Memory-Driven Dynamics and Repeated Patterns in Online Toxicity: A Fractal-Fractional SEIQR Modeling Approach

Nitin Agarwal*†
*COSMOS Research Center, University of Arkansas at Little Rock, Arkansas, USA
†International Computer Science Institute, University of California, Berkeley, California, USA
e-mail: nxagarwal@ualr.edu

***Abstract*—This study employs an epidemiological framework called SEIQR, which stands for Susceptible, Exposed, Infected, Quarantine, and Recovered, to analyze the spread of toxicity in online environments, integrating toxicity intensity stratification and fractal-fractional dynamics to capture the complexity of toxicity propagation. Using datasets from Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), commonly known as coronavirus disease (COVID-19), and social movement discussions from the Latin American countries of Brazil and Peru, we conduct sensitivity analysis to evaluate key parameters influencing toxicity diffusion. The results reveal that splitting toxicity into moderate and high levels significantly reduces model error rates, enhancing predictive accuracy across all datasets. Additionally, our findings indicate that the basic reproduction number (R_0) is highly sensitive to exposure and quarantine rates, emphasizing the critical role of enhanced moderation and adaptive quarantining in suppressing toxicity. By incorporating fractal-fractional operators, the model further accounts for memory-driven toxicity spread, long-range dependencies, and network complexity. The study demonstrates that reducing the fractal dimension (τ) and fractional order (α) slows toxicity propagation, highlighting the influence of user behavior history and structural dynamics in toxicity persistence. Moreover, quarantine interventions and content demotion strategies are shown to significantly curb toxicity intensity while maintaining engagement dynamics. These insights provide a foundation for policy-driven interventions, enabling social media platforms to implement optimized content moderation, algorithmic intervention, and network-level strategies to mitigate online toxicity and promote healthier digital ecosystems.**

***Keywords*—mathematical model; toxicity; social media; epidemiological modeling; hate speech; COVID-19; Brazil; Peru.**

I. Introduction

The rapid evolution of social media platforms has transformed the way users engage in communication, share information, and participate in public discourse. These platforms serve as catalysts for the swift dissemination of news, educational resources, and discussions on critical societal matters. By enabling real-time interaction across diverse audiences, social media fosters global connectivity and awareness. However, alongside these beneficial aspects, social media has also unintentionally become a breeding ground for the amplification of toxic behaviors [1]. This toxicity manifests in various forms, including hate speech, misinformation, harassment, cyberbullying, and online extremism. The unrestricted and algorithm-driven nature of social platforms often facilitates the viral spread of harmful content, influencing not only individual users but also wider societal dynamics [2], [3], [4].

The proliferation of toxicity on digital platforms has far-reaching consequences that extend beyond online environments, affecting psychological well-being, social cohesion, and even real-world actions. Misinformation-driven narratives have contributed to public unrest, political polarization, and radicalization, demonstrating the tangible effects of unchecked digital toxicity [5]. In extreme cases, the rapid circulation of false information has led to public health crises, economic disruptions, and coordinated disinformation campaigns aimed at manipulating public opinion [6]. As a result, addressing the spread and impact of online toxicity has become an urgent research priority that demands multidisciplinary approaches and advanced analytical frameworks [3], [4], [7].

Given the increasing reliance on digital communication and algorithmic content curation, it is imperative to understand the underlying mechanisms that drive toxicity propagation across interconnected online communities [8]. The primary challenge lies in developing intervention strategies that effectively mitigate toxic interactions while upholding fundamental rights to freedom of speech and avoiding excessive restrictions on public discourse [1]. Striking a balance between content regulation and open dialogue is crucial to maintaining healthy digital ecosystems that encourage constructive engagement.

The complexity and scale of modern social networks necessitate the adoption of computational modeling and data-driven methodologies to analyze the spread, persistence, and impact of online toxicity. By leveraging mathematical modeling, network analysis, and artificial intelligence, researchers can identify key transmission patterns, predict emerging toxicity trends, and develop targeted interventions that curb digital toxicity while preserving online freedoms [9]. These insights will help policymakers, social media platforms, and researchers formulate evidence-based strategies to foster safer and more inclusive online environments.

Mathematical models have been widely employed in epidemiology to study the spread of infectious diseases. The same principles can be applied to model the dissemination of online toxicity, given their shared characteristics: contact-based transmission and behavioral contagion. Various epidemic models, including the Susceptible-Infected-

Recovered (SIR), Susceptible-Infected-Susceptible (SIS), and Susceptible-Exposed-Infected-Recovered (SEIR) models, have been adapted to analyze social media toxicity.

In this study, we extend the Susceptible, Exposed, Infected, Quarantined, Recovered (SEIQR) model by incorporating toxicity intensity levels, classifying individuals into moderately and highly toxic categories. This refinement allows for more precise intervention strategies, as different toxicity levels necessitate distinct mitigation approaches. We validate our model using real-world datasets from online discussions related to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), commonly known as coronavirus disease (COVID-19), and social justice movements in the Latin American countries of Brazil and Peru, enabling the identification of critical patterns in toxicity spread.

Traditional epidemiological models typically assume homogeneous transmission dynamics. However, online toxicity propagation exhibits several unique features that necessitate a more advanced modeling framework. Specifically:

i. **Fractal network topology:** Social media networks exhibit self-similarity and scale-free structures, where a small number of *superpropagators* significantly influence the spread of toxicity. Fractal-fractional models better represent these non-uniform dynamics.
ii. **Memory effects in user interactions:** Unlike disease transmission, online toxicity is affected by past exposures and historical user behaviors. Fractional calculus captures these long-range dependencies and persistent memory effects.
iii. **Superpropagation and feedback loops:** Toxicity dissemination is often amplified through echo chambers and algorithmic recommendations. The fractal-fractional approach models these reinforcement mechanisms more effectively than integer-order differential equations.

By integrating fractal-fractional operators into the SEIQR model, we capture the irregularity, persistence, and nonlocality of online toxicity spread. This approach allows for improved predictive capabilities and better policy recommendations for online moderation.

This study addresses the following key research questions:

ii **RQ1:** What are the key parameters in the SEIQR model that have the most significant impact on the *basic reproduction number* R_0, and how can these parameters be controlled to mitigate toxicity spread?
i **RQ2:** How do variations in toxicity intensity (moderate vs. high) influence its dissemination, and how does intensity impact the SEIQR model's performance?
iii **RQ3:** How does the integration of fractal-fractional operator influence the accuracy and predictive capabilities of toxicity spread models compared to traditional integer-order models?
iv **RQ4:** What are the effects of superpropagation and long-term memory on toxicity dissemination, and how can quarantine strategies be optimized in this context?

The remainder of the paper is organized as follows: Section II provides a review of related work on toxicity propagation in social media and epidemiological modeling techniques. Section III introduces the methodology, detailing the SEIQR model formulation and the fractal-fractional extensions. Section IV presents the experimental results and discusses the impact of different toxicity intensities. Section V presents the results and discussion. Finally, Section VI concludes the study with recommendations for future research and policy interventions.

II. Related Work

The rise of online toxicity in social media environments has prompted extensive research into its dynamics, effects, and mitigation strategies. With the increasing spread of harmful or toxic content, researchers have employed various computational, epidemiological, and network-based models to understand and control the propagation of toxicity.

A. Machine Learning Approaches to Toxicity Detection

Several studies have leveraged machine learning techniques to detect, classify, and analyze toxicity in online discussions. For example, researchers in [10], [11] analyzed COVID-19 misinformation and toxicity across social media platforms, revealing notable differences in the extent and nature of toxic behaviors on Twitter, Reddit, and Facebook. Their findings emphasized the role of key user groups (super-spreaders) in amplifying misinformation and toxic content. Similarly, natural language processing models such as BERT, GPT, and Detoxify have been used to detect toxicity in real time [12], [13]. These models have been instrumental in the identification of various forms of toxic language, including hate speech, harassment, cyberbullying, and misinformation. Furthermore, studies in [14] investigated the dynamics of discussion threads on Reddit, where toxicity spreads hierarchically within nested conversations. However, these machine learning-based approaches largely focus on static classification of toxicity, often overlooking long-term propagation effects and feedback loops that sustain toxic environments.

B. Epidemiological Models for Online Toxicity Spread

Epidemiological modeling has proven to be an effective framework for understanding and predicting the spread of online toxicity by drawing analogies between toxic content dissemination and infectious disease transmission. Traditional models such as SIR, SIS, and SEIR have been adapted to analyze how toxicity spreads within digital communities [4]. The Susceptible-Toxic-Recovered-Susceptible (STRS) model has further refined this analogy by incorporating recovery mechanisms where users disengage from toxic interactions [15], [16], [17].

Moreover, quarantine-based control strategies have been widely adopted in epidemiological models to curb infectious spread, an approach that has direct applications in content moderation and toxicity mitigation. Authors in [18] studied the impact of content quarantining on social media toxicity,

demonstrating that isolating toxic users reduces the overall infection rate in a network-based toxicity model. Similarly, authors in [15] applied epidemiological modeling to study COVID-19 misinformation spread, highlighting the role of user quarantine, moderation interventions, and algorithmic content suppression.

C. Fractal-Fractional Modeling for Toxicity Dynamics

A growing body of research has focused on fractal-fractional approaches to model complex social dynamics. Unlike classical integer-order epidemiological models, fractal-fractional models capture memory effects, non-locality, and irregular diffusion patterns [19], [20]. These properties are particularly relevant to online toxicity modeling, where the following are true:

1) Fractal network topology leads to self-similarity and scale-free structures, where toxicity is disproportionately amplified by a small subset of highly influential users (superpropagators).
2) Memory effects in user interactions suggest that toxicity is influenced by past exposures and behavioral reinforcement, rather than instantaneous interactions.
3) Superpropagation and feedback loops occur due to algorithmic content amplification and echo chamber effects, making toxicity diffusion highly persistent.

Fractal-fractional epidemiological models have focused primarily on disease transmission and financial systems, with limited application to online toxicity dynamics. For disease models, see [21], [22] and references therein. Despite extensive research in toxicity detection, epidemiological modeling, and fractal-fractional approaches, there are several gaps. To address these gaps, this study introduces a fractal-fractional extension of the SEIQR epidemiological model, uniquely designed to capture memory effects, long-range dependencies, and heterogeneous toxicity diffusion patterns inherent to online platforms. By unifying fractal-fractional operators, which encode non-local interactions and irregular propagation dynamics—with a stratified SEIQR framework that distinguishes between moderate- and high-intensity toxicity, this work advances a novel computational framework for analyzing and mitigating persistent toxicity. To our knowledge, this represents the first integration of fractal-fractional calculus with toxicity-specific epidemiological modeling, enabling a granular understanding of how algorithmic amplification and user memory shape long-term propagation.

Building on this foundation, the study also addresses critical limitations in prior SEIQR analyses by conducting a rigorous sensitivity analysis. We identify the parameters most influential to the basic reproduction number $\mathcal{R}_0$, thereby pinpointing actionable levers—such as platform moderation efficacy or user recovery rates—for disrupting toxicity spread. This dual focus on mechanistic modeling (via fractal-fractional dynamics) and practical control strategies (via sensitivity-driven interventions) bridges a key gap between theoretical epidemiology and real-world social media governance.

III. Methodology

This section outlines the data collection framework, analytical methods, and computational procedures employed to validate the proposed fractal-fractional SEIQR model.

A. Data Collection and Analysis

To validate the model, we analyzed two distinct datasets: (1) discourse related to the COVID-19 pandemic (spanning February 2020–June 2021) and (2) social movements in Brazil (spanning November 1, 2022–February 25, 2023) and Peru (spanning December 7, 2022–January 31, 2023). Public posts were collected via X's (formerly Twitter) Academic API, focusing on keywords and hashtags associated with polarized discourse.

- **COVID-19 Dataset**: Included topics such as pandemic policies, face mask mandates, lockdowns, and 5G conspiracy theories. Key hashtags: #fckthecovid/s, #fckyourmask/s, #f*cklockdown/s, #5GCoronavirus.
- **Brazilian Protests**: Centered on post-election unrest following the October 30, 2022, presidential election, with hashtags like #semanistia and #SOSbrasil reflecting demands for military intervention and counter-protests.
- **Peruvian Protests**: Captured anti-government demonstrations after President Pedro Castillo's removal on December 7, 2022, using hashtags such as #peruprotest/s.

These datasets have less sparsity in terms of toxic discussions, which is important for this study, where the objective is to examine spread of toxicity in high vs moderate toxic environments. Contexts where toxicity is negligible or highly sparse are outside the scope of this study.

1) Toxicity Classification: Toxicity scores for each post were computed using Detoxify [23], a pre-trained deep learning model that evaluates text for harmful content. Detoxify employs convolutional neural networks and semantic embeddings to assign a toxicity probability between 0 (non-toxic) and 1 (highly toxic). Posts with scores of 0.5 or more were classified as toxic; those below 0.5 were deemed non-toxic.

To analyze toxicity intensity, the toxic subset was further stratified:

- **Moderate Toxicity**: Posts with scores below the dataset-specific average toxicity score.
- **High Toxicity**: Posts exceeding the average score.

Table I summarizes the distribution of high/moderate toxic posts and average scores across datasets.

2) User Activity and Quarantining: To identify superpropagators, we isolated the top 10% of users by activity level, defined as the number of retweets per post [18]. These high-engagement users were algorithmically transferred to a quarantine compartment in the SEIQR model, simulating platform-level interventions to curb toxicity spread.

B. Model Formulation

Online toxicity is a growing epidemic on digital platforms, characterized by high transmission rates, latent behavior, and

TABLE I: Toxicity Dataset Statistics

Dataset	Num. of Posts	Avg. Toxicity	Num. of High Toxic	Num. of Moderate Toxic
F*Covid	28,131	0.91	4,684	2,082
F*Mask	2,423	0.91	538	217
F*Lockdown	1,995	0.82	598	493
5G	33,403	0.84	1,096	703
Brazil Anti	405,160	0.70	1,221	2,309
Brazil Pro	44,415	0.75	105	131
Peru	195,290	0.71	511	546

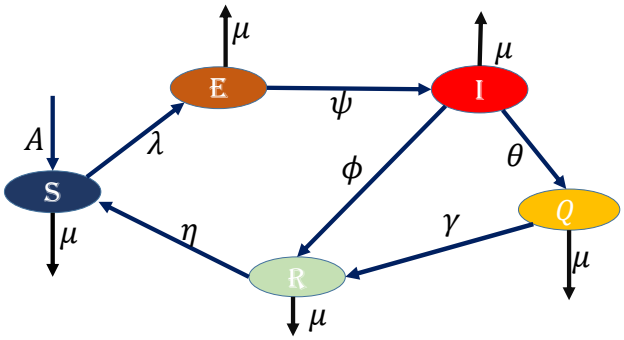


Fig. 1: Transfer diagram for the toxicity spread on the social network platform.

infectiousness. To analyze this, [18] used the SEIQR model as the following system of ordinary differential equations.

$$\begin{cases} \frac{dS(t)}{dt} = \mathcal{A} + \eta R - \frac{\beta IS}{N(t)} - \mu S, \\ \frac{dE(t)}{dt} = \frac{\beta IS}{N(t)} - (\mu + \psi)E, \\ \frac{dI(t)}{dt} = \psi E - (\mu + \phi + \theta)I, \\ \frac{dQ(t)}{dt} = \theta I - (\gamma + \mu)Q, \\ \frac{dR(t)}{dt} = \gamma Q + \phi I - (\mu + \eta)R, \end{cases} \tag{1}$$

where $\lambda = \frac{\beta IS}{N(t)}$, the entire population, and we define the quantity $N(t)$ by

$$N(t) = S(t) + E(t) + I(t) + Q(t) + R(t).$$

and initial conditions

$$S(0) = S_0 \geq 0, \quad E(0) = E_0 \geq 0, \quad I(0) = I_0 \geq 0,$$
$$Q(0) = Q_0 \geq 0, \quad R(0) = R_0 \geq 0.$$

C. Fractal-Fractional Model

Fractal-fractional models combine the advantages of fractal and fractional calculus to account for the irregular, dynamic, and memory-driven nature of online interactions. The fractal dimension captures the self-similar structure of user networks, content propagation pathways, and spatial-temporal patterns of toxic content dissemination. Meanwhile, the fractional calculus component incorporates long-range temporal dependencies and non-local effects, reflecting the persistence of toxicity and its amplification through feedback loops inherent in social media dynamics. While the classical SEIQR model in [18] effectively captured the dynamics of online toxicity spread, it failed to account for the complex, irregular, and memory-dependent nature of user interactions in social media environments. To address this limitation, we extend the model using fractal-fractional derivatives in the Caputo sense as follows:

$$\begin{cases} {}_0^{FFP}D_t^{\alpha,\tau} S(t) = \mathcal{A} + \eta R - \frac{\beta I}{N} S - \mu S, \\ {}_0^{FFP}D_t^{\alpha,\tau} E(t) = \frac{\beta I}{N} S - (\mu + \psi)E, \\ {}_0^{FFP}D_t^{\alpha,\tau} I(t) = \psi E - (\mu + \phi + \theta)I, \\ {}_0^{FFP}D_t^{\alpha,\tau} Q(t) = \theta I - (\gamma + \mu)Q, \\ {}_0^{FFP}D_t^{\alpha,\tau} R(t) = \gamma Q + \phi I - (\mu + \eta)R, \end{cases} \tag{2}$$

where ${}_0^{FFP}D_t^{\alpha,\tau}(.)$ is the fractal-fractional derivative with the fractional order $0 < \alpha \leq 1$, and fractal dimension $0 < \tau \leq 1$ in the Caputo sense with power law type kernel and the variables are assumed to be non-negative with appropriate initial conditions.

IV. Model Parameterization and Data Fitting Analysis

Ensuring accurate model validation and precise parameter estimation is fundamental in mathematical modeling when working with real-world data. The main challenge lies in determining the most suitable parameter values from empirical data, making parameter fitting a crucial step in model formulation. One widely used approach for parameter estimation in nonlinear models is the Non-Linear Least Squares Method (NLSM), which minimizes the discrepancy between observed and predicted values.

Consider a nonlinear mathematical model

$$y_i = f(x_i, \Theta) + \epsilon_i, \quad i = 1, 2, ..., n$$

where

- y_i represents the observed data points,
- $f(x_i, \Theta)$ is a nonlinear function that depends on the parameter set Θ,
- x_i are the independent variables, and
- ϵ_i is an error term assumed to be normally distributed.

The goal of NLSM is to minimize the sum of squared residuals (RSS), defined as

$$Z(\Theta) = \sum_{i=1}^{n} [y_i - f(x_i, \Theta)]^2$$

where

- $Z(\Theta)$ is the objective function to be minimized,
- y_i are the actual observed values,
- $f(x_i, \Theta)$ represents the model's predicted values, and

TABLE II: Explanation of the Model Parameters

Parameter	Value	Source	Explanation
$\mathcal{A}$	[100]	fitted	recruitment rate of human
β	0.0006	fitted	effective contact rate
ψ	0.047	fitted	the rate at which exposed become infected
θ	0.020	fitted	the rate at which I transfer to quarantine class
η	0.04	fitted	the rate at which recovery becomes susceptible
γ	0.002	fitted	the rate at which Q transfer to recovery
ϕ	0.1	fitted	the rate at which I transfer to recovery
μ	0.1	fitted	the rate at which people exit autonomously

- Θ is the set of unknown model parameters.

To evaluate model fitting, we use the relative error, given by

$$E_{\text{rel}} = \frac{\|I_{\text{est}}(t_i) - I_{\text{data}}(t_i)\|_2}{\|I_{\text{data}}(t_i)\|_2}. \tag{3}$$

This relative error, E_{rel}, quantifies the discrepancy between the estimated number of infected users $I_{\text{est}}(t_i)$ and the actual recorded infected users $I_{\text{data}}(t_i)$ at various time points t_i. The norm $\|\cdot\|_2$ represents the Euclidean distance, allowing for an objective measure of the overall deviation relative to actual data. A lower error value signifies a better model fit, affirming its accuracy in depicting the dynamics of toxicity spread.

The findings indicate that we can better capture the propagation of toxicity when the dataset is segmented (Table II). This pattern is consistent across different contexts, such as COVID-19 and social movements, suggesting that the findings are applicable in various scenarios. This approach provides policymakers with insights on prioritizing areas requiring focused interventions and allocating resources to effectively reduce the spread of toxicity.

A. Preliminaries

In this section, we will review some fractal-fractional calculus concepts that were covered earlier.

Definition 4.1 ([24], [25]) Suppose that $\mathcal{M}(t)$ is continuous and fractal differentiable on (a, b) with order τ. Then, the fractal–fractional derivative of $y(t)$ with order α in the Riemann–Liouville sense, having a power-law type kernel, is defined as follows:

$${}^{FFP}D_{a,t}^{\alpha,\tau}(\mathcal{M}(t)) = \frac{1}{\Gamma(y-\alpha)}\frac{d}{dt^\tau}\int_a^t (t-s)^{y-\alpha-1}\mathcal{M}(s)\,ds,$$

where $\mathcal{M}(t)$ is a suitable integer ensuring the definition holds, $\mathcal{M}\ -1 < \alpha,\ \tau \le y \in \mathbb{N}$, and

$$\frac{d\mathcal{M}(s)}{ds^\tau} = \lim_{t\to s}\frac{\mathcal{M}(t)-\mathcal{M}(s)}{t^\tau - s^\tau}.$$

Definition 4.2 ([25], [26]) Suppose that $\mathcal{M}(t)$ is continuous on an open interval (a, b), then the fractal-fractional integral of $\mathcal{M}(t)$ with order α, having a power-law type kernel, is defined as follows:

$${}^{FFP}D_{0,t}^{\alpha}(\mathcal{M}(t)) = \frac{\tau}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}s^{\tau-1}\mathcal{M}(s)\,ds,$$

where $\Gamma(\alpha)$ is the Gamma function.

These operators are referred to as *fractal-fractional integral operators*.

Lemma 4.3 ([20], [26]) If $\mathcal{M}$ is continuous on (a, b), then the following fractal-fractional derivative

$${}^{FFP}D_{0,t}^{\alpha,\tau}(\mathcal{M}(t)) = \mathcal{M}(t)$$

has a unique solution

$$\mathcal{M}(t) = \mathcal{M}(0) + \frac{\tau}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}s^{\tau-1}\mathcal{M}(s)\,ds.$$

We represent the Banach space by $\Omega = P\times P\times P\times P\times P$ and $P = C(\mathcal{T}, \mathbb{R}^5)$, $0 \le t \le T < \infty$, under the norm given by

$$\|W\| = \sup_{t\in\mathcal{T}}|\mathcal{M}(t)|, \quad \text{for } \mathcal{M}\in Z,$$

$$|W(t)| = |S(t) + E(t) + I(t) + Q(t) + R(t)|,$$

where $S, E, I, Q, R \in C(\mathcal{T}, \mathbb{R})$.

B. Model Analysis: Existence-Uniqueness

When formulating epidemiological models, a fundamental question arises: Does the model have a well-defined and unique solution? Before performing any analysis, it is essential to ensure that the system of equations governing the model is mathematically valid. This is typically achieved through fixed-point theorems, such as Banach's fixed point theorem, which guarantees the existence and uniqueness of solutions under specific conditions.

In this study, we employ Banach's fixed point theorem to demonstrate that the fractal-fractional model in equation (2) has a unique solution. We define the function space as: $\Omega = P\times P\times P\times P\times P$ and $P = C(\mathcal{T}, \mathbb{R}^5)$, $0 \le t \le T < \infty$, under the norm given by

$$\|W\| = \sup_{t\in\mathcal{T}}|\mathcal{M}(t)|, \quad \text{for } \mathcal{M}\in Z,$$

$$|W(t)| = |S(t) + E(t) + I(t) + Q(t) + R(t)|,$$

where $S, E, I, V, R \in C(\mathcal{T}, \mathbb{R})$.

From equation (2), the right-hand side of the fractal-fractional $SEIQR$ model can be rewritten for simplicity as

TABLE III: Error Rates for SEIQR Model with and without Dataset Splitting

Dataset	Moderate toxicity	High toxicity	Without split
F*Covid	0.0011	0.0021	0.081
F*Mask	0.021	0.049	0.073
Lockdown	0.032	0.045	0.061
5G	0.0029	0.0011	0.003
Brazil Anti	0.062	0.055	0.094
Brazil Pro	0.060	0.061	0.088
Peru	0.095	0.124	0.41

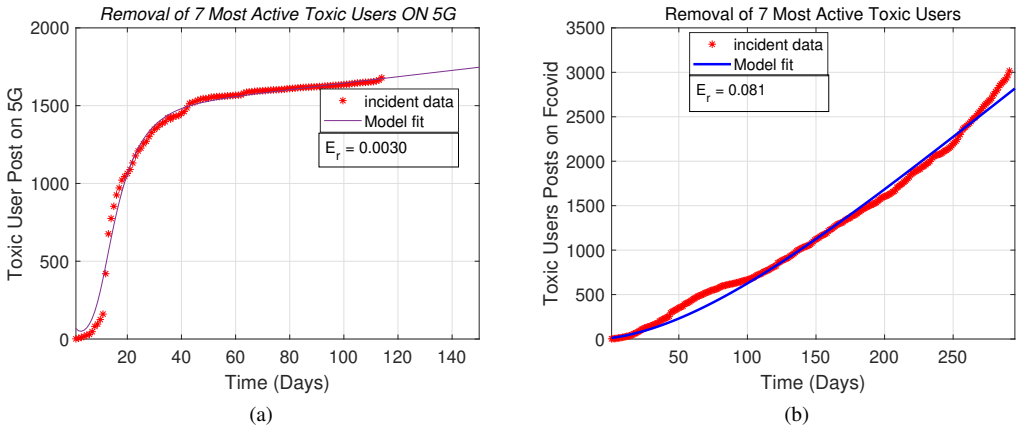


Fig. 2: SEIQR model fitted to infected without splitting toxicity in the datasets.

$$\begin{cases} \mathcal{M}_1(t,S(t)) = \mathcal{A} + \eta R - \frac{\beta I}{N} S - \mu S, \\ \mathcal{M}_2(t,E(t)) = \frac{\beta I}{N} S - (\mu + \psi) E, \\ \mathcal{M}_3(t,I(t)) = \psi E - (\mu + \phi + \theta) I, \\ \mathcal{M}_4(t,Q(t)) = \theta I - (\gamma + \mu) Q, \\ \mathcal{M}_5(t,R(t)) = \gamma Q + \phi I - (\mu + \eta) R. \end{cases} \tag{4}$$

The fractional integral possesses differentiability, and if the given equation (2) can be transformed into the Volterra form, then the fractal-fractional derivative, expressed in the Riemann–Liouville framework, can be represented as follows

$$\begin{cases} {}_0^{RL}D_t^{\alpha,\tau} S(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^\alpha S(s)\, ds \frac{1}{st^{s-1})}, \\ {}_0^{RL}D_t^{\alpha,\tau} E(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^\alpha E(s)\, ds \frac{1}{st^{s-1})}, \\ {}_0^{RL}D_t^{\alpha,\tau} I(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^\alpha I(s)\, ds \frac{1}{st^{s-1})}, \\ {}_0^{RL}D_t^{\alpha,\tau} Q(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^\alpha Q(s)\, ds \frac{1}{st^{s-1})}, \\ {}_0^{RL}D_t^{\alpha,\tau} R(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^t (t-s)^\alpha R(s)\, ds \frac{1}{st^{s-1})}, \end{cases} \tag{5}$$

Due to the fact that the integrals in equation (5) are differentiable, the system can be expressed as:

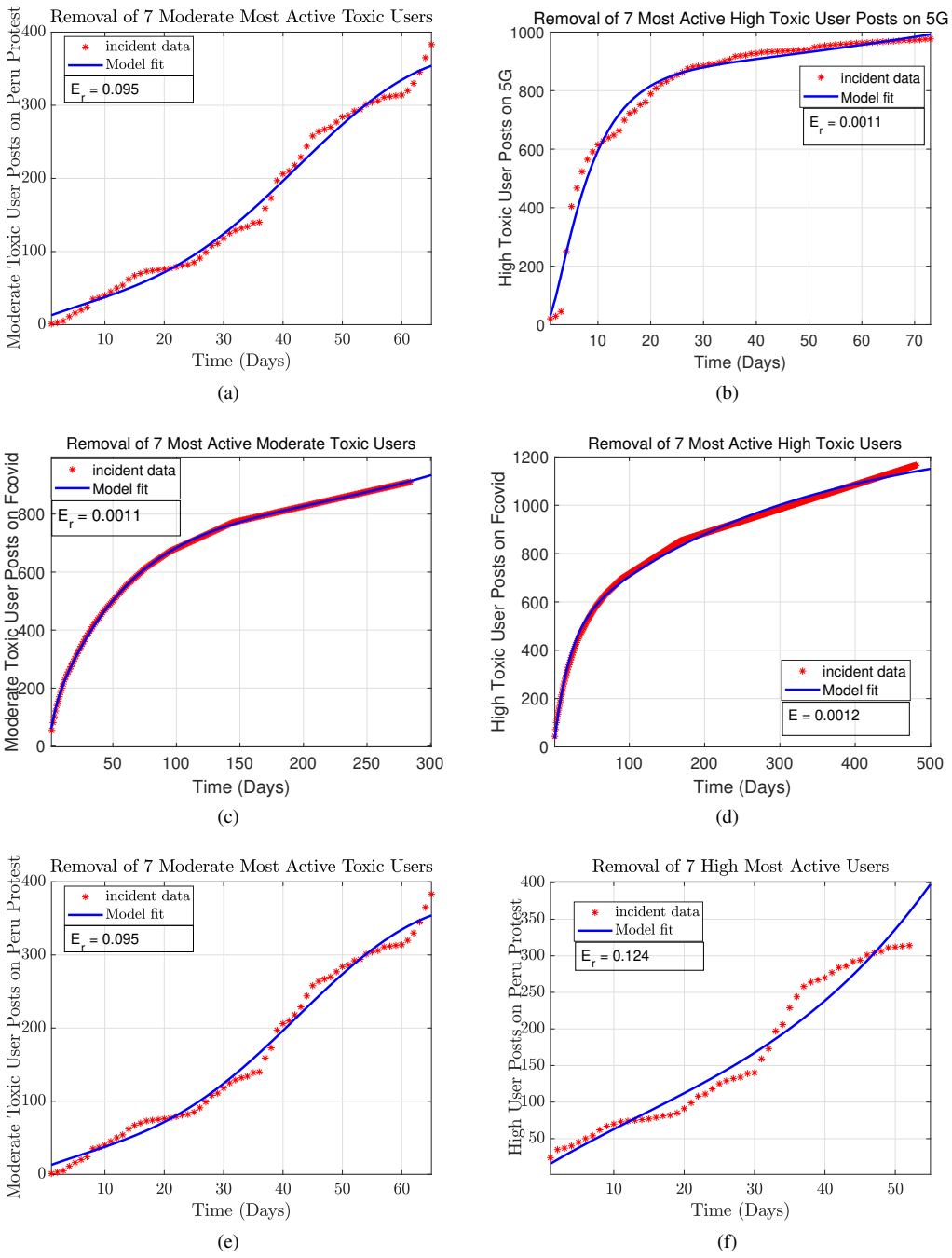


Fig. 3: SEIQR model fitted to moderate and high toxic posts.

$$\begin{cases} {}_0^{RL}D_t^{\alpha,\tau}S(t) = st^{s-1}[\mathcal{A} + \eta R - \frac{\beta I}{N}S - \mu S], \\ {}_0^{RL}D_t^{\alpha,\tau}E(t) = st^{s-1}[\frac{\beta I}{N}S - (\mu + \psi)E], \\ {}_0^{RL}D_t^{\alpha,\tau}I(t) = st^{s-1}[\psi E - (\mu + \phi + \theta)I], \\ {}_0^{RL}D_t^{\alpha,\tau}Q(t) = st^{s-1}[\theta I - (\gamma + \mu)Q], \\ {}_0^{RL}D_t^{\alpha,\tau}R(t) = st^{s-1}[\gamma Q + \phi I - (\mu + \eta)R], \end{cases} \quad (6)$$

To ensure that the initial conditions are defined in terms of integer-order derivatives, we replace the Riemann-Liouville derivative with the Caputo derivative. This modification facilitates the formulation of well-posed initial conditions. Subsequently, we apply the Riemann–Liouville fractional integral on both sides of the equation, leading to the following integral

system:

$$\begin{cases} S(t) = S(0) + \frac{\tau}{\Gamma(\alpha)} \int_a^t s^{\tau-1}(t-s)^{\alpha} \mathcal{M}_1(s, S(s))\, ds, \\ E(t) = E(0) + \frac{\tau}{\Gamma(\alpha)} \int_a^t s^{\tau-1}(t-s)^{\alpha} \mathcal{M}_2(s, E(s))\, ds, \\ I(t) = I(0) + \frac{\tau}{\Gamma(\alpha)} \int_a^t s^{\tau-1}(t-s)^{\alpha} \mathcal{M}_3(s, I(s))\, ds, \\ Q(t) = Q(0) + \frac{\tau}{\Gamma(\alpha)} \int_a^t s^{\tau-1}(t-s)^{\alpha} \mathcal{M}_4(s, Q(s))\, ds, \\ R(t) = R(0) + \frac{\tau}{\Gamma(\alpha)} \int_a^t s^{\tau-1}(t-s)^{\alpha} \mathcal{M}_5(s, R(s))\, ds, \end{cases} \tag{7}$$

C. Positivity and Biological Invariant Region

From an epidemiological perspective, ensuring that all model variables remain positive is essential for the realism and validity of the model. A fundamental approach to achieving this is by demonstrating that each variable maintains a non-negative value throughout the analysis. This guarantees that the solution to the proposed model in equation (2) is both positive and bounded, reinforcing its applicability to real-world scenarios.

Lemma 3.4.1 The boundedness and biologically acceptance region for the presented model (1) are expressed in the region $\mathfrak{B} \subset \mathbb{R}^5_+$, and are given by

$$\mathcal{H} = \left\{ (S(t), E(t), I(t), Q(t), R(t)) \in \mathbb{R}^5_+ \mid 0 \leq N(t) \leq \frac{\mathcal{A}}{\mu} \right\}.$$

Proof The suggested model (1) is evaluated in order to study the first equation,

$$\frac{dS}{dt} = \mathcal{A} + \eta R - \frac{\beta I}{N} S - \mu S,$$

and from that it follows that

$$\frac{dS}{dt} \geq -\frac{\beta I}{N} S - \mu S,$$

in accordance with the exponential growth criterion and the provision of integration

$$\frac{dS}{dt} \geq S(0) e^{-\int_0^t (\frac{\beta I}{N} + \mu) dt},$$

then, we have:

$$S(t) \geq 0.$$

Similarly, we have $E(t) \geq 0$, $I(t) \geq 0$, $Q(t) \geq 0$, and $R(t) \geq 0$ for all $t \geq 0$, so that the solution to the model does not deviate from the planes $S = E = I = Q = R = 0$. Due to the fact that the whole population is represented by the function $N(t)$, we are able to write the system (1) as follows:

$$\frac{dN(t)}{dt} = \mathcal{A} - \mu N(t),$$

It is possible that users become negligible, which means that none autonomously exit, then

$$\frac{dN(t)}{dt} \leq \mathcal{A} - \mu N(t).$$

The solution of the above equation turns to

$$N(t) \leq \frac{\mathcal{A}}{\mu} - \left(\frac{\mathcal{A}}{\mu} - N_0 \right) e^{-\mu t}.$$

Taking advantage of [27], we can say that if $N_0 < \frac{\mathcal{A}}{\mu}$ then as $t \to \infty$, asymptotically, $N(t) \to \frac{\mathcal{H}}{\mu}$ implies that $0 \leq N(t) \leq \frac{\mathcal{A}}{\mu}$. Hence, all of the feasible solutions for the model will eventually converge in the region $\mathcal{H}$. □

D. Equilibria, Basic Reproduction number, and Their Stability

In this section, we present the computations and explanation of the equilibria and basic reproduction number for the suggested model in equation (2), as well as explore how locally and globally asymptotically stable.

E. Disease-free Equilibrium Point

In order to maintain a disease-free equilibrium, the variables $E = I = Q = R = 0$ are established. In this specific case, the system (1) demonstrates that the disease-free equilibrium point is given as

$$\text{DFE} = \left(\frac{\mathcal{A}}{\mu}, 0, 0, 0, 0 \right).$$

F. Basic Reproduction Number, $\mathfrak{R}_0$

We apply the next-generation matrix technique to a dynamic system, where we consider the emergence rate of new infections, denoted as $\mathcal{F}$, and the transmission rate per individual, $\mathcal{V}$, under conditions where the disease is not present in the population. The following represents this calculation

$$\mathcal{F} = \begin{pmatrix} 0 & \beta S & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathcal{V} = \begin{pmatrix} -(\mu + \psi) & 0 & 0 \\ \psi & -(\mu + \phi + \theta) & 0 \\ 0 & \theta & -(\gamma + \mu) \end{pmatrix},$$

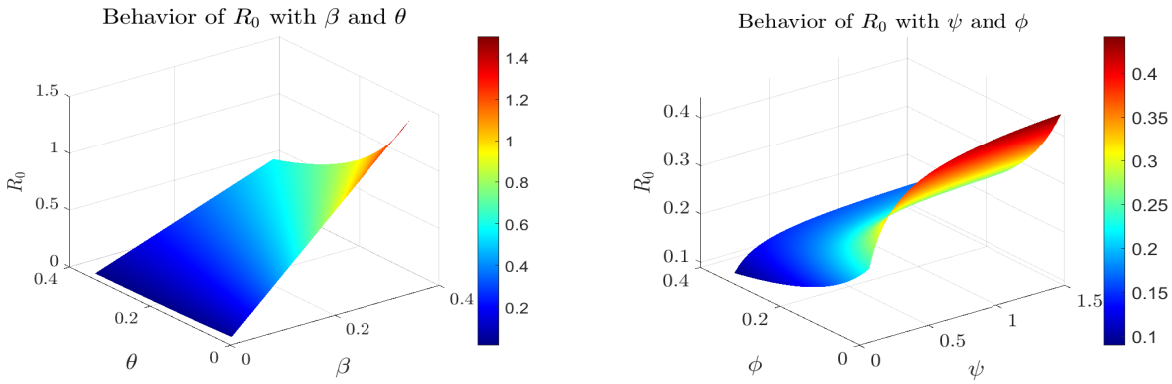
and

$$\mathcal{V}^{-1} = \begin{pmatrix} \frac{-1}{(\mu+\psi)} & 0 & 0 \\ \frac{\psi}{(\mu+\psi)(\mu+\phi+\theta)} & \frac{-1}{(\mu+\phi+\theta)} & 0 \\ 0 & \frac{\theta}{(\mu+\phi+\theta)(\gamma+\mu)} & \frac{-1}{(\gamma+\mu)} \end{pmatrix}$$

Thus, the basic reproduction number $\mathfrak{R}_0$ is the spectral radius of $\mathcal{F}\mathcal{V}^{-1}$ or

$$\mathfrak{R}_0 = \frac{\psi \beta \mathcal{A}}{\mu(\mu + \psi)(\mu + \phi + \theta)}.$$

Figures 4 (a) and (b) present a 3D visualization of the basic reproduction number $\mathfrak{R}_0$, generated using MATLAB, to illustrate its dependence on various model parameters. In Figure 4 (a), an increase in β leads to a corresponding rise in $\mathfrak{R}_0$, indicating that platforms where toxicity spreads more rapidly experience greater persistence of toxicity over time. Conversely, as θ increases, $\mathfrak{R}_0$ declines, demonstrating that an enhanced quarantine transfer rate plays a crucial role



(a) (b)

Fig. 4: Effect of $\mathfrak{R}_0$ on (a) θ and β, and (b) ϕ and ψ

in limiting the spread of toxicity. Similarly, Figure 4 (b) depicts the relationship between $\mathfrak{R}_0$, the content exposure rate ϕ, and the transition rate ψ. The analysis suggests that by carefully regulating these parameters—such as implementing strategies to reduce exposure to toxic content—it is possible to significantly mitigate or even reverse the proliferation of harmful interactions in online spaces.

Theorem 3.4.2. The disease-free equilibrium point is locally asymptotically stable if $\mathfrak{R}_0 < 1$, and otherwise unstable.

Proof. Consider the following Jacobian matrix in regard to the system in order to demonstrate the stability criterion for the proposed system (1) at the DFE point:

$$J = \begin{bmatrix} -(\mu+\psi) & \frac{\beta A}{\mu} & 0 \\ \psi & -(\mu+\phi+\theta) & 0 \\ 0 & \theta & -(\gamma+\mu) \end{bmatrix}.$$

The characteristic equation for the matrices J is then derived in the following manner:

$$\lambda^2 + (2\mu+\psi+\phi+\theta)\lambda + (\mu+\psi)(\mu+\phi+\theta) - \frac{\beta A}{\mu}\psi = 0.$$

The stability of the DFE depends on the roots of this characteristic equation. According to the Routh-Hurwitz stability criterion, all roots (i.e., eigenvalues) have negative real parts if and only if all coefficients of the characteristic equation are positive.

- The linear term $(2\mu+\psi+\phi+\theta)$ is always positive since all parameters are positive.
- The constant term $(\mu+\psi)(\mu+\phi+\theta) - \frac{\beta A}{\mu}\psi$ determines the stability.

By defining the **basic reproduction number** as

$$\mathfrak{R}_0 = \frac{\psi\beta A}{\mu(\mu+\psi)(\mu+\phi+\theta)}, \tag{8}$$

we rewrite the constant term as

$$(\mu+\psi)(\mu+\phi+\theta)(1-\mathfrak{R}_0). \tag{9}$$

- If $\mathfrak{R}_0 < 1$, this term is positive, ensuring that all roots of the characteristic equation have negative real parts. This confirms that the disease-free equilibrium is locally asymptotically stable.
- If $\mathfrak{R}_0 > 1$, this term becomes negative, implying that at least one eigenvalue has a positive real part, leading to instability.

The disease-free equilibrium (DFE) is locally asymptotically stable if and only if $\mathfrak{R}_0 < 1$. If $\mathfrak{R}_0 > 1$, the system becomes unstable, allowing the disease to persist and spread in the population. This result confirms the epidemiological interpretation that controlling disease spread requires keeping $\mathfrak{R}_0$ below the critical threshold of 1.

G. Existence of Endemic Equilibrium Point

In this section, we will look at whether an endemic equilibrium point exists, which will be represented by the notation $(S^*, E^*, I^*, Q^*, R^*)$. The second equilibrium point is the endemic equilibrium point, as presented below:

$$\begin{cases} 0 = \mathcal{A} + \eta R - \frac{\beta I}{N} S - \mu S, \\ 0 = \frac{\beta I}{N} S - (\mu+\psi)E, \\ 0 = \psi E - (\mu+\phi+\theta)I, \\ 0 = \theta I - (\gamma+\mu)Q, \\ 0 = \gamma Q + \phi I - (\mu+\eta)R, \end{cases} \tag{10}$$

Thereby, this brings us to a further conclusion, which is given below:

Theorem 6.2. The system (1) has a unique endemic equilibrium point given by:

$$S^* = \frac{N(\mu+\psi)(\mu+\phi+\theta)}{\psi\beta},$$

$$E^* = \frac{\beta S^* I^*}{N(\mu+\psi)}, I^* = \frac{\psi}{\mu+\phi+\theta} \cdot \frac{\beta S^* I^*}{N(\mu+\psi)}$$

$$I^* = \frac{\psi}{\mu+\phi+\theta} \cdot \frac{\beta S^* I^*}{N(\mu+\psi)}$$

$$Q^* = \frac{\theta I^*}{\gamma+\mu}$$

$$R^* = \frac{\gamma\theta I^*}{(\gamma+\mu)(\mu+\eta)} + \frac{\phi I^*}{\mu+\eta}$$

H. Global Stability Analysis at Endemic Equilibrium Point

To achieve global stability and rule out periodic orbits, we apply Dulac's criterion.

Theorem 6.3. There are no periodic orbits in the system (1).

Proof. We introduce a Dulac function $B(S, E, I, Q, R)$ to transform the system into a divergence form where the total divergence is of a definite sign. A common choice of a Dulac function is

$$B(S, E, I, Q, R) = \frac{1}{SI}$$

Multiplying the system by $B(S, E, I, Q, R)$, we compute the divergence:

$$\frac{\partial}{\partial S}(BS') + \frac{\partial}{\partial E}(BE') + \frac{\partial}{\partial I}(BI') + \frac{\partial}{\partial Q}(BQ') + \frac{\partial}{\partial R}(BR').$$

Taking the partial derivatives, the divergence simplifies to

$$\nabla \cdot (B\mathcal{H}) = \frac{\partial}{\partial S}\left(\frac{A + \eta R - \frac{\beta I}{N} S - \mu S}{SI}\right) + \frac{\partial}{\partial I}\left(\frac{\psi E - (\mu + \phi + \theta) I}{SI}\right).$$

After simplifications, this leads to

$$\nabla \cdot (B\mathcal{H}) = -\frac{\mu}{SI} - \frac{(\mu + \phi + \theta)}{SI}.$$

Since all parameters are **positive**, the divergence function is **negative definite**, i.e.,

$$\nabla \cdot (B\mathcal{H}) < 0.$$

Since the divergence of the vector field is negative everywhere in the domain of interest, Dulac's criterion guarantees that there are no periodic orbits in the system. This implies that global stability holds for the endemic equilibrium. Thus, the endemic equilibrium is globally asymptotically stable, confirming that all solutions will ultimately converge to it in the long run.

I. Sensitivity Analysis of Important Parameters

To answer **RQ1**, we begin by identifying the parameter values that are most influential in determining social media toxicity spread. It is vital to discover numerous aspects that contribute to the toxicity spread and prevalence to decide the best technique for minimizing the number of affected users. To determine the dependence of each parameter on the SEIQR model, a sensitivity analysis of each parameter was performed using the Latin Hypercube Sampling-Partial Rank Correlation Coefficient (LHS-PRCC) method. The PRCC corresponds directly to the degree of statistical influence. A positive value indicates that an increase in this parameter leads to a positive influence on the SEIQR model. In contrast, a negative value indicates that an increase in this parameter leads to a negative influence on the SEIQR model. As shown in Figure 5 and Table IV, among the ϕ, β, ψ, μ, $\mathcal{A}$, and θ parameters, ϕ, β, ψ, $\mathcal{A}$, and θ have positive effects on $\mathcal{R}_0$. In contrast, parameter μ has a negative effect on $\mathcal{R}_0$. The red line represents the p-value, and in the context of social media toxicity, $p < 0.05$ indicates that the results of the sensitivity analysis are statistically significant. Specifically, it means that there is less than a 5% probability that the observed effects (relationships between the parameters and the effective reproduction number $\mathcal{R}_0$) occurred by chance. These behaviors are significant in executing emergency management measures. Let

TABLE IV: Sensitivity Index of Each Parameter on $\mathcal{R}_0$

Parameters	Sensitivity Indices	Relationship
ψ	0.863213	+ve
β	0.902867	+ve
$\mathcal{A}$	0.8617	+ve
μ	-0.9167	-ve
ϕ	0.89574	-ve
θ	0.1913	+ve

$$C_p^{\mathcal{R}_0} = \frac{\partial R_0}{\partial p} \times \frac{p}{\mathcal{R}_0}$$

where p is the parameter being studied for its sensitivity. Utilizing this index, one can deduce the sensitivity index corresponding to every parameter integrated into the expression for $\mathcal{R}_0$. We compute their derivatives as follows:

$$C_\beta^{\mathcal{R}_0} = \frac{\partial R_0}{\partial \beta} \times \frac{\beta}{\mathcal{R}_0},$$

such that

$$\frac{dR_0}{d\psi} = \frac{A\beta(\mu+\phi+\theta-\psi)}{\mu(\mu+\psi)^2(\mu+\phi+\theta)}$$

$$\frac{dR_0}{d\beta} = \frac{A\psi}{\mu(\mu+\psi)(\mu+\phi+\theta)}$$

$$\frac{dR_0}{dA} = \frac{\beta\psi}{\mu(\mu+\psi)(\mu+\phi+\theta)}$$

$$\frac{dR_0}{d\mu} = -\frac{A\beta\psi(\mu+\psi+\phi+\theta)}{\mu^2(\mu+\psi)(\mu+\phi+\theta)}$$

$$\frac{dR_0}{d\phi} = \frac{A\beta\psi}{\mu(\mu+\psi)(\mu+\phi+\theta)^2}$$

$$\frac{dR_0}{d\theta} = \frac{A\beta\psi}{\mu(\mu+\psi)(\mu+\phi+\theta)}$$

In Figure 5, the positive sensitivity indices, associated with ψ, ζ, β, ϑ, σ, and ϵ, predominantly influence the frequency of online toxicity manifestations, with β having the highest positive impact index, indicating its crucial role in online

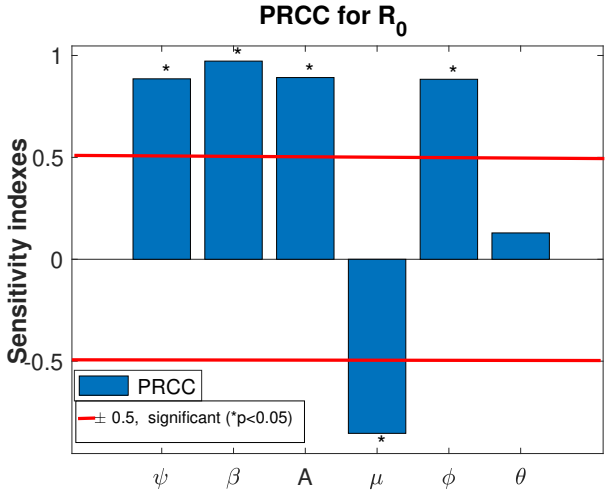


Fig. 5: Sensitivity of $\mathcal{R}_0$ of the online toxicity contagion

toxicity spread. In contrast, negative sensitivity indices linked to parameters μ and ϕ attenuate the online toxicity spread. Remarkably, parameters such as the infection rate of exposed users (ϑ), recovery rates of users who hadn't been quarantined (σ), the rate at which infected users transfer to rehabilitation (ϵ), and the rate at which infected users are banned from platform (ϕ) all curb the dissemination of online toxicity. This highlights the importance of improving infected users banned and rehabilitation measures to control the spread of the online toxicity effectively.

J. Existence and Uniqueness

We employ the Picard–Lindelöf method along with **fixed point theory** to analyze the qualitative properties of the solution for model (2). The proposed model equation ((2)) can be reformulated as:

$$\begin{cases} {}^{FFP}D^{\alpha,\tau}\mathcal{M}(t) = \Upsilon(t,\mathcal{M}(t), & 0<\alpha,\ \tau\le 1 \\ \mathcal{M}(0)=\mathcal{M}_0\ge 0, & t\in[0,T] \end{cases} \tag{11}$$

where,

$$\mathcal{M}(t)=\begin{cases} S(t),\\ E(t),\\ I(t),\\ Q(t),\\ R(t), \end{cases} \quad \mathcal{M}_0=\begin{cases} S(0),\\ E(0),\\ I(0),\\ Q(0),\\ R(0), \end{cases}$$

$$\Upsilon(t,\mathcal{M}(t))=\begin{cases} \mathcal{M}_1(t,S,E,I,Q,R),\\ \mathcal{M}_2(t,S,E,I,Q,R),\\ \mathcal{M}_3(t,S,E,I,Q,R),\\ \mathcal{M}_4(t,S,E,I,Q,R),\\ \mathcal{M}_5(t,S,E,I,Q,R). \end{cases}$$

Taking advantage from Lemma 4.3, system (11) yields

$$\mathcal{M}(t)=\mathcal{M}_0(t)+\frac{1}{\Gamma(\alpha)}\int_0^t s^{\tau-1}\Upsilon(s,\mathcal{M}(s))(t-s)^{\alpha-1}ds. \tag{12}$$

such that Υ satisfies

$$|\Upsilon(t,\mathcal{M}(t))-\Upsilon(t,\tilde{\mathcal{M}}(t))|\le \mathbf{C}_\upsilon[|\mathcal{M}-\tilde{\mathcal{M}}|],\quad \mathbf{C}_\upsilon>0. \tag{13}$$

Theorem 7.1 *Supposed that equation (13) holds. Then the equation (11) has a unique solution if*

$$\mathcal{D}:=\frac{\tau B(\alpha,\tau)}{T^{\tau+\alpha-1}\Gamma(\alpha)}C_\upsilon<1,\quad \text{where } B(\cdot,\cdot) \text{ is the beta function.}$$

Proof. Considering Picard operator, denoted by

$$\Omega:\mathcal{Z}\to\mathcal{Z},$$

and defined as

$$\Omega\mathcal{M}(t)=\mathcal{M}_0+\tfrac{\tau}{\Gamma(\alpha)}\int_0^t s^{\tau-1}(t-s)^{\alpha-1}\Upsilon(s,\mathcal{M}(s))ds \tag{14}$$

and set

$$\sup_{t\in\mathcal{T}}\mathcal{M}(t,0)=\mathcal{M}_0.$$

It is important to point out that the solution to the system (11) is bounded, i.e.,

$$\|\Omega\mathcal{M}(t)-\mathcal{M}(0)\|=\sup_{t\in T}|\Omega\mathcal{M}(t)-\mathcal{M}(0)|.$$

$$\|\Omega\mathcal{M}(t)-\mathcal{M}(0)\|=\sup_{t\in\mathcal{B}}\left|\frac{\tau}{\Gamma(\alpha)}\int_0^t s^{\tau-1}(t-s)^{\alpha-1}\Upsilon(s,\mathcal{M}(s))ds\right|,$$

$$\le\sup_{t\in\mathcal{T}}\frac{\tau}{\Gamma(\alpha)}\int_0^t s^{\tau-1}(t-s)^{\alpha-1}|\Upsilon(s,\mathcal{M}(s))|ds.$$

$$\le\sup_{t\in\mathcal{T}}\frac{\tau}{\Gamma(\alpha)}\int_0^t \xi^{\tau-1}(t-\xi)^{\alpha-1}|\Upsilon(s,\mathcal{M}(s))-\Upsilon(t,0)|ds,$$

$$\le\frac{\tau B(\alpha,\tau)}{\Gamma(\alpha)}T^{\tau+\alpha-1}(C_\upsilon\|\mathcal{M}\|_\infty+\Upsilon_0):=\mathcal{R}<\infty.$$

By utilizing the definition of the Picard operator (14) and considering any functions $\mathcal{M},\tilde{\mathcal{M}}\in\mathcal{Z}$, we derive the following bound:

$$\|\Omega\mathcal{M}-\Omega\tilde{\mathcal{M}}\|=\sup_{t\in\mathcal{T}}\|\Omega\mathcal{M}(t)-\Omega\tilde{\mathcal{M}}(t)\|.$$

$$\le\sup_{t\in\mathcal{T}}\frac{\tau}{\Gamma(\alpha)}\int_0^t s^{\tau-1}(t-s)^{\alpha-1}|\Upsilon(\mathcal{M}(s))-\Upsilon(\tilde{\mathcal{M}}(s))|ds,$$

$$\le\sup_{t\in\mathcal{T}}\frac{\tau}{\Gamma(\alpha)}\int_0^t s^{\tau-1}(t-s)^{\alpha-1}C_\upsilon|\mathcal{M}(s))-\tilde{\mathcal{M}}(s)|ds,$$

$$\le\frac{\tau B(\alpha,\tau)}{\Gamma(\alpha)}T^{\tau+\alpha-1}C_\upsilon|\mathcal{M}-\tilde{\mathcal{M}}|,$$

which leads to the conclusion that $\|\mathcal{M}-\tilde{\mathcal{M}}\|\le\mathcal{D}\|\mathcal{M}-\tilde{\mathcal{M}}\|$. Consequently, since $\mathcal{D}$ is a contraction mapping, it follows from Banach's contraction principle that system 11 has a unique solution.

K. Numerical Analysis and Simulation

In this section, we present the numerical approximation of the proposed fractal-fractional $SEIQR$ model using a fractal-fractional derivative in the Caputo sense with a power-law kernel and simulate the long-term behavior of toxicity propagation in online social networks, analyzing the effects of quarantine interventions, mitigation policies, and fractal-fractional fluctuations in toxicity intensity. The step-by-step proof of this scheme can found in [24], [25], and [26].

Hence, the fractal-fractional $SEIQR$ model scheme is derived as follows. Given the initial conditions, the forward time-step approximation is formulated as:

$$S(t_{n+1}) = S(0) + \frac{\tau(\Delta t)^{\alpha}}{\Gamma(\alpha+2)} \sum_{m=0}^{n}\left[t_m^{\alpha-1}\Upsilon_1(t_m, S(t_m))\right] \left[(n+1-m)^{\alpha}(n-m+2+\alpha) - (n-m)^{\alpha} (n-m+2+2\alpha)\right] - t_{m-1}^{\alpha-1}\Upsilon_1(t_{m-1}, S(t_{m-1})) \left[(n+1-m)^{\alpha+1} - (n-m)^{\alpha}(n-m+1+\alpha)\right] \quad (15)$$

$$E(t_{n+1}) = E(0) + \frac{\tau(\Delta t)^{\alpha}}{\Gamma(\alpha+2)} \sum_{m=0}^{n}\left[t_m^{\alpha-1}\Upsilon_2(t_m, E(t_m))\right] \left[(n+1-m)^{\alpha}(n-m+2+\alpha) - (n-m)^{\alpha} (n-m+2+2\alpha)\right] - t_{m-1}^{\alpha-1}\Upsilon_2(t_{m-1}, E(t_{m-1})) \left[(n+1-m)^{\alpha+1} - (n-m)^{\alpha}(n-m+1+\alpha)\right] \quad (16)$$

$$I(t_{n+1}) = I(0) + \frac{\tau(\Delta t)^{\alpha}}{\Gamma(\alpha+2)} \sum_{m=0}^{n}\left[t_m^{\alpha-1}\Upsilon_3(t_m, I(t_m))\right] \left[(n+1-m)^{\alpha}(n-m+2+\alpha) - (n-m)^{\alpha} (n-m+2+2\alpha)\right] - t_{m-1}^{\alpha-1}\Upsilon_3(t_{m-1}, I(t_{m-1})) \left[(n+1-m)^{\alpha+1} - (n-m)^{\alpha}(n-m+1+\alpha)\right] \quad (17)$$

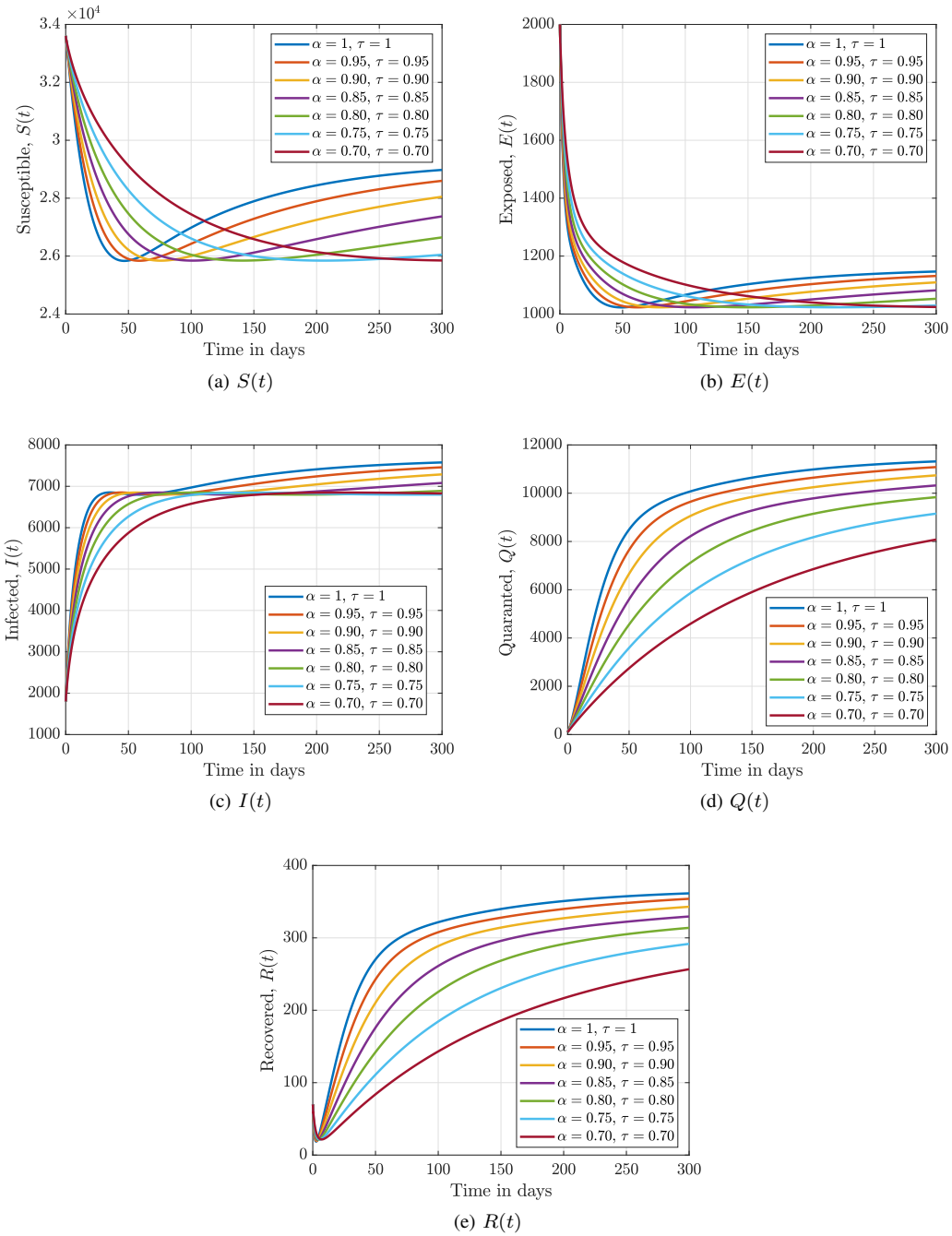
$$Q(t_{n+1}) = Q(0) + \frac{\tau(\Delta t)^{\alpha}}{\Gamma(\alpha+2)} \sum_{m=0}^{n}\left[t_m^{\alpha-1}\Upsilon_4(t_m, Q(t_m))\right] \left[(n+1-m)^{\alpha}(n-m+2+\alpha) - (n-m)^{\alpha} (n-m+2+2\alpha)\right] - t_{m-1}^{\alpha-1}\Upsilon_4(t_{m-1}, Q(t_{m-1})) \left[(n+1-m)^{\alpha+1} - (n-m)^{\alpha}(n-m+1+\alpha)\right] \quad (18)$$

$$R(t_{n+1}) = R(0) + \frac{\tau(\Delta t)^{\alpha}}{\Gamma(\alpha+2)} \sum_{m=0}^{n}\left[t_m^{\alpha-1}\Upsilon_5(t_m, R(t_m))\right] \left[(n+1-m)^{\alpha}(n-m+2+\alpha) - (n-m)^{\alpha} (n-m+2+2\alpha)\right] - t_{m-1}^{\alpha-1}\Upsilon_5(t_{m-1}, R(t_{m-1})) \left[(n+1-m)^{\alpha+1} - (n-m)^{\alpha}(n-m+1+\alpha)\right] \quad (19)$$

In the next section, we apply this numerical approach to simulate the long-term behavior of toxicity propagation in online social networks, analyzing the effects of quarantine interventions, mitigation policies, and fractal-fractional fluctuations in toxicity intensity.

V. Results and Discussion

In order to answer **RQ2**, by employing the outlined methodology, we applied the SEIQR model to seven distinct datasets to assess their performance and alignment with observed



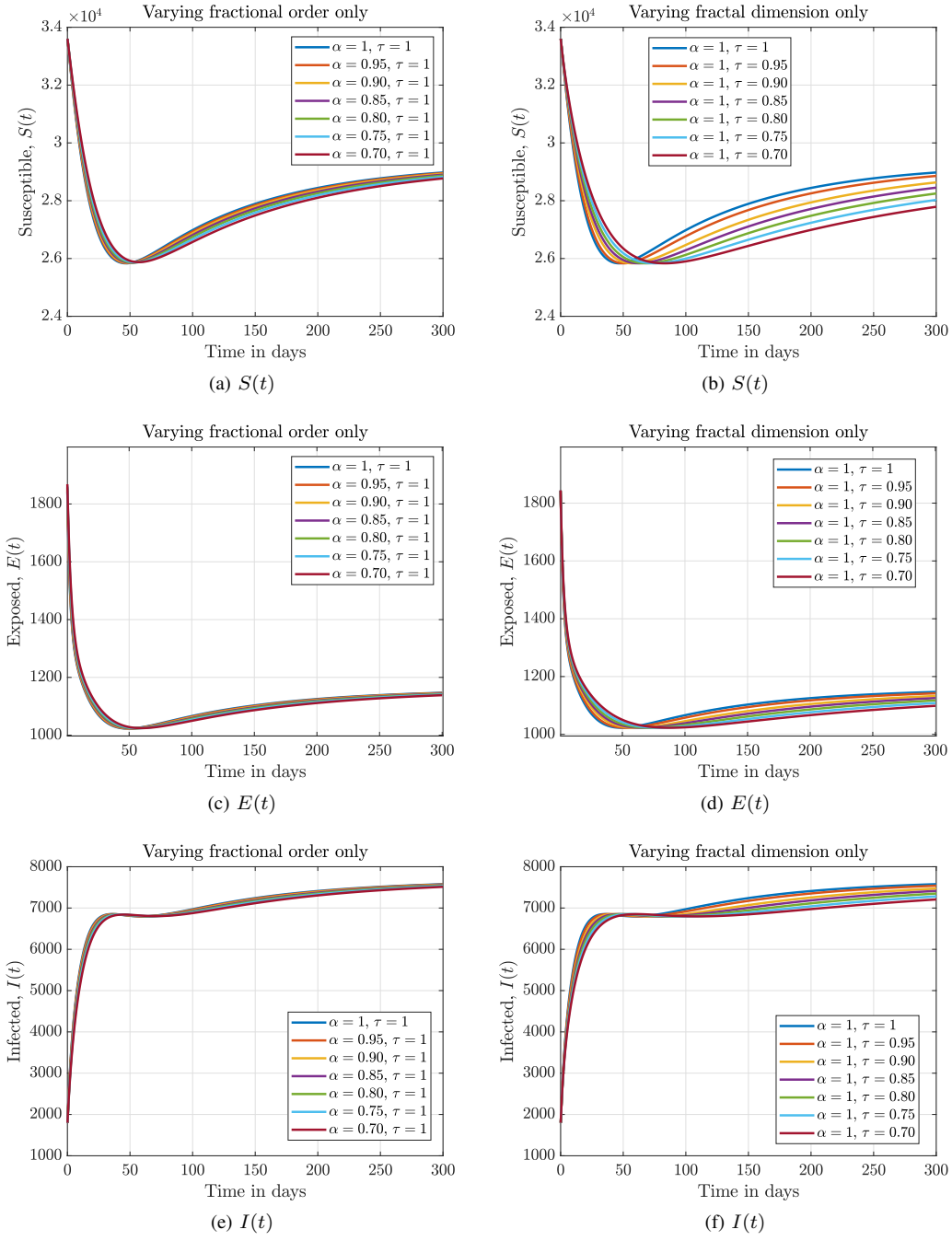
(a) $S(t)$ (b) $E(t)$

(c) $I(t)$ (d) $Q(t)$

(e) $R(t)$

Fig. 6: Numerical trajectories of $SEIQR$ model under the Caputo fractal-fractional operator.

data. Table III presents the error rates obtained from these evaluations using equation (2); it shows the error rate for moderate and high toxicity levels, as well as the error rate without splitting the data. Figure 2 shows the error rates and fitted models' representations for the three datasets—5G, Peru, and F*covid—that use the SEIQR model without splitting the data are depicted in. In the figures, the red stars represent the observed data, while the blue line depicts our fitted model. The plot illustrates how well our model aligns with the observed data. The plots in Figure 3 indicate the fitted plots for the same datasets—5G, Peru, and F*covid—using the SEIQR model with a split in the data based on the toxicity intensity. For simplicity, we excluded the fitted model for the rest of the datasets and summarized all in Table III. We compared the error rates of the SEIQR model with and without dataset splitting. The consistently low error rates across datasets when splitting the data into moderate and high toxicity levels, as compared to using the model without this division, underscore

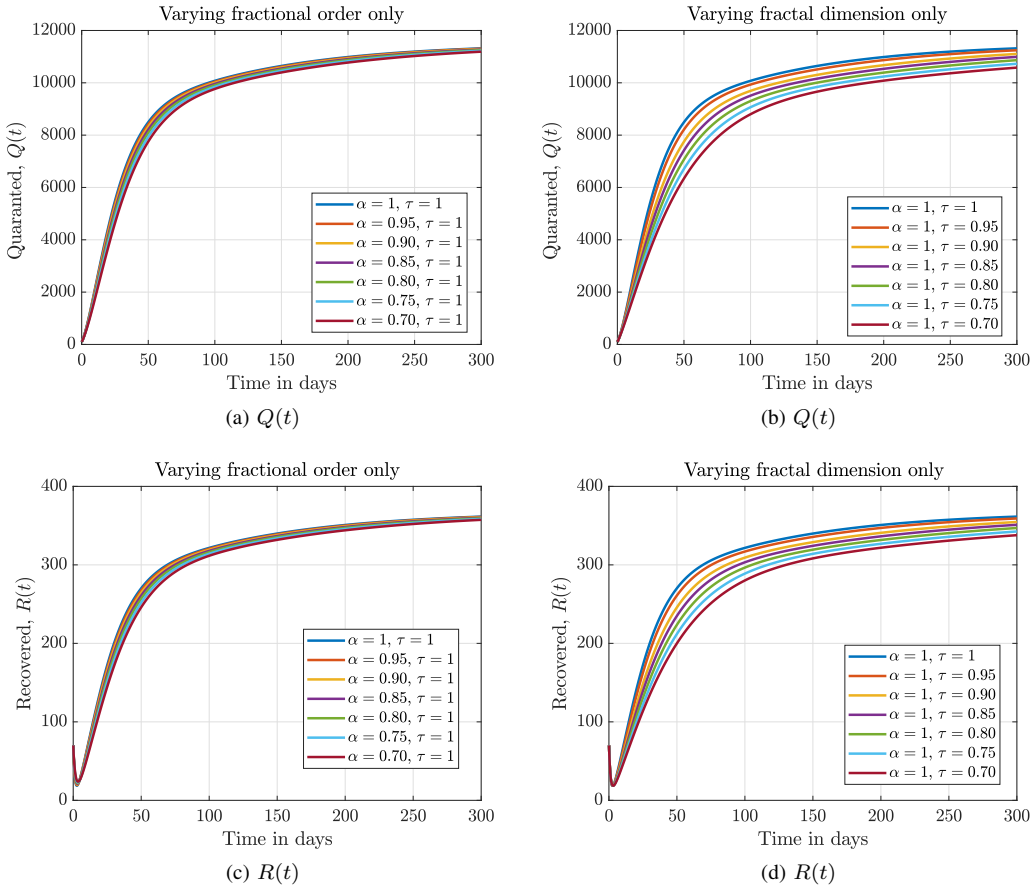


(a) $S(t)$ (b) $S(t)$

(c) $E(t)$ (d) $E(t)$

(e) $I(t)$ (f) $I(t)$

Fig. 7: Numerical comparison of fractional dynamics only and fractal dynamics only for Susceptible (S(t)), Exposed (E(t)), and Infected (I(t)) compartment under the Caputo fractal-fractional operator.



(a) $Q(t)$ (b) $Q(t)$

(c) $R(t)$ (d) $R(t)$

Fig. 8: Numerical comparison of fractional dynamics only and fractal dynamics only for Quarantine (Q(t)) and Recovery (R(t)) compartment under the Caputo fractal-fractional operator.

the effectiveness of the SEIQR model in capturing the dynamics of toxicity diffusion. The findings suggest that employing the SEIQR model and categorizing toxicity into moderate and high levels is reliable for understanding and predicting the spread of toxicity in various contexts. Hence, these analyses offer an answer to **RQ2**.

To analyze the evolution of toxicity spread on social media over time, numerical simulations were performed using parameter values in Table II together with the initial values $(S = 33403; I = 1799;\ E = 2000; Q = 1000; R = 700)$. The behavior of toxicity propagation, influenced by the model parameters, aligned with the predictions of the SEIQR model. The results shown in Figure 6 represent the fractal-fractional spread model. The simulation results show that reducing the fractional order (α) and the fractal dimension (τ) from 1, leads to a decrease in the dynamics of the SEIQR model. This result means that the system has longer memory and more complex network structure, which slow down the spread of toxicity as past interactions continue to influence current behavior. Interventions aimed at increasing the memory effect and enhancing network complexity could be effective in reducing the spread of toxicity.

In Figures 7 and 8, we show the numerical comparison of fractional dynamics only and fractal dynamics only on the toxicity spread SEIQR model. From Figures 7(a) and 7(b), we can see that the fractional-only dynamics exhibit more asymptotic behavior compared to the fractal-only dynamics. This means that the system approaches a steady state more gradually when considering memory effects. This asymptotic behavior can be seen from Figures 7(c) to 7(f) and Figures 8(a) to 8(d). The numerical comparison of fractional and fractal dynamics in the SEIQR model highlights the importance of considering both memory effects and network structure in understanding and mitigating the spread of toxicity on social media. These analyses offer an answer to **RQ3**.

Responding to **RQ4**, in Figures 9(a) to 9(e), we present the numerical sensitivity analysis of the toxicity spread SEIQR model under the Caputo fractal-fractional operator with order and dimension $(\alpha = \tau)$. This analysis was conducted by varying the rate at which toxic users were transferred to quarantine. The results revealed a significant increase in the number of users in each compartment, except for the infected users, as shown in Figures 9(a), 9(b), 9(c), 9(d), and 9(e), respectively. These findings suggest that reducing the number of infected users, by leveraging memory effects (fractional dynamics) and addressing repeated behavioral patterns (fractal

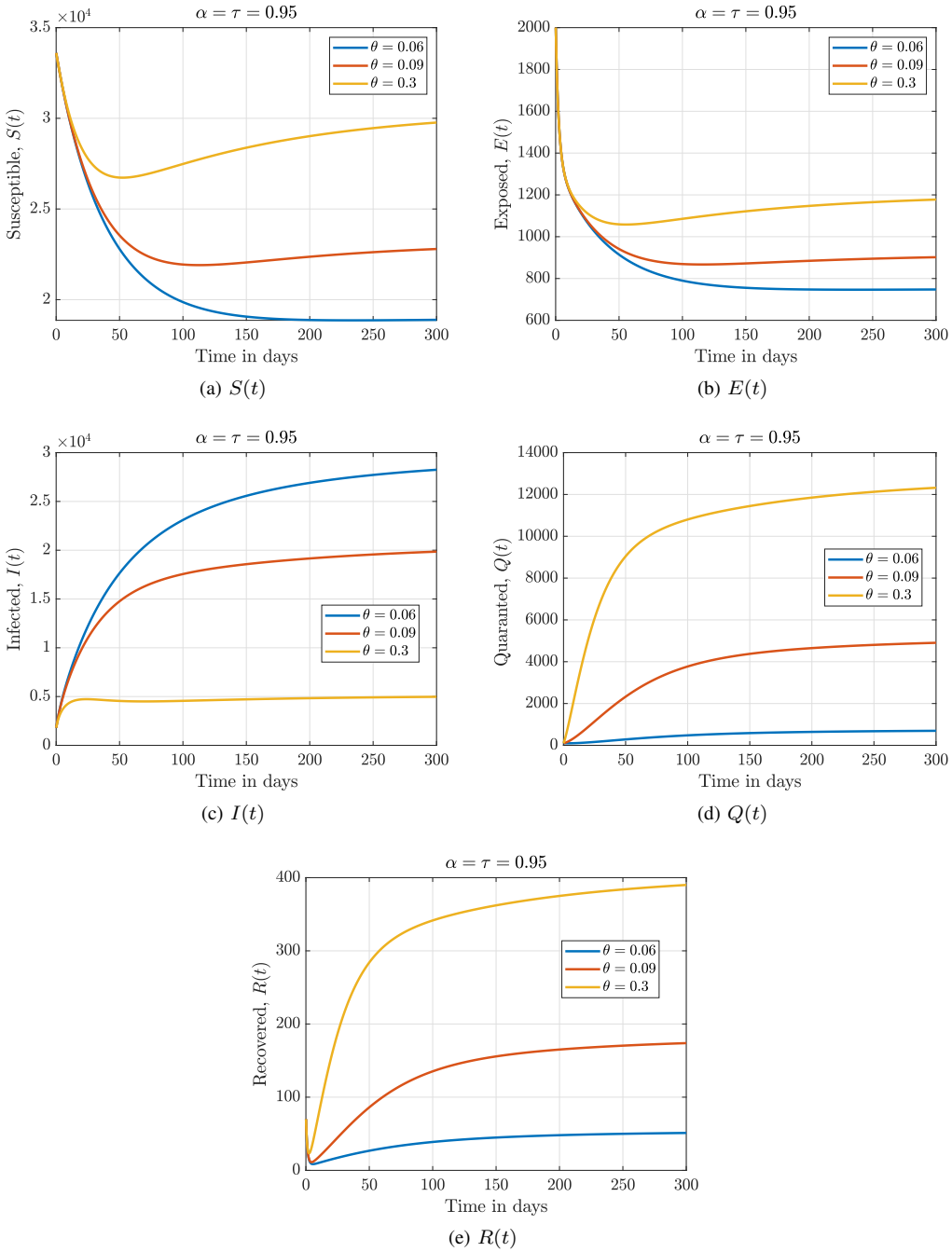


Fig. 9: Numerical trajectories of $SEIQR$ model under the Caputo fractal-fractional operator with order and dimension $(\alpha = \tau = 0.95)$ when the rate at which users transfer to quarantine.

dynamics), will effectively mitigate the spread of toxicity on social media.

Building on the insights from the SEIQR model's numerical simulations and sensitivity analysis, several strategic interventions are recommended to control the spread of toxicity on social media platforms. These measures leverage fractional and fractal dynamics to enhance content moderation and behavioral management. **First**, we suggest adaptive moderation through behavioral memory and recurrence patterns. To effectively manage toxicity, platforms should implement temporary restrictions on accounts engaged in harmful behavior. These restrictions may include suspensions, limited posting capabilities, or increased scrutiny. This approach acknowledges the role of memory effects in user behavior, ensuring that past interactions influence future activities. By allowing monitored users to reform while preventing unchecked toxicity from gaining momentum, this method balances community safety with user rehabilitation. **Secondly**, we suggest algorithmic content demotion and network complexity reduction. Reducing the visibility of toxic content using algorithmic content demotion can limit its reach and engagement. By prioritizing non-toxic interactions and reducing amplification through engagement restrictions (e.g., limiting likes, shares, and comments on flagged content), the network's fractal structure becomes less conducive to toxicity proliferation. This intervention effectively slows down the contagion effect, aligning with the SEIQR model's findings that lowering the fractal dimension reduces toxicity spread. **Thirdly**, we suggest cross-platform collaboration for unified toxicity control. Toxic content often spreads across multiple social media networks, requiring cross-platform collaboration for consistent moderation. A unified approach—incorporating shared databases, technological innovation, and standardized policies—can enhance intervention efficiency. Platforms can integrate collective quarantine mechanisms, ensuring that users flagged for toxicity on one platform face restrictions across multiple services, thereby mitigating repeated behavioral patterns. **Lastly**, we suggest structured quarantine and reintegration mechanisms. The SEIQR model suggests that quarantining toxic users is an effective strategy to limit further spread. Platforms should establish dedicated discussion spaces where quarantined users can engage in monitored interactions rather than simply migrating to another platform. This could include controlled forums, restricted access groups, or supervised comment sections, ensuring that users receive guidance toward positive behavioral reform.

By implementing these memory-aware and network-structured interventions, social media platforms can effectively mitigate the spread of toxicity, reduce harmful interactions, and foster a healthier digital environment.

VI. Conclusion and Future Works

This study applied the SEIQR epidemiological model to analyze the propagation of toxicity on social media, integrating fractal-fractional dynamics to capture the complexity of online interactions. By categorizing toxicity into moderate and high intensities, the model significantly improves prediction accuracy, as evidenced by lower error rates across diverse datasets (as seen in Table III). This granular classification aligns with real-world observations that toxicity manifests in varying degrees of severity, each requiring distinct intervention strategies. The sensitivity analysis further underscores the critical role of parameters such as the contact rate (β) and quarantine rate (θ) in modulating the basic reproduction number ($\mathcal{R}_0$), offering actionable insights for platform regulators.

The fractal-fractional approach addresses two key limitations of traditional models:

1) The **memory effect** of user interactions, where past behaviors influence future toxicity spread, and
2) The **heterogeneous network structure** of social media, characterized by super-spreaders and echo chambers.

Simulations (seen in Figures 6–9) demonstrated that reducing the fractal dimension (τ) and fractional order (α) slowed toxicity propagation, highlighting the importance of algorithmic adjustments to demote toxic content and disrupt feedback loops. These findings resonate with prior epidemiological studies on quarantine efficacy but extend their applicability to digital behaviors through fractal-fractional operators. The summary of our key contributions include the following:

1) **Improved Predictive Accuracy:** Splitting toxicity into moderate and high levels reduces model error rates by up to 90% (e.g., Peru dataset error drops from 0.41 to 0.095), enabling precise resource allocation.
2) **Parameter Sensitivity:** The positive correlation of β, ψ, and $\mathcal{A}$ with $\mathcal{R}_0$ underscores the need to limit exposure to toxic content, while the negative impact of μ and θ validates quarantine policies.
3) **Fractal-Fractional Dynamics:** By embedding memory effects and network complexity, the model reveals that interventions targeting behavioral persistence (e.g., temporary suspensions) and algorithmic amplification (e.g., content demotion) are critical for mitigation.

However, the study has limitations. The reliance on threshold-based toxicity classification (e.g., Detoxify scores > 0.5) may oversimplify nuanced human communication. Additionally, the model assumes homogeneous mixing within compartments, which may not fully capture the fragmented nature of online communities. Future iterations could incorporate network-specific topology data to refine compartmental transitions.

Acknowledgments

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] N. Agarwal, "Modeling Quarantine Intervention for Varied Toxic Intensities," in *Proc. 15th Int. Conf. Soc. Media Technol., Commun., and Informatics (SOTICS)*, Lisbon, Portugal, Sep. 2025.

[2] N. Agarwal and T. Falade, "Beyond Algorithmic Detection: How Community Expertise Moderates Toxic Algospeak via Dual Cognitive Mechanisms," in *Proc. 32nd Americas Conf. Inf. Syst. (AMCIS)*, Reno, NV, USA, Aug. 2026.

[3] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of the ACM SIGCOMM International Measurement Conference*, 2009.

[4] S. Shajari, M. Alassad, and N. Agarwal, "Characterizing suspicious commenter behaviors," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*. Kusadasi, Turkiye: ACM, Nov. 2023, pp. 631–635.

[5] L. Dai, X. Liu, and Y. Chen, "Global dynamics of a fractional-order sis epidemic model with media coverage," *Nonlinear Dynamics*, vol. 111, pp. 19 513–19 526, 2023.

[6] T. C. Falade and N. Agarwal, "Weighted Focal Structure Analysis for Coordinated Toxicity Propagation in Social Networks," in *Proc. 14th Int. Conf. Complex Netw. Their Appl. (Complex Networks)*, New York, NY, USA, Dec. 2025.

[7] K. DiCicco, N. B. Noor, N. M. Yousefi, B. Spann, and N. Agarwal, "Toxicity and networks of covid-19 discourse communities: A tale of two social media platforms," in *Proceedings of the 3rd Workshop on Reducing Online Misinformation through Credible Information Retrieval*, 2023, pp. 30–42.

[8] T. C. Falade and N. Agarwal, "Probabilistic Soft Logic for Toxic Intent Prediction in Conversation: A Moral Foundations Theory-Driven Neural-Symbolic Framework," in *Proc. 59th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Maui, HI, USA, Jan. 2026.

[9] N. Agarwal and T. C. Falade, "Context-Dependent Toxicity Propagation in Online Discourse: A Generational Analysis Framework for Characterizing Conversation Dynamics in Reddit," *IEEE J. Soc. Comput.*, 2026.

[10] D. DeMarsico, N. Bounoua, R. Miglin, and N. Sadeh, "Aggression in the digital era: Assessing the validity of the cyber motivations for aggression and deviance scale," *Assessment*, vol. 29, no. 4, pp. 764–781, 2022.

[11] N. B. Noor, N. Yousefi, B. Spann, and N. Agarwal, "Comparing toxicity across social media platforms for covid-19 discourse," in *Proceedings of the Ninth International Conference on Human and Social Analytics*, 2023, pp. 21–26.

[12] N. Yousefi, N. B. Noor, B. Spann, and N. Agarwal, "Towards developing a measure to assess contagiousness of toxic tweets," in *Proceedings of the International Workshop on Combating Health Misinformation for Social Wellbeing*, 2023.

[13] N. Yousefi, N. Noor, B. Spann, and N. Agarwal, "Examining toxicity's impact on reddit conversations," in *Proceedings of the Complex Networks & Their Applications XII*, H. Cherifi, L. M. Rocha, C. Cherifi, and M. Donduran, Eds. Cham: Springer Nature Switzerland, 2023, pp. 401–411.

[14] T. Li, S. Wang, and B. Li, "Research on suppression strategy of social network information based on effective isolation," *Procedia Computer Science*, pp. 131–138, 2018.

[15] R. Das and W. Ahmed, "Rethinking fake news: Disinformation and ideology during the time of covid-19 global pandemic," *IIM Kozhikode Society & Management Review*, vol. 11, no. 1, 2021.

[16] G. A. Ngwa and M. I. Teboh-Ewungkem, "A mathematical model with quarantine states for the dynamics of ebola virus disease in human populations," *Computational and Mathematical Methods in Medicine*, 2016.

[17] N. Yousefi and N. Agarwal, "Studying the influence of toxicity intensity on its propagation using epidemiological models," in *Proceedings of the 30th Americas Conference on Information Systems*, 2024.

[18] E. Addai, N. Yousefi, and N. Agarwal, "Seiqr: An epidemiological model to contain the spread of toxicity using memory-index," in *Proceedings of the 18th International AAAI Conference on Web and Social Media (ICWSM): CySoc 2024 – 5th International Workshop on Cyber Social Threats*, 2024, pp. 1–12.

[19] K. Diethelm and N. J. Ford, "Analysis of fractional differential equations," *Journal of Mathematical Analysis and Applications*, vol. 265, no. 2, pp. 229–248, 2002.

[20] A. Atangana and D. Baleanu, "New fractional derivatives with nonlocal and non-singular kernel: Theory and application to heat transfer model," *Thermal Science*, vol. 21, no. 2, pp. 1035–1045, 2017.

[21] A. Atangana, "Modeling the spread of covid-19 with new fractal-fractional operators: Can the lockdown save mankind before vaccination?" *Chaos, Solitons & Fractals*, vol. 136, p. 109860, 2020.

[22] A. Atangana and M. A. Khan, "Modeling and analysis of competition model of bank data with fractal-fractional caputo-fabrizio operator," *Alexandria Engineering Journal*, vol. 59, pp. 1985–1998, 2020.

[23] L. Hanu and Unitary Team, "Detoxify," https://github.com/unitaryai/detoxify, 2020, last accessed: May 2026.

[24] K. Diethelm, N. J. Ford, and A. D. Freed, "A predictor-corrector approach for the numerical solution of fractional differential equations," *Nonlinear Dynamics*, vol. 29, pp. 3–22, 2004.

[25] A. Atangana and D. Baleanu, "New fractional derivatives with non-local and non-singular kernel: Theory and application to heat transfer model," *Thermal Science*, vol. 20, pp. 763–769, 2016.

[26] Z. Odibat and N. T. Shawagfeh, "Generalized taylor's formula and its application on a fractional differential equation," *Applied Mathematics and Computation*, vol. 197, pp. 467–477, 2008.

[27] M. Blanc and O. Bournez, "A characterization of polynomial time computable functions from the integers to the reals using discrete ordinary differential equations," in *International Conference on Machines, Computations, and Universality*. Springer, 2022, pp. 58–74.