Grenze International Journal of Engineering and Technology, June Issue

Information Diffusion on Instagram using SIR and SEIR Models

Manoj Kumar Srivastav[1], Somsubhra Gupta[2] and Subhranil Som[3]
[1-2]School of Computer Science, Swami Vivekananda University, Barrackpore, West Bengal, India
Email: mksrivastav2015@gmail.com, gsomsubhra@gmail.com
[3]Department of Computer Science, Bhairab Ganguly College, Kolkota, West Bengal, India
Email: subhranil.som@gmail.com

Abstract*— Social Network Analysis (SNA) studies how users interact with each other on online platforms. It helps in understanding how information moves from one user to another. Information diffusion explains how content spreads among users over time. In this study, the spread of information on social media is analyzed using epidemic models, namely the Susceptible–Infected–Recovered (SIR) model and the Susceptible–Exposed–Infected–Recovered (SEIR) model. The dataset is based on Instagram influencer data and includes variables such as the number of followers, average likes, total likes, and number of posts, influence score, and engagement rate. Since social media datasets do not directly provide SIR components, proxy variables are defined. Followers represent susceptible users, average likes represent actively engaged users (infected), and total likes per post represent recovered users. All variables are normalized to make them comparable and stable. The SIR and SEIR models are simulated for 30 days. The comparison is based on two measures: peak infection and the time taken to reach the peak. The results show that the SIR model gives higher and faster peaks. The SEIR model shows a more realistic pattern because it includes a delay in user engagement through the exposed stage. This study uses a proxy-based approach due to limited observation of actual user states. The mapping of followers, average likes, and total likes per post to SIR compartments is treated as an approximation of aggregated engagement behavior rather than exact user states. A sensitivity study is performed by varying key parameters such as the infection rate (β), recovery rate (γ), and incubation rate (α). In addition, a basic empirical validation is carried out by examining the consistency between simulated diffusion patterns and observed engagement trends. The results provide a constrained but practical interpretation of information diffusion in large-scale social media environments.

***Keywords*— Social Network Analysis, Information Diffusion, SIR Model, SEIR Model.**

I. INTRODUCTION

Social network platforms have grown very fast in recent years and are now important for communication and sharing information [1]. Platforms like Facebook and Instagram allow users to share ideas, opinions, and content with many people in a short time. These platforms create networks [2] where users are connected through interactions such as likes, comments, and shares. Because of this, information spreads quickly from one user to another; similar to how people share information in real life. Information diffusion is the process by which ideas, knowledge, or influence pass from one user to another through social connections [3].

Grenze ID: 01.GIJET.12.2.293_26

It usually starts with a small group of users and gradually spreads to a larger audience. Over time, the spread slows down as fewer new users remain to engage. The diffusion process depends on several factors [4]. Social networks are used in many areas such as marketing, public awareness, and controlling misinformation. The spread of information depends on the relationships and interactions between users. Some users play a key role in spreading information, especially highly connected influencers.From a network point of view, information diffusion has three main parts. The sender starts the spread of information. The receiver gets the information and may share it further. The medium is the channel through which information is shared. These parts work together to create different diffusion patterns, which are often difficult to predict. On online platforms, information can spread very quickly. Sometimes it becomes viral and reaches many users, but not all content spreads widely. This shows the need to understand how diffusion works. Mathematical models help to study information diffusion. Epidemic models are commonly used for this purpose[5]. The SIR model divides users into three groups: susceptible, infected, and recovered.

This study extends earlier work based on the SIR model. It also includes the SEIR model. This helps to capture delayed user engagement in social networks [6]. The SEIR model is an extension of the SIR model. It adds one more stage called exposed, which shows the delay between seeing content and interacting with it. Applying these models to social media data is difficult because real data does not directly provide these categories. Traditional methods study relationships between variables like followers and likes, but they do not show how engagement changes over time. Epidemic models such as SIR and SEIR help to study this process step by step. The SEIR model is more realistic because it includes the exposed stage. These models show how engagement grows, reaches a peak, and then stabilizes over time.

Real-world social media data does not directly show categories such as susceptible, infected, or recovered users. To solve this problem, proxy variables are used in this study. Followers are treated as susceptible users, while average likes represent actively engaged users. Total likes per post represent users who have interacted in the past. This mapping is based on earlier studies. Earlier studies use only one diffusion model or depend mainly on theoretical data. In this study, both SIR and SEIR models are compared using Instagram influencer data. Followers and likes are used as proxy variables to represent different diffusion stages. This helps to study information diffusion using real engagement data in a simple and practical way.

The main objective of this study is to analyze and compare the SIR and SEIR models using real Instagram influencer data. The study focuses on important aspects such as peak engagement and the time taken to reach peak diffusion. The comparison helps to understand which model fits real data better.

Epidemic models such as the SIR model are widely used to study information diffusion. However, most studies assume that user-level transition data is available. This type of data is not present in real social media datasets. On platforms like Instagram, only overall engagement data such as followers and likes are available. It is not known which users have seen or interacted with the content. Because of this, there is a gap between theoretical models and real data.

To deal with this problem, this study uses a proxy-based approach. Engagement variables are used as approximate values for different states. The aim is not to track exact user behavior, but to understand overall engagement patterns. The study compares the SIR and SEIR models to examine the effect of delayed engagement. A sensitivity analysis and a basic validation are also performed to check the stability of the results.

The main contributions of this study are as follows:

- A proxy-based mapping of social media engagement variables to epidemic model states under data constraints.
- A comparative evaluation of SIR and SEIR models using real Instagram influencer data.
- A sensitivity analysis is done by changing the main parameters (β, γ, α) to see how they affect the spread.
- A basic comparison is done between simulated results and real engagement data to check if the pattern is similar.

II. Literature Review

A. Social Network Analysis and Information Diffusion

Social Network Analysis (SNA) studies how users interact on online platforms. It explains user connections and how information flows between them. Information diffusion means the spread of ideas, news, or influence in a network. Li et al. (2017) explained that diffusion depends on many factors[4]. These include network structure, user behavior, and content type. Therefore, the diffusion process is complex and dynamic. Everett M. Rogers (2003) defined diffusion as the spread of new ideas over time among people in a social system[3]. In modern social media platforms, information spreads rapidly due to high user connectivity and continuous activity.

B. Epidemic Models for Information Diffusion (SIR Model)

Epidemic models are used to study information diffusion in social networks. The Susceptible–Infected–Recovered (SIR) model is a simple and commonly used model. In this model, users are divided into three groups: susceptible, infected, and recovered. Hilmi and Sharif (2022) applied the SIR model to study Instagram-based marketing [7].

These models treat information diffusion similar to disease transmission. However, the SIR model assumes immediate user response after exposure, which may not reflect real user behavior on social media [8]. Earlier studies used the SIR model for social network analysis [6].

C. SEIR Model and Extended Epidemic Models

To overcome the limitations of the SIR model, the SEIR model adds one more state. This stage represents users who have seen the content but have not yet interacted. This helps to show delay in user response. Jiang et al. (2021) showed that the SEIR model explains user engagement better [9]. It is useful for modeling multi-stage interaction processes in social media [10]. Other extensions such as SEI and SEIR models further improve realism by incorporating repeated exposure and interaction dynamics.

D. Network Structure and Neighbor Influence

Network structure is important in information diffusion. It affects both the speed and reach of information spread. Connections between users matter. Clustering of users also matters. Interaction strength also affects diffusion. Xiao et al. (2019) explained that dense networks spread information quickly in a local area [11]. Sparse networks help information reach more users. Bioglio and Pensa (2017) showed that users are influenced by their neighbors [12]. This influence affects how information spreads in the network.

E. Platform Behavior and Social Media Diffusion

Social media platforms affect how information spreads. Their algorithms and engagement features decide what users see. Recommendation systems and ranking algorithms often promote popular content. This allows the content to reach more people quickly. Cinelli et al. (2020) showed that this can lead to very fast spread of information [13]. The spread is often uneven across users. It is similar to an infodemic. An infodemic refers to the rapid spread of large amounts of information, both accurate and misleading, making it difficult to identify reliable content. This shows that platform behavior plays an important role in information spread in social networks.

F. Mathematical Perspective of Diffusion Models

Epidemic models such as the SIR and SEIR models use mathematical equations. The models use parameters such as infection rate and recovery rate. These parameters control how the spread happens. In social networks, these parameters can be linked to engagement data. For example, likes, followers, and interaction rates can be used. This helps the models represent real information spread more clearly[5].

G. Limitations of Existing Studies

Despite many useful contributions, existing studies have some limitations. Many models assume uniform user behavior and constant diffusion rates, which is unrealistic in real-world scenarios. Most studies focus on a single diffusion model and do not provide comparative analysis using real-world social media datasets. In addition, user states such as susceptible, exposed, infected, and recovered are usually not directly available in social media data, making practical implementation difficult.

H. Research Gap and Proposed Work

There is limited research comparing SIR and SEIR models using real social media datasets. Key aspects such as peak engagement and time to peak are not analyzed in depth in existing studies. Real social media datasets usually provide only engagement measures such as followers and likes rather than direct user-state information. Because of this, diffusion behavior cannot be observed directly. Very few studies use engagement variables as proxy values for diffusion analysis using real social media data.

Proposed Work: This study compares the SIR and SEIR models using real Instagram influencer data. Engagement-based variables such as followers, average likes, and total likes are used to represent the susceptible, infected, and recovered states. The study analyzes how information spreads over time and examines peak engagement and the time taken to reach peak diffusion. This approach helps to understand diffusion patterns in social networks in a more realistic and practical way.

III. Methodology

A. Data Description

The dataset used in this study is based on Instagram influencer data [14]. The dataset contains information on Instagram influencers, including the number of followers, average likes, total likes, and number of posts, influence score, and 60-day engagement rate. These variables are used to analyze user engagement and information diffusion on social media platforms.

B. Data Preprocessing

Before applying the model, data preprocessing is performed. Missing values are replaced with zero. Some numerical values are written in short forms such as ‘k’ (thousand), ‘m’ (million), and ‘b’ (billion). These values are converted into proper numeric form.

The 60-day engagement rate is given in percentage form and is converted into decimal form. These steps make all variables uniform for the simulation.

C. Variables Used in the Model

Only selected variables from the dataset are used in the model. The number of followers is used to represent the Susceptible (S) population, while average likes (avg_likes) are used to represent the Infected (I) population. Total likes are used in the calculation of the Recovered (R) population and are combined with the number of posts to estimate long-term engagement.

The influence score is used to calculate the infection rate (β), and the 60-day engagement rate is used to calculate the recovery rate (γ).

The variable new_post_avg_like is not included in the model because it provides information similar to average likes and may introduce redundancy and instability into the analysis.

D. Model Mapping

Epidemic models are mainly used to study the spread of diseases. To use them for social networks, some changes are required. Here, the idea is applied to information diffusion.In this study, users are divided into three groups:

- Susceptible (S): These are followers who have not interacted with the content.
- Infected (I): These are users who are currently active. They like or interact with the content.
- Recovered (R): These are users who interacted earlier but are not active now.

This mapping does not show exact user-level transitions. So, S, I, and R are treated as overall engagement levels, not exact categories. The same users may appear in more than one group. Because of this, the mapping is only an approximation to study diffusion at a general level.

The model in table 1 should be seen as a simple representation of overall diffusion, not a strict epidemiological model. The table 2 shows the difference between average likes and total likes per post. These variables are used to represent current and past user activity.

TABLE I: MODEL MAPPING (SIR REPRESENTATION)

Variable	Definition	Time Perspective	Interpretation	Model Mapping
Followers	Total number of followers	Potential reach	Users who have not yet interacted	Susceptible (S)
Average Likes	Likes on recent posts	Present (short-term)	Current active engagement	Infected (I)
Total Likes / Posts	Total likes divided by number of posts	Past (long-term)	Historical engagement	Recovered (R)

TABLE II: FEATURE COMPARISON (ENGAGEMENT ANALYSIS)

Aspect	Average Likes	Total Likes / Posts
Definition	Average likes on recent posts	Total likes divided by posts
Data Type	Average value	Overall value
Time	Present activity	Past activity
Users	Users active now	Users active before
Overlap	May be same users	May be same users
Values	Can be similar	Can be similar
User Identity	Exact users not known	Exact users not known
Meaning	Current engagement	Overall engagement
Role in Model	Active users	Previously active users
SIR Model	Infected (I)	Recovered (R)

E. Interpretation of Engagement Variables

In this study, user engagement is represented using two variables: average likes and total likes per post.The average likes (avg_likes) value is taken from the dataset. It shows the average likes per post. New posts usually get more likes. Old posts get fewer likes. So, this value shows the current engagement level. Therefore, it is used to represent the active stage and is taken as Infected (I).

The total likes per post are calculated by dividing total likes by the total number of posts. It uses all posts, not just recent ones. Because of this, it shows the overall or long-term engagement. The total likes per post are calculated as:

$$R = \frac{\text{total\_likes}}{\text{posts}}$$

This value uses total likes and total number of posts. It includes all posts of the account. So, it represents the full posting history. It is more stable than average likes. User behavior changes over time. At the beginning, users like and interact more. After some time, interaction decreases..Because of this, engagement becomes stable. So, total likes per post represents this stable stage. It is used as the Recovered (R) category. The values of average likes and total likes per post are close. But they are not exactly the same .Average likes are more affected by recent posts. Total likes per post includes all posts, both old and new. Also, values are rounded. Likes are not equal for every post. Because of this, small differences appear.

TABLE III: CONCEPTUAL DIFFERENCE

Aspect	Average Likes (I)	Total Likes per Post (R)
Source	Direct dataset value	Calculated
Posts used	Mostly recent influence	All posts
Time focus	Present	Long-term
Nature	Dynamic	Stable
Role in model	Active stage	Saturation stage

The dataset also provides new_post_avg_like, but it is not used because:

- It is based on very few recent posts
- It is highly unstable
- Can change due to one viral or low-performing post

In contrast, avg_likes gives a more balanced and consistent measure of engagement.

F. Normalization

Normalization is applied to make values comparable across influencers. All variables are adjusted using the maximum follower count in the dataset.

The normalization process is defined as:

$$S^N = \frac{\text{followers}}{\text{max\_followers}}, \quad I^N = \frac{\text{avg\_likes}}{\text{max\_followers}}, \quad R^N = \frac{\text{total\_likes / posts}}{\text{max\_followers}}$$

Where max(followers) represents the maximum follower count in the dataset used for normalization.All variables are normalized by dividing them by the maximum number of followers, and are denoted as S^N **,**I^N and R^N**.**

G. Model Representation (SIR Mapping)

The variables are mapped into the SIR model as follows:

TABLE IV: MODEL VARIABLE

Model Variable	Formula Used	Meaning
Susceptible (S^N)	followers / max_followers	Users who have not interacted
Infected (I^N)	avg_likes / max_followers	Users currently active
Recovered (R^N)	(total_likes / posts) / max_followers	Users active in the past

Important Assumption

Average likes are used as a proxy for current engagement, while total likes normalized by the number of posts are used as a proxy for past engagement. The new post average like variable is not used to avoid duplication, as it represents similar information to average likes

H. Parameters

The model parameters are defined based on influencer characteristics and engagement behavior:
Infection Rate (β): The infection rate (β) represents how quickly information spreads among users. It is defined as:

$$\boldsymbol{\beta} = \frac{\text{influence\_score}}{\mathbf{100}} \ \mathbf{X\ 0.1}$$

- The influence score reflects the ability of an account to reach and affect a large audience.
- Dividing by 100 normalizes the value.
- Multiplying by 0.1 scales it to a realistic range for diffusion.
- β shows the spreading power of the influencer.
- A scaling factor of 0.1 is used to keep β in a proper range.
- If a smaller value like 0.01 or 0.05 is used, β becomes too low. This makes the spread very slow.
- If a larger value like 0.2 or 0.5 is used, β becomes too high.This makes the spread too fast and unrealistic.
- So, 0.1 is chosen as a balanced value. It keeps the model stable and realistic.

Recovery Rate (γ): The recovery rate (γ) represents the rate at which active users stop engaging with content. It is defined as:

$$\gamma = \frac{\text{engagement\_rate}}{100}$$

- The 60-day engagement rate reflects how users interact with content over time.
- It indicates how quickly engagement stabilizes.
- Dividing by 100 converts' percentage into proportion.

Incubation Rate (α): The incubation rate (α) represents the transition from exposed to active users.
$\alpha = 0.2$

- A constant value is used to represent a moderate transition speed.
- It assumes that users take some time before actively engaging after seeing content.
- The value of α is taken as 0.2. It shows a moderate transition from seeing a post to liking it.
- If α is 0.1, the change is very slow. Users take more time to interact. This is not realistic for social media.
- If α is 0.4 or 0.5, the change is very fast. Users interact almost immediately. This is also not realistic.
- So, α = 0.2 is chosen as a balanced value. It keeps the model stable and realistic.

These parameter values are taken as initial assumptions. Their effects on diffusion are tested using sensitivity analysis. The choice of parameters is heuristic due to the absence of direct diffusion observations. The infection rate (β) is assumed proportional to the influence score, reflecting the ability of an influencer to spread information. The recovery rate (γ) is derived from engagement decay patterns, approximated using the 60-day engagement rate. While these assumptions simplify real-world dynamics, they allow the model to approximate aggregated diffusion behavior under data constraints."

I. Models

SIR Model -- The SIR model is defined by the following discrete-time equations:
$S_{t+1} = S_t – \beta * (S_t * I_t)$
$I_{t+1} = I_t + \beta * (S_t * I_t) - \gamma * I_t$
$R_{t+1} = R_t + \gamma * I_t$
SEIR Model--- The SEIR model is defined by the following discrete-time equations:
$S_{t+1} = S_t – \beta * (S_t * I_t)$
$E_{t+1} = E_t + \beta * (S_t * I_t) - \alpha * E_t$
$I_{t+1} = I_t + \alpha * E_t – \gamma * I_t$
$R_{t+1} = R_t + \gamma * I_t$
S^N , I^N, and R^N are used only for normalization, and these values are taken as initial conditions (S_0, I_0, R_0) in the model, while St , It , and Rt are used in the equations during simulation. Where:

- St: susceptible users at time t
- Et: exposed users at time t
- It: actively engaged users at time t
- Rt: stabilized users at time t
- β: diffusion (infection) rate
- γ: recovery (stabilization) rate

- α: transition rate from exposure to engagement

Exposed Compartment (E) in SEIR Model

In the SEIR model used in this study, the Exposed (E) compartment represents users who have seen the content but have not interacted with it. Users are aware of the content but have not liked, commented, or shared it. This stage represents the gap between seeing and reacting, as immediate response is not common on social media.

This makes the exposure stage important in modeling information spread. The exposed compartment is not directly available in the dataset. Therefore, E is introduced as a latent (unobserved) state in the model to improve behavioral realism.

Since no direct measurement of exposure is available, the initial value of the exposed compartment is assumed as a proportion of the infected state (I). The initial exposed value is defined as $E_0=0.5*I_0$, where I_0 is the normalized initial infected value.$E_0 = 0.5 \times I_0$ is used as a balanced assumption to represent delayed engagement. A very small value (such as 0.1 × I) would make the exposed group too small and ignore delayed engagement. A very large value (such as 0.8 × I) would make the exposed group too large and affect the model results. Therefore, 0.5 is taken as a middle value to keep the model balanced and stable, and to reflect both immediate and delayed user engagement.

During the simulation, the value of E changes over time based on the SEIR model equation.

$$E_{t+1} = E_t + \beta^* (S_t * I_t) - \alpha * E_t$$

Where $\beta^* (S_t * I_t)$ represents the flow of users entering the exposed state due to interaction between susceptible and infected users, and $\alpha * E_t$ represents the transition from exposed to active engagement.

The parameter α controls how fast exposed users become active users. It shows how quickly a user decides to interact after seeing the content. A moderate value is used to reflect that users usually take some time before reacting. The exposed group acts as a middle stage between users who do not interact and those who actively engage. This makes the SEIR model more realistic than the SIR model. The exposed variable is not taken from the dataset. It is introduced in the model to represent delayed user response.

J. Simulation

Both SIR and SEIR models are simulated for each influencer. The simulation is performed for 30 days. At each step, model variables are updated according to the corresponding equations. The results are used to compare engagement dynamics and diffusion behavior between the SIR and SEIR models.

IV. Results

A. Dataset Overview

The dataset shown in Table 5 contains 200 Instagram influencers [14] ranked by influence score. It includes variables such as followers, average likes, total likes, number of posts, and engagement rate.

Before simulation, preprocessing is done. Missing values are replaced with zero. Values like k, m, and b are converted into numeric form. Engagement rate is converted into decimal form. Normalization is applied using the maximum follower count to make values comparable. The SIR and SEIR models are run for 30 days.

TABLE V: TOP 200 INSTAGRAM INFLUENCERS

rank	Influence_score	posts	followers	avg_likes	60_day_eng_rate	new_post_avg_like	total_likes	country
1	92	3.3k	475.8m	8.7m	1.39%	6.5m	29.0b	Spain
2	91	6.9k	366.2m	8.3m	1.62%	5.9m	57.4b	United States
3	90	0.89k	357.3m	6.8m	1.24%	4.4m	6.0b	
200	80	4.2k	32.8m	232.2k	0.30%	97.4k	969.1m	Indonesia

B. Data Distribution

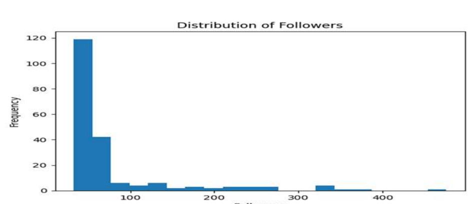

Figure 1: Distribution of Followers

The distribution of followers is not equal across influencers. A few influencers have very high followers, while most have lower followers. This affects how information spreads.
Figure 1 shows that follower distribution is right-skewed. A small number of influencers have very large audiences. These influencers play a major role in spreading information.

C. Followers and Engagement Relationship

Followers and average likes have a positive relationship. The relationship is weak to moderate. Influencers with more followers usually get more likes. Figure 2 shows the relationship between followers and average likes. There is a positive relationship, but it is not very strong. This means higher followers do not always give proportionally higher engagement. Other factors like content quality also affect engagement.

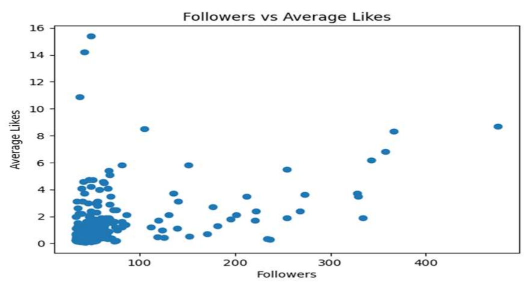

Figure 2: Followers vs Average Likes

D. Model Simulation Results

SIR Model: In the SIR model, the number of infected users increases quickly. This shows that users start interacting immediately. The number of susceptible users decreases fast. The number of recovered users increases slowly over time.

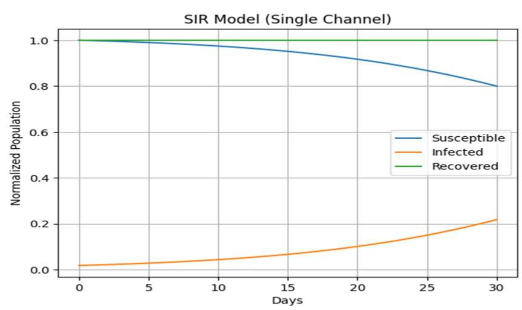

Figure 3: SIR Model Simulation

Figure 3 shows the SIR model results. The number of infected users increases quickly in the beginning. This means users start interacting immediately. The susceptible users decrease fast. The recovered users increase steadily over time.
SEIR Model: The SEIR model adds an exposed stage. Because of this, the number of infected users increases slowly .There is a delay before users become active. This makes the spread smoother. The peak is lower compared to the SIR model.

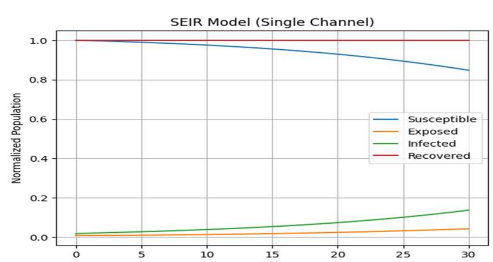

Figure 4: SEIR Model Simulation

Figure 4 shows the SEIR model results. The increase in infected users is slower compared to the SIR model. This is because users first move to the exposed stage. After that, they become active. This creates a delay in engagement.

E. Peak Infection Analysis

The SIR model gives higher peak values than the SEIR model. This shows stronger and faster diffusion in the SIR model. The time to reach peak is similar in both models.

TABLE VI: COMPARISON IF SIR AND SEIR RESULTS

Channel	SIR Peak	SIR Time	SEIR Peak	SEIR Time
1	0.218001	30	0.137018	30
2	0.120853	30	0.094923	30
3	0.094473	30	0.075514	30
4	0.085172	30	0.068377	30
5	0.025123	30	0.020409	30

F. Comparison of Peak Infection

Peak infection values are compared for 10 influencers. This shows how engagement spreads in both models.

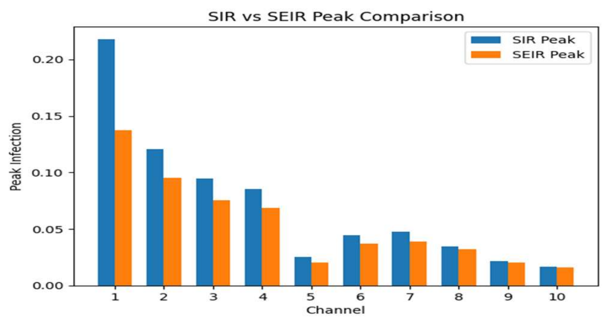

Figure 5: Peak Infection Comparison (SIR vs SEIR)

Figure 5 shows that the SIR model always gives higher peak infection values. This is because it assumes immediate user interaction. The SEIR model gives lower peaks due to delay in engagement.

G. Time to Peak Comparison

Time to reach peak is compared for 10 influencers. This shows how fast engagement reaches its maximum in both models.

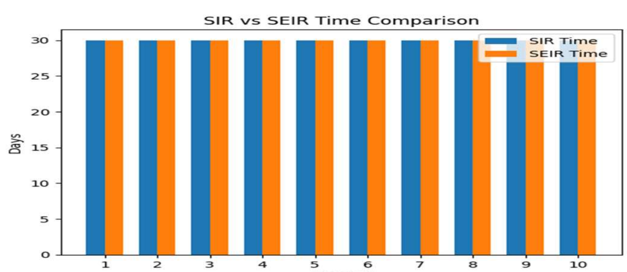

Figure 6: Time-to-Peak Comparison

Figure 6 shows that both models reach the peak at almost the same time. The SEIR model has a delay in the initial stage, but the overall simulation time remains similar for both models.

H. Large-Scale Analysis

The comparison is done for 200 influencers. The SIR model gives higher peak values. The SEIR model gives lower peak values. This pattern is seen for most influencers. The difference is larger for top influencers.
It becomes smaller for lower-ranked influencers. This shows that the SEIR model giving more balanced and realistic result.

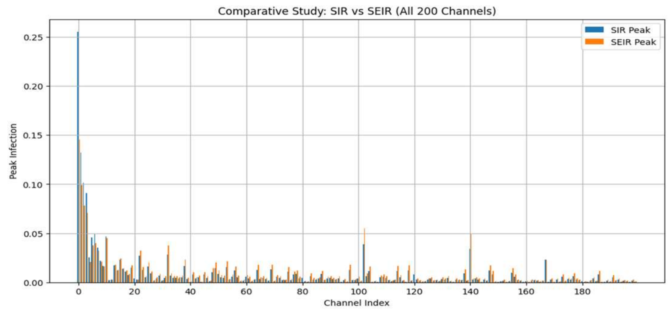

Figure 7: Peak Infection for 200 Influencers

Figure 7 shows results for all influencers. The SIR model gives higher peaks for most influencers. The difference is larger for top influencers and smaller for lower-ranked ones.

I. Empirical Validation and Consistency Analysis

Since user-level diffusion data is not available, direct validation is not possible. So, an indirect method is used to check whether the simulated results match real engagement patterns. A comparison is made between real engagement values (such as average likes and followers) and simulated peak infection values. Influencers with higher engagement usually show higher peak values in both SIR and SEIR models. This shows a positive relationship between real data and model results. The overall diffusion pattern shows fast growth followed by stabilization. This is similar to common social media behavior. It suggests that the model can capture general diffusion trends even with proxy variables. However, this validation is limited. It shows general trends but not exact prediction accuracy. The lack of time-based user interaction data is still a limitation.

TABLE VII: SENSITIVITY ANALYSIS RESULTS

Beta (β)	Alpha (α)	Average Peak
0.05	0.1	0.0067
0.05	0.2	0.0073
0.05	0.3	0.0075
0.1	0.1	0.0082
0.1	0.2	0.0097
0.1	0.3	0.0105
0.2	0.1	0.0121
0.2	0.2	0.0168
0.2	0.3	0.0199

Table 7 shows how changes in β and α affect the average peak values. Higher values of β and α showing in higher diffusion peaks. Table 8 shows the correlation between simulated peak values and real engagement data. The positive value indicates consistency between model results and actual data.

TABLE VIII: VALIDATION RESULT

Metric	Value
Correlation (SIR Peak vs Avg Likes)	0.589

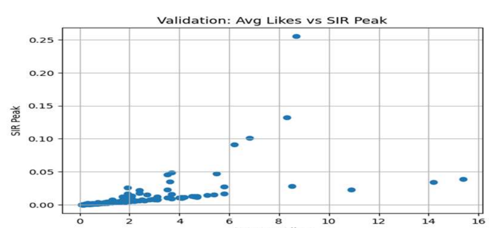

FIGURE 8: VALIDATION PLOT

Figure 8 shows the relationship between average likes and SIR peak values. The upward trend indicates that higher engagement is associated with higher diffusion peaks.

The results show that the exposed stage helps to represent delayed user engagement more clearly. The study also shows that engagement variables can be used to study diffusion behavior when direct user-level data is not available. Therefore, the approach is useful for practical social media analysis.

J. Overall Interpretation

The results show clear differences between the two models. The SIR model produces faster diffusion and higher peak engagement because users are assumed to interact immediately. In contrast, the SEIR model includes an exposed stage, creating a delay before users become active. As a result, the SEIR model generates lower and smoother diffusion peaks. Therefore, the SEIR model provides a more realistic representation of social media engagement by capturing delayed user interaction.

V. CONCLUSION AND FUTURE SCOPE

The SIR and SEIR models are applied to the Instagram influencer dataset. The simulation is performed for 30 days for each influencer. The results show how user engagement changes over time. At the beginning, the number of susceptible users is high, while active users are low. Over time, active users increase. After reaching the peak, active users decrease and recovered users increase. The SIR model shows faster information diffusion and higher peak engagement because users are assumed to interact immediately. In contrast, the SEIR model includes an exposed stage in which users see the content before becoming active. As a result, diffusion occurs more slowly and the peak remains lower. The comparison indicates that delayed engagement plays an important role in online information diffusion and that the SEIR model provides a more realistic representation of social media behavior. Followers, engagement, and influence score affect the spread of information. The study demonstrates a simple and practical approach for applying diffusion models to real social media data using engagement-based variables.

There are some limitations in this study. The model uses overall engagement values and does not track individual users. The mapping of followers, average likes, and total likes per post is only an approximate representation of user behavior. Time-based interaction data is also not available. Therefore, exact transitions between user states cannot be observed directly. The parameters (β, γ, α) are selected based on assumptions, and the results may vary when different parameter values are used. Therefore, parameter selection remains an important aspect of diffusion analysis. A basic validation is performed using real engagement data. A positive relationship is observed between engagement values and peak diffusion values, indicating that the model captures general diffusion trends. However, the model is intended to explain overall diffusion behavior rather than provide exact prediction accuracy.

Future work may incorporate time-based interaction data, content characteristics, and user behavior analysis. Machine learning techniques may also be integrated to improve model performance. In addition, the approach can be extended to other social media platforms such as Facebook and YouTube, while network structure and user connections may be included to provide a more detailed analysis of information diffusion.

REFERENCES

[1] M. K. Srivastav and A. Nath, "Study on mathematical modeling of social networks," *International Journal of Emerging Technology and Advanced Engineering*, vol. 5, no. 3, pp. 611–618, Mar. 2015.

[2] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge University Press, 1994.

[3] M. Rogers, *Diffusion of Innovations*, 5th ed. New York, NY, USA: Free Press, 2003.

[4] M. Li, X. Wang, K. Gao, and S. Zhang, "A survey on information diffusion in online social networks: Models and methods," *Information*, vol. 8, no. 4, Art. no. 118, 2017, doi: 10.3390/info8040118.

[5] J. Cifuentes-Faura, U. Faura-Martínez, and M. Lafuente-Lechuga, "Mathematical modeling and the use of network models as epidemiological tools," *Mathematics*, vol. 10, no. 18, Art. no. 3347, 2022, doi: 10.3390/math10183347.

[6] M. K. Srivastav, S. Gupta, and S. Som, "Modeling friendship dynamics in social networks: A graph-theoretic and computational approach," *Journal of Information Systems Engineering and Management*, vol. 10, no. 37s, Art. no. 90, 2025, doi: 10.52783/jisem.v10i37s.6382.

[7] N. S. A. Mohd Hilmi and N. Sharif, "Analyzing online marketing for contemporary snacks through Instagram using SIR model," *Journal of Computing Research and Innovation*, vol. 7, no. 2, pp. 23–34, 2022, doi: 10.24191/jcrinn.v7i2.280.

[8] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on Twitter," in *Proc. 7th Workshop on Social Network Mining and Analysis (SNAKDD '13)*, New York, NY, USA: Association for Computing Machinery (ACM), 2013, pp. 1–9, doi: 10.1145/2501025.2501027.

[9] J.-G. Jiang, G.-M. Meng, and S.-S. Li, "Research on product diffusion in social commerce based on improved SEIR model," in *Proc. International Conference on Electronic Communication and E-commerce (ICECEM)*. Piscataway, NJ, USA: IEEE, 2021, pp. 171–175, doi: 10.1109/ICECEM54757.2021.00039.

[10] S. Kumar, M. Saini, M. Goel, and N. Aggarwal, "Modeling information diffusion in online social networks using SEI epidemic model," *Procedia Computer Science*, vol. 171, pp. 672–678, 2020, doi: 10.1016/j.procs.2020.04.073.

[11] Y. Xiao, Z. Wang, Q. Li, and T. Li, "Dynamic model of information diffusion based on multidimensional complex network space and social game," *Physica A: Statistical Mechanics and Its Applications*, vol. 521, pp. 578–590, 2019, doi: 10.1016/j.physa.2019.01.117.

[12] L. Bioglio and R. G. Pensa, "Impact of neighbors on the privacy of individuals in online social networks," *Procedia Computer Science*, vol. 108, pp. 28–37, 2017, doi: 10.1016/j.procs.2017.05.030.

[13] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The COVID-19 social media infodemic," *Scientific Reports*, vol. 10, no. 1, Art. no. 16598, Oct. 2020, doi: 10.1038/s41598-020-73510-5.

[14] Chitaxiang, "Instagram influencer score analysis" [Dataset]. Kaggle, 2024. Available: https://www.kaggle.com/code/chitaxiang/instagram-influencer-data-analysis/input