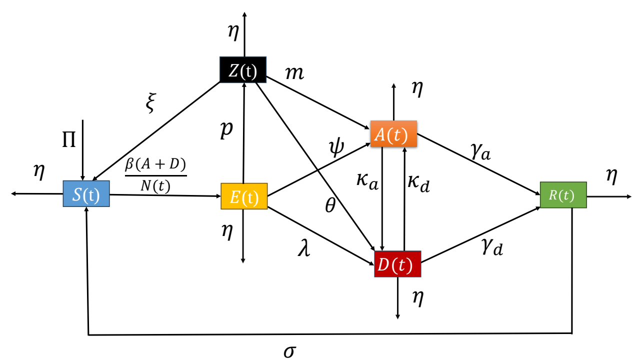



**Fig. 1** Transfer diagram for the Anti-dis/disinformation spread on the social network platform

### 3.2 Data Collection and Analysis

To evaluate the robustness and cross-platform adaptability of the proposed SEDAZR model, we curated and analyzed disinformation-related datasets from

**Table 1** Interpretation of parameters in the model

| Parameter | value | source | Interpretation |
|---|---|---|---|
| $\Pi$ | 1000 | Fitted | Recruitment rate which new users joining the platform |
| $\xi$ | 0.0021 | Fitted | Rate at which skeptics user revert become susceptible after exposed |
| $\beta_1$ | 0.0014 | Fitted | Effective contact rate |
| $\eta$ | 0.000233 | Fitted | the rate at which users naturally leave the platform. |
| $p$ | 0.000375 | Assumed | the rate at which exposed users become skeptic |
| $m$ | 0.000375 | Assumed | the rate at which skeptic users disagree with disinformation |
| $\theta$ | 0.000375 | Assumed | the rate at which skeptic users agree with disinformation |
| $p$ | 0.000375 | Assumed | the rate at which exposed users become skeptic |
| $\lambda$ | 0.01 | Assumed | the rate at which $E(t)$ transition to $D(t)$ |
| $\psi$ | 0.30 | Assumed | the rate at which $E(t)$ transition to $A(t)$ |
| $\kappa_a$ | 0.000375 | Assumed | the rate at which $(A(t))$ become $(D(t))$ |
| $\kappa_d$ | 0.000375 | Assumed | the rate at which $(D(t))$ become $(A(t))$ |
| $\sigma$ | 0.30 | Assumed | the rate at which $(r(t))$ users become $(S(t))$ |
| $\gamma_a$ | 0.001 | Fitted | Rate at which $(A(t))$ move to recovery |
| $\gamma_d$ | 0.001 | Fitted | Rate at which $(D(t))$ move to recovery |

**X (formerly Twitter)**, **Telegram**, and **TikTok**. Each dataset corresponds to a distinct real-world case study characterized by polarized discourse and coordinated narrative campaigns. Below is a summary of each dataset and its collection process:

- **COVID-19 Vaccine (Twitter):** Tweets were collected between *December 22, 2020 and June 14, 2021*, focusing on conspiracy narratives linking COVID-19 vaccines to adverse outcomes. We used keyword filters such as `"covid," "coronavirus," "vaccine," "microchip," "5G," "infertility,"` and `"Bill Gates"`, coupled with Boolean operators and hashtag matching (e.g., #DoNotComply, #VaccineHoax). Tweets were retrieved via the Twitter Academic API and manually labeled as *pro-disinformation* (n = 1,673) or *anti-disinformation* (n = 176). This dataset captures both virality potential and resistance within the Twitter ecosystem.
- **Russia–Ukraine War (Telegram):** Telegram messages were sourced from high-subscriber public channels aligned with *Pro-Kremlin* and *Pro-Ukraine* stances. Over *4.75 million messages spanning 120 days* were collected using automated web crawlers with topic-specific filters such as `"Donbas," "NATO," "invasion," "neo-Nazi," "Kyiv regime,"` and `"biolabs"`. Manual annotation was conducted by trained coders fluent in Russian and Ukrainian, with high inter-rater agreement (Cohen's $\kappa \geq 0.87$). This dataset reflects cross-border disinformation strategies and ideological framing.
- **Taiwan Election (TikTok):** Short videos were collected from TikTok between *January 13 and 27, 2024*, surrounding Taiwan's presidential election. We used a hybrid strategy combining *keyword-based crawling* (e.g., `"Taiwan election," "disinformation," "DPP," "KMT," "US-China-Taiwan"`) with *unsupervised topic clustering* to extract relevant narratives. Videos were an-

notated into *anti-disinformation* (n = 130) and *disinformation* (n = 58) based on content veracity, source type, and narrative framing. This case highlights how algorithmically amplified short-form content can influence electoral integrity.

## 4 Model Analysis and Result

In this section, we conduct a qualitative analysis of the proposed SEDAZR model to better understand the dynamics of disinformation and anti-disinformation propagation in online networks. The analysis focuses on key dynamical properties such as the basic reproduction number $\mathcal{R}_0$ sensitivity analysis, existence and uniqueness of solutions, stability analysis, data fitting analysis, and numerical analysis and simulation.

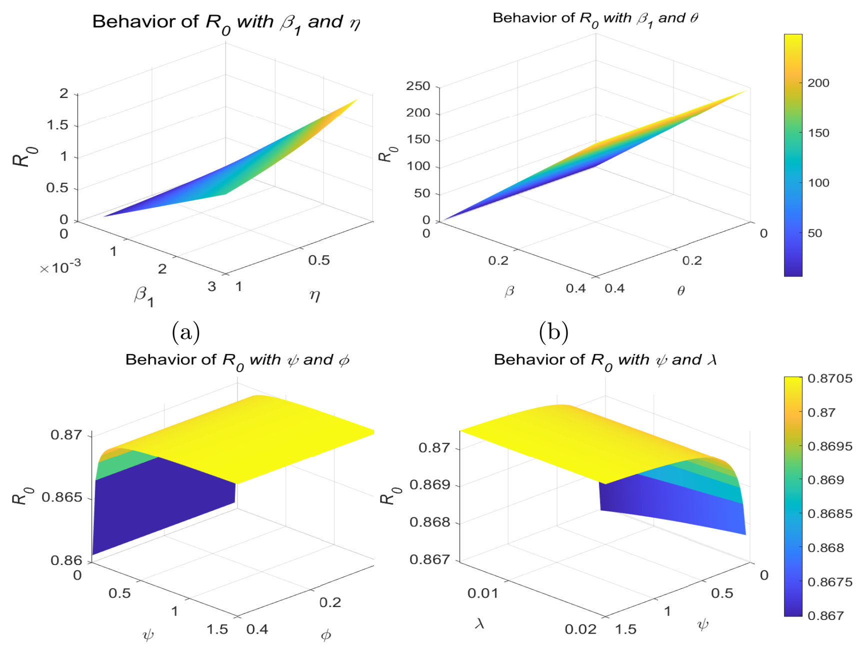
### 4.1 Basic reproduction number, $\mathfrak{R}_0$

The basic reproduction number $\mathcal{R}_0$ measures the expected number of secondary disinformation cases generated by a single spreader in a fully susceptible network. Using the next-generation matrix approach, $\mathcal{R}_0$ is derived from the infection-related compartments $E(t)$, $D(t)$, and $A(t)$, capturing both disinformation and anti-disinformation dynamics.

A simplified form is:

$$\mathcal{R}_0 = \frac{\beta_1}{\eta + \lambda + \psi} \left( \frac{\lambda}{\gamma_d + \kappa_d + \eta} + \frac{\psi}{\gamma_a + \kappa_a + \eta} \right)$$

Here, $\beta_1$ is the contact rate; $\lambda$, $\psi$, $\kappa_d$, and $\kappa_a$ are transition rates; and $\gamma_a$, $\gamma_d$, $\eta$ denote recovery and attrition. The condition $\mathcal{R}_0 > 1$ signals potential disinformation proliferation.




(a) (b) (c) (d)

**Fig. 2** Effect of $\mathfrak{R}_0$ on (a) $\eta$ and $\beta_1$, (b) $\beta_1$ and $\theta$, (c) $\phi$ and $\psi$, and (c) $\lambda$ and $\psi$

### 4.2 Sensitivity Analysis of $\mathcal{R}_0$

To identify the most influential parameters driving the spread of disinformation, we employed the Latin Hypercube Sampling (LHS) method combined with Partial Rank Correlation Coefficient (PRCC) analysis. This approach quantifies the sensitivity of the basic reproduction number $\mathcal{R}_0$ to model parameters, providing insights into which variables should be targeted to suppress disinformation.

The resulting PRCC values, summarized in Fig. 3, indicate the degree of monotonic influence each parameter exerts on $\mathcal{R}_0$. Parameters with high absolute PRCC values are most impactful.

**Key Findings:**

- $\beta_1$ (effective contact rate) exhibited the strongest positive correlation with $\mathcal{R}_0$, implying that reducing exposure to disinformation content significantly lowers disinformation spread.
- $\psi$ and $\lambda$ (transition rates from exposed to anti-/disinformation compartments) also had strong positive correlations, indicating the importance of early interventions post-exposure.
- Recovery parameters $\gamma_a$, $\gamma_d$ and user attrition rate $\eta$ showed negative correlations with $\mathcal{R}_0$, suggesting that increased user resistance or disengagement mitigates spread.

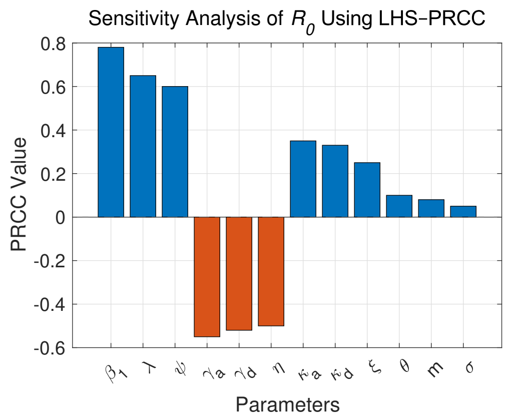



**Fig. 3** Sensitivity of $\mathcal{R}_0$ of the online disinformation and anti-disinformation contagious

- Parameters such as $\kappa_a$, $\kappa_d$, and $\xi$ had moderate impacts, while others like $m$, $\theta$, and $\sigma$ were less influential.

These results highlight critical levers for disinformation control—most notably reducing contact/exposure, increasing recovery or skepticism, and enhancing anti-disinformation efforts. Hence these results give a clear response to our **RQ1**.

### 4.3 Existence and Uniqueness of Solution

To validate the well-posedness of the proposed fractal-fractional $SEDAZR$ model, we establish the existence and uniqueness of solutions using fixed-point theory in the context of non-integer order derivatives with fractal dimensions. The proposed model 1 has generalized form denoted as

$$ {}_0^{FFP}D_t^{\alpha,\beta}\mathcal{H} = \frac{\beta}{\Gamma(1-\alpha)} \int_0^t (t-s)^{-\alpha} s^{\beta-1} \mathcal{H}\, ds, \quad 0 < \alpha \leq 1,\ 0 < \beta \leq 1 $$

We transform the system into a Volterra-type integral equation:

$$ \mathcal{H}(t) = \mathcal{H}_0 + \frac{\beta}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} s^{\beta-1} \Upsilon(\mathcal{H}(t))\, ds $$

where $\Upsilon(t) \in B$, a Banach space of continuous functions on $[0, T]$, and $\mathcal{H}$ is a nonlinear functional representing the model dynamics.

We show:

(i) $\Upsilon$ satisfies a Lipschitz condition: $\|\Upsilon(\mathcal{H}_1) - \Upsilon(\mathcal{H}_2)\| \leq K\|\mathcal{H}_1 - \mathcal{H}_2\|$ with $K = \frac{\beta T^{\beta+\alpha-1}\mathcal{B}(\beta,\alpha)}{\Gamma(\alpha+1)} < 1$

(ii) The integral operator $\Upsilon$ is compact and continuous
(iii) $\Upsilon(\mathcal{H}) \subseteq B$ is uniformly bounded

Using conditions of Schauder's and Banach's fixed-point theorems, we conclude that there exists a unique continuous solution $\mathcal{H}$ to the system over $t \in [0, T]$. This confirms that the model is analytically consistent and suitable for further analysis. For analytical proof see [17] and [18]. From here we say our model is well-posed which answer **RQ2.**

### 4.4 Ulam-Hyers Stability

To assess the robustness of the proposed fractal-fractional model under perturbations, we investigate Ulam–Hyers stability using nonlinear functional analysis. This concept ensures that small deviations in initial conditions or parameters lead only to small deviations in the solution, thereby validating the model's reliability for numerical approximations.

The model is said to be Ulam–Hyers stable if for any approximate solution $\tilde{\mathcal{H}}(t)$ satisfying:

$$\left\| {}_{0}^{FFP}D_t^{\alpha,\beta}\tilde{\mathcal{H}}(t) - \mathcal{F}(t, \tilde{\mathcal{H}}(t)) \right\| \leq \epsilon,$$

there exists a true solution $\mathcal{H}(t)$ such that

$$\|\tilde{\mathcal{H}}(t) - \mathcal{H}(t)\| \leq C\epsilon, \quad \forall t \in [0, T],$$

where $C > 0$ is a constant dependent on the Lipschitz structure of the operator.

For analytic proof, see [15, 16] since we use the same approach, hence we omit the proof. The proof utilizes fixed-point theory, showing that the nonlinear operator defined by the model satisfies both Lipschitz continuity and contraction conditions within a Banach space. These ensure the existence of a unique fixed point that attracts all nearby approximate trajectories.

This result confirmed that the proposed system exhibits Ulam-Hyers stability, guaranteeing that the model's solutions are structurally stable under small perturbations, and hence suitable for numerical simulation.

### 4.5 Numerical Analysis and Simulation

To answer our **RQ3** and **RQ4**, we simulated the proposed fractal-fractional SEDAZR model using the Caputo derivative with a power-law kernel to capture memory effects and non-local dynamics in online disinformation ecosystems. The numerical approximation is carried out using an Adams-Bashforth scheme adapted for fractal-fractional systems.

Our analysis focuses on: (1) the influence of the fractal-fractional orders $(\alpha, \beta)$ on user transitions, and (2) the impact of the disinformation transmission rate $(\beta_1)$ on compartmental behavior. In Fig. 4(a), the higher values of $(\alpha, \beta)$ lead to a faster reduction in the susceptible population, indicating quicker user engagement with disinformation. Lower values reflect stronger memory, delaying user conversion and mimicking content fatigue or selective attention. In Fig. 4(b)–(f), we observe that lower values of ($\alpha$) and ($\beta$) correspond to slower growth across all compartments, indicating that stronger memory effects dampen user transitions. Specifically, the Exposed, Disinformed, Anti-disinformed, Skeptics, and Recovered populations grow more gradually when $(\alpha, \beta)$ are small. This suggests that users in high-memory environments engage more cautiously with both disinformation and counter-narratives. Overall, lower $(\alpha, \beta)$ values capture fewer users across all dynamic states, reflecting delayed response and reduced information spread. From a practical standpoint, memory-aware modeling offers actionable insights: network service providers can design recommender systems and mitigation strategies that incorporate user history, limit redundant exposure, and mitigate content virality—ultimately dampening the spread of both disinformation and anti-disinformation spread. From Fig. 5(b)–(f), we observe that increasing the disinformation transmission rate $\beta_1$ results in higher peaks across all compartments, except susceptible. This indicates that stronger transmission accelerates user transitions and expands the overall reach, resembling the viral nature of highly shareable or sensational disinformation. As disinformation spreads more aggressively, it also triggers faster counter-responses and a quicker rise in user skepticism.

### 4.6 Model Fitting and Parameter Estimation

To capture the dynamics of disinformation and anti-disinformation spread, precise parameter estimation is vital for predictive accuracy and effective intervention design. We adopted the *Non-Linear Least Squares Method (NLSM)* to estimate model parameters by minimizing the sum of squared errors (SSE) between observed data and model predictions. The error formula is defined as

$$\text{SSE} = \sum_{i=1}^{n} \left(y_i - f(x_i; \theta)\right)^2$$

where $y_i$ are observations, $f(x_i; \theta)$ are model outputs, and $\theta$ denotes the parameter vector.

This process tailors the model to platform-specific data, enabling strategic disinformation and anti-disinformation spread control. Figures in Fig. 6 illustrate how well the model fits observed disinformation and anti-disinformation data across different topics and platforms. In Fig. 6a disinformation tweets show a sigmoidal growth pattern typical in social contagion, while Fig. 6b anti-disinformation tweets exhibit a more linear and steady rise. We found anti-disinformation efforts tended to respond more consistently but perhaps less virally than disinformation. Figs. 6c and 6d show two dynamics of disinformation exchange from different ideological groups. These figures exhibit more complex waves in disinformation propagation. And the last Figs. 6e and 6f for the Taiwan election on TikTok have low error values, reflecting the effectiveness of model calibration even on short-video platforms.

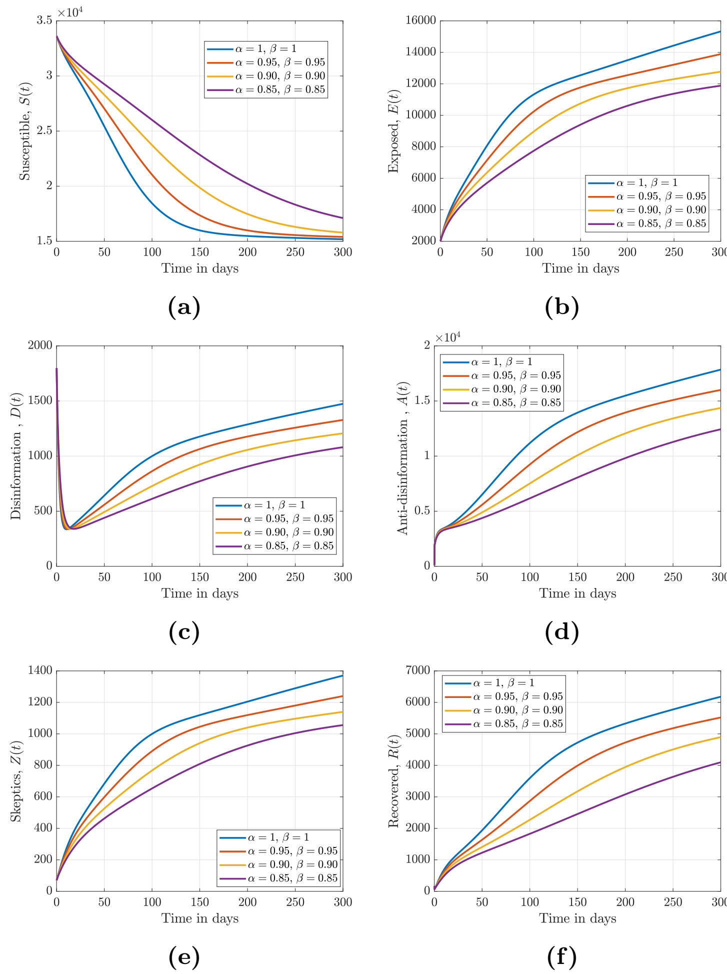



**Fig. 4** Numerical trajectories of $SEIQR$ model under the Caputo fractal-fractional operator.

## 5 Conclusion and Future Works

This study introduces a novel SEDAZR epidemiological model, enhanced with fractal-fractional operators, to understand the intertwined dynamics of disinformation and anti-disinformation campaigns on social media platforms. Through a rigorous framework incorporating data from Twitter, Telegram, and TikTok, the model successfully captures the temporal evolution and memory-dependent behaviors of users engaging with misleading or corrective narratives.

Key findings include the derivation of the basic reproduction number $\mathcal{R}_0$, which quantifies the spread potential of disinformation, and its sensitivity to key transmission and recovery parameters. The Ulam-Hyers stability and existence-

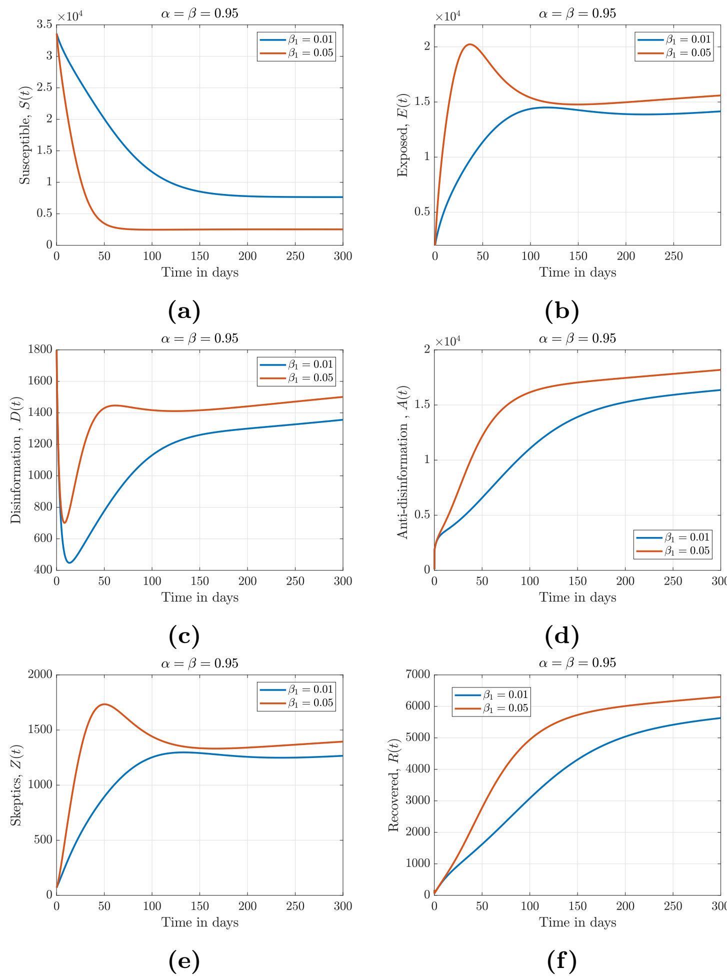
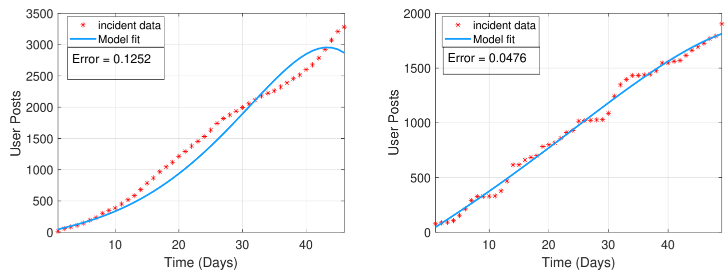



**Fig. 5** Numerical trajectories of $SEDAZR$ model under the Caputo fractal-fractional operator with order and dimension ($\alpha = \beta = 0.95$) when one varies the rate of transmission.

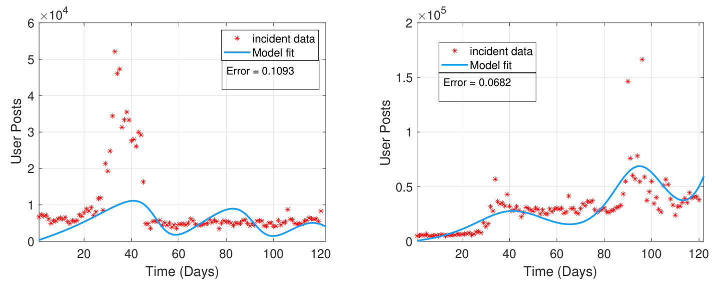
uniqueness analysis affirm the model's robustness under perturbations. Numerical simulations demonstrate that both memory effects and transmission rates play crucial roles in shaping the spread dynamics. Model fitting across real-world case studies revealed strong alignment with empirical data, underscoring the model's applicability for platform-specific interventions.

Looking forward, future work will explore adaptive control strategies leveraging real-time data to dynamically adjust platform responses. Integration with agent-based modeling and reinforcement learning could enable scenario testing for disinformation mitigation policies. Additionally, extending the model to multi-platform interaction networks and incorporating user demographic and be-



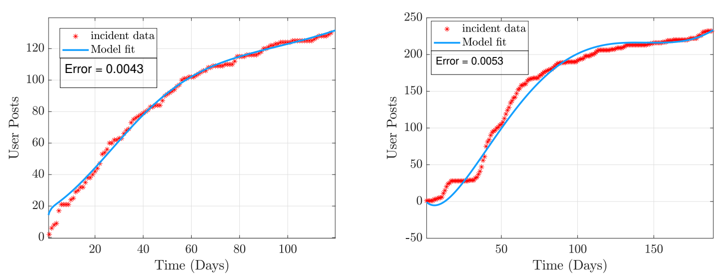

**(a)** Model fitting for users sharing disinformation tweets about COVID-19 vaccine

**(b)** Model fitting for users sharing Anti-disinformation tweets about COVID-19 vaccine




**(c)** Model fitting for users sharing disinformation about Ukraine/Russia war on Telegram (Pro-Ukraine)

**(d)** Model fitting for users sharing disinformation about Ukraine/Russia war on Telegram (Pro-Russia)




**(e)** Model fitting for users sharing disinformation content related to Taiwan's election campaign on TikTok

**(f)** Model fitting for users sharing anti-disinformation content related to Taiwan's election campaign on TikTok.

**Fig. 6** The $SEDAZR$ model fitted to disinformation and anti-disinformation posts on different platform.

havioral heterogeneity will offer more granular insights for combating the evolving landscape of digital disinformation.

**Acknowledgements** This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army

Research Office (W911NF-23-1-0011, W911NF-24-1-0078, W911NF-25-1-0147), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, the Australian Department of Defense Strategic Policy Grants Program, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Donaghey Foundation at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

## References

1. Cinelli, M., Quattrociocchi, W., Galeazzi, A., et al.: The COVID-19 social media infodemic. Scientific Reports 10, 16598 (2020).
2. Wang, Y., et. al.: Systematic literature review on the spread of health-related disinformation on social media. Social Science & Medicine, 240, 112552 (2019).
3. Cinelli, M., et al.: The echo chamber effect on social media. Proceedings of the National Academy of Sciences 118(9) (2021).
4. Roth, Y., Achuthan, K.: Algorithmic amplification and the limits of content moderation. Brookings Institution TechStream (2021).
5. Nguyen, T.T., et al.: Echo chambers and information spread on social media: a case study on Telegram during geopolitical crises.
6. Velásquez, N., et al.: Hate multiverse spreads malicious COVID-19 content online beyond individual platforms. Scientific Reports 11, 1–13 (2021).
7. Wilson, S.L., Wiysonge, C.: Social media and vaccine hesitancy. BMJ Global Health, 5(10):e004206 (2020).
8. Ballard, A.O., et al.: Stop the Steal: Tracking the Rise of Election disinformation Narratives. Brookings Institution, (2021).
9. Vosoughi, S., et. al.: The spread of true and false news online. Science, 359(6380), 1146–1151 (2018).
10. Wardle, C., Derakhshan, H.: Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Council of Europe report, DGI(2017)09, (2017).
11. MediaWise: Helping people of all ages identify disinformation online. The Poynter Institute.
12. United Nations: #ThinkBeforeYouShare Campaign.
13. Twitter Transparency Center: Political content policy updates (2022).
14. Shu, K., Sliva, A., et al.: Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36, 2017.
15. Wang, L., Chen, X., et al.: Epidemic spreading on complex networks with general degree and weight distributions, Scientific Reports, 9, 12520, 2019.
16. Tambuscio, M., et al.: Fact-checking effect on viral hoaxes: A model of disinformation spread in social networks, Proceedings of the 24th International Conference on World Wide Web (WWW), 2015.
17. Baleanu, D., et al.: On a fractional-order SIR epidemic model with Atangana–Baleanu derivatives. Chaos, Solitons & Fractals, 117, 409–417, 2019.
18. Khan, M. A., et. al.: A new fractal–fractional SEIR model with non-singular kernel: Application to COVID-19 data. Results in Physics, 21, 103817, 2021.